On the testing hypothesis of equality distribution densities

Babilua P., Nadaraya E., Sokhadze G.

Ivane Javakhishvili Tbilisi State University, Faculty of Exact and Natural Sciences, Department of Mathematics, 3 Chavtchavadze Ave, Tbilisi 0179, Georgia

Let $X^{(i)} = \left(X_1^{(i)}, \dots, X_{n_i}^{(i)}\right)$, $i = 1, \dots, p$ independent samples with sizes $n_1, \dots, n_p$, $p \geq 2$ with densities $f_1(x), \dots, f_p(x)$ it is required based on $X^{(i)}$, $i = 1, \dots, p$ samples testing two hypotheses: homogeneity hypothesis

$$H_o: \ f_1(x) = \cdots = f_p(x)$$

and goodness-of-fit test

$$H_o': \ f_1(x) = \cdots = f_p(x) = f_o(x)$$

where $f_o(x)$ is known density function. In case of H_o $f_o(x)$ is unknown.

In this investigation we will construct criteria for testing hypotheses H_o and H_o' against sequence of "close" alternative ([1],[2]):

$$H_1: \ f_i(x) = f_0(x) + \alpha(n_0)\varphi_i\left(\frac{x - l_i}{\gamma(n_0)}\right) + o\big(\alpha(n_0)\gamma(n_0)\big)$$

$$(\alpha(n_0), \gamma(n_0) \to 0),$$

$$\int \varphi_i\,(x)dx = 0, \ \ n_0 = \min\big(n_1, \dots, n_p\big) \to \infty.$$

We will consider criteria for testing hypotheses H_o and H_o' based on statistics

$$T(n_1, \dots, n_p) = \sum_{i=1}^{p} N_i \int \left[\hat{f}_i(x) - \frac{1}{N}\sum_{j=1}^{p} N_j\,\hat{f}_j(x)\right]^2 r(x)dx$$

where $\hat{f}_i(x)$ is kernel estimator of Rosenblatt-Parzen of density $f_i(x)$:

$$\hat{f}_i(x) = \frac{a_i}{n_i}\sum_{j=1}^{n_i} K\left(a_i\left(x - X_j^{(i)}\right)\right), \qquad N_i = \frac{a_i}{n_i}, \qquad N = N_1 + N_2 + \cdots N_p.$$

References:

1. Rosenblatt, M. A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** (1975), 1--14.
2. Nadaraya, È. A. Nonparametric estimation of probability densities and regression curves. Translated from the Russian by Samuel Kotz. Mathematics and its Applications (Soviet Series), 20. *Kluwer Academic Publishers Group, Dordrecht,* 1989.