INVESTIGATION OF THE STABILITY OF STATISTICAL HYPOTHESES TESTING PROCEDURES USED IN QUALITY MANAGEMENT PROBLEMS[1]

B.Yu. Lemeshko, S.B. Lemeshko, S.S. Pomadin, E.P. Mirkin
Novosibirsk State Technical University
Novosibirsk, Russia
e-mail: `headrd@fpm.ami.nstu.ru`

Abstract

Statistic distributions and the power of normality tests have been investigated by statistical modeling methods. Advantages and disadvantages of different criteria have been shown. The considered criteria have been compared by power with goodness-of-fit criteria for testing normality. It has been shown that the best normality criterion is the D'Agostino test, which wasn't included to the standard ISO 5479-97.

The classical statistic distributions used for testing hypotheses on mathematical expectations and variances have been investigated by means of statistical simulation methods. It has been shown that for testing a hypothesis on mathematical expectations the classical results are valid even when an observed law essentially differs from a normal distribution. The tables of percentage points which can be used for correct testing of hypotheses about variances have been obtained when observed laws are described with the exponential distribution family.

The Bartlett, Cochran and F- test statistics have been investigated for different observed distribution laws.

1 Investigation of the normality tests

Normality testing is an obligatory procedure during measuring, monitoring and testing. The international standard ISO 5479-97 "Statistical methods. Test for departure of the probability distribution from the normal distribution" doesn't answer such questions as: which of the criteria is preferable, which of them is the most powerfull and against what alternatives, for what sample sizes some certain test has an advantage or disadvantage?

The distributions of the following test statistics, included into the standard ISO 5479-97 have been investigated at the paper by means of statistic simulation technique. These are the test on symmetric property, the test on kurtosis, Shapiro-Wilk, Epps-Pulley and the modified Shapiro-Wilk tests. The test power against close alternative hypotheses has been investigated; merits and demerits of the criteria have been shown; the comparison with goodness-of-fit tests [1,2] by power has been carried out. A number of criteria which were not included to the standard, have been investigated. It has been

[1]The research is supported by the Ministry of Education of the Russian Federation (projects No. T02-3.3-3356 and No. A03-2.8-280)

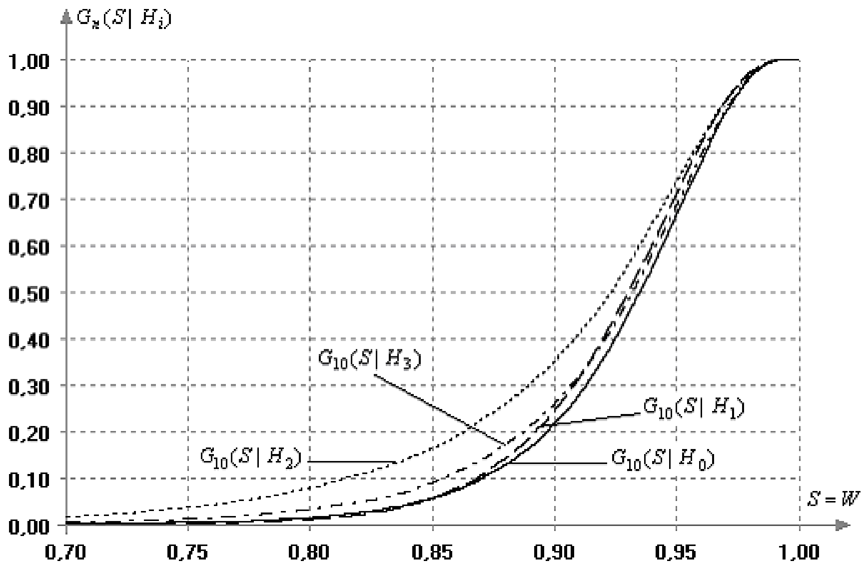

Figure 1: The conditional distributions $G(S|H_i)$ of the Shapiro-Wilk test statistic for the sample size $n = 10$

shown that the Shapiro-Wilk, Epps-Pulley and some other tests have the following disadvantage. These tests have a low power with respect to the flat-topped distributions in comparison with the normal distribution (they cannot distinguish such distributions). For example, in figure 1 the conditional distributions $G(S|H_i)$ of the Shapiro-Wilk test statistic are represented for the sample size $n = 10$, when the H_0, H_1, H_2, H_3 hypotheses are true. The exponential family of distributions with the density function

$$f(x) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \exp\left\{-\left(\frac{|x-\theta_0|}{\theta_1}\right)^{\theta_2}\right\}, \tag{1}$$

has been considered as H_0, H_1, H_2 hypotheses. The normal distribution with $\theta_1 = 1$, $\theta_0 = 0$, $\theta_2 = 1$ corresponds to H_0; the law with parameters $\theta_1 = 1, \theta_0 = 0, \theta_2 = 4$ to H_1 and H_2 hypothesis is corresponded to the Laplace distribution with $\theta_1 = 1, \theta_0 = 0, \theta_2 = 1$; the logistic distribution with the density function

$$f(x) = \frac{\pi}{\theta_1\sqrt{3}} \exp\left\{-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}}\right\} \Big/ \left[1 + \exp\left\{-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}}\right\}\right]^2, \tag{2}$$

and parameters $\theta_1 = 1, \theta_0 = 0$ corresponds to H_3 hypothesis.

It has been shown that distributions of the modified Shapiro-Wilk test statistic, included to the standard ISO 5479-97, converge to the limiting law very poorly. So using of this test results in increasing of alpha error probability.

The test proposed by D'Agostino [3], which wasn't included to ISO 5479-97, has been shown to be the most preferable normality test (by the power). This criterion doesn't have the mentioned disadvantage.

2 The investigation of the criteria for testing hypotheses on mathematical expectations and variances

The hypothesis of mathematical expectations value $H_0 : \mu = \mu_0$ (with known or unknown variance) or the hypothesis of variances value $H_0 : \sigma^2 = \sigma_0^2$ (with known or unknown mathematical expectation) are often needed of testing in applications.

The numerical investigations of classical statistic distributions used for testing hypotheses on mathematical expectations have confirmed their high stability to the deviation of the observed distribution from the normal law. The empirical distributions of corresponding statistics are in a good agreement with the limiting laws obtained on the basis of the assumption of the observed distribution normality. This enables using classical results correctly in practice, when the laws under observation essentially differ from the normal distribution. Obtained results underscore the common regularity: the tests concerning hypotheses of mathematical expectations are robust to deviations of the observations from the normal distribution. This was shown in [4] during investigation of statistics used for testing hypotheses on mathematical expectation vector and correlation coefficient for multidimetional distribution laws.

On the other hand, the distributions of statistics, which concern testing of the hypotheses $H_0 : \sigma^2 = \sigma_0^2$, essentially depend on the observed law. If observed distribution is significantly different from the normal law, then the classical results are not valid as using of them will inevitably result in incorrect conclusions.

The distributions of statistics, used in criteria of testing hypotheses $H_0 : \sigma^2 = \sigma_0^2$, have been investigated with different distributions under observation. The tables of percentage points have been constructed for given test statistics. The tables are valid for the observed distribution laws, described by the family (1).

3 The investigation of Bartlett, Cochran and F-tests robustness

The Bartlett [5] and Cochran [6] tests are intended for testing of the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2$. The limiting distribution law of the Bartlett statistic for normal random variables is the χ^2_{m-1}-distribution [6]. The conditional distributions $G\left(\chi^2|H_0\right)$ of the Bartlett statistic have been numerically shown to be strongly dependent on the distribution under observation (figure 2). The Cochran statistic distributions have been investigated in a similar way. As a result of investigations the percentage points for the Bartlett and Cochran statistics have been made in case when observed variables submit the law (1) with different values of θ_2.

The investigations of F-test statistic distributions used for testing of the hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_m$ have shown that the observed data distribution practically doesnt influence the statistic distributions.

The more detailed information on the investigation results is represented on the

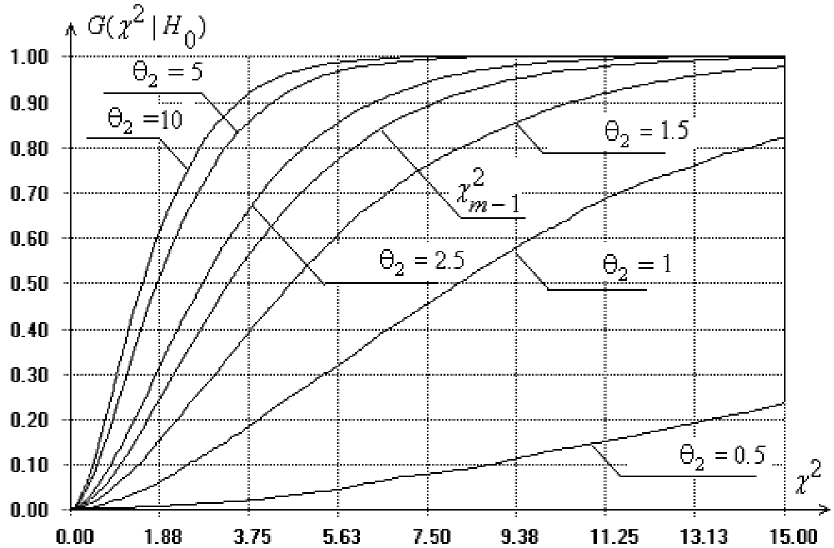

Figure 2: The Bartlett statistic distributions for $m = 5$ when observations submitting the law (1) with different values of θ_2

website `http://www.ami.nstu.ru/~headrd/seminar/Kontrol_Q/krit_zad.htm`.

References

[1] R 50.1.033-2001. (2002). *Recommendations for standardization. Applied statistics. Rules of check of experimental and theoretical distribution of the consent. Part I. Goodness-of-fit tests of a type chi-square*. Publishing house of the standards, Moscow. (in Russian)

[2] R 50.1.037-2002. (2002). *Recommendations for standardization. Applied statistics. Rules of check of experimental and theoretical distribution of the consent. Part II. Nonparametric goodness-of-fit test*. Publishing house of the standards, Moscow. (in Russian)

[3] D'Agostino R.B. (1970). Transformation to normality of the null distribution of g_1. *Biometrika*. Vol. **57**, pp. 679-681.

[4] Lemeshko B.Yu., Pomadin S.S. (2002). The correlation analysis of multidimensional random variables observations in the failure of normality assumption. *Sibirskiy journal industrialnoi mathematiki*. Vol. **5**, N3, pp. 115-130. (in Russian)

[5] Bartlett M.S. (1937). Properties of sufficiency of statistical tests. *Proc. Roy. Soc.* A160, pp. 268-282.

[6] Cochran W.G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. of Eugenics*. Vol. **11**, pp. 47-52.