COLLOQUIUM
Department of Statistics and Probability
Michigan State University

Qunhua Li
University of California, Berkeley

Measuring Reproductibility of High-Throughput Biological Experiments

Tuesday, February 8, 2011
A405 Wells Hall
10:20 a.m. - 11:10 a.m.
Refreshments: 10:00 a.m.

Abstract

Reproducibility is essential to reliable scientific discovery in large-scale high-throughput biological studies. In this talk, I will present a unified approach to measure reproducibility of findings identified from replicate experiments and select discoveries using reproducibility between replicates.

Unlike the usual scalar measures of reproducibility, our approach views reproducibility as when the findings are no longer consistent across replicates. To measure the pairwise consistency between replicates, we develop a graphical statistic based on empirical copulas and a copula mixture model to quantitatively describe the change of consistency in the decreasing significance of findings. Based on the copula mixture procedure, we define a quantity, called "irreproducible discovery rate", in a fashion analogous to the false discovery rate. This quantity, which describes the lack of reproducibility for the identifications selected at each threshold, provides a reproducibility criterion for selecting reliable signals and assessing the overall reproducibility of findings. Our approach can be applied to both probabilistic- and heuristic-based significance scores, and permits principled setting of selection thresholds.

This method has been adopted by ENCODE consortium for selecting ChIP-seq signal identification algorithms and monitoring the performance of their experimental facility. I will illustrate the effectiveness of our method using some ENCODE examples.