Change-point Problems
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

A NONPARAMETRIC TEST FOR HOMOGENEITY: APPLICATIONS TO PARAMETER ESTIMATION

BY K. GHOUDI AND D. MCDONALD
Université Laval and University of Ottawa

Testing for homogeneity has many applications in statistical analysis. For example, regression analysis may be viewed as determining the set of parameters that makes the residuals homogeneous. Assume for each $i = 1, \cdots, q$ a small sample $n(i)$ observations is collected. Let $N = \sum_{i=1}^{q} n(i)$, let F^i be the empirical distribution function of the ith sample and let the empirical distribution function of all the samples taken together be $\bar{F}$. The problem is to test if these samples are homogeneous. Lehmann (1951) considered the problem of testing the equality of the distributions of q samples. He proposed the statistic

$$\sum_{i=1}^{q} \int (F^i(s) - \bar{F}(s))^2 d\bar{F}(s).$$

The asymptotic properties of Crámer-Von Mises statistics like the above were studied by Kiefer (1959). He considered the case where $n(i) \to \infty$ while q stayed fixed. McDonald (1991) considered the situation where $q \to \infty$ and $n(i)$ stays fixed for univariate observations for a more general family of statistics called *randomness statistics*. In case of multivariate observations similar asymptotics are discussed in Ghoudi (1992).

Here we present an application of the above statistics to the estimation of the parameters of regression models with independent additive errors. The main novelty of our approach is the use of blocking to contrast the empirical distribution of the residuals of observations whose independent variables are in the same block with the empirical distribution of all the residuals taken together. Our estimated regression surface is the one whose residuals minimize a randomness statistic like Lehmann's. Confidence intervals and a test for the model follow without assumptions on the error distribution.

Introduction. Consider a multivariate linear regression model. Observations of a dependent vector $\mathbf{y} = (y_1, y_2, \cdots, y_d)$ and independent variables $\mathbf{x} = (x_1, x_2, \cdots, x_p)$ are indexed by $t = 1, \cdots, N$ and are governed by the model

$$\mathbf{y}_t = (\mathbf{x}_t)'(\beta) + \epsilon_t$$

AMS 1991 Subject Classification: Primary 62G10; Secondary 62J02.

Key words and phrases: Homogeneity tests, nonparametric regression.