Application of Nonparametric Goodness-of-Fit Tests: Problems and Solution

B. Yu. Lemeshko

Applied Mathematic Department, Novosibirsk State Technical University, Novosibirsk, Russia

E-mail: Lemeshko@ami.nstu.ru
www.ami.nstu.ru/~headrd/

In this paper, the problems of application of nonparametric goodness-of-fit tests in the case of composite hypotheses have been considered. The factors influencing test statistic distributions have been discussed. We have prepared a manual on application of nonparametric tests. Proposed recommendations would reduce errors in statistical inference when using considered tests in practice.

Keywords: Composite hypotheses of goodness-of-fit; Anderson-Darling test, Cramer–Von Mises–Smirnov test, Kolmogorov test, Kuiper test, Watson test, Zhang tests.

Introduction

In applications of statistical data analysis, there are a lot of examples of incorrect usage of nonparametric goodness-of-fit tests (Kolmogorov [1], Cramer–von Mises–Smirnov [2], Anderson–Darling [3,4], Kuiper [5], Watson [6,7], Zhang [8] tests). The most common errors in testing composite hypotheses are associated with using classical results obtained for simple hypotheses.

There are simple and composite goodness-of-fit hypotheses. Simple hypothesis tested has following form \( H_0 : F(x) = F(x, \theta) \), where \( F(x, \theta) \) is the distribution function, which is tested for goodness-of-fit with observed sample, and \( \theta \) is known value of parameter (scalar or vector). Composite hypotheses tested has form \( H_0 : F(x, \theta) \in \{ F(x, \theta), \theta \in \Theta \} \), where \( \Theta \) is definition domain of parameter \( \theta \). If estimate \( \hat{\theta} \) of scalar or vector parameter of tested distribution was not found by using the sample, for which goodness-of-fit hypothesis is tested, then application of goodness-of-fit test for composite hypothesis is similar to application the test in the case of testing simple hypothesis.

The problems arise in testing composite hypothesis, when estimate \( \hat{\theta} \) of distribution parameter was found by using the same sample by which goodness-of-fit hypothesis is tested.

Problems of application of tests for composite hypotheses

In the case of testing composite hypotheses, all nonparametric goodness-of-fit tests lose their property of being distribution free, if parameters estimation is based on the same sample, by which the hypothesis is tested. Statistic distributions \( G(S[H_0]) \) of these tests depend on:

- distribution \( F(x, \theta) \) corresponding to tested hypothesis \( H_0 \);
- the type and the number of parameters estimated;
- the estimation method used;
- in some cases, the particular value of parameter (for example, in the case of gamma-distribution);

The statistic distributions for simple hypotheses and the distributions of the same statistics for composite hypotheses are quite different. Therefore, it is unacceptable to disregard this difference.

Moreover, the greatest problem consists in the dependence of test statistic distributions on specific value of distribution shape parameter.

Application of tests for composite hypotheses: solution of problems

The investigation of limiting statistic distributions of nonparametric goodness-of-fit tests for composite hypotheses was initiated in Ref. [9].

Various approaches has been used for solving problems in this area: limiting statistic distributions have been studied by analytical and numerical methods. In particular cases, the tables of percentage points for limiting statistic distributions of nonparametric tests have been obtained by using statistical simulation methods.

Apparently, the first papers, in which Monte-Carlo method and computer simulation appeared to be efficient method of development of applied mathematical statistic, were Refs. [10]. In these papers, percentage points for the Kolmogorov test statistic (without Bolshev’s correction) were obtained for testing composite hypotheses relative to normal distribution law.

---

1 This work is supported by the Russian Ministry of Education and Science (project № 2.541.2014K)
In a number of our papers, the analytically simple models, approximating the limiting statistic distributions of nonparametric tests in the case of testing composite hypothesis and using maximum likelihood estimation method, have been constructed by using computer simulation of statistic distributions relative to various distributions corresponding to hypothesis $H_0$. Recommendations for standardization P 50.1.037-2002 have been published on the base of these studies (Ref. [11]). Later, results presented in Ref. [11] have been clarified and extended in Refs. [12-17]. At present, the manual (Ref. [18]) based on obtained results has been prepared and intended to replace recommendations in Ref. [11].

The manual Ref. [18] includes the tables of percentage points and models of limiting statistic distributions of nonparametric tests (altogether 63 tables), which can be used for testing various composite hypotheses. Moreover, the manual includes description of computer simulation techniques for research of probabilistic regularities, which can be used for investigation of test statistic distributions.

In the cases, when statistic distributions of nonparametric tests depend on specific values of shape parameter(s) of tested distribution, the statistic distribution cannot be found in advance (before computing corresponding estimates). In such situations, it is recommended to find test statistic distribution by using interactive mode in statistical analysis process, see Ref. [19], and then, to use this distribution for testing composite hypothesis.

References