Confidence intervals for the length of the ROC curve based on a smooth estimator

Pablo Martinez-Camblor[1],

[1]Pablo.Martinez-Camblor@Hitchcock.org, Department of Anaesthesiology and Biomedical Data Sciences Department, Geisel School of Medicine at Dartmouth

A good diagnostic test should show different behaviour on both the positive and the negative populations. However, this is not enough for having a good classification system. The binary classification problem is a complex task, which implies to define decision criteria. The knowledge of the level of dissimilarity between the two involved distributions is not enough. We also have to know how to define those decision criteria. The length of the receiver-operating characteristic, ROC, curve has been proposed as an index of the optimal discriminatory capacity of a biomarker. It is related not with the actual but with the optimal classification capacity of the considered diagnostic test. One particularity of this index is that its estimation should be based on parametric or smoothed models. We explore here the behaviour of a kernel density estimator-based approximation for estimating the length of the ROC curve. We prove the asymptotic distribution of the resulting statistic, propose a parametric bootstrap algorithm for confidence intervals construction, discuss the role that the bandwidth parameter plays in the quality of the provided estimations and, via Monte Carlo simulations, study its finite-sample behaviour considering four different criteria for the bandwidth selection. The practical use of the length of the ROC curve is illustrated through two real-world examples.

Keywords: Asymptotic distribution; Binary classification problem; Length of the ROC curve.