

Somnath Datta, PhD
Professor
Department of Bioinformatics and Biostatistics
School of Public Health and Information Sciences
University of Louisville

Multi-Sample Reweighted U-Statistics that Account for Confounding Covariates

Multi-sample U-statistics encompass a wide class of test statistics that allow the comparison of two or more distributions. U-statistics are especially powerful because they can be applied to both numeric and non-numeric (e.g., textual) data. However, when comparing the distribution of a variable across two or more groups, observed differences may be due to confounding covariates. For example, in a case-control study, the distribution of exposure in cases may differ from that in controls entirely because of variables that are related to both exposure and case status and are distributed differently among case and control participants. We propose to use individually-reweighted data (using the propensity score for prospective data or the stratification score for retrospective data) to construct adjusted U-statistics that can test the equality of distributions across two (or more) groups in the presence of confounding covariates. Asymptotic normality of our adjusted U-statistics is established and a closed form expression of their asymptotic variance is presented. The utility of our procedures is demonstrated through simulation studies as well as an analysis of genetic data.