

**RUTGERS UNIVERSITY  
DEPARTMENT OF STATISTICS AND BIOSTATISTICS  
CENTER FOR INTEGRATIVE PROTEOMICS RESEARCH**

[www.stat.rutgers.edu](http://www.stat.rutgers.edu)

**Seminar**

Speaker: **Yuekai Sun**  
**Institute for Computational and Mathematical Engineering**  
**Stanford University**

Title: **Distributed estimation and inference for sparse regression**

Time: **12:00pm – 1:00pm, Tuesday, February 24, 2015**

Place: **552 Hill Center**

**Abstract**

We address two outstanding challenges in sparse regression: (i) computationally efficient estimation in distributed settings (ii) valid inference for the selected coefficients. The main computational challenge in a distributed setting is harnessing the computational capabilities of all the machines while keeping communication costs low. We devise an approach that requires only a single round of communication among the machines. We show the approach recovers the convergence rate of the (centralized) lasso as long as each machine has access to an adequate number of samples. Turning to the second challenge, we devise an approach to post-selection inference by conditioning on the selected model. In a nutshell, our approach gives inferences with the same frequency interpretation as those given by data/sample splitting, but it is more broadly applicable and more powerful. The validity of our approach also does not depend on the correctness of the selected model; i.e. it gives valid inferences even when the selected model is incorrect.

Joint work with Jason Lee, Qiang Liu, Dennis Sun, Jonathan Taylor

**\*\* Refreshments will be served @11:40am in Room 502 Hill Center**