THE UNIVERSITY OF CHICAGO

Department of Statistics

PHD THESIS PRESENTATION

YUNDA ZHONG

Department of Statistics

The University of Chicago

A Large Scale Multiple Testing for Data with Spatial Signals

FRIDAY, February 15, 2013, at 1:30 PM

110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

This thesis consists of three projects. The abstracts for each project are listed below.

1. We consider parameter estimation for stationary processes that exhibit both long-range dependence and heavy tails. Processes with such features exist widely in network traffic, geology, telecommunications, finance and other areas. The long-range dependence parameter and the heavy tail index play a fundamental role in understanding such processes. Consistency of our estimators of both parameters is established under mild conditions. Our approach is illustrated by a simulation study and is applied to a real internet traffic data collected from an email server.

2. We consider large scale multiple testing for clustered signals, where the data exhibit spatial dependence structure. A change point boundary detection procedure is proposed to make use of the spatial information for hypothesis testing. It is assumed that the number of clusters for alternative hypotheses is finite and the clusters are well separated. We show that by exploiting the spatial structure, the precision of a multiple testing procedure can be improved substantially. Simulation studies evidence that the methods perform well with realistic sample sizes and demonstrate the improved detection ability compared with competing methods. The practical utility of the method is illustrated on a DNA copy number variation data.

3. In modern scientific applications statisticians often need to consider a large number of tests at the same time. Most previous researches are based on the implicit assumption that data are homogeneous. However, there are many examples in which this assumption is not valid and heterogeneity of data has to be taken into account. Our goal is to explore asymptotic properties of false discovery proportions and numbers of rejected hypotheses and to develop large-sample theories under a mean-nonstationary framework of hypotheses.
