Ann Inst Stat Math (2014) 66:33–61
DOI 10.1007/s10463-013-0404-2

A U-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and Behrens–Fisher setting

M. Rauf Ahmad

Received: 11 July 2012 / Revised: 12 February 2013 / Published online: 20 April 2013

Abstract A two-sample test statistic is presented for testing the equality of mean vectors when the dimension, p, exceeds the sample sizes, n_i, $i = 1, 2$, and the distributions are not necessarily normal. Under mild assumptions on the traces of the covariance matrices, the statistic is shown to be asymptotically Chi-square distributed when n_i, $p \to \infty$. However, the validity of the test statistic when p is fixed but large, including $p > n_i$, and when the distributions are multivariate normal, is shown as special cases. This two-sample Chi-square approximation helps us establish the validity of Box's approximation for high-dimensional and non-normal data to a two-sample setup, valid even under Behrens–Fisher setting. The limiting Chi-square distribution of the statistic is obtained using the asymptotic theory of degenerate U-statistics, and using a result from classical asymptotic theory, it is further extended to an approximate normal distribution. Both independent and paired-sample cases are considered.

Keywords High-dimensional multivariate inference · Box's approximation · Behrens–Fisher setting · Degenerate U-statistics