Mathematics, Statistics, and Computer Science **@ UIC**

Statistics and Data Science Seminar

A Super Scalable Algorithm for Short Segment Detection

Yue Niu (University of Arizona)

Abstract: In many applications such as copy number variant (CNV) detection, the goal is to identify short segments on which the observations have different means or medians from the background. Those segments are usually short and hidden in a long sequence, and hence are very challenging to find. We study a super scalable short segment (4S) detection algorithm in this paper. This nonparametric method clusters the locations where the observations exceed a threshold for segment detection. It is computationally efficient and does not rely on Gaussian noise assumption. Moreover, we develop a framework to assign significance levels for detected segments. We demonstrate the advantages of our proposed method by theoretical, simulation, and real data studies.

Wednesday, March 6 at 4:00 PM in 636 SEO