Faculty of Natural, Mathematical, and Engineering Sciences
Department of Mathematics

KING'S College LONDON

PhD Projects in Mathematics

Applied Mathematics Research: Disordered Systems/Financial Mathematics/Probability/Theoretical Physics MPhil/PhD Projects

Project: Kasteleyn theory
Supervisor: Sam Johnston
Group: Probability

Summary: Kasteleyn theory is a tool used by probabilists to obtain explicit solutions for the partition functions and correlation functions for models of random surfaces and random tilings. In this project we use a new coordinate system to study generalisations of these models in higher dimensions, making connections with areas of probability, algebra, and statistical physics.

Project: Microstructure of the Kerr black hole
Supervisor: Dionysios Anninos
Group: Theoretical Physics

Summary: We will apply tools from modern theoretical physics to assess mathematical properties of highly spinning black holes. Such highly spinning black holes have been observed in our own Universe and display remarkable geometric features.

Project: Cylindridcal Levy processes
Supervisor: Markus Riedle
Group: Probability/ Financial Mathematics

Summary: Cylindridcal Lévy processes are a new approach to model random perturbations of complex dynamical systems such as partial differential equations. It has already been observed that these cause completely different behaviour of the dynamical system than if perturbed by a classical stochastic process. In this project we investigate one of the many aspects which are unknown for cylindrical Levy processes, such as stability, ergodicity etc.

Project Title: Equilibrium between providers and takers of liquidity in order driven markets
Supervisor: Leonardo Sanchez Betancourt
Group: Financial Mathematics

Summary: We search for optimal trading strategies between takers and makers of liquidity with transaction costs. We aim to contribute to the literature around stochastic games for algorithmic trading.

Project: Lorentzian and Euclidean Conformal Field Theory
Supervisor: Petr Kravchuk
Group: Theoretical Physics

Summary: Conformal field theories have many facets, and appear in physics in both Lorentzian (or realtime), and Euclidean (or statistical) versions. From the theory perspective the Lorentzian point of view brings powerful analytical insights, while the Euclidean one allows for a precise numerical analysis. You will study the interplay between these two worlds to obtain new results relevant both in condensed-matter and high-energy physics

---

Project: Entanglement and universality in inhomogeneous and out-of-equilibrium many body quantum systems
Supervisor: Paola Ruggiero & Benjamin Doyon
Group: Disordered Systems

Summary: The project lies at the intersection of several different fields: from statistical physics to quantum information, from high energy to condensed matter physics. The common denominators are the phenomenon of quantum entanglement and the out-of-equilibrium behaviour of many-body quantum systems, particularly in low dimensional quantum many body systems. While the emergence of universality is rather well established for systems which are homogeneous and are at equilibrium, the same is not true for inhomogeneous systems, both in and out of equilibrium. Those, on the other hand, are the rule rather than the exception in the context of quantum experiments. Different new theories and tools are currently under study to investigate what remains of the aforementioned universality in such more complicated situations, and the projects aims at going deeper in this direction, mainly using the “lens” of quantum entanglement.

Project: Random models for arithmetic correlations
Supervisor: Aled Walker
Group: Number Theory

Summary: The project lies at the intersection of several different fields: number theory, harmonic analysis, probability, and additive combinatorics, with even some side connections to mathematical physics. The goal is to investigate the gap distributions of certain types of arithmetic sequences (dilated squares, dilated primes, dilated integers, etc.), with the aim to understand how closely these distributions agree with a simple random model, formed by replacing the arithmetic sequence with independent uniform random variables. It has been conjectured, for example, that the dilated squares can indeed be modelled by a random sequence in this way (having so-called 'poissonian gap distribution'). This is turn is connected to the phenomenon of quantum chaos, which concerns the gap distribution of the eigenvalues of certain Hamiltonians. Additive combinatorics also plays a role: recent work has found new connections between correlation functions (certain objects associated to gap distributions) and the so-called 'additive energy' of the underlying sequence. The exact direction of the project would be determined by the student, but a possible first move concerns establishing a 'density threshold' for the triple correlation function: given a randomly chosen sequence of integers (with a certain density), can one establish when the triple correlations will be poissonian?

Project: The mathematics of adaptive, stable and robust Artificial Intelligence
Supervisor: Ivan Tyukin
Group: Disordered Systems

Summary: Recent years have seen explosive growth in the applications of Artificial Intelligence (AI). Notwithstanding significant successes of the new technology, empirical evidence points to numerous examples of errors in the decisions of state-of-the-art AI systems. Current theoretical works confirm that modern design practices and classical mathematical frameworks supporting these could suffer from various fundamental shortcomings including the typicality of adversarial attacks, susceptibility to backdoors and stealth attacks, instabilities, and hallucinations. At the same time, there is a plethora of under-explored and seemingly controversial phenomena, such as learning from scarce yet high-dimensional information, which has been broadly observed experimentally in large-scale AI systems but whose complete mathematical understanding is yet to be developed. Other factors hindering the performance of current AI systems include the need to adapt to concept drifts and unforeseen changes in operational conditions. The project aims to explore and systematically assess major causes of AI errors and instabilities and develop appropriate mathematical foundations enabling the creation of stable and robust Artificial Intelligence capable of dealing with errors and concept drifts. The successful candidate will work alongside a team of early career researchers funded by the UKRI Turing AI Acceleration Fellowship (EP/V025295/2) and will benefit from the team's established relationships with industrial partners, academic networks (Turing AI Fellows, Trustworthy Autonomous Systems Node in Verifiability EP/V026801/2), and other collaborators in academia, healthcare, defence & security.

Pure Mathematics Research MPhil/PhD Projects

Project: Foliations of 3-dimensional manifoldsSupervisor: Mehdi Yazdi
Group: Geometry

Summary: The student is expected to work in low-dimensional topology; in one of the areas of foliations, 3-dimensional manifolds, knot theory, mapping class groups of surfaces, or related subjects. Below some of these terms are briefly explained: A foliation of a 3-dimensional manifold is a decomposition of it into disjoint surfaces that locally fit together like a stack of papers. Examples of singular foliations in one dimension lower are marbleised papers and tree rings. The student can for example study an important class of foliations called taut foliations, which has applications to knot theory. A knot is an embedding of the circle in a 3-dimensional manifold, considered up to isotopy. In other words, we are allowed to deform the knot without the knot crossing itself. Knots have been extensively studied both for their own beauty, and as an important class of 3-dimensional manifolds (the complement of the knot). The mapping class group of a surface is the group of homeomorphisms of the surface considered up to isotopy. This is an infinite group and has applications in various parts of geometry and topology including hyperbolic geometry of surfaces and the topology of 3-dimensional manifolds.

Project: Analytic aspects of L-functions and automorphic forms
Supervisor: Stephen Lester
Group: Number Theory

Summary: L-functions and automorphic forms appear prominently in modern number theory and also arise in other areas of mathematics. The analytic properties of these functions are closely related to arithmetic problems involving the distribution of the prime numbers, lattice points, and structure of elliptic curves. For instance, Millennium prize problems such as the Generalised Riemann Hypothesis and Birch and Swinnerton-Dyer Conjecture are directly related to the theory of these functions. This project aims to further understand the analytic behaviour of L-functions and automorphic forms, with applications to problems arising in number theory or mathematical physics. Specific problems will follow the interests of the student and possible topics include: the distribution of zeros of Hecke cusp forms, moment estimates for families of L-functions, and the behaviour of Fourier coefficients of automorphic forms.

Project: Birational Geometry and foliations
Supervisor: Calum Spicer
Group: Geometry (pure mathematics)

Summary: The goal of this project is to explore cutting edge applications and interactions between birational geometry and holomorphic foliation theory. There is a specific focus on the use of the Minimal Model Program and related techniques as applied to questions in local and global geometry of algebraic varieties and foliations. The study of the Minimal Model Program is one of the main research directions in modern algebraic geometry, and in recent years has seen many significant developments. Among these is a new focus on relations between the Minimal Model Program and foliations, which has resulted in applications of ideas and techniques of the Minimal Model Program to a wide range of novel situations.

Statistics Research MPhil/PhD Projects

Project: Multi-Objective Optimal Design for Big Data
Supervisor: Kalliopi Mylona and Steven Gilmour
Group: Statistics, Data Analysis, Machine Learning

Big data sets typically arise from observational data and contain many more observations than are needed to draw inferential conclusions about important research questions. Analysing full datasets can be challenging due to their size and unrewarding due the vast amount of redundant information available. One approach to dealing with such datasets, which has received some attention recently, is to use the theory of optimal design of experiments to select maximally informative subsamples of the data (Deldossi & Tommasi, 2021). Like most work on optimal design, this emphasises D-optimality, which implies that point estimation of the parameters of a given model is the main inferential objective. Of course, many studies have more complex and diverse objective than this. There is currently a large EPSRC-funded grant at King's on Multi-Objective Optimal Design of Experiments, on which 3 academic staff and 2 post-doctoral researchers are working. The aim of this PhD project is to adapt some of the methods of multi-objective optimal design being explored to the subsampling of big data sets. One of the challenges will be in adapting methods which require estimation of "pure error", which can be ensured in experimental studies by replication, to observational data where there might be no exact replicates.

Project: Hierarchical models for function-valued traits in quantitative genetics
Supervisor: Davide Pigoli
Group: Statistics, Mathematical Modelling, Applied Statistics.

Summary: Some biological traits, such as growth trajectories, have a functional nature and are known as function-valued traits (Kingsolver et al., 2001). These traits can be assessed for continuous genetic variation using a quantitative genetic approach. In quantitative genetics, one of the goals is to decompose the genetic (i.e. due to the genes of the individuals) and the environmental (i.e. due to the influence from the external environment). While some works exist on the analysis of function-valued traits using mixed effects models (Meyer, 1998; Vazquez et al., 2010), the field could potentially benefit from the use of modern functional data analysis techniques, such as the ones in Behseta et al. (2005), Scarpa & Dunson (2009) or Scheipl et al. (2015). The goal of the project is to extend these techniques to the case of quantitative genetics experiments, where individuals have a known relationship matrix given by their genealogy.

Project: Supersaturated multi-stratum experiments
Supervisor: Kalliopi Mylona
Group: Statistics, Applied Statistics, Engineering Mathematics, Agricultural Sciences

Summary: The objective of this project is to develop a new general methodology for setting up and analysing informative experiments with both restricted randomisation and a large number of factors. As designs with restricted randomisation are often much more cost-efficient than completely randomised designs, they are extremely popular in industry for quality improvement experiments and for experimenting with new products or processes. Multi-stratum designs with more than two strata are very effective in reducing the cost of an experiment in the presence of different levels of hard-to-change factors and/or of multi-stage processes. In addition, supersaturated designs (SSDs) is a large class of factorial designs which can be used for screening out the important factors from a large set of potentially active variables. The huge advantage of these designs is that they reduce the experimental cost drastically, but their critical disadvantage is the confounding involved in the statistical analysis. The goal is to combine the multi-stratum designs (more than two strata) with the supersaturated designs. This will give the flexibility to the experimenters to use the cost-efficient SSDs under restricted randomised situations that appear very often in industrial experimentation. The combination of these two well-known classes of designs is a relatively unexplored research area (see, Koh et al. [2013], Lee et al. [2009], Chatterjee et al. [2018]). General construction and analysis methods for supersaturated multi-stratum designs with more than two strata will allow experiments that can extract the maximum information from the data with the minimum time and cost constraints.

Project: Statistical analysis of spatially and/or temporally dependent manifold-valued data.
Supervisor: Davide Pigoli
Group: Statistics, Stochastic Processes, Mathematical Modelling.

Summary: There is a growing interest in the statistical analysis of data that take value on smooth manifolds (3D rotations, shapes, directions, positive definite matrices, points on a circle or a sphere, constrained curves to mention just a few examples). The goal of the project is to develop a framework to analyse manifold-valued data that are observed over time and/or space, example including molecular dynamics simulations of protein folding in medicinal chemistry (temporally dependent), permeability tensors in mining engineering (spatially dependent) or coordinates (latitude and longitude) of earthquakes aftershocks in earth science (spatio-temporally dependent). Possible lines of research include (but are not limited to): - modelling of stochastic dependence in these non-linear spaces, where usual concepts of co-variability are not valid. - extension to dependent samples of existing techniques for dimensional reduction (Principal Geodesic Analysis, Principal Flows) and centerpoint estimation (Fréchet mean) for manifold-valued data. - multiresolution analysis and prediction for spatio-temporal manifold-valued processes.

Project title : The geometry of probability distributions in computational statistics and machine learning
Supervisor: Nikolas Nüsken
Group: Statistics

The process of learning can be understood as transforming one probability measure into another, guided by available data. In order to analyse and construct computational learning routines, three perspectives haven proved especially powerful: First, such transitions can be described by the dynamics of interacting particle systems, often using stochastic and/or partial differential equations. Second, one can (heuristically) view the set of probability distributions as an infinite-dimensional Riemannian manifold, in which the learning transition evolves as a curve. Algorithmic efficiency can then directly be linked to curvature properties and their bearings on Riemannian gradient flows and geodesics. Third, desirable transitions can be obtained as solutions to suitable optimal control or optimal transport problems. The aim of this project is to contribute to a deeper understanding of the connections between these three approaches and to leverage those to develop novel and innovative methodology for data science. On the theoretical side, a possible focus could be on the formulation of evolution equations of second order or with memory (akin to Hamiltonian dynamics). On the applied side, a major challenge is the seamless combination of large amounts of data with models coming from science, economics or engineering.

Project: Changepoint detection in Non-Euclidean spaces
Supervisor: Davide Pigoli
Group: Statistics, Stochastic Processes, Applied Statistics.

Summary: This project is concerned with the development and analysis of cumulative sums statistics (Barnard, 1959) to detect changepoints in sequences where the observations belong to a non-Euclidean space (Chen, 2019). Examples include time series of directional data (e.g., wind directions), time series of correlation matrices (e.g., correlation matrices between electroencephalogram signals) or dynamical networks (e.g., social networks or trade networks). Cumulative sum statistics can be defined based on a metric in the non-Euclidean space and there are already some results for data on a circle (Lombard et al., 2020). The purpose of this project is to extend this approach to more general non-Euclidean spaces, and analyse the performance of changepoint detection procedures based on these statistics both theoretically and in practice.

Project: Statistical analysis of supersaturated split-plot experiment

Supervisor: Kalliopi Mylona
Group: Statistics, Applied Statistics, Data Analysis, Mathematical Modelling, Engineering Mathematics, Agricultural Sciences

Summary: The objective of this project is to develop a new general methodology for analysing informative experiments with both restricted randomisation and a large number of factors. Split-plot designs are very effective in reducing the cost of an experiment in the presence of hard-to-change factors and/or of two-stage processes. In addition, supersaturated designs (SSDs) is a large class of factorial designs which can be used for screening out the important factors from a large set of potentially active variables. The huge advantage of these designs is that they reduce the experimental cost drastically but their critical disadvantage is the confounding involved in the statistical analysis. The combination of split-plot designs with supersaturated designs is a relatively unexplored research area, and the statistical analysis of data from this type of experiments is still not efficient. In this project, the use of Bayesian analysis methods, will be studied. Gilmour and Goos (2009) recommended a Bayesian analysis, using an informative prior distribution for the whole plot variance component and implement this by using Markov chain Monte Carlo sampling. The designs considered in this project have an additional complication compared to the designs used in Gilmour and Goos (2009) and this is supersaturated property. We are planning to study and adapt several statistical analysis methods that had been used in the literature of supersaturated designs, such as Dantzig selector (Candes and Tao, 2007), the LARS (Efron, Hastie, Johnstone and Tibshirany,2004) methods and the Bayesian methods: SSVS (Chipman, Hamada and Wu, 1997) and Bayesian LASSO (Park and Casella, 2008).

Project: Stein thinning in practice, with application to cardiac modelling
Supervisor: Marina Riabiz
Group: Statistics

Summary: Markov Chain Monte Carlo (MCMC) methods form the main approach to sample from probability distributions with intractable normalizing constants, such as the parameter posterior in Bayesian computation. Whilst MCMC samples converge in the limit to the target distribution, in many realworld applications it is possible to draw only a finite number of samples, due to restrictions in computing budget. In this framework, Stein thinning is a recent method developed to optimally select a set of samples from an MCMC output of fixed length, by minimizing a measure of discrepancy between the empirical approximation produced and the target. This PhD project aims at exploring different research questions that are still open, in order to produce a robust version of the algorithm, including the selection of a weighted kernel function that jointly guarantees weak convergence and convergence of certain moments of the distribution, based on related work on control variates and optimization of the power of goodness of fit test. The project will be applied to challenging Bayesian inverse problems arising in cardiac electrophysiology.

Project Title: Bayesian learning for partially specified models
Supervisor: Yu Luo
Group: Statistics

Summary: Bayesian inference shifts from unknown deterministic quantities to studying distributions, which has been proven increasing powerful in a large spectrum of applications over the decades. The posterior inference is often thought as a procedure to update prior beliefs in light of the data. However, in the usual Bayesian setting where a full probabilistic model is required. However, recently some researchers in statistical learning have investigated extensions of the Bayes paradigm to the data-driven procedure, which one does not make any assumption on the probabilistic model, but rather relies on an agnostic measure between the model prediction and outcome, resulting a partially specified model. These extensions no longer abide by the canonical Bayesian rules and may be harder to interpret by practitioners. Therefore, the overall goal of this project is to help develop a mature general Bayesian decision-making machinery for models which are partially specified. This methodology advancement has the potential to apply in various fields, such as machine learning and causal inference.

Project Title: Estimation and testing for extremes of spherical data
Supervisor: Claudia Neves
Group: Statistics

Summary: Extreme value theory (and related statistical methodology) is concerned with the definition, characterisation and estimation of tale-telling characteristics of extreme and rare events, whilst accounting for the possible dependence between them. These could be, for example, strong wind and heavy rainfall, or strong demand for electric energy and little solar radiation. It is the primary aim of this project to develop statistical inference methodology for assessing and estimating changes in the intensity of extreme events on a surface area, thus extending current methods considerably, in an actionable way.

Project Title: The geometry of probability distributions in computational statistics and machine learning
Supervisor: Nikolas Nüsken
Group: Statistics, Disordered Systems, Analysis, Geometry

Summary: The process of learning can be understood as transforming one probability measure into another, guided by available data. In order to analyse and construct computational learning routines, three perspectives haven proved especially powerful: First, such transitions can be described by the dynamics of interacting particle systems, often using stochastic and/or partial differential equations. Second, one can (heuristically) view the set of probability distributions as an infinite-dimensional Riemannian manifold, in which the learning transition evolves as a curve. Algorithmic efficiency can then directly be linked to curvature properties and their bearings on Riemannian gradient flows and geodesics. Third, desirable transitions can be obtained as solutions to suitable optimal control or optimal transport problems. The aim of this project is to contribute to a deeper understanding of the connections between these three approaches and to leverage those to develop novel and innovative methodology for data science. On the theoretical side, a possible focus could be on the formulation of evolution equations of second order or with memory (akin to Hamiltonian dynamics). On the applied side, a major challenge is the seamless combination of large amounts of data with models coming from science, economics or engineering.
