

## The Cramér Society autumn meeting 2017:

### Big Data and High-Dimensional Statistical Inference: Recent Advances and Challenges



You are warmly invited to attend the Cramér Society autumn meeting 2017, Big Data and High-Dimensional Statistical Inference: Recent Advances and Challenges

**Date:** Friday the 27th of October, 2017 09:45 AM-- 17:30 PM

**Location:** KTH Royal Institute of Technology, Stockholm. Sal K1, Teknikringen 56, Kemi, floor 3, KTH Campus. See the link below

<https://www.kth.se/places/room/id/19814134-e285-49bc-9683-2f5c2edadbcf>

#### **Meeting summary:**

Statistical methodologies to handle Big Data and, especially, high-dimensional data, have gained importance in many applied scientific fields. Examples include information technology, finance, astrophysics, functional genomics and proteomics, to name just a few. To reflect the importance of bridging state-of-the-art research with practical applications, the goal of this meeting is two-fold: to highlight novel research addressing the high dimensionality in statistical inference, and to demonstrate examples of exploiting statistical methods for Big Data from a variety of applied areas. The meeting features a number of speakers from the Swedish statistical community as well as international researchers.

The conference is hosted by the Department of Mathematics, KTH Royal Institute of Technology, Stockholm. It is financially supported by the Swedish Statistical Society (Svenska Statistikfrämjandet).

Please register no later than October 13th , by writing an e-mail to [pavlenko@math.kth.se](mailto:pavlenko@math.kth.se)

*Schedule:*

- 9.45 Registration, welcome, coffee
- 10.15 Mattias Villani, Linköping University  
*Hamiltonian Monte Carlo with Energy Conserving Subsampling for Big Data Problems.*
- 11.15 Rolf Sundberg, Stockholm University  
*Inference when dimension (much) exceeds sample size.  
A 20th century controversy – Routine today?*
- 12.00 Lunch
- 13.15 Raazesh Sainudiin, Uppsala University  
*Seeded by Hate? Characterizing the Twitter Networks of Prominent Politicians and Hate Groups in the 2016 US Election.*
- 14.15 Sofia Olhede, University College, London  
*Methods of network comparison*
- 15.00 Break
- 15.30 Malgorzata Bogdan, Wroclaw University  
*Sorted L-One Penalized Estimation (SLOPE) for high dimensional statistical inference.*
- 16.30 Stepan Mazur, Örebro University  
*Tangency portfolio weights in small and large dimensions: estimation and test theory.*

The talks will last about 45 minutes, including discussion, followed by a 15 minute break.

Please note that lunch will not be provided during the conference, but a map of KTH with a list of local restaurants can be found by clicking

[https://www.yelp.com/search?cflt=restaurants&find\\_near=kth-stockholm](https://www.yelp.com/search?cflt=restaurants&find_near=kth-stockholm)

Warmly welcome!

For the Cramér Society,

Tatjana Pavlenko  
Rolf Larsson

Speakers, titles and abstracts:

**Mattias Villani, Linköpings Universitet.**

Hamiltonian Monte Carlo with Energy Conserving Subsampling for Big Data Problems

*Hamiltonian Monte Carlo (HMC) has recently received considerable attention in the literature due to its ability to overcome the slow exploration of the parameter space inherent in random walk proposals. While data subsampling has been successful in traditional MCMC algorithms such as Metropolis-Hastings, it has been demonstrated to be unsuccessful in the context of HMC, both in terms of poor sampling efficiency and in producing highly biased inferences. We propose an efficient HMC-within-Gibbs algorithm that utilizes data subsampling to speed up computations. Contrary to previous unsuccessful approaches, we perform subsampling in a way that conserves energy but for a modified Hamiltonian. We can therefore maintain high acceptance rates even for distant proposals. We apply the method for simulating from the posterior distribution of a high-dimensional spline model for bankruptcy data and document speed ups of several orders of magnitude compare to standard HMC and, moreover, demonstrate a negligible bias.*

**Rolf Sundberg, Stockholms Universitet.**

Inference when dimension (much) exceeds sample size.

A 20th century controversy – Routine today?

Abstract: TBA

**Raazesh Sainudiin, Uppsala Universitet**

Seeded by Hate? Characterizing the Twitter Networks of Prominent Politicians and Hate Groups in the 2016 US Election.

*During the 2016 US presidential election, there was considerable debate about the unique appeal of Donald Trump's candidacy to hate groups. Though some vehemently argued for Trump's ability to rile up hateful ideologies, others actively denied such claims. In the present work, we characterize the Twitter networks of both major presidential candidates, Donald Trump and Hillary Clinton, with various American hate groups. We further examined the Twitter networks for Bernie Sanders, Ted Cruz, and Paul Ryan, for 9 weeks around the 2016 election (4 weeks prior to the election and 4 weeks post-election). By carefully accounting for the observed heterogeneity in the Twitter activity levels across individuals, our data of 21,749,868 Tweets revealed that there were significantly more people linked to leaders and followers of American hate groups belonging to six ideologies (Anti-Government, Anti-Immigrant, Anti-LGBT, Anti-Muslim, Alt-Right, Neo-Nazi, and White-Nationalist) concurring with Donald Trump than with the other four politicians. We also found that significantly more individuals in the Twitter network had the fewest degrees of separation simultaneously from Donald Trump and each one of these six hateful ideologies relative to the other four politicians.*

**Sofia Olhede, University College, London**

Methods of network comparison

*The topology of any complex system is key to understanding its structure and function. Fundamentally, algebraic topology guarantees that any system represented by a network can be understood through its closed paths. The length of each path provides a notion of scale, which is vitally important in characterizing dominant modes of system behavior. Here, by combining topology with scale, we prove the existence of universal features which reveal the dominant scales of any network. We use these features to compare several canonical network types in the context of a social media discussion which evolves through the sharing of rumors, leaks and other news. Crucially, our results allow networks to be quantified and compared in a purely model-free way that is theoretically sound, fully automated, and inherently scalable.*

**Malgorzata Bogdan, Wroclaws Universitet.**

Sorted L-One Penalized Estimation (SLOPE) for high dimensional statistical inference.

*Sorted L-One Penalized Estimation (SLOPE) is a novel statistical method for reduction of dimensionality in large data bases. It can be considered as extension of popular LASSO, which apart from inducing sparsity additionally reduces the dimension by clustering (averaging) similar values of model parameters. We will briefly introduce the idea of the method and present its applications in the context of identification of important predictors, constructing gaussian graphical models and portfolio optimization.*

**Stepan Mazur, Örebro Universitet.**

Tangency portfolio weights in small and large dimensions: estimation and test theory.

*In this talk, we consider the weights of the tangency portfolio (TP). It is assumed that the sample size is smaller than the dimension, the underlying covariance matrix is singular, and the vectors of returns are independent and normally distributed. A stochastic representation for the estimated TP weights is derived which is then used to obtain their asymptotic distribution under the high-dimensional asymptotic regime. Furthermore, we suggest several exact tests of general linear hypothesis about the elements of the TP weights. The distributions of the test statistics are derived under the null and the alternative hypothesis. The results are illustrated using actual stock returns and a discussion of practical relevance of the model is presented.*

*(Jointly with Taras Bodnar, Krzysztof Podgorski and Joanna Tyrcha)*