



THE UNIVERSITY OF CHICAGO

Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

CHRISTOPHER MCKENNAN

Department of Statistics
The University of Chicago

Latent Variables in 'omic' Data

FRIDAY, June 7, 2019, at 12:30 PM
Jones 111, 5747 S. Ellis Avenue

ABSTRACT

Nearly all high-through put 'omic' data are influenced by technical and biological factors unknown to the researcher, which, if unaccounted for, can severely obfuscate estimation of and inference on the effects of interest. While the importance of this problem has precipitated the development of many methods that attempt to correct for these latent factors, most are designed for gene expression data and are not amenable for modern, complex experimental designs. In this thesis, we develop novel and provably accurate methodology to estimate and perform inference on the coefficients of interest in a multivariate linear model in the presence of latent covariates. This includes correcting for latent cellular heterogeneity in DNA methylation data, the first methods amenable to data with complex sample correlation structures and the first method to account for latent covariates in mass spectrometry experiments with non-random missing data.