XVII Congreso Galego de Estatística e Investigación de Operacións
Ourense, 23 - 25 de outubro de 2025

Kolmogorov–Smirnov and Cramér–von Mises tests for the k-sample problem with left-truncated and right-censored data

Adrián Lago[1], Juan Carlos Pardo-Fernández[1] and Jacobo de Uña-Álvarez[1]

[1]SiDOR Research Group, Department of Statistics and Operations Research, CITMAga, Universidade de Vigo, Campus Universitario Lagoas-Marcosende, Vigo, 36310, Spain

ABSTRACT

Truncation and censoring are two phenomena that one frequently encounters when sampling survival data. They both appear for instance, in astronomy, clinical trials and reliability studies, among many other applied fields. Left truncation occurs when individuals who experience the event of interest before observation cannot be sampled. Under right censoring, the event of interest cannot be observed for all of the individuals in the sample, for instance due to dropout or the end of follow-up. In such cases, some of them are known to experience it beyond a certain time point. Both truncation and censoring induce biased estimators if ignored or not treated properly, thus adequate statistical techniques may be used. The comparison of populations is a classical problem in statistics which, in addition to being of interest in itself, has provided practitioners with many different statistical methods. Despite this, and truncation and censoring frequently appearing when collecting data, the literature on the comparison of populations for left-truncated and right-censored data is rather scarce. The main tests are the so-called rank-based tests, which, despite being optimal under some circumstances, they are not omnibus, meaning they may not correctly detect the alternative hypothesis in some scenarios. Moreover, as discussed previously, the use of techniques that ignore truncation and censoring yield inconsistent tests.

In this work, we adapt the k-sample versions of the Kolmogorov–Smirnov and Cramér–von Mises tests proposed in Kiefer (1959) to data subject to both left truncation and right censoring. The test statistics are based on the difference between the estimators of the distribution functions under left truncation and right censoring (Tsai et al., 1987) and an estimator of the common distribution function under the null hypothesis. The asymptotic null distributions and the consistency of the proposed tests will be established under mild assumptions, which commonly arise in the context of left truncation and right censoring. A bootstrap resampling plan will be proposed to approximate the null distributions of the proposed test statistics. The validity of the method will be proved both theoretically and via simulations. The power of the test will be tackled in a simulation study, where the log-rank test is also included for comparison. The results will expose the weakness of the log-rank test under nonproportional hazards and the good performance of the new proposals in different scenarios. A real dataset regarding unemployment times will also be discussed to show the performance of the new tests in practice. This real application will also show that ignoring truncation and censoring might lead to wrong conclusions.

Keywords: Censoring, Cramér–von Mises, k-Sample problem, Kolmogorov–Smirnov, Truncation

REFERENCES

Kiefer, J. (1959) k-Sample analogues of the Kolmogorov–Smirnov and Cramér–von Mises tests. Annals of Mathematical Statistics, 30, 420-447.

Tsai, W. Y., Jewell, N. P., and Wang, M. C. (1987) A note on the product-limit estimator under right censoring and left truncation. Biometrika, 74, 883-886.