Integrative Analysis of Cancer Genomic Data

Shuangge (Steven) MA
School of Public Health, Yale University

Abstract

In cancer genomic studies, markers identified from the analysis of single datasets often suffer from a lack of reliability because of the small sample sizes. A cost-effective remedy is to pool data from multiple comparable studies and conduct integrative analysis. Integrative analysis of multiple datasets is challenging because of the high dimensionality of markers and, more importantly, because of the heterogeneity among studies. We consider penalized approaches for marker selection in the integrative analysis of multiple datasets. The proposed approaches can effectively identify markers with consistent effects across multiple studies and automatically accommodate the heterogeneity among studies. We establish the asymptotic consistency properties, conduct simulations, and analyze pancreatic and liver cancer studies.