A SIMPLE GOODNESS-OF-FIT TEST FOR LINEAR MODELS UNDER A RANDOM DESIGN ASSUMPTION

HOLGER DETTE AND AXEL MUNK

Fakultät für Mathematik, Ruhr-Universität Bochum,
Universitätstr. 150, D-44780 Bochum, Germany

(Received February 6, 1997; revised June 9, 1997)

Abstract. Let (X,Y) denote a random vector with decomposition $Y = f(X) + \varepsilon$ where $f(x) = E[Y \mid X = x]$ is the regression of Y on X. In this paper we propose a test for the hypothesis that f is a linear combination of given linearly independent regression functions $g_1, \ldots, g_d$. The test is based on an estimator of the minimal L^2-distance between f and the subspace spanned by the regression functions. More precisely, the method is based on the estimation of certain integrals of the regression function and therefore does not require an explicit estimation of the regression. For this reason the test proposed in this paper does not depend on the subjective choice of a smoothing parameter. Differences between the problem of regression diagnostics in the nonrandom and random design case are also discussed.

Key words and phrases: Nonparametric regression check, validation of goodness of fit, L^2-distance, equivalence of regression functions, random design.