



Department of Biostatistics

BIOSTATISTICS SEMINAR

Optimal Sparse Signal Identification with Applications in Copy Number Variation Analysis

Jessie Jeng

Postdoctoral Researcher, Department of Biostatistics and Epidemiology
University of Pennsylvania

Abstract

Motivated by DNA copy number variation (CNV) analysis based on high-density single nucleotide polymorphism (SNP) data, we consider two problems arising from the need to identify sparse and short CNV segments in long sequences of genome-wide data. The first problem is to identify the CNVs utilizing a single sample. An efficient likelihood ratio selection (LRS) procedure is developed, and its asymptotic optimality is presented for identifying short and sparse CNVs. The second problem aims to identify recurrent CNVs based on a large number of samples from a population. We propose a proportion adaptive segment selection (PASS) procedure that automatically and optimally adjusts to the unknown proportions of CNV carriers.

In these problems, An innovative statistical framework is introduced for developing optimal procedures for CNV analysis. We study fundamental properties for signal identification by characterizing the detectable and the undetectable regions. Only in the detectable region, it is possible to consistently separate the CNV signals from noise. Such demarcations can provide deep insights for the identification problems and serve as benchmarks for method evaluation. We prove that the LRS and PASS are consistent in the interiors of each of their respective detectable regions, thus, implying asymptotic optimalities of the proposed methods.

The proposed methods are demonstrated with simulations and analysis of a family trio dataset and a Neuroblastoma dataset. The results show that the LRS procedure can yield greater gain in power for detecting short CNVs than some popular CNV identification procedures and PASS significantly improves the power for CNV detection by pooling information from multiple samples and efficiently identifying both rare and common CNVs carried by neuroblastoma patients.

**The Johns Hopkins Bloomberg School of Public Health
Department of Biostatistics, Monday, January 23, 2012
Room W2030, 12:00-1:00pm, School of Public Health, (Refreshments:)**

For disability access information or listening devices, please contact the [Office of Support Services](#) at 410-955-1197 or on the Web at www.jhsph.edu/SupportServices. EO/AA