Fall 2015 – Spring 2016

Colloquium

Unless stated otherwise, colloquia are scheduled for Thursdays 4:30-5:30pm in WH-100E with refreshments served from 4:00-4:25 pm in WH-102.

Organizers: Michael Dobbins, Andrey Gogolev and kargin

- **October 15, 4:30pm**
  Speaker: **Wushi Goldring** (University of Washington in Saint Louis)
  Topic: **The limits of algebraicity: The case of automorphic representations**
  Abstract:
  This talk will explore the general theme I like to call “algebraicity” — that objects which seem purely analytic in nature actually admit a deep algebraic interpretation — in the context of the Langlands correspondence. The latter predicts that a specific, distinguished subset of all automorphic representations should have deep algebraic properties, such as having algebraic Hecke eigenvalues, admitting a system of associated Galois representations and, ultimately, corresponding to a motive in the sense of Grothendieck. Two approaches have been used to verify Langlands' prediction: (1) Finding and exploiting a direct link with algebraic geometry and (2) Using Langlands' Functoriality Principle. I will discuss the possibilities and limitations of the two approaches and report on recent work on each approach. The results using the geometric approach are joint work with Jean-Stefan Koskivirta, see arXiv:1507.05032; closely related results were obtained independently around the same time by Pilloni-Stroh, and in a more restricted setting announced also by Boxer.
- **October 22, 4:30pm**
  Speaker: **Gary Greaves** (Tohoku)
  Topic: **Equiangular Lines in Euclidean Space**
  Abstract:
  Given some dimension d, what is the maximum number of lines in $\mathbb R^d$ such that the angle between any pair of lines is constant? (Such a system of lines is called “equiangular”.) This classical problem was initiated by Haantjes in 1948 in the context of elliptic geometry. In 1966, Van Lint and Seidel showed that graphs could be used to study equiangular line systems. Recently this area has enjoyed a renewed interest due to the current attention the quantum information community is giving to its complex analogue. I will give an introduction to the area and report on some new developments in the theory of equiangular lines in Euclidean space. Among other things, I will present a new construction using real mutually unbiased bases (orthonormal bases such that the angle between two elements of different bases is $arccos(\frac{1}{\sqrt{d}})$, as well as improvements to two long-standing upper bounds for equiangular lines in dimensions 14 and 16.
- **October 29, 4:30pm**
  Speaker: **Alice Medvedev** (City College of New York)
  Topic: **Model Theory and Algebraic Dynamics.**
  (Joint work with Thomas Scanlon.)
  Abstract: Model Theory, a branch of mathematical logic, has arcane abstract definitions of geometric concepts like “dimension.” They specialize to interesting and useful notions in many settings around arithmetic geometry and algebraic number theory. The goal of the talk is to present the intuitive meaning of these notions through the

relatively simple example of coordinate-wise polynomial discrete dynamical systems. Consider a (discrete) dynamical system $F(x, y, z) := (f(x), g(y), h(z))$ for polynomials f, g, and h, acting on the three-dimensional space over complex numbers. What subsets S are invariant under F, in the sense that $F(S)$ is a subset of S? In particular, what algebraic sets, that is solution sets of systems of polynomial equations, are invariant under F? I will describe the tools from modern model theory, a branch of mathematical logic, that reduce this question to understanding composition of one-variable polynomials, and an old theorem of Ritt that supplies this understanding.

- **November 12, 4:30pm**
  Speaker: **Xiangjin Xu** (Binghamton University)
  Topic: **Sharp Li–Yau type estimates and new heat kernel estimates on negative curved manifolds**
  Abstract: In this talk, we firstly discuss some background and history related to the Gaussian type upper bound of the heat kernel and the sharp Li–Yau type estimates for the positive solution $u(x,t)$ of the heat equations $u_t-\Delta u=0$ on a complete manifold. Then we obtain some new almost sharp Li–Yau type Harnack inequalitieson a complete manifold with $Ricci(M)\ge-k$. As applications, new parabolic Harnack inequalities are derived, and monotonicity of Perelman type entropy for the heat kernel and the positive solutions are achieved. And we are able to obtain the Gaussian type upper bound of the heat kernel on a complete manifold with $Ricci(M)\ge-k$. At the end, we discuss some open questions related to the sharp Li–Yau type estimates. Part of talk is joint with Junfang Li.

- **December 2, 4:40pm**
  Speaker: **Dehan Kong** (University of North Carolina at Chapel Hill)
  Topic: **High-dimensional Matrix Linear Regression Model**
  Time: **4:40–5:40pm**
  Abstract: We develop a high-dimensional matrix linear regression model (HMLRM) to correlate matrix responses with high-dimensional scalar covariates when coefficient matrices have low-rank structures. We propose a fast and efficient screening procedure based on the spectral norm to deal with the case that the dimension of scalar covariates is ultra-high. We develop an efficient estimation procedure based on the nuclear norm regularization, which explicitly borrows the matrix structure of coefficient matrices. We systematically investigate various theoretical properties of our estimators, including estimation consistency, rank consistency, and the sure independence screening property under HMLRM. We examine the finite-sample performance of our methods using simulations and a large-scale imaging genetic dataset collected by the Alzheimer's Disease Neuroimaging Initiative study.

- **December 4, 4:40pm**
  Speaker: **Ni Zhao** (Fred Hutchinson Cancer Research Center)
  Topic: **Statistical Inference Methods for High Dimensional “Omics” Data**
  Time: **4:40–5:40pm**
  Abstract: Recent advances in high-throughput biotechnology have enabled multiple platform “omics” profiling of biological samples. In this talk, I will present two statistical inference methods for testing the association between high-dimensional “omics” data and a phenotype of interest. The first method is designed to analyze large-scale methylation changes using high-resolution CpG data [Zhao et al, Genetic Epidemiology, 2015]. The second method is a powerful likelihood ratio test via the composite kernel machine regression in Genome Wide Association Studies [Zhao et al, Biometrics, Prepare for Submission]. The method tests the association between multiple SNPs with the phenotype, considering possible gene-environment interaction. The utility of the methods will be illustrated within the Norwegian Mother and Child Cohort Study: a large prospective cohort study of pregnant Norwegian women and their children.

- **December 7, 4:40pm**
  Speaker: **Jyotishka Datta** (Duke University and SAMSI)
  Topic: **Priors for Sparse High-Dimensional Discrete or Continuous Data**
  Time: **4:40–5:40pm**
  Abstract: Sparse signal detection has been one of the most important challenges in the analysis of large-scale data-sets arising from many different disciplines, e.g. Genomics, finance and astronomy. In this talk, I will focus on two key aspects of inference on a high-dimensional sparse mean vector: (1) how to provide theoretical justifications for existing methods that perform strongly, and (2) how to use this theoretical insight to develop new approaches that can outperform the current methods in the `ultra-sparse' regime. In the first half of the talk, I will discuss multiple testing optimality for continuous data, and prove Oracle properties of the popular `Horseshoe' prior [1]. I will then develop a novel prior called the 'Horseshoe+' prior [2] that sharpens the 'Horseshoe' prior's signal detection abilities. I will illustrate that the Horseshoe+ prior outperforms the existing methods both in theory and practice and correctly identifies the `differentially expressed' genes from microarray data. In the second half, I will briefly discuss inference on high dimensional sparse count data which is fundamentally different from the high-dimensional Gaussian case. I will present the 'Gauss-Hypergeometric' prior for sparse Poisson means [3], motivated by the growing interest in analyzing sparse count data and end with an application to detect mutational hotspots in whole exome sequencing data. References: [1] Datta, J. and Ghosh, J. K. (2013). Asymptotic properties of Bayes risk for the horseshoe prior. Bayesian Analysis, 8(1):111–131. [2] Bhadra, A., Datta, J., Polson, N. G., and Willard, B. (2015). The Horseshoe+ Estimator of Ultra-Sparse Signals. arXiv preprint arXiv:1502.00560. [3] Datta, J. and Dunson, D. B. (2015). Priors for High-Dimensional Sparse Poisson Means. arXiv preprint arXiv:1510.04320. (Biometrika, under revision)

- **January 25, 4:40pm**
  Speaker: **Tianyi Zheng** (Stanford University)
  Topic: **Random walk parameters and the geometry of groups**
  Time: **4:40–5:40pm**
  Abstract: The first characterization of groups by an asymptotic description of random walks on their Cayley graphs dates back to Kesten's criterion of amenability. I will first review some connections between the random walk parameters and the geometry of the underlying groups. I will then discuss a flexible construction that gives solution to the inverse problem (given a function, find a corresponding group) for large classes of speed, entropy and return probability and Hilbert compression functions of groups of exponential volume growth. Based on joint work with Jeremie Brieussel.

- **January 28, 4:30pm**
  Speaker: **David Renfrew** (University of Colorado)
  Topic: **Spectral properties of large Non-Hermitian Random Matrices**
  Time: **4:30–5:30pm**
  Abstract: The study of the spectrum of Non-Hermitian random matrices with independent, identically distributed entries was introduced by Ginibre and Girko. I will present two generalizations of the iid model when the independence and identically distribution assumptions are relaxed and discuss applications to modeling Neural Networks.

- **January 29, 4:40pm**
  Speaker: **Sanjeena Dang** (University of Guelph)
  Topic: **Model-based clustering: some recent work and biological applications**
  Time: **4:30–5:30pm**
  Abstract: With advances in high throughput technologies, massive amounts of data can be generated in an

increasingly shorter period of time. Challenges and approaches to dealing with high-dimensional data such as RNA-Seq data, microbiome data, and microarray data will be discussed in a model-based clustering context. The talk will provide an overview of different frameworks for clustering increasingly complex biological data using both mixtures of Gaussian distributions and mixtures of non-Gaussian distributions. A number of families of mixture models will be considered and the talk will conclude with a discussion of future trends.

- **February 1, 3:30-4:30pm**
  Speaker: **Subhro Ghosh** (Princeton University)
  Topic: **Rigidity phenomena in random point sets**
  Time: **3:30-4:30pm**
  Abstract: In several naturally occurring (infinite) random point processes, we establish that the number of the points inside a bounded domain can be determined, almost surely, by the point configuration outside the domain. This includes key examples coming from random matrices and random polynomials. We further explore other random processes where such “rigidity” extends to a number of moments of the mass distribution. The talk will focus on particle systems with such curious “rigidity” phenomena, and their implications. We will also talk about applications to natural questions in stochastic geometry and harmonic analysis.

- **February 2, 4:30-5:30pm**
  Speaker: **Martin Slawski** (Rutgers University)
  Topic: **High-dimensional regression with sign constraints and quantized observations**
  Time: **4:30-5:30pm**
  Abstract: In this talk, I will present several results on two specific aspects in the area of high-dimensional statistics and compressed sensing. The first part of the talk is concerned with the potential usefulness of sign constraints in sparse high-dimensional linear regression, which, along with some extensions, is put into the broader context of regularization-free estimation. The second part of the talks is dedicated to the trade-off between the number of samples and the bit depth per sample in compressed sensing with quantized observations. I will outline an analysis of a popular approach to signal recovery due to Plan and Vershynin in this regard, with the conclusion that a small number of bits (one or two) is optimal. If time permits I will also provide a brief overview on the related problem of similarity estimation for massive data sets based on quantized random projections.

- **February 4, 4:40-5:40pm**
  Speaker: **Cheng Li** (Duke University)
  Topic: **PIE: Simple, Scalable and Accurate Posterior Interval Estimation**
  Time: **4:40-5:40pm**
  Abstract: Scalable Bayes for big data is a rapidly growing research area, but most existing methods are either highly complex to implement for practitioners, or lack theoretical justification for uncertainty quantification. Bayesian methods quantify uncertainty through posterior and predictive distributions. For massive datasets, it is difficult to efficiently estimate summaries of these distributions, such as posterior quantiles and credible intervals. In small scale problems, posterior sampling algorithms such as Markov chain Monte Carlo (MCMC) remain the gold standard, but they face major problems in scaling up to big data. We propose a very simple and general Posterior Interval Estimation (PIE) algorithm to evaluate the posterior distributions of one-dimensional (1-d) functionals, which are typically the focus in many applications. The PIE algorithm consists of three steps. First, full data are partitioned into computationally tractable subsets. Second, sampling algorithms such as MCMC are run in parallel across every subset. Finally, PIE approximates the full posterior by simply averaging posterior quantiles estimated from each subset. This allows standard Bayesian algorithms such as MCMC to be trivially scaled up to big data. We provide strong theoretical guarantees for PIE on its posterior uncertainty quantification, and compare its empirical performance with variational Bayes and the recent WASP algorithm for mixed effects models and nonparametric

Bayesian mixture models.

- **April 7, 4:30-5:30pm**
  Speaker: **David Spivak** (MIT)
  Topic: **Calculating steady states of nonlinear dynamical systems using matrix arithmetic**
  Abstract: Open dynamical systems are mathematical models of machines that take input, change their internal state, and produce output. For example, one may model anything from neurons to robots in this way. Several open dynamical systems can be arranged in series, in parallel, and with feedback to form a new dynamical system—this is called compositionality—and the process can be repeated in a fractal-like manner to form more complex systems of systems. I will discuss a technique for calculating the steady states of an interconnected system of systems, in terms of the steady states of the component dynamical systems. The steady states, or equilibria, are organized into “steady state matrices” which generalize bifurcation diagrams. I'll show that the compositionality structure of dynamical systems fits with familiar operations on matrices: serial, parallel, and feedback compositions correspond to multiplication, Kronecker product, and partial trace operations on matrices. Thus we can calculate the steady states of a system of dynamical systems by doing matrix arithmetic on the individual steady state matrices. This talk will be aimed at an undergraduate level.

- **April 15, 3:30-4:30pm**
  Dean's Lecture in Geometry.
  Speaker: **Stephen Preston** (Brooklyn College CUNY)
  Title: **The Geometric Approach to Partial Differential Equations** *Abstract:*
  I will give a survey of those PDEs that can be viewed as geodesics on infinite-dimensional spaces. Vladimir Arnold observed in 1966 that an ideal fluid can be viewed as a geodesic on the group of volume-preserving diffeomorphisms of a domain, and he computed some of the sectional curvatures, showing that many of them are negative. Since then many other equations have found interpretation as geodesics, including the equation for inextensible whips, the Korteweg-de Vries equation, and other conservative PDEs. I will describe some of the finite-dimensional models (including the equations for a rigid body) along with the general aspects of finite-dimensional Riemannian geometry and what still works in infinite dimensions. Finally I will show how a new one-dimensional model of the Euler equation shares many of the same properties and also ties into the Teichmuller theory in complex analysis.

- **May 3**
  Dean's Lecture in Geometry.
  Speaker: **Melvyn Nathanson** (CUNY)
  Title: **Every Finite Subset of an Abelian group is an Asymptotic Approximate Group** *Abstract:*
  If A is a nonempty subset of an additive abelian group G, then the *h-fold sumset* is \[hA = \{x_1 + \cdots + x_h : x_i \in A_i \text{ for } i=1,2,\ldots, h\}.\]
  We do not assume that A contains the identity, nor that A is symmetric, nor that A is finite. The set A is an *(r,ℓ)-approximate group in G* if there exists a subset X of G such that $|X| \leq \ell$ and $rA \subseteq XA$. The set A is an *asymptotic (r,ℓ)-approximate group* if the sumset hA is an (r,ℓ)-approximate group for all sufficiently large $h.$ It is proved that every polytope in a real vector space is an asymptotic (r,ℓ)-approximate group, that every finite set of lattice points is an asymptotic (r,ℓ)-approximate group, and that every finite subset of every abelian group is an asymptotic (r,ℓ)-approximate group.

- **May 10**

Dean's Lecture in Geometry.

Speaker: **Alexandru Buium** (University of New Mexico)

Title: **The differential geometry of Spec Z** *Abstract:*

The aim of this talk is to show how one can develop an arithmetic analogue of classical differential geometry. In this new geometry the ring of integers Z will play the role of a ring of functions on an infinite dimensional manifold. The role of coordinate functions on this manifold will be played by the prime numbers.

The role of partial derivatives of functions with respect to the coordinates will be played by the Fermat quotients of integers with respect to the primes. The role of metrics (respectively 2-forms) will be played by symmetric (respectively antisymmetric) matrices with coefficients in Z. The role of connections (respectively curvature) attached to metrics or 2-forms will be played by certain adelic (respectively global) objects attached to matrices as above. One of the main conclusions of our theory will be that Spec Z is ``intrinsically curved;" the study of this curvature will then be one of the main tasks of the theory.