Speakers

Abstracts

Stein's Method: The Golden Anniversary

(13 Jun–8 Jul 2022)

1 Krishna Balasubramanian

University of California, Davis, USA
Regularized Stein variational gradient descent and Langevin dynamics

Abstract

In this talk, I will present a regularized formulation of SVGD algorithm and present several results in the mean-filed limit setting. We start with a motivation for the proposed regularization and highlight how the regularized SVGD formulation relates to the Langevin diffusion. Next, we provide convergence results for the regularized SVGD in both the continuous-time and discrete-time setting (with the space being continuous in both cases).

2 Alessandro Barp

University of Cambridge and Alan Turing Institute, UK
Geometry, characteristicness, and weak convergence control of Stein kernels

Abstract

In this talk we discuss the intrinsic geometry of the canonical and diffusion Stein operator, highlighting its connection with the curl/rotationnel operator of de Rham and Koszul. Using ideas from Schwartz's theory of distributions, we then discuss its application to reproducing kernel Hilbert spaces, and derive conditions under which the induced Stein kernel is characteristic, and controls weak convergence to the target distribution - a required property for many machine learning applications of Stein's method.

3 Carina Betken

Ruhr-Universität Bochum, Germany
Central limit theory and variance asymptotics for Poisson cylinder processes

Abstract

We introduce the model of a Poisson process of cylinders in $\mathbb{R}^n$, which jointly generalises the concepts of a Boolean model and a Poisson hyperplane or m-flat process. Using techniques from Malliavin-Stein method we develop a quantitative central limit theory for a broad class of geometric functionals, including volume, surface area and intrinsic volumes. In this context we analyze the asymptotic variance constant, which in contrast to the Boolean model leads to a new degeneracy phenomenon.

This talk is based on joint work with Matthias Schulte and Christoph Thäle.

4 Chinmoy Bhattacharjee

Université du Luxembourg, Luxembourg
Gaussian approximation for region-stabilizing functionals and its applications

Abstract

In this talk, I will consider the Gaussian approximation for functionals of a Poisson process that are expressible as sums of region-stabilizing (determined by the points of the process within some specified regions) score functions and provide bounds on the rate of convergence in the Wasserstein and the Kolmogorov distances. While similar results have previously been shown in Lachièze-Rey, Schulte and Yukich (2019), we extend the applicability by working with stabilization regions that may differ from balls of random radii commonly used in the literature concerning stabilizing functionals.
I will discuss three main applications, namely, the Gaussian approximation of number of minimal points and total length of rooted edges in a *minimal directed spanning tree* on a homogeneous Poisson process in $[0, 1]^d$, and in the context of *Johnson-Mehl tessellations*, and provide presumably optimal rates of convergence.
The talk is based of joint works with Ilya Molchanov and Riccardo Turin.

5 Louis Chen

National University of Singapore, Singapore
Some recollections and reflections on Stein's method of normal approximation

Abstract

This talk gives a short exposition of Stein's method of normal approximation from my personal perspective with anecdotes of historical interest.

6 Laure Coutin

Université Paul Sabatier, France
Normalized Poisson martinagle Vs brownian motion in Wassertein 1 distance

Abstract

We will give a rate of convergence of a sequence of normalized Poisson martingale converge to Brownian motion in Wasserstein 1 distance. This talk is based on a join work with E. Besancon, L. Decreusefond and P. Moyal

`https://arxiv.org/abs/2107.05339`

7 Laurent Decreusefond

Telecom Paris - LINCS, France
Applications of the functional Stein's method

Abstract

Continuous Time Markov Chains, Hawkes processes and many other interesting processes can be described as solution of stochastic differential equations driven by Poisson measures. Previous works, using the Stein's method, give the convergence rate of a sequence of renormalized Poisson measures towards the Brownian motion in several distances, constructed on the model of the Kantorovitch-Rubinstein (or Wasserstein-1) distance. We show that many operations (like time change, convolution) on continuous functions are Lipschitz continuous to extend these quantified convergences to diffusive limits of Markov processes and long-time behavior of Hawkes processes.
Joint work with L. Coutin, E. Besançon and P. Moyal.

8 Hanna Döring

Osnabrück University, Germany
Asymptotics of dynamic Boolean models

Abstract

One way to model telecommunication networks are static Boolean models. However, dynamics such as node mobility have a significant impact on the performance evaluation of such networks. Consider a Boolean model in the plane and a random direction movement scheme. Take time as the third dimension, we model these movements via cylinders. Applying Stein's method, we derive central limit theorems for functionals of the union of these cylinders. The volume and the number of isolated cylinders and the Euler characteristic of the random set are considered and give an answer to the achievable throughput, the availability of nodes, and the topological structure of the network. Furthermore, I would like to present and discuss the asymptotics of a model of movement that allows for a recurring centre of motion like a home. This is joint work with Stephan Bussmann as well as with Carina Betken.

9 Murat A. Erdogdu

University of Toronto, Canada
Representation learning in two-layer neural networks

Abstract

We study the first gradient step on the first-layer parameters of a width-N two-layer neural network, where the training objective is the empirical MSE loss. In the proportional asymptotic limit, and an idealized student-teacher setting, we show that the first gradient update contains a rank-1 "spike", which results in the alignment between the first-layer weights and the linear component of the teacher model. When the learning rate is small, based on a recently established Stein-CLT in the feature space, we prove a Gaussian equivalence property for the trained feature map; this allows us to show that the learned kernel improves upon the initial random features model, but cannot defeat the best linear model on the input. Whereas for sufficiently large learning rate, we prove that trained features can go beyond this "linear regime" and outperform a wide range of random features models.

10 Max Fathi

Université Paris Cité, France

Stability of the sharp spectral gap bound for positively curved manifolds via Stein's method

Abstract

The Bonnet-Myers theorem is a classical result in Riemannian geometry, which gives the maximal value for the diameter of a manifold with positive curvature. In this talk, I will explain how we can use Stein's method for beta distributions to study the structure of such a manifold when the diameter is almost maximal (which is equivalent to having an almost minimal spectral gap). Joint work with Ivan Gentil and Jordan Serres.

11 Anum Fatima

University of Oxford, UK

Stein's method for Poisson-Exponential distributions

Abstract

The distribution of the maximum of a zero truncated Poisson number of i.i.d. exponentially distributed random variables is known as a Poisson-Exponential distribution. This distribution arises for example as a model for failure rates. In this paper, we develop Stein's method for Poisson-Exponential distributions. We provide upper bounds on the approximation errors in total variation distance when approximating Poisson-Exponential distributions and a Generalized Poisson-Exponential distribution by different Poisson-Exponential distributions. Moreover, employing standardized Stein equations we obtain upper bounds on the bounded Wasserstein distance when using a Poisson-Exponential distribution to approximate the distribution of maxima of a zero truncated Poisson number of i.i.d. geometric random variables.

The results are applied to obtain a bound on the bounded Wasserstein distance between a distribution of the maximum waiting time of the occurrence of Bernoulli sequence patterns and a Poisson-Exponential distribution. We also illustrate the results numerically when using a Poisson-Exponential distribution to approximate a Generalized Poisson-Exponential distribution on a real data set.

12 Han Liang Gan

Waikato University, New Zealand

Stein's method and moment duality for two-island model approximations

Abstract

Two-island Wright-Fisher models are used to model genetic frequencies and variability for subdivided populations. The two key components of the model are the rates of migration between the two islands and the rates of mutation within the two islands. Using Stein's method, we show that as the population size increases, the appropriate approximation and limit for the stationary distribution of a two-island Wright-Fisher Markov chain depends on the rate of migration relative to the rate of mutation. If migration is of a higher order than mutation, then a beta approximation is appropriate. If migration is of the same order as mutation, then the stationary distribution of the Wright-Fisher two island diffusion model is the better approximating distribution. The former is derived using results from Gan, Rollin and Ross (2017), and the latter via combining Stein's method and moment duality of the two island diffusion.

13 Arthur Gretton

Gatsby Computational Neuroscience Unit, UK

Relative goodness-of-fit testsfor models with latent variables

Abstract

I will describe a kernel-based nonparametric test of relative goodness of fit, where the goal is to compare two models, both of which may have unobserved latent variables, such that the marginal distribution of the observed variables is intractable. Given the premise that "all models are wrong," the goal of the test is to determine whether one model significantly outperforms the other in respect of a reference data sample. The test generalises earlier kernel Stein discrepancy (KSD) tests to the case of latent variable models, a much more general class than the fully observed models treated previously. The new test, with a properly calibrated threshold, has a well-controlled type-I error. In the case of models with low-dimensional latent structure and high-dimensional observations, our test significantly outperforms the relative maximum mean

discrepancy test, which is based on samples from the models, and does not exploit the latent structure. We illustrate the test on probabilistic topic models of arXiv articles.

14 Christian Houdré

Georgia Institute of Technology, USA
Covariance representation, Stein's kernels and high dimensional CLTs

Abstract

In this continuing joint work with Benjamin Arras, we explore connections between covariance representations and Stein's method. In particular, via Stein's kernels we obtain quantitative high-dimensional CLTs in 1-Wasserstein distance when the limiting Gaussian probability measure is anisotropic. The dependency on the parameters is completely explicit and the rates of convergence are sharp.

15 Arturo Jaramillo

Center of Research in Mathematics, Mexico
Quantitative Erdös-Kac theorem for additive functions

Abstract

The talk will take as a starting point the celebrated Erdös-Kac theorem; a result of great importance in probabilistic number theory, which establishes that the fluctuations of the number of prime factors in a uniform sample over 1,..., n are asymptotically Gaussian. Naturally, after the publication of this result, many quantitative versions of it have been studied. LeVeque conjectured that the optimal rate was asymptotically equivalent to $loglog(n)^{(-1/2)}$. This was later proved Turan y Rényi by means of an ingenious manipulation of the underlying characteristic function. Unfortunately, up to this day, all the perspectives for solving LeVeque's conjecture are based on the use of non-trivial complex analysis tools, while the probabilistic perspective has been only successfully applied to obtain suboptimal rates of convergence. In this talk, we will give a purely probabilistic proof of LeVeque's conjecture which will allow us to address the problem in a general fashion by means of Stein's method techniques.

The paper can be found in the Arxiv link https://arxiv.org/abs/2102.05094

16 Wittawat Jitkrittum

Google Research, USA
Testing goodness of fit of conditional density models with kernels

Abstract

We propose a nonparametric statistical test of goodness of fit for conditional distributions: given a conditional probability density function $p(y|x)$ and a joint sample, decide whether the sample is drawn from $p(y|x)r(x)$ for some density r. Our test, formulated with a Stein operator, can be applied to any differentiable conditional density model, and requires no knowledge of the normalizing constant. We show that 1) our test is consistent against any fixed alternative conditional model; 2) the statistic can be estimated easily, requiring no density estimation as an intermediate step; and 3) a slight modification of the test results in an interpretable test that provides insight on where the conditional model does not fit well in the domain of the covariate. We demonstrate the interpretability of our test on a task of modeling the distribution of New York City's taxi drop-off location given a pick-up point. To our knowledge, our work is the first to propose such conditional goodness-of-fit tests that simultaneously have all these desirable properties.

17 Mikolaj Kasprzak

Université du Luxembourg, Luxembourg
How good is your Bayesian CLT? Finite-sample error bounds for a variety of useful divergences

Abstract

The Bayesian Central Limit Theorem (BCLT) for finite-dimensional models is a primary motivation for the widely-used Laplace approximation. But currently the BCLT is expressed only asymptotically and only in terms of total variation (TV) distance. To understand the quality of an actual Laplace approximation, we need finite-sample bounds. And to understand the quality

of posterior mean and variance estimates, we need bounds on alternative divergences. Our work provides the first closed-form, finite-sample bounds for the BCLT that do not require strong log concavity of the posterior. We bound not only the TV distance but also (A) the Wasserstein-1 distance, which controls error in a posterior mean estimate, and (B) an integral probability metric that controls the error in a posterior variance estimate. We compute exact constants in our bounds for a variety of standard models, including logistic regression, and numerically investigate the utility of our bounds. And we provide a framework for analysis of more complex models.

18 Sumit Mukherjee

Columbia University, USA

Motif counting via subgraph sampling

Abstract

Consider the subgraph sampling model, where we observe a random subgraph of a given (possibly non random) large graph G_n, by choosing vertices of G_n independently at random with probability p_n. In this setting, we study the question of estimating the number of copies $N(H, G_n)$ of a fixed motif/small graph (think of H as edges, two stars, triangles) in the big graph G_n. We derive necessary and sufficient conditions for the consistency and the asymptotic normality of a natural Horvitz-Thompson (HT) type estimator.

As it turns out, the asymptotic normality of the HT estimator exhibits an interesting fourth-moment phenomenon, which asserts that the HT estimator (appropriately centered and rescaled) converges in distribution to the standard normal whenever its fourth-moment converges to 3. The proof technique depends on a careful analysis of the error terms arising from Stein's method for dependency graphs, and gives a normal approximation for general multilinear forms which is of possible independent interest.

This talk is based on joint work with Bhaswar B. Bhattacharya and Sayan Das.

19 Daniel Paulin

The University of Edinburgh, Scotland

Efron-Stein inequalities for random matrices

Abstract

We establish new concentration inequalities for random matrices constructed from independent random variables. These results are analogous with the generalized Efron–Stein inequalities developed by Boucheron et al. The proofs rely on the method of exchangeable pairs. This is joint work with Lester Mackey and Joel A. Tropp.

20 Guillaume Poly

Université de Rennes 1, France
On central limit theorems of "Salem-Zygmund" type

Abstract

The central limit Theorem of Salem-Zygmund in its classic form focuses on the almost sure central limit behavior of random trigonometric series which are evaluated at some point X uniformly distributed according to the Lebesgue measure. In this talk, we will introduce and explain this phenomenon as well as present some generalizations in higher dimensions or for random series of more general functions. We will specifically focus on the role played by the Stein's method in providing quantitative proofs of this phenomenon and on the importance of the rate of convergence for further applications which notably concern the study of local geometric functional such that the volume of the zero set of random functions for instance.

21 Nicolas Privault

Nanyang Technological University, Singapore
Berry-Esseen bounds for functionals of independent random variables

Abstract

We derive Berry-Esseen approximation bounds for general functionals of independent random variables, based on a continuous-time integration by parts setting and discrete chaos expansions methods. Our approach improves on related results obtained in discrete-time integration by parts settings and applies to U-statistics satisfying the weak assumption of decomposability in

the Hoeffding sense. It also yields Kolmogorov distance bounds instead of the Wasserstein bounds previously derived in the special case of degenerate U -statistics. Linear and quadratic functionals of arbitrary sequences of independent random variables are considered as particular cases, with new fourth moment bounds, and applications are given to Hoeffding decompositions, weighted U-statistics, quadratic forms, and random subgraph weighing. In the case of quadratic forms, our results recover and improve the bounds available in the literature, and apply to matrices with non-empty diagonals. (joint work with Grzegorz Serafin, Wroclaw)

22 Adrian Röllin

National University of Singapore, Singapore
Higher order fluctuations in dense random graph models

Abstract

Dense graph limit theory is mainly concerned with law-of-large-number type of results. We propose a corresponding central limit theorem — or rather fluctuation theory — based on Janson's theory of Gaussian Hilbert Spaces and generalised U-statistics from the 1990s. Our approach provides rates and allows for proper statistical inference based on subgraph counts.

23 Nathan Ross

The University of Melbourne, Australia
Gaussian process approximation using Stein's method, with applications to queues

Abstract

Gaussian processes are used as limits or approximations in a wide variety of application areas, such as finance and queuing. In this talk, we discuss a general approach via Stein's method to bound the error when approximating càdlàg random processes by a real continuous Gaussian process. For processes that are representable as integrals of a certain general form against an underlying point process, the bound is in terms of couplings of the original process to processes generated from the reduced Palm measures associated with the

point process. We apply the approach to some GI/GI/Infinity queues in the "heavy traffic" regime.
Based on joint work with A. D. Barbour and Guangqu Zheng

24 Adil Salim

Microsoft Research, USA
Gaussian variational inference with Wasserstein gradient flows (Lambert et al. 2022)

Abstract

In statistics, variational inference is the task of approximating a target distribution by an element of a parametric family of distributions. We consider Gaussian variational inference where this element is defined as the closest, in the sense of the KL divergence, to the target among Gaussian distributions. I will present some recent results by Lambert, Chewi, Bach, Bonnabel and Rigollet (2022) where they study Gaussian variational inference using optimal transport techniques.

The goal of this talk is to make a parallel between the algorithm of Lambert et al. (2022) and the Stein Variational Gradient Descent (Liu and Wang 2016), a widely used sampling algorithm that can also be seen as an optimization algorithm for the KL divergence. More precisely, the algorithm of Lambert et al. (2022) can be seen as an implementable version of SVGD in the population limit, for a particular kernel.

25 Adrien Saumard

ENSAI Bruz, France
Stein kernels, functional inequalities and applications in statistics

Abstract

We will present the notion of Stein kernel, which provides generalizations of the integration by parts Stein's formula for the normal distribution (which has a constant Stein kernel, equal to its covariance). We will first focus on dimension one, where under good conditions the Stein kernel has an explicit formula. We will see that the Stein kernel appears naturally as a weighting

of a Poincaré type inequality and that it enables precise concentration inequalities, of the Mills' ratio type. In a second part, we will work in higher dimensions, using in particular Max Fathi's construction of a Stein kernel through the so-called "moment maps" transportation. This will allow us to describe the performance of some shrinkage and thresholding estimators, beyond the classical assumption of Gaussian (or spherical) data. This presentation is mostly based on joint works with Max Fathi, Larry Goldstein, Gesine Reinert and Jon Wellner.

26 Uwe Schmock

Vienna University of Technology, Austria
Probabilistic interpretation of a theorem of Kolmogorov using the zero bias transformation

Abstract

A theorem of Kolmogorov gives a necessary and sufficient condition on the characteristic function of a mean zero, finite variance random variable X to be infinitely divisible. We prove the equivalent probabilistic condition that the zero bias distribution X^* has the stochastic representation

$$X^* \stackrel{\mathrm{d}}{=} X + UY,$$

where U, X, Y are independent and U is uniformly distributed on the unit interval. A comparison of these relationships with corresponding results for infinitely divisible distributions on the positive half line, Steutel's theorem and size biasing are discussed. Several examples illuminate our theorem and the probability measure appearing in Kolmogorov's condition. (Joint work with Larry Goldstein.)

27 Matthias Schulte

Hamburg University of Technology, Germany
Multivariate normal approximation of stabilising functionals of Poisson processes

Abstract

In this talk functionals of Poisson processes are studied which are sums of scores of the underlying points. It is assumed that the scores stabilise in the sense that the score of a point is determined by the points in a random neighbourhood given by a so-called radius of stabilisation. We consider a vector of sums of stabilising scores. For the situation that the radii of stabilisation decay exponentially fast and some moment assumptions are satisfied, quantitative bounds for the multivariate normal approximation are established. The results deal with several distances including a multivariate generalisation of the Kolmogorov distance and lead to rates of convergence that are in general unimprovable. Several examples concerning spatial random graphs will be discussed. The proofs of the main results rely on the Malliavin-Stein method for multivariate normal approximation and a careful analysis of the covariance structure. This talk is based on joint work with J.E. Yukich (Lehigh University).

28 Yvik Swan

Université libre de Bruxelles, Belgium
Stein's method for extreme value distributions

Abstract

We present some Stein operators for extreme value distributions (Weibull, Gumbel, Fréchet), discuss the properties of the solutions to the corresponding Stein equations, and provide some applications to computing rates of convergence in various instantiations of the Extreme Value Theorem.

29 Tadas Temcinas

University of Oxford, UK
Multivariate central limit theorems for random clique complexes

Abstract

Motivated by open problems in applied and computational algebraic topology, we establish multivariate normal approximation theorems for three random vectors which arise organically in the study of random clique complexes. These are:

1. the vector of critical simplex counts attained by a lexicographical Morse matching,
2. the vector of simplex counts in the link of a fixed simplex, and
3. the vector of total simplex counts.

The first of these random vectors forms a cornerstone of modern homology algorithms, while the second one provides a natural generalisation for the notion of vertex degree, and the third one may be viewed from the perspective of U-statistics. To obtain distributional approximations for these random vectors, we extend the notion of dissociated sums to a multivariate setting and prove a new central limit theorem for such sums using Stein's method.

30 Anna Paola Todino

Università di Milano-Bicocca, Italy

Stein's method for Gaussian approximations of random spherical Eigenfunctions

Abstract

Recently, lot of efforts have been devoted to the analysis of the high-frequency behaviour of geometric functionals (Lipschitz-Killing Curvatures) for the excursion sets of random eigenfunctions on the unit sphere (spherical harmonics) and on its subdomains. The asymptotic behavior of their expected values and variances have been investigated and quantitative central limit theorems have been established in the high energy limits, after exploiting Wiener chaos expansions and Stein-Malliavin techniques. Another interesting issue concerns the Gaussianity hypothesis of the random field. In this direction we introduce a model of Poisson random waves in S^2 and we study Quantitative Central Limit Theorems when both the rate of the Poisson process and the energy (i.e., frequency) of the waves (eigenfunctions) diverge to infinity. We consider finite-dimensional distributions, harmonic coefficients and convergence in law in functional spaces, and we investigate carefully the interplay between the rates of divergence of eigenvalues and Poisson governing measures.

31 Xin Tong

National University of Singapore, Singapore
Sampling with constraints using Stein variational gradient descent and Langevin dynamics

Abstract

Sampling-based inference and learning techniques, especially Bayesian inference, provide an essential approach to handling uncertainty in machine learning (ML). As these techniques are increasingly used in daily life, it becomes essential to safeguard the ML systems with various trustworthy-related constraints, such as fairness, safety, interpretability. We propose a family of constrained sampling algorithms which generalize Langevin Dynamics (LD) and Stein Variational Gradient Descent (SVGD) to incorporate a moment constraint or a level set specified by a general nonlinear function. By exploiting the gradient flow structure of LD and SVGD, we derive algorithms for handling constraints, including a primal-dual gradient approach and the constraint controlled gradient descent approach. We investigate the continuous-time mean-field limit of these algorithms and show that they have $O(1/t)$ convergence under mild conditions.

32 Nathakhun Wiroonsri

King Mongkut's University of Technology Thonburi, Thailand
Normal approximation for associated point processes with applications to fire incident simulation using permanental Cox processes

Abstract

Estimating the number of natural disasters benefits the insurance industry in terms of risk management. However, the estimation process is complicated due to the fact that there are many factors affecting the number of such incidents. In this work, we propose a Normal approximation technique for associated point processes for estimating the number of natural disasters under the following two assumptions: 1) the incident counts in any two distinct areas are positively associated and 2) the association between these counts in two distinct areas decays exponentially with respect to distance outside some small local neighborhood. Under the stated assumptions, we

use Stein's method to provide non asymptotic L^1 bounds (also known as Wasserstein bounds) to the normal for functionals of associated point processes. Then we apply the main result to permanental Cox processes that are known to be positively associated. Finally, we apply our Normal approximation results for permanental Cox processes to Thailand's fire data from 2007 to 2020, which was collected by the Geo-Informatics and Space Technology Development Agency of Thailand. **Keywords**: Correlation inequality, Cox process, Local dependence, Random fields, Natural disaster, Positive association **AMS 2010 subject classifications**: 60F05, 65C20

33 Wenkai Xu

University of Oxford, UK
A unifying view on kernelised Stein discrepancy tests for goodness-of-fit

Abstract

Non-parametric goodness-of-fit testing procedures based on kernel Stein discrepancies (KSD) are promising approaches to validate general unnormalised distributions in various scenarios. Existing works focused on studying kernel choices to boost test performances. However, the choices of (non-unique) Stein operators also have a considerable effect on the test performances. This talk starts by showing KSD approaches in different testing scenarios beyond $\mathbb{R}^d$, e.g. testing censored data. Then we present a unifying framework for the set of existing KSD tests via standardisation-function. In addition, we show that how the proposed framework can be used as a guide to develop novel kernel-based non-parametric tests on complex data scenarios, e.g. truncated distributions or compositional data.

34 Xiaochuan Yang

Brunel University London, UK
Random coverage of a manifold with boundary

Abstract

The random coverage problem asks whether a set of n balls in a metric space A, with centers placed independently at random according to some probabil-

ity distribution on A, fully covers a specified subset B (possibly B=A). Restricting attention to balls of equal radius, one may ask what is the minimum radius required to achieve coverage. This is a random variable, sometimes called the coverage threshold. We establish a strong law and a weak law for the coverage thresholds in the large n limit, when A is a smooth manifold with boundary. The weak law is closely related to a Poisson convergence result for the number of vacant regions in A, which is proved by the Chen-Stein method. Joint work with Mathew Penrose (Bath).