Computers and Mathematics with Applications 56 (2008) 1617–1625

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

A power comparison and simulation study of goodness-of-fit tests

Barış Sürücü

Department of Statistics, Middle East Technical University, 06531 Ankara, Turkey

ARTICLE INFO

Article history:
Received 7 February 2007
Received in revised form 25 February 2008
Accepted 5 March 2008

Keywords:
Combined statistic
Goodness-of-fit
Normal
Simulation
Spacings

ABSTRACT

We give the results of a comprehensive simulation study of the power properties of prominent goodness-of-fit tests. For testing the normal $N(\mu, \sigma^2)$, we propose a new omnibus goodness-of-fit statistic C which is a combination of the Shapiro–Wilk statistic W and the correlation statistic R. We show that the test of normality based on C is overall more powerful than other prominent goodness-of-fit tests and is effective against both symmetric as well as skew alternatives. We also show that the null distribution of C can be approximated by a four-moment F. For the exponential $E(\theta, \sigma)$, Tiku statistic Z (using sample spacings) and modified Anderson–Darling A are the most powerful. For testing other distributions, the statistics based on generalized sample spacings and the modified Anderson–Darling statistic provide the most powerful tests.

1. Introduction

A parametric procedure usually hinges on the assumption of a particular distribution. It is, therefore, of utmost importance to assess the validity of the assumed distribution. This is accomplished by doing a goodness-of-fit test. A galaxy of omnibus goodness-of-fit tests are available; see, for example, [1] and the references in [2, Chapters 1–7]. In this paper, we report the simulation results for the tests we found to be overall most powerful: (i) the combined statistic C for testing normal $N(\mu, \sigma^2)$, (ii) the statistics Z (using exponential sample spacings) and the modified Anderson–Darling statistic for testing exponential $E(\theta, \sigma)$, (iii) the statistic Z^* (using generalized sample spacings) for testing a skew distribution, and for testing a symmetric distribution against skew alternatives, (iv) the modified Anderson–Darling statistic and the correlation statistic R for testing uniform $(\theta_1,\ \theta_2)$. A four-moment approximation is available to the null distribution of C. The null distributions of Z^* and U are effectively normal for all sample sizes $n \geq 7$ and their power functions can also be derived analytically. The tests based on the EDF statistics are generally difficult to implement since they involve parameter estimation.

This work can be good material for software developers because it identifies the most powerful goodness-of-fit statistics and gives their approximate null distributions. The latter can be used for calculating p-values.

2. Testing for normality

Consider normal density $N(\mu, \sigma^2)$, μ and σ being unknown. A large number of statistics are mentioned in [1] and [2] but we will consider only the prominent ones which are convenient from a computational point of view.

Shapiro–Wilk statistic: Let

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)} \tag{2.1}$$

be the order statistics of a random sample $y_1, y_2, \ldots, y_n$. The Shapiro–Wilk statistic is

$$W = \left(\sum_{i=1}^{n} a_i y_{(i)}\right)^2 \Big/ \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad (0 < W < \infty), \quad \sum_{i=1}^{n} a_i^2 = 1. \tag{2.2}$$

E-mail address: sbaris@metu.edu.tr.

0898-1221/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2008.03.010

The values of the coefficients a_i $(1 \le i \le n)$, and the percentage points of the null distribution of W, are given in [3, Tables 15 and 16]. Shapiro and Wilk [4] give a part-theoretical and part-empirical method of deriving the null distribution of W.

Correlation statistic: The correlation statistic R is due to Filliben [5] and Smith and Bain [6], and it can be used for testing any assumed density of the type $(1/\sigma)f_0((y-\mu)/\sigma)$. Let

$$t_{(i)} = F_0^{-1}\left(i/(n+1)\right), \quad 1 \le i \le n, \tag{2.3}$$

be the quantiles of the density f_0, $F_0(z) = \int_{-\infty}^{z} f_0(z)\mathrm{d}z$ being the cumulative density function. The correlation statistic simply is

$$R = 1 - \hat{\rho}^2 \quad (0 < R < 1), \tag{2.4}$$

where $\hat{\rho}$ is the ordinary correlation coefficient between $y_{(i)}$ and $t_{(i)}$ $(1 \le i \le n)$. IMSL subroutines are available for determining $t_{(i)}$ for numerous distributions. The null distribution of R is not known for any density f_0. Its percentage points have to be determined empirically by Monte Carlo simulation.

Combined statistic: There is no omnibus test of normality which is the most powerful against all kinds of alternatives. Therefore, linear combinations of directional tests are mostly used instead of omnibus tests to achieve higher power. However, these tests are powerful only against a specific class of alternatives. In this section, we propose a new test statistic C which is a linear combination of two prominent statistics, W and R; W is overall the most powerful against skew and short-tailed symmetric alternatives and R is overall the most powerful against long-tailed symmetric alternatives. In spite of the fact that W and R are correlated with one another, a good improvement from the perspective of power is achieved with the new statistic. The new statistic C is a weighted sum of W and R with positive weights adding to 1:

$$C = 1 - \{[1 + a_1(a_2 - 1)]\, W + a_1(1 - a_2)(1 - R)\}, \tag{2.5}$$

where $0 < a_1, a_2 < 1$; C is location and scale invariant. Large values of C lead to the rejection of normality. The coefficients a_1 and a_2 are calculated from the equations

$$a_1 = \exp(-(b_1/0.6)^5), \qquad a_2 = \exp(-(b_2/3.5)^5). \tag{2.6}$$

$\sqrt{b_1}$ and b_2 being the sample skewness and kurtosis, respectively. For b_1 less than 0.6 and b_2 not more than 3.5, W automatically receives a dominant weight. The coefficients a_1 and a_2 were determined empirically to achieve the highest power overall. In essence, C is obtained by using the preliminary information inherent in sample skewness and sample kurtosis. The asymptotic null distribution of C can be obtained by a four-moment F approximation.

Null distribution of C: Asymptotically, $\sqrt{b_1}$ and b_2 converge to their expected values which makes the C statistic a linear combination of W and R. Since the distributions of W and R are not known, we are unable to find the null distribution of C (even asymptotically). However, we are able to obtain a four-moment F approximation to the null distribution of C as follows:

Let a positive random variable X have skewness $\sqrt{\beta_1} = \mu_3/\mu_2^{3/2}$ and kurtosis $\beta_2 = \mu_4/\mu_2^2$. The F-region in (β_1, β_2)- plane is bounded by the χ^2 line and the reciprocal of the χ^2 line [7]. That is, (β_1, β_2) satisfy the conditions

$$\beta_1 > \frac{32(v_2-4)}{(v_2-6)^2} \quad \text{and} \quad \beta_2 > 3 + 1.5\beta_1. \tag{2.7}$$

If the (β_1, β_2) values of X lie within the F-region, one can successfully use the four-moment F approximation to the distribution of X as follows [8].

Write

$$F_{v_1,v_2} = \frac{C+f}{h}, \quad v_1, v_2 > 0, \tag{2.8}$$

where F has a central-F distribution with v_1 and v_2 degrees of freedom, and the coefficients f and h are determined by equating the first four moments on both sides of (2.8). That gives

$$h = \sqrt{\mu_2 \frac{v_1(v_2-2)^2(v_2-4)}{2v_2^2(v_1+v_2-2)}}, \qquad f = \frac{v_2}{v_2-2}h - \mu_1', \quad v_2 = 2\left[3 + \frac{\beta_2+3}{\beta_2-(3+1.5\beta_1)}\right]$$

and

$$v_1 = \frac{1}{2}(v_2-2)\left(-1 + \sqrt{1 + \frac{32(v_2-4)/(v_2-6)^2}{\beta_1 - 32(v_2-4)/(v_2-6)^2}}\right), \tag{2.9}$$

where μ_1' and μ_2 are the mean and the variance of C, respectively. Thus, the distribution of $\frac{C+f}{h}$ is approximated by central F with v_1 and v_2 degrees of freedom.

We noticed that the simulated values (β_1, β_2) of C satisfy the condition (2.7). Therefore, the percentage points of C can be obtained from (2.8) and (2.9). Table 1 shows the percentage points of C so obtained and the corresponding type I errors.