

Trimmings and goodness-of-fit. Some asymptotic results.

P.C. Alvarez-Esteban,^a J.A. Cuesta-Albertos^b and C. Matrán.^a

^a Departamento de Estadística e Investigación Operativa
Facultad de Ciencias, Universidad de Valladolid

^b Departamento de Matemáticas, Estadística y Computación
Facultad de Ciencias, Universidad de Cantabria

Abstract

Impartial trimming methods as developed in [1] and [2] can be considered not only as a way to robustify statistical procedures but also as a method to discard a part of the data to achieve the best possible fit between a sample and a theoretical distribution or between two given samples. This point of view is adopted in this work with respect to the Wasserstein distance between probability measures.

Trimming methods in the goodness of fit framework were introduced in [3]. In this paper, the authors use the classical trimming procedure based on deleting the same proportion of observations in the tails of the distribution. However the procedure inherits the problems associated to the arbitrariness in the choice of the zones to discard the data. This becomes even more apparent when we want to relate or compare two data sets obtained from the same distribution through two different types of contamination. Then the trimming procedure should be based on the data, discarding from each data set those points which highly contribute to the dissimilarity of the distributions. One approach to these ideas could be obtained if we do not fix the proportion to be trimmed in each tail of the distribution. Some asymptotic results derived from this way of trimming will be presented.

References

- [1] Cuesta-Albertos, J.A.; Gordaliza, A. and Matrán, C. (1997), "Trimmed k -Means: An Attempt to Robustify Quantizers," *The Annals of Statistics*, 25, 553-576.
- [2] García-Escudero, L.A.; Gordaliza, A. and Matrán, C. (1999) "A Central Limit Theorem for Multivariate Generalized Trimmed k -means" *The Annals of Statistics*, 27, 1061-1079.
- [3] Munk, A.; Czado, C. (1998) "Nonparametric Validation of Similar Distributions and Assessment of Goodness of Fit", *J.R. Statist. Soc. B*, 60, 223-241.