A density-based test for the $k-$sample problem with left-truncated data

Adrián Lago[1], Ingrid van Keilegom[2], Juan Carlos Pardo-Fernández[3] and Jacobo de Uña-Álvarez[1]

[1]CINBIO, Universidade de Vigo, Department of Statistics and Operations Research, SiDOR research group, Vigo, Spain
[2]Research Centre for Operations Research and Statistics (ORSTAT), KU Leuven, Leuven, Belgium
[3]CITMAga, Universidade de Vigo, Department of Statistics and Operations Research, SiDOR research group, Vigo, Spain

ABSTRACT

Left truncation arises in many applied fields due to the way an experiment is designed or limitations in measurement instruments. It causes observational bias, which yields bias on the estimation. It is a frequent issue to determine whether the target variables from k independent populations follow the same distribution. From the kernel density estimator for left-truncated data in Zhou (1999), a test based on an integral distance between the estimator of the density function in each population and the one of the pooled sample is proposed. The asymptotic distribution is studied and, due to the difficulty of its application in practice, a bootstrap resampling plan is proposed to approximate the null distribution of the test statistic. The choice of the bandwidth will be addressed via Monte Carlo simulations and the proposed test will be compared to the Kolmogorov-Smirnov and the log-rank tests for left-truncated data. The performance of the test will be exemplified with real data regarding pregnancy times.

REFERENCES

Zhou, Y. (1999) Asymptotic representations for kernel density and hazard function estimators with left truncation. Statistica Sinica, 9, 521–533.