The Screening and Ranking Algorithm for Change-Points Detection in Multiple Samples

Heping Zhang*
Yale University School of Public Health, New Haven, USA, heping.zhang@yale.edu

Chi Song
Yale University School of Public Health, New Haven, USA, chi.song@yale.edu

Xiaoyi Min
Yale University School of Public Health, New Haven, USA, xiaoyi.min@yale.edu

The chromosome copy number variation (CNV) is the deviation of genomic regions from their normal copy number states, which may associate with many human diseases. Current genetic studies usually collect hundreds to thousands of samples to study the association between CNV and diseases. CNVs can be called by detecting the change-points in the means from sequences of measurements with noise. Although multiple samples are of interest, the majority of the available CNV calling methods are single sample based. Only a few multiple sample methods were proposed. They all used scan statistics similar to the circular binary segmentation (CBS) algorithm that is computationally expensive, and were designed toward either common or rare change-points detection. In this paper, we propose a novel multiple sample method by adaptively combining the scan statistic of the screening and ranking algorithm (SaRa), which is computationally efficient and able to detect both common and rare change-points. We prove that asymptotically this method can find the true change-points with certainty and show in theory that multiple sample methods are superior to single sample methods when shared change-points are of interest. Additionally, we give extensive simulation studies and a real data application to examine the performance of our proposed method.

Keywords: change-point detection, multi-sample inference, adaptive Fisher's method