A Multisample Change-Point Model for DNA Copy Number Analysis

Nancy R. Zhang
Stanford University, Stanford, USA

The DNA copy number of an individual can be viewed as a change-point process along the chromosome, with a "normal" level at 2 and "aberrations" being locations where the copy number deviates from normal. Chromosomal aberrations occur naturally in the human population, and is a common source of genetic variation. High throughput genomic profiling technologies have been developed to measure DNA copy number at a fine scale along the chromosome. Given this data for a sample of individuals from the population, how do we statistically detect locations of shared aberration across individuals?

We discuss the properties of this type of data and propose a mixture model for its analysis, where at each change-point, the sample is composed of a mixture of individuals who have the change and those who do not. We have experimented with several statistics for detection of shared change-points. For some of the statistics, large sample tail approximations for significance evaluation can be derived. We compare the performance of these statistics in the context of DNA copy number detection using replicate samples from the same individual and from parent-child trios.

This is joint work with David Siegmund and Jun Li.

[Nancy R. Zhang, 390 Serra Mall, Stanford University Stanford, CA 94305-4065, USA; nzhang@stanford.edu]