Smooth $k-$sample tests under left truncation

Adrián Lago[1], *Ingrid Van Keilegom*[2], *Juan Carlos Pardo-Fernández*[3], *Jacobo de Uña-Álvarez*[4]

[1]adrian.lago@uvigo.es, Department of Statistics and Operations Research, Universidade de Vigo
[2]ingrid.vankeilegom@kuleuven.be, Department of Decision Sciences and Information Management, KU Leuven
[3]juancp@uvigo.es, Department of Statistics and Operations Research, Universidade de Vigo
[4]jacobo@uvigo.es, Department of Statistics and Operations Research, Universidade de Vigo

Left truncation arises in many different applied fields due to the impossibility of observation of every individual that experiments the event of interest, frequently as a result of the way a study is designed or limitations on the measurement instruments. Truncation causes an observational bias which also induces bias in the estimators of different population quantities, such as the survival function and, as a consequence, on the estimation of the density function. This implies the necessity to adapt the density function estimator for complete data to left-truncated data. Let us now consider k different populations in which the target variable is left truncated. A common applied problem is to determine whether these target variables follow the same distribution in each of the k populations. To address this problem, a test based on a L_2 distance involving the estimator of the density function in every sample and in the pooled sample is proposed. Its asymptotic null distribution is studied and, due to the difficulty to apply it in practice, a bootstrap resampling plan is proposed to approximate the null distribution of the test statistic. As the test is based on the estimation of the density function, the bandwidth plays an important role on its performance. This leads to propose a choice of the smoothing parameter, based on a double bootstrap algorithm, to maximize the power of the test. The performance of the bootstrap and the choice of the bandwidth will be studied through Monte Carlo simulations. The proposed test will be compared to other test in the literature for left-truncated data, such as the Kolmogorov-Smirnov and the log-rank, under different simulation scenarios to determine under which conditions each one is more adequate to be used. A dataset regarding pregnancy times will be employed to exemplify the performance of those three tests, which will all determine that a drug called coumarin does not have an effect on the pregnancy time until an spontaneous abortion.

Keywords: Left truncation, $k-$sample problem, bootstrap.