

**NYU Stern School of Business**  
**Department of Information, Operations & Management Sciences**  
**STATISTICS RESEARCH SEMINAR**

**TOPIC:** Identification of Informative Features for Clustering in High Dimensions

**SPEAKER:** Wen Zhou (Colorado State University)

**DATE:** Friday, March 25<sup>th</sup>, 2016

**TIME:** 11:30 AM - 12:30 PM

**PLACE:** KMC 3-60

**ABSTRACT**

Identifying important features for separating unlabeled observations into homogeneous groups plays a critical role in dimension reduction and modeling data with complex structures. This problem is directly related to selecting informative variables in cluster analysis, where a small fraction of features is identified for separating observed feature vectors  $X_i \in \mathbb{R}^p$ ,  $i = 1, \dots, n$ , into  $K$  possible classes. Utilizing the framework of model-based clustering, we introduce a PAirwise Reciprocal fuSE (PARSE) procedure based on a new class of penalization functions that imposes infinite penalties on features with small differences across clusters, which effectively avoids selecting overly dense number of features for separating observations in cluster analysis. We establish the consistency of the proposed procedure for identifying informative features for cluster analysis. The PARSE procedure is shown to enjoy certain optimality properties as well. We developed a backward selection algorithm, in conjunction with the EM algorithm, to implement PARSE. Simulation studies show that PARSE has competitive performance compared to other popular model-based clustering methods. PARSE is shown to select a sparse set of features and to produce accurate clustering results. We apply PARSE to a microarray experiment on human asthma and discuss the biological implications of the results.

**BIO**

Wen Zhou is an Assistant Professor in the [Department of Statistics](#) at the [Colorado State University](#). He is also a core faculty member of the interdisciplinary program of [GAUSSI](#) at CSU. Before joining CSU, Professor Zhou received my Ph.D. in [Statistics](#) at [Iowa State University](#) in May 2014 under Professor [Stephen Vardeman](#) and Professor [Huaiqing Wu](#)'s advisorships. He also received my Ph.D. in Applied Mathematics at ISU in Aug 2010.