
STATISTICS SEMINAR

UW-Department of Statistics

www.stat.wisc.edu



Abstract: Improvements in biomedical technology and a surge in other data-driven sciences lead to the collection of increasingly large amounts of data. In this affluence of data, contamination is ubiquitous but often neglected, creating substantial risk of spurious scientific discoveries. Especially in applications with high-dimensional data, for instance proteomic biomarker discovery, the impact of contamination on methods for variable selection and estimation can be profound yet difficult to diagnose.

In this talk I present a method for variable selection and estimation in high-dimensional linear regression models, leveraging the elastic-net penalty for complex data structures. The method is capable of harnessing the collected information even in the presence of arbitrary contamination in the response and the predictors. I showcase the method's theoretical and practical advantages, specifically in applications with heavy-tailed errors and limited control over the data. I outline efficient algorithms to tackle computational challenges posed by inherently non-convex objective functions of robust estimators and practical strategies for hyper-parameter selection, ensuring scalability of the method and applicability to a wide range of problems.

TITLE:

**Detecting the Signal
Among Noise and
Contamination in
High Dimensions**

Speaker:

**David
Kepplinger**

PhD Student
University of British
Columbia

Time & Place:

Monday, February

17, 2020 **4pm**,

Room 133 SMI

Cookies & Coffee @

3:30, Rm 1210 MSC

