

Smoothing estimation of the functional ROC Curve

*Graciela Estévez-Pérez*¹

¹graciela.estevez.perez@udc.es, Departamento de Matemáticas, Universidade da Coruña

Technical development over the last few decades has resulted in the emergence of complex data, in many cases functional data (FD). This type of data can emerge in many medical studies which are geared towards detecting diseases, predicting their course or evaluating the response to a therapy, to name a few. Thus, it is very useful to have statistical methods enabling us to evaluate diagnostic tests based on functional biomarkers. Estévez and Vieu (2021) developed recently a diagnostic test that use functional variables as biomarkers and proposed an empirical estimate of the functional ROC curve. In order to improve this methodology, the present paper proposes a procedure to obtain a smooth version of nonparametric estimator of ROC curve.

In addition, a comprehensive simulation study was carried out to investigate the discriminatory and predictive abilities of the new functional diagnostic tests and the effect of the bandwidth parameter over them. Special attention was paid to the improvement achieved using the kernel estimators versus the empirical one. The empirical study was undertaken to examine the impact of estimation method of ROC curve in the functional context and their interaction with the other parameters involved in diagnostic test: type of representative curves and semi-metric for the projection. This effect is measured by on the discriminatory power of the test, through numerical indexes like the *AUC*, and on its predictive ability through misclassification rates.

Finally, to illustrate the methodology of functional diagnostic test and to complete the comparison among kernel estimation methods and the empirical ROC curve, the analysis of two real medical data sets has been carried out: one deals with gene expression levels for tumoral/normal samples of prostate cancer (data published by Singh et al. (2002) and included in the R package `depthTools`); the other dataset is about white matter structures in the brain in multiple sclerosis patients (dataset included in the R package `refund`, (Goldsmith et al., 2016)).

Keywords: Functional biomarker; ROC curve; Kernel smoothing.

REFERENCES

- Estévez-Pérez, G. & Vieu, P. (2021). A new way for ranking functional data with applications in diagnostic test. *Computational Statistics*, 36(1), 127-154.
- Goldsmith, J., Huang, L., Scheipl, F., Gellar, J., Harezlak, J., McLean, M. W., ... & Reiss, P. (2016). `refund`: Regression with Functional Data. R package version 0.1-14
- Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., ... & Sellers, W. R. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer cell*, 1(2), 203-209.