A bootstrap test for the k-sample problem

M.D. Jiménez-Gamero[1], *M.V. Alba-Fernández*[2], *P. Jodrá*[3], *I. Barranco-Chamorro*[4]

[1]dolores@us.es, Departamento de Estadística e I.O., Universidad de Sevilla
[2]mvalba@ujaen.es, Departamento de Estadística e I.O., Universidad de Jaén
[3]pjodra@unizar.es, Departamento de Métodos Estadísticos, Universidad de Zaragoza
[4]chamorro@us.es, Departamento de Estadística e I.O., Universidad de Sevilla

Let X and Y be two random vectors taking values in $\mathbb{R}^d$, for some fixed $d \in \mathbb{N}$, with distribution functions F_X and F_Y, respectively. We consider the problem of testing the equality of both distributions from samples of the populations, which is usually known as the two-sample problem. A class of tests for this problem whose test statistic is an L_2 norm of the difference of the empirical characteristic functions of the samples is considered. The null distribution can be estimated by means of bootstrap or permutation procedures. Although very easy to implement, such procedures can become very computationally expensive as the sample size or the dimension of the data increase. This work proposes to approximate the null distribution through a weighted bootstrap. The method is studied both theoretically and numerically. It provides a consistent estimator of the null distribution. The method can easily be extended to the k-sample problem. In the numerical examples carried out, the estimated type I errors are close to the nominal values. The asymptotic properties are similar to those of the bootstrap and permutation estimators but, from a computational point of view, it is more efficient.

Keywords: characteristic function; k-sample problem; weighted bootstrap.