THE SCREENING AND RANKING ALGORITHM FOR CHANGE-POINTS DETECTION IN MULTIPLE SAMPLES

Abstract

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 mean for sequences of array-based intensity measurements. Although multiple samples are of interest, the majority of the available CNV calling methods are single sample based. Only a few multiple sample methods have been proposed using scan statistics that are computationally intensive and 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 is able to detect both common and rare change-points. We prove that asymptotically this method can find the true change-points with almost certainty and show in theory that multiple sample methods are superior to single sample methods when shared change-points are of interest. Additionally, we report extensive simulation studies to examine the performance of our proposed method. Finally, using our proposed method as well as two competing approaches, we attempt to detect CNVs in the data from the Primary Open-Angle Glaucoma Genes and Environment study, and conclude that our method is faster and requires less information while our ability to detect the CNVs is comparable or better. This is a joint work with Chi Song and Xiaoyi Min.