2ndCGM: Session of Probability, Statistics & Applications

PROGRAM

(To view the webpage of the speaker or the abstract click on the affiliation or the title, respectively)

1st Day: Monday – July 4, 2022
(Room 001)

Statistics topics

16:30 – 17:30
Omiros Papaspiliopoulos (Bocconi University, Milano)
From multiway ANOVA to Alon's conjecture: complexity of coordinate-wise inference algorithms for crossed random effect models

17:30 – 18:00
Georgios Fellouris (University of Illinois at Urbana-Champaign)
Anomaly identification under sequential sampling and sampling constraints

18:00 – 18:30
Eftychia Solea (ENSAI)
Nonparametric and high-dimensional functional graphical models

18:30 – 19:00
Giorgos Vasdekis (University of Warwick)
Piecewise Deterministic Monte Carlo and the Speed Up Zig-Zag sampler

2nd Day: Tuesday – July 5, 2022
(Room 001)

Probability topics

9:00 – 10:00
Grigorios A. Pavliotis (Imperial College, London)
Mean field limits for weakly interacting diffusions: phase transitions, multiscale analysis and inference

10:00 – 10:30
Antonios Papapantoleon (Technical University, Delft)
Model-free bounds in finance: a journey through probability, statistics and optimization

10:30 – 11:00
Konstantinos Spiliopoulos (Boston University)
Mean field analysis and scaling limits for neural networks: typical events and fluctuations

---

BREAK

Statistics topics

16:30 – 17:30
Marianthi Markatou (University at Bufallo)
Statistical Distances: A Global Framework for Inference

17:30 – 18:00
George Deligiannidis (University of Oxford)
Quantitative Uniform Stability of the Iterative Proportional Fitting Procedure

18:00 – 18:30
Nikos Ignatiadis (Stanford University)
Noise-Induced Randomization in Regression Discontinuity Designs

3rd Day: Wednesday – July 6, 2022
(Room 001)

Probability topics

16:30 – 17:30
Takis Konstantopoulos (University of Liverpool)
Longest and heaviest paths in random directed graphs

17:30 – 18:00
Dimitrios Tsagkarogiannis (University of L'Aquila)
Density functionals for inhomogeneous systems

18:00 – 18:30
Aggeliki Menegaki (Institut des Hautes Études Scientifiques, Paris)
Sharp spectral gap for regular and disordered chain of oscillators

18:30 – 19:00
Konstantinos Kavvadias (University of Cambridge)
Some recent progress on the properties of Schramm-Loewner Evolution (SLE)

Organizers
G. Afendras; M. Loulakis; K. Zografos.

Second Congress of Greek Mathematicians SCGM-2022

Session: ***Probability, Statistics & Applications***

From multiway ANOVA to Alon's conjecture: complexity of coordinate-wise inference algorithms for crossed random effect models

Omiros Papaspiliopoulos

Bocconi University, Milano

Abstract: High-dimensional crossed random effect models is the modern canonical framework for understanding the relative importance of different sources of variation in a data set, with applications from political science to recommendation systems. Popular inference algorithms in this context are backfitting (aka generalized least squares), Gibbs sampling and variational inference using coordinate-wise ascent (known as CAVI). All these schemes involve deterministic or random coordinate-wise updates. For modern applications it is critical that the complexity of such schemes (cost per iteration times the number of iterations) scales linearly with the size of the data set and the total number of parameters. We establish conditions under which suitable implementations of such schemes are scalable and provide a common theoretical framework for the analysis of the three aforementioned schemes. Interestingly, our theory leverages very recent results on the fast mixing time of random walks on random graphs, a set of results related to the so-called Alon¢s conjencture.

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Anomaly identification under sequential sampling and sampling constraints

Georgios Fellouris

University of Illinois at Urbana-Champaign

Abstract: In this talk we will consider a setup where multiple data sources are monitored in real time and the goal is to identify the "anomalous" ones among them, if any, when it is not possible to sample all of them at all sampling instances. To this end, we need to specify a policy that consists of not only a stopping rule and a decision rule, but also a sampling rule. We will propose a novel formulation, in which the number of anomalous sources is not necessarily known in advance and the number of sampled sources per time instance is not necessarily fixed. In addition to the sampling constraint, we will require control of the false positive and false negative rates below user-specified tolerance levels. We will present various policies that have minimal computational and memory requirements and achieve the minimum expected time until stopping to a first-order asymptotic approximation as both familywise error rates go to zero. We will further discuss comparisons with alternative policies and extensions to other error metrics.

(This talk will be based on joint work with Aristomenis Tsopelakos (UIUC)).

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Nonparametric and high-dimensional functional graphical models

Eftychia Solea
École Nationale de Statistique et Analyse de l'Information (ENSAI)

Abstract: We consider the problem of constructing nonparametric undirected graphical models for high-dimensional functional data. Most existing statistical methods in this context assume either a Gaussian distribution on the vertices or linear conditional means. In this article, we provide a more flexible model which relaxes the linearity assumption by replacing it by an arbitrary additive form. The use of functional principal components offers an estimation strategy that uses a group lasso penalty to estimate the relevant edges of the graph. We establish statistical guarantees for the resulting estimators, which can be used to prove consistency if the dimension and the number of functional principal components diverge to infinity with the sample size. We also investigate the empirical performance of our method through simulation studies and a real data application.

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Piecewise Deterministic Monte Carlo and the Speed Up Zig-Zag sampler

Giorgos Vasdekis
University of Warwick

Abstract: Piecewise Deterministic Markov Processes (PDMPs) have recently drawn the attention of the Markov Chain Monte Carlo (MCMC) community. The main reason is that these processes have a natural notion of momentum, which sometimes leads to better exploration of the state space and faster mixing. In the first half of this talk, we will give a brief introduction to how one can use these processes in MCMC and we will introduce the state of the art PDMP algorithms. In the second half of the talk, we will introduce a new PDMP algorithm called the Speed Up Zig-Zag sampler, we will study its properties and explain why it can be used to efficiently sample from heavy tailed distributions.

(This is joint work with Gareth O. Roberts.)

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Mean field limits for weakly interacting diffusions: phase transitions, multiscale analysis and inference

Grigorios A. Pavliotis

Imperial College, London

Abstract: We consider a system of N weakly interacting particles driven by white noise. The mean field limit of this system is described by the (nonlinear and nonlocal) McKean-Vlasov-Fokker-Planck PDE. We present a detailed analysis of continuous and discontinuous phase transitions for the McKean-Vlasov PDE on the torus. We study the combined diffusive/mean field limit of systems of weakly interacting diffusions with a periodic interaction potential. We show that, in the presence of phase transitions, the two limits do not commute. We then show the equivalence between uniform propagation of chaos, a uniform-in-N Logarithmic Sobolev inequality, the absence of phase transitions for the mean field limit, and of Gaussian fluctuations around the McKean-Vlasov PDE. Finally, we develop inference methodologies for estimating parameters in the drift of the McKean SDE using either the stochastic gradient descent algorithm or eigenfunction martingale estimators.

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Model-free bounds in finance: a journey through probability, statistics and optimization

Antonios Papapantoleon
Technical University, Delft

Abstract: Academics, practitioners and regulators have understood that the classical paradigm in mathematical finance, where all computations are based on a single "correct" model, is flawed. Model-free methods, were computations are based on a variety of models, offer an alternative. In this talk, we will discuss model-free methods and bounds, starting from the improved Fréchet-Hoeffding bounds and their applications in option pricing and risk management, and will present how ideas from probability, statistics, optimal transport and optimisation can be applied in this field.

Mean field analysis and scaling limits for neural networks: typical events and fluctuations

Konstantinos Spiliopoulos
Boston University

Abstract: Machine learning, and in particular neural network models, have revolutionized fields such as image, text, and speech recognition. Today, many important real-world applications in these areas are driven by neural networks. There are also growing applications in finance, engineering, robotics, and medicine. Despite their immense success in practice, there is limited mathematical understanding of neural networks. Our work shows how neural networks can be studied via stochastic analysis and develops approaches for addressing some of the technical challenges which arise. We analyze both multi-layer and one-layer neural networks in the asymptotic regime of simultaneously (A) large network sizes and (B) large numbers of stochastic gradient descent training iterations. In the case of single layer neural networks, we rigorously prove that the empirical distribution of the neural network parameters converges to the solution of a nonlinear partial differential equation. In addition, we rigorously prove a central limit theorem, which describes the neural network's fluctuations around its mean-field limit. The fluctuations have a Gaussian distribution and satisfy a stochastic partial differential equation. For multilayer neural networks we rigorously derive the limiting behavior of the neural networks output. We also prove convergence to the global minimum under appropriate conditions. We demonstrate the theoretical results in the study of the evolution of parameters in the well known MNIST and CIFAR10 data sets.

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Statistical Distances: A Global Framework for Inference

Marianthi Markatou
University at Bufallo

Abstract: Statistical distances play a ubiquitous role in statistical theory and statistical thinking. However, within the statistical literature the extensive role of distances has too often been played out behind the scenes, with other aspects of the statistical problems being viewed as more central, more interesting, or more important. Distances are prominent in goodness of fit, but the usual question we ask is: *"how powerful is this method against a set of interesting alternatives"* not *"what aspect of the difference between the hypothetical model and the alternative are we measuring?"*

Statistical distances provide a very natural way to create a global framework to inference for parametric and semiparametric models. In this talk, we will present an approximation framework based on distances, in which statistical distances are used as loss functions. This is accomplished by examining in detail the properties of a number of distances, with emphasis on determining the sense in which we can give them meaningful interpretations as measures of statistical loss. Furthermore, we discuss a special class of distances, the class of quadratic distances, and connect goodness of fit with the kernel literature by showing that we can write classical goodness of fit statistics as functions of specific kernels. We present tests of goodness of fit based on estimates of quadratic distances and exemplify their performance in testing for normality and uniformity on the high dimensional sphere. If time permits, we will describe model building methods and methods of assessment of model adequacy.

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Density functionals for inhomogeneous systems

Dimitrios Tsagkarogiannis
University of L'Aquila

Abstract: We compute free energy functionals as power series in infinite dimensional spaces. We provide a rigorous framework to prove validity of density functionals for inhomogeneous, non-translation invariant systems with applications in classical density functional theory, liquid crystals, molecules with various shapes or other degrees of freedom. We use cluster expansions and compare them to large deviations techniques. A key technical tool is a combinatorial identity for trees which allows us to obtain convergence estimates in situations where Banach inversion fails.

(This is joint work with Sabine Jansen and Tobias Kuna.)

Noise-Induced Randomization in Regression Discontinuity Designs

Nikos Ignatiadis
Stanford University

Abstract: Regression discontinuity designs are used to estimate causal effects in settings where treatment is determined by whether an observed running variable crosses a pre-specified threshold. While the resulting sampling design is sometimes described as akin to a locally randomized experiment in a neighborhood of the threshold, standard formal analyses do not make reference to probabilistic treatment assignment and instead identify treatment effects via continuity arguments. Here we propose a new approach to identification, estimation, and inference in regression discontinuity designs that exploits measurement error, or other noise, in the running variable. Under an assumption that the measurement error is exogenous, we show how to consistently estimate causal effects using a class of linear estimators that weight treated and control units so as to balance a latent variable of which the running variable is a noisy measure. We find this approach to facilitate inference for familiar estimands from the literature, as well as policy-relevant estimands that correspond to the effects of realistic changes to the existing treatment assignment rule. We demonstrate the method with a study of retention of HIV patients, and evaluate its performance using both simulated data and a regression discontinuity design artificially constructed from test scores in early childhood.

(Joint work with Dean Eckles, Stefan Wager, and Han Wu.)

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: *Probability, Statistics & Applications*

Longest and heaviest paths in random directed graphs

Takis Konstantopoulos

University of Liverpool

Abstract: I will give an overview of research in the area of random directed graphs with possibly random edge weights. We are interested in longest paths between two vertices (or heaviest paths if there are weights). Typically, the longest path satisfies a law of large numbers and a central limit theorem (which gives a non-normal distribution in the limit if the vertex set is not one-dimensional). The constant in the law of large numbers as a function of the graph parameters and weight distributions cannot be computed explicitly except, perhaps, in very simple cases. A lot of work has been done in obtaining bounds and in studying its behavior. For example, is it a smooth function of the connectivity parameter p? These kinds of graphs appear in several areas: in computer science, in statistical physics, in performance evaluation of computer systems and in mathematical ecology. They originated in a paper by Barak and Erdos but have also been studied independently, in connection with the applications above. The questions asked are related to the so-called last passage percolation problems because we can interpret "longest" in a time sense (what?s the worst case road that will take us from a point to another point?). As such, it is not surprising that in some cases, the limiting behavior is related to limits of large random matrices. However, the complete picture is not understood and so open problems will also be presented. This is a long-standing ongoing project, based on several papers (completed or in progress) with Sergey Foss, Denis Denisov, Katja Trinajstic, Artem Pyatkin, Bastien Mallein and Sanjay Ramassamy.

Density functionals for inhomogeneous systems

Dimitrios Tsagkarogiannis
University of L'Aquila

Abstract: We compute free energy functionals as power series in infinite dimensional spaces. We provide a rigorous framework to prove validity of density functionals for inhomogeneous, non-translation invariant systems with applications in classical density functional theory, liquid crystals, molecules with various shapes or other degrees of freedom. We use cluster expansions and compare them to large deviations techniques. A key technical tool is a combinatorial identity for trees which allows us to obtain convergence estimates in situations where Banach inversion fails.

(This is joint work with Sabine Jansen and Tobias Kuna.)

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Sharp spectral gap for regular and disordered chain of oscillators

Aggeliki Menegaki

Institut des Hautes Études Scientifiques, Paris

Abstract: We consider one-dimensional chains and multi-dimensional networks of harmonic oscillators coupled to two Langevin heat reservoirs at different temperatures. Each particle interacts with its nearest neighbours by harmonic potentials and all individual particles are confined by harmonic potentials, too. In this talk, I will present the sharp N-particle dependence of the spectral gap of the associated generator in different physical scenarios and for different spatial dimensions. Our method of proof relies on a new approach to analyse a non self-adjoint eigenvalue problem involving low-rank non-hermitian perturbations of discrete Schrödinger operators. We will also discuss possible extensions of this method to oscillator chains with longer-range interactions.

(This is a joint work with Simon Becker.)

Second Congress of Greek Mathematicians SCGM-2022

104 YEARS HMS
GOLD MEDAL 2018
ACADEMY OF ATHENS

Session: ***Probability, Statistics & Applications***

Some recent progress on the properties of Schramm-Loewner Evolution

Konstantinos Kavvadias
University of Cambridge

Abstract: The Schramm-Loewner Evolution (SLE_κ) is a one parameter family ($\kappa > 0$) of curves which connect two boundary points of a simply connected domain. It was introduced by Schramm in 1999 as a candidate to describe the scaling limit of the interfaces that arise in discrete models at criticality from statistical mechanics on planar lattices, such as the loop erased random walk and the percolation model. In my talk, I will discuss about the intuition behind the definition of SLE_κ and I will introduce some of its basic properties obtained during the last twenty years. I will also discuss about some recent results obtained in a series of recent research works jointly with Jason Miller and Lukas Schoug. Finally, if time permits, I will discuss about some ongoing research results.