London meeting on Computational Statistics
11-13 June 2024

Location: G01 Lankester Lecture Theatre, Medawar Building, University College London (lectures). North Cloisters (poster session). Roberts Foyer G02 (lunches/coffee breaks)

Sponsors: EPSRC (via New Investigator Award EP/V055380/1), UCL Institute for Mathematical and Statistical Sciences (IMSS).

Contact: Samuel Livingstone (samuel.livingstone@ucl.ac.uk).

Program

		Tuesday		Wednesday		Thursday
			10:00	Arthur Gretton	10:00	Sergio Bacallado
			10:40	Cecilia Balocchi	10:40	Emiko Dupont
			11:20	BREAK	11:20	BREAK
			11:50	Nick Whiteley	11:50	Murray Pollock
12:30		LUNCH	12:30	2*20 Luke & Joanna	12:30	2*20 Yuga & Kevin
13:20		Welcome (Sam)	13:10	LUNCH	13:10	LUNCH
13:25		Remi Bardenet				
14:05		BREAK				
14:30		Kolyan Ray	14:30	Matteo Fasiolo	14:30	Elizaveta Semenova
15:10		Purvasha Chakravarti	15:10	Alice Corbella	15:10	Valentin De Bortoli
15:50		BREAK	15:50	BREAK	15:50	BREAK
16:20		2*20 Max & Alicja	16:20	Chris Sherlock	16:20	Giorgos Vasdekis
17:00		Giacomo Zanella	17:00	Bob Carpenter		
			18:00	Poster session		

Titles and Abstracts

Tuesday 11th June

13:25 Remi Bardenet

Title: Variance reduction with the repelled Poisson point process

Abstract: Linear statistics of point processes yield Monte Carlo estimators of integrals. While the simplest approach relies on a homogeneous Poisson point process, more regularly spread point processes, such as scrambled low-discrepancy sequences or determinantal point processes, can yield Monte Carlo estimators with fast-decaying mean square error. Following the intuition that more regular configurations result in lower integration error, we

introduced the repulsion operator, which reduces clustering by slightly pushing the points of a configuration away from each other. Our main theoretical result is that, for a quadratic cost, applying the repulsion operator to a homogeneous Poisson point process yields an unbiased Monte Carlo estimator with lower variance than under the original point process.

14:30 Kolyan Ray

Title: A variational Bayes approach to debiased inference in high-dimensional linear regression

Abstract: We consider statistical inference for a single coordinate of a high-dimensional parameter in sparse linear regression. It is well-known that high-dimensional procedures such as the LASSO can provide biased estimators for this problem and thus require debiasing. We propose a scalable variational Bayes method for this problem based on assigning a mean-field approximation to the nuisance coordinates and carefully modelling the conditional distribution of the target given the nuisance. We investigate the numerical performance of our algorithm and establish accompanying theoretical guarantees for estimation and uncertainty quantification. Joint work with I. Castillo, A. L'Huillier, L. Travis

15:10 Purvasha Chakravarti

Title: Robust Signal Detection in Particle Physics using a Classifier Decorrelated via Optimal Transport

Abstract: New physics searches are usually done by training a supervised classifier to separate a signal model from known Standard Model physics (also called the background model). However, even when the signal model is correct, systematic errors in the background model can influence supervised classifiers and might adversely affect the signal detection procedure. To tackle this problem, our approach is to use the (possibly misspecified) classifier only to perform a signal-enrichment step, and then to carry out a model-independent search in the signal-rich region using only the real experimental data. In order for this procedure to work, we show that the classifier used for

signal enrichment must be independent of the variables (called protected variables) used for the signal detection step. So we need this classifier constrained to be decorrelated with one or more protected variables. We carry out this decorrelation by considering an optimal transport map of the classifier output that makes it independent of the protected variable for the background. We then perform signal detection using the protected variable by fitting a semi-parametric mixture model. We compare and contrast this decorrelation method with previous approaches, show that the decorrelation procedure is robust to background misspecification, and analyze the power of the signal detection test. The talk will also briefly mention how our decorrelation algorithm can be used to address fairness in classification algorithms.

16:20 Max Hird

Title: Quantifying the effectiveness of linear preconditioning in MCMC

Abstract: Linear transformation of the state variable (linear preconditioning) is a common technique that often drastically improves the practical performance of a Markov chain Monte Carlo algorithm. Despite this, however, the benefits of linear preconditioning are not well-studied theoretically, and rigorous guidelines for choosing preconditioners are not always readily available. Mixing time bounds for various samplers have been produced in recent works for the class of strongly log-concave and Lipschitz target distributions and depend strongly on a quantity known as the condition number. We study linear preconditioning for this class of distributions, and under appropriate assumptions we provide bounds on the condition number after using a given linear preconditioner. We provide bounds on the spectral gap of RWM that are tight in their dependence on the condition number under the same assumptions. Finally we offer a review and analysis of popular preconditioners. Of particular note, we identify a surprising case in which preconditioning with the diagonal of the target covariance can actually make the condition number *increase* relative to doing no preconditioning at all.

16:40 Alicja Polanska

Title: Learned harmonic mean estimation of the marginal likelihood with normalizing flows

Abstract: TBA

17:00 Giacomo Zanella

Title: Zero-order parallel sampling

Abstract: Finding effective ways to exploit parallel computing in order to speed up MCMC convergence is an important problem in Bayesian computation and related disciplines. Here we consider the zero-order (aka derivative-free) version of the problem, where we assume that (a) the gradient of the target distribution is unavailable (either for theoretical, practical or computational reasons) and (b) we can evaluate the (expensive) target distribution in parallel at K different locations and use those evaluations to speed up MCMC convergence. We provide two main contributions in this respect. First, we show that any method falling within a quite general "multiple-proposal framework" can only speed up convergence by log(K) factors in high dimensions. The fundamental limitation of such framework, which includes multiple-try MCMC as well as many other previously proposed methods, lies in restricting possible moves to the support of the K evaluation points. We state our results in terms of upper bounds on the Spectral Gap of the resulting scheme. Second, we discuss two ways (one based on gradient estimators and the other based on factorized proposals) which make better use of parallel computing and achieve polynomial speed-ups in K. Some of the methods share similarities, but also notable differences, with classical zero-order optimization methods.

Wednesday 12th June

10:00 Arthur Gretton

Title: Adaptive two-sample testing

Abstract: I will address the problem of two-sample testing using the Maximum Mean Discrepancy (MMD). The MMD is an integral probability

metric defined using a reproducing kernel Hilbert space (RKHS), with properties determined by the choice of kernel. For good test power, the kernel must be chosen in accordance with the properties of the distributions being compared. I will assume that the distributions being tested have densities, and the difference in densities lies in a Sobolev ball. The MMD test is then minimax optimal with a specific kernel depending on the smoothness parameter of the Sobolev ball. In practice, this parameter is unknown: to overcome this issue, I describe an aggregated test, called MMDAgg, which is adaptive to the smoothness parameter. The test power is maximised over the collection of kernels used, without requiring held-out data for kernel selection (which results in a loss of test power). MMDAgg controls the test level non-asymptotically, and achieves the minimax rate over Sobolev balls, up to an iterated logarithmic term. Guarantees hold for any product of one-dimensional translation invariant characteristic kernels.

10:40 Cecilia Balocchi

Title: Improving uncertainty quantification in Bayesian cluster analysis

Abstract: The Bayesian approach to clustering is often appreciated for its ability to provide uncertainty in the partition structure. However, summarizing the posterior distribution over the clustering structure can be challenging. Wade and Ghahramani (2018) proposed to summarize the posterior samples using a single optimal clustering estimate, which minimizes the expected posterior Variation of Information (VI). In instances where the posterior distribution is multimodal, it can be beneficial to summarize the posterior samples using multiple clustering estimates, each corresponding to a different part of the space of partitions that receives substantial posterior mass. In this work, we propose to find such clustering estimates by approximating the posterior distribution in a VI-based Wasserstein distance sense. An interesting byproduct is that this problem can be seen as using the k-means algorithm to divide the posterior samples into different groups, each represented by one of the clustering estimates. Using both synthetic and real datasets, we show that our proposal helps to improve the understanding of uncertainty,

particularly when the data clusters are not well separated, or when the employed model is misspecified.

11:50 Nick Whiteley

Title: Statistical Exploration of the Manifold Hypothesis

Abstract: TBA

12:30 Luke Hardcastle

Title: Averaging polyhazard models using Piecewise deterministic Monte Carlo with applications to data with long-term survivors

Abstract: Polyhazard models are a class of flexible parametric models for modelling survival over extended time horizons. Their additive hazard structure allows for flexible, non-proportional hazards whose characteristics can change over time while retaining a parametric form, which allows for survival to be extrapolated beyond the observation period of a study. Significant user input is required, however, in selecting the number of latent hazards to model, their distributions and the choice of which variables to associate with each hazard. The resulting set of models is too large to explore manually, limiting their practical usefulness. We extend the standard polyhazard model through a prior structure allowing for joint inference of parameters and structural quantities. Posterior sampling is achieved via a bespoke MCMC scheme based on Piecewise Deterministic Markov Processes. This is joint work with Samuel Livingstone and Gianluca Baio.

12:50 Joanna Tobiasz

Title: Machine learning in support of breast cancer subtyping

Abstract: Even though breast cancer (BC) demonstrates large diversity in clinical and molecular portraits, this heterogeneity is not entirely reflected by the commonly used classifications, which remain unchanged for decades. The well-established clinical classification relies on several markers, while molecular classification is based on

gene expression profiling. However, the progress in molecular biology and data mining allows a more accurate characterization of BC cases. In this study, we aimed to develop machine-learning methods for proteomic-based identification and molecular and clinical evaluation of novel BC patient subpopulations.

We tested various combinations of feature engineering and clustering approaches for patients' division based on their proteomic profiles. Moreover, we proposed metrics for the comparison of the clustering outcomes. To characterise the revealed subpopulations, we supported the classic statistical testing with the appropriate effect size estimates and machine learning algorithms. This allowed the comparison of groups with different, sometimes insufficient, sizes.

Six subpopulations were identified, including three novel subgroups corresponding to the luminal A transcriptomic-based subtype. Revealed subpopulations demonstrated different survival and proteomic and transcriptomic profiles. Novel luminal subpopulations varied in prognosis but occurred less differentiated at the transcriptomic level than in proteomic space.

The obtained profiles of the revealed subpopulations, especially the proteomic one, may complement the well-established BC classifications and support the search for novel targeted therapies. However, further investigation and independent validation are still required for the potential clinical applications.

14:30 Matteo Fasiolo

Title: Additive covariance matrix models

Abstract: Dynamic covariance matrix models for multivariate normal data are a widely-applicable class of statistical models, meant to capture the covariate-dependent nature of the variance and dependence parameters. However, such models are rarely used in practice, partly due to the computational difficulties involved in model fitting. In particular, it is challenging to ensure the positive definiteness of the covariance matrix while guaranteeing computational scalability for even moderate dimension of the outcome vector. In this talk we will present methods for fitting multivariate Gaussian regression models where each parameter of the mean vector and of (an unconstrained

parametrisation of) the covariance matrix can be modelled additively, via parametric or spline-based smooth effects. We will focus particularly on the modified Cholesky decomposition and we will show how the sparsity of the corresponding derivative system aids scalability w.r.t. the dimension of the response vector. The usefulness of the new models will be illustrated on a UK regional electrical net-demand forecasting application.

15:10 Alice Corbella

Title: The Lifebelt Particle Filter for robust estimation from low-valued count data

Abstract: Particle filtering methods are well developed for continuous state-space models. When dealing with discrete spaces on bounded domains, particle filters can still be used to sample from and marginalise over the unknown hidden states. Nevertheless, problems such as particle degradation can be even more severe in this context: proposed particles can easily be incompatible with the data and the discrete system could often result in all particles having weight of zero. However, if the boundaries of the discrete hidden space are known, then these could be used to prevent particle collapse. In this talk I will introduce the Lifebelt Particle Filter (LBPF), a novel method for robust likelihood estimation when low-valued count data arise. The LBPF combines a standard particle filter with one (or more) lifebelt particles which, by construction, lie within the boundaries of the discrete random variables, and therefore are compatible with the data. A mixture of resampled and non-resampled steps allows the preservation of the lifebelt particle, which, together with the remaining particle swarm, provides samples from the filtering distribution, and can be used to generate estimates of the likelihood. The LBPF can be used within a pseudo-marginal scheme to draw inference on static parameters, governing a discrete state-space model with low-valued counts. We present here the applied case estimating a parameter governing probabilities and timings of deaths and recoveries of hospitalised patients during an epidemic.

16:20 Chris Sherlock

Title: Tuning pseudo-marginal Metropolis-Hastings: a vase or two faces?

Abstract: The general applicability and ease of use of the pseudo-marginal Metropolis-Hastings (PMMH) algorithm, and particle Metropolis-Hastings in particular, makes it a popular method for inference on discretely observed Markovian stochastic processes. It substitutes realisations of unbiased estimators of the posterior into both the numerator and denominator of the Metropolis-Hastings acceptance ratio. The more accurate the estimator (enacted, for example, by increasing the number of particles in an underlying particle filter), the better the algorithm mixes. However, increased accuracy comes at an increased computational cost, and tuning the algorithm involves a trade off between these factors. Two independent papers, both published in 2015, suggested choosing the number of particles so that the variance of the logarithm of the estimator of the posterior at a fixed sensible parameter value is approximately 1. This advice does not hold for a single, simple importance sampling estimator, but it has been widely and successfully adopted when using products of such estimators or using particle filters. We provide simple examples which demonstrate that the 2015 advice cannot be correct for a general particle filter or a product. We explore the asymptotic variance associated with the PMMH algorithm and provide a remarkably simple bound which leads to alternative advice. In most situations our guidance and the early advice closely coincide - the picture is both a vase and two faces. We then extend the analysis to the correlated pseudo-marginal method.

17:00 Bob Carpenter

Title: GIST: Gibbs self-tuning for locally adaptive Hamiltonian Monte Carlo

Abstract: We present a novel and flexible framework for localized tuning of Hamiltonian Monte Carlo samplers by sampling the algorithm's tuning parameters conditionally based on the position and momentum at each step. For adaptively sampling path lengths, we show that randomized Hamiltonian Monte Carlo, the No-U-Turn Sampler, and the Apogee-to-Apogee Path Sampler all fit within this unified framework as

special cases. The framework is illustrated with a simple alternative to the No-U-Turn Sampler for locally adapting path lengths.

Thursday 13th June

10:00 Sergio Bacallado

Title: Sketching the Jaccard kernel

Abstract: The Jaccard kernel is a widely used positive-definite kernel for binary vectors, which is known in the chemical informatics literature as the Tanimoto coefficient. We discuss two extensions of this kernel to Euclidean spaces, one of which is novel, and introduce two different sketches or random features approximations which allow us to scale kernel methods to large datasets. One of sketches is based on a power series approximation and is differentiable in the input. For this sketch, we provide error bounds on the spectral norm of the Gram matrix which scale optimally with the stable rank of the matrix up to log factors. Experimentally, we show that these random features are effective at approximating the Tanimoto coefficient in real-world chemical datasets and are useful for molecular property prediction and optimization tasks.

10:40 Emiko Dupont

Title: Demystifying spatial confounding

Abstract: Spatially indexed data, i.e. data that have been collected at different geographical locations, are common in many areas of applied statistics. Regression models for such data use spatial random effects to approximate unmeasured spatial variation in the response variable. However, as spatial random effects are typically not independent of the covariates in the model, this can lead to significant bias in covariate effect estimates, making the estimation unreliable. This fundamental problem is referred to as spatial confounding. In recent years, there has been considerable interest in spatial confounding, not least because the most established methods for dealing with the problem were proven to be problematic. However,

spatial models are generally complex, and research into the topic has sometimes led to puzzling and seemingly contradictory results. Here, we develop a broad theoretical framework that brings mathematical clarity to the mechanisms of spatial confounding, providing explicit and interpretable analytical expressions for the resulting bias. From these, we see that it is a problem directly linked to spatial smoothing, and we can identify exactly how the features of the model and the data generation process affect the size and occurrence of bias. Our framework can also be used to understand and generalise some of the existing results on spatial confounding, including suggested methods for bias adjustment.

11:50 Murray Pollock

Title: Fusion: New Developments

Abstract: "Fusion" is a Monte Carlo approach for unifying distributed analyses and inferences on shared parameters from multiple sources, into a single coherent inference. In this talk, I will outline the general approach, along with new areas of application.

12:30 Yuga Iguchi

Title: Parameter estimation of hypo-elliptic SDEs

Abstract: Over the past few decades, there has been significant research into the statistical inference of Stochastic Differential Equations (SDEs). Earlier works mainly focused on elliptic SDEs, where all components are driven by Brownian motion (BM). However, recently, more attention has shifted to hypo-elliptic SDEs, where some components are not directly driven by BM. This talk will focus on the latter case and explain why conventional likelihood-based methods for elliptic SDEs fail under the hypo-elliptic setting. Then, I will deliver several results from my recent papers, which aim to bridge the gaps between elliptic and hypo-elliptic SDEs in the context of statistical inference.

12:50 Kevin Huang

Title: Gaussian universality beyond the classical asymptotic regime

Abstract: Gaussian universality is the phenomenon when a general (typically non-linear) function of independent random vectors can be approximated in distribution by the same function applied to Gaussian vectors. The flexibility of such results has attracted a lot of attention in machine learning, statistics and probability; some notable applications include random matrix theory and characterisation of double-descent risk. This short talk discusses some results on how universality can be applied to setups with strong correlations and high-dimensional data, and when universality starts to fail.

14:30 Elizaveta Semenova

Title: Deep generative modelling aiding spatial statistics

Abstract: Recent advances have demonstrated the use of deep generative models, like variational autoencoders (VAEs), to encode Gaussian Process (GP) priors or their finite realisations. These learned generators can replace original priors within Markov Chain Monte Carlo (MCMC) as a drop-in replacement, enabling efficient inference. However, this approach loses information about the original priors' hyperparameters, rendering hyperparameter inference impossible and the learned priors less distinct. To address this issue, a new method called PriorCVAE can be used. It conditions the VAE on stochastic process hyperparameters, allowing joint encoding and inference of hyperparameters and GP realisations. Importantly, PriorCVAE is model-agnostic, making it applicable to various domains, including encoding solutions of ordinary differential equations (ODEs). This method offers a practical tool for approximate inference with promising applications in spatial and spatiotemporal contexts. I will begin this talk with an overview of spatial statistics and then introduce the PriorVAE method for encoding prior realisations. By discussing the advantages and limitations of PriorVAE, I will introduce PriorCVAE, demonstrate its applications, and explore potential future research directions.

15:10 Valentin De Bortoli

Title: Diffusion Schrödinger Bridge Matching

Abstract: Solving transport problems, i.e. finding a map transporting one given distribution to another, has numerous applications in machine learning. Novel mass transport methods motivated by generative modeling have recently been proposed, e.g. Denoising Diffusion Models (DDMs) and Flow Matching Models (FMMs) implement such a transport through a Stochastic Differential Equation (SDE) or an Ordinary Differential Equation (ODE). However, while it is desirable in many applications to approximate the deterministic dynamic Optimal Transport (OT) map which admits attractive properties, DDMs and FMMs are not guaranteed to provide transports close to the OT map. In contrast, Schrödinger bridges (SBs) compute stochastic dynamic mappings which recover entropy-regularized versions of OT. Unfortunately, existing numerical methods approximating SBs either scale poorly with dimension or accumulate errors across iterations. In this work, we introduce Iterative Markovian Fitting (IMF), a new methodology for solving SB problems, and Diffusion Schrödinger Bridge Matching (DSBM), a novel numerical algorithm for computing IMF iterates. DSBM significantly improves over previous SB numerics and recovers as special/limiting cases various recent transport methods. We demonstrate the performance of DSBM on a variety of problems.

16:20 Giorgos Vasdekis

Title: Skew-symmetric sampling schemes for SDEs and where to find them

Abstract: Locally balancing algorithms are a new class of MCMC algorithms, recently introduced in (Livingstone and Zanella, 2022). One of these algorithms, the Barker algorithm, has been shown to be robust to heteroskedasticity of the posterior target and the step size of the algorithm. At the same time, the algorithm seems to preserve high dimensional properties of state-of-the-art MCMC, making it an interesting alternative to the existing literature. It turns out that in order to sample from the Barker algorithm, one can use ideas of sampling from skew-symmetric distributions. We will transfer these ideas in the context of simulating diffusion processes and we will suggest a new class of unadjusted MCMC algorithms, which are robust with respect to the step size. This is joint work with S. Livingstone, N. Nusken and R. Zhang.

Poster Presenters

- Gulzar Ahmad
- Brett Arnold
- Rancy Chepchirchir
- Sam Duffield
- Mauro Camara Escudero
- Alexandre Galashov
- Max Hird
- Sam Holdstock
- Aziz Lmakri
- Yann McLatchie
- Alexander Modell
- Chris Oates
- Luke Tandjung
- Congye Wang
- Lennie Wells
- Joanna Zyla