

STATISTICS SEMINAR

Patrick Breheny, Ph.D.

University of Iowa

Thursday, October 9, 2014

Dickens Hall, Room 207, 4:00-5:00 pm

Refreshments: Dickens 108, 3:30-4:00 pm



Estimating False Inclusion Rates in Penalized Regression Models

Abstract: Penalized regression methods are an attractive tool for feature selection with many appealing properties, although their widespread adoption has been hampered by the difficulty of applying inferential tools. In particular, the question "How reliable is the selection of those features?" has proved difficult to address, partially due to the complexity of defining a false discovery in the penalized regression setting. Here we define a false inclusion as a variable that is independent of the outcome regardless of whether other variables are conditioned on. We show that this definition permits straightforward estimation of the number of false inclusions in near-independence conditions. We also develop a permutation-based approach and show that it yields more accurate estimation of the false inclusion rate in highly correlated settings.