Regression model for surrogate data in high dimensional statistics and the PaO_2 prediction

Abstract. This work deals with the problem of estimating the regression of a surrogated scalar response variable given a functional random one.
We construct an estimator of the regression operator by using, in addition to the available (true) response data, a surrogate data. We then establish some asymptotic properties of the constructed estimator in terms of the almost-complete and the quadratic mean convergences. Notice that the obtained results generalize a part of the results obtained in the finite dimensional framework. An illustration on the applicability of our results on both simulated data and a real dataset was realized. We have thus shown the superiority of our estimator on classical estimators when we are lacking complete data. An application is finally realized for the prediction of the PaO2 given the thoracic volume.

Keywords and phrases*:* Functional data analysis (FDA) ; Small ball probability ; Surrogate data ; Mean square convergence ; Almost-complete convergence ; Regression estimation; Classical kernel estimator ; k nearest neighbors (kNN).