INVESTIGATION OF SOME CRITERIA FOR DETECTION OF VARIANCE SHIFT AND NONRANDOMNESS IN A SEQUENCE OF OBSERVATIONS UNDER VARIOUS DISTRIBUTION LAWS AND SAMPLE SIZES

V. Volkova, A. Tronina, T. Pogorelova
Novosibirsk State Technical University
Novosibirsk, Russia
e-mail: `vik-ky@yandex.ru`

Abstract

The distribution of the Hsu test statistic has been investigated in case when distributions of observed random variables differ from the normal law by methods of statistical simulation. The limiting statistic distributions have been approximated for a number observation distribution laws. The investigation of Bartels and Wald-Wolfowitz test statistic distributions has been carried out in the case of the limited sample sizes.

1 Introduction

A wide range of statistical methods are based on the assumption of observed random variables normality. Under real conditions normality and often some other assumptions are not satisfied. The use of classical methods of mathematical statistics in such situations can turn out to be incorrect.

Many classical results have an asymptotical nature. At the same time in practice one usually works with samples of a limited size. The application of asymptotical results is not always valid for limited sample sizes.

As a rule revealing fundamental statistic regularities in nonstandard conditions is a complicated problem for researchers. The best way out is to use the numerical approach that is computer modeling of statistical regularities under conditions which simulate some real situations of measurement taking.

2 The Hsu test investigation

One of the main assumptions which should be taken into account while constructing the classical tests for homogeneity of variances is the observed random variables normality. The Hsu [1] test is intended for testing of the hypothesis $H_0 : \sigma_1^2 = ... = \sigma_n^2 = \sigma_0^2$. And the competing hypothesis about variance shift at an unknown time point is

$$H_1 : \sigma_1^2 = ... = \sigma_k^2 = \sigma_0^2, \quad \sigma_{k+1}^2 = \sigma_{k+2}^2 = ... = \sigma_n^2 = \sigma_0^2 + \delta, \quad |\delta| > 0, \quad (1 \le k \le n-1)$$

As sequence of observations in this paper we consider set of element $\{x_1, ..., x_n\}$ where $\{x_i\}$ is the sample of observation of a one-dimensional continuous random value in order of their appearance. Let m_x is the median of $\{x_i\}$. Then the Hsu statistic has the form

$$Q = \left(\frac{\sum\limits_{i=1}^{n} (i-1)(x_i - m_x)^2}{(n-1)\sum\limits_{i=1}^{n} (x_i - m_x)^2} - 0,5 \right) \Bigg/ \sqrt{\frac{n+1}{6(n-1)(n+2)}}$$

The limiting distribution law of the Hsu statistic Q for normal random variables is the standart normal distribution ($N(0,1)$). The exponential family of distributions with the density function $De(\theta_0) = \frac{\theta_0}{2\theta_1 \Gamma(1/\theta_0)} \exp\left\{ -\left(\frac{|x-\theta_2|}{\theta_1} \right)^{\theta_0} \right\}$ have been considered as symmetric alternatives of normal distribution. The special cases of this law are the Laplace distribution (with $\theta_0 = 1$), the normal distribution ($\theta_0 = 2$); the limiting cases are the Cauchy distribution ($\theta_0 \longrightarrow 0$) and the uniform distribution ($\theta_0 \longrightarrow +\infty$).

The conditional distributions $G(Q\,|H_0)$ of the Hsu statistic have been numerically shown to be strongly dependent on the distribution under observation (figure 1). Basing on the obtained empirical statistic distributions we have constructed approximate analytical models of the statistic distribution laws for $n = 30$. Table 1 contains results of test simple goodness-of-fit hypotheses of the constructed limiting statistic distributions approximations to the empirical one for differ observation sample sizes.

Table 1: Results of goodness-of-fit test of constructed distribution statistical models to corresponding empirical distributions under various observation distributions and sample sizes

n	20	30	60	100	constructed models for $n = 30$
xi∼Norm	0,1941	0,4457	0,1465	0,1331	$De(2), \theta_1 = 1, \theta_2 = 0 (Norm(0,1))$
xi∼De(1)	0,1585	0,71	0,1559	0,1091	$De(1.84), \theta_1 = 1.74, \theta_2 = -0.01$
xi∼De(10)	0,2321	0,28	0,0037	<0,0001	$De(1.94), \theta_1 = 1.13, \theta_2 = -0.01$
xi∼Max	0,0014	0,39	0,01	0,005	$De(1.56), \theta_1 = 1.41, \theta_2 = -0.01$

Presented in table sinificance levels achieved show that the use obtained analytical models of the statistic distribution laws will not lead to considerable errors in the statistical conclusion for a range number of n, if observation under symmetrical laws $De(\theta_0)$. At the same time It has shown that Hsu statistic distribution strongly depend on the sample sizes n when observed variables submit the asymmetrical Extreme-value law.

3 The tests for randomness investigation

The hypothesis of observation randomness is often needed of testing in applications. Wald and Wolfowitz [2] proposed test for randomness in the non-parametric case based on serial correlation. Let $R_i = rank(x_i)$. Then the statistic has the form

$$R = \left(\sum_{i=1}^{n-1} \left(R_i - \frac{n+1}{2} \right) \left(R_{i+1} - \frac{n+1}{2} \right) \right) \Bigg/ \sqrt{\frac{n^2 (n+1)(n-3)(5n+6)}{720}}$$

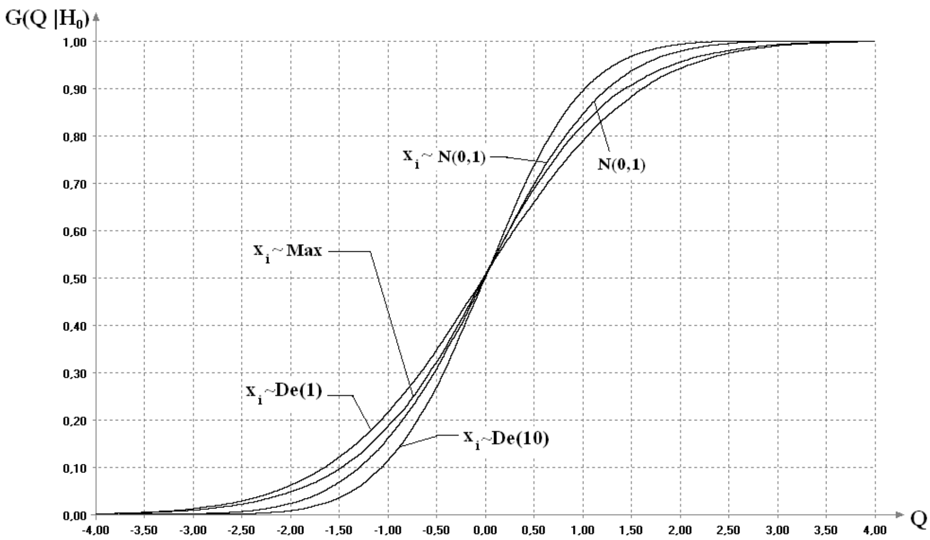

Figure 1: The Hsu statistic distributions when observations submitting the Extreme-value (maximum) law(Max in figure) and the $De(\theta_0)$ with different values of θ_0, $n = 30$

The authors supposed that distribution of R is asymptotically normal for $n > 20$. In our research statistic distributions of this test have been investigated by the methods of statistical simulating. Figure 2 illustrates the dependence of Wald-Wolfowitz test statistic distribution upon the number of observations. The statistic distribution converges slowly (from the left) to the limiting normal distribution. Distribution of R statistic is in a good agreement with the limiting law only for $n > 500$. Hereupon the use of the normal distribution as the limiting law when sample sizes are limited lead to the overrated values of significance level achieved and, hence, to increasing the number of beta errors.

Bartels [3] proposed two forms of test for randomness: test with B statistic for $n \le 100$ and test with B^* statistic for $n > 100$:

$$B = \frac{\sum_{i=1}^{n-1}\left(R_i - R_{i+1}\right)^2}{\sum_{i=1}^{n-1}\left(R_i - \overline{R}\right)^2}, \quad B^* = \frac{B-2}{2\sqrt{\frac{5}{5n+7}}}, \quad \overline{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$$

The distribution of B^* is asymptotically normal. By means of the numerical investigations It has been shown that B^* statistic can be used either for $n \le 100$ and $n > 100$ cases because the corresponding empirical distributions of statistic are in a good agreement with the limiting normal distribution (see Table 2).

Table 2: Results of goodness-of-fit test of B statistic empirical distribution to $N(0,1)$ distribution for differ sample sizes

n	20	30	50	70	100	200	500	1000
sinificance level achieved	0,183	0,302	0,327	0,416	0,59	0,477	0,642	0,805

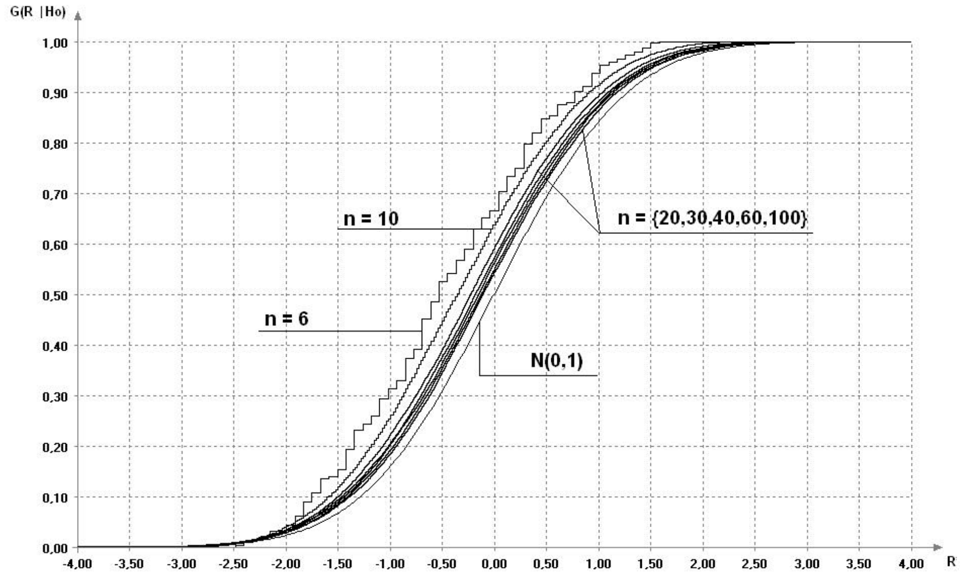

Figure 2: The Wald-Wolfowitz statistic distributions depending on the sample size n when the null hypothesis is true

Tabulated by Bartels confidence intervals for B statistic are wider than actual one obtained on the basis of the B statistic empirical distributions. For example, when theoretical interval is $[0.56, 3.44]$, the actual one is $[1.41, 2.59]$ (for $n = 30$). Therefore using of B^* statistic is more preferable to using of B statistic in case of small sample sizes n.

4 Acknowledgements

These investigations were supported by the Russian Foundation for Basic Research (Grant No. 09-01-00056-a) and the federal targeted program of the Ministry of Education and Science of the Russian Federation "Scientific and scientific-educational personnel of innovation Russia" (Project NK-178P/1).

References

[1] Hsu D.A. (1977) Tests for variance shift at an unknown time point. *Appl. Statist.*. Vol. **26**, number 3, pp. 279-284.

[2] Wald A., Wolfowitz J. (1943) An exact test for randomness in the non-parametric case based on serial correlation. *AMS*. Vol. **14**, pp. 378-388.

[3] Bartels R. (1982) The rank version of von Neimann's ratio test for randomness. *JASA*. Vol. **77**, number 377, pp. 40-46.