MATHEMATICAL MODELING AND INVESTIGATION OF THE STATISTIC DISTRIBUTIONS FOR MULTIDIMENSIONAL RANDOM VARIABLES[1]

S.S. Pomadin, B.Yu. Lemeshko
Novosibirsk State Technical University
Novosibirsk, Russia
e-mail: ser@fpm.ami.nstu.ru

Abstract

A number of statistics used in testing hypotheses concerning multidimensional variables observed has been considered. The limiting statistic distributions have been shown to change insignificantly if the observable law differs from the normal distribution in a rather wide range (more peaked or more flat-topped distributions). The empirical statistic distributions has proved to be in good agreement with the corresponding limiting distributions obtained in the classical correlation analysis under the assumption of observed data normality. With the results obtained being used, methods of classical correlation analysis can be valid in a wider area of applications.

1 Introduction

The correlation analysis problems take a very important place among the applications of multidimensional random variables analysis. These problems should be solved to reveal the presence and type of correlation as well as to test different statistical hypotheses on multidimensional distribution parameters and correlation rates.

The classical technique of correlation analysis is based on the assumption that the observed random vector submits to the multidimensional normal distribution law. The limiting distributions of statistics used in correlation analysis are obtained according to this assumption. And at the same time in practice of correlation analysis the assumptions made are known not to be valid in many cases. However the problems of correlation analysis are still urgent to be solved. That is why it is necessary to analyse the correctness of conclusions obtained for random variables that may not follow a normal distribution.

The purpose of the paper is to investigate the distributions of different correlation analysis statistics when an observed multidimensional law is not normal. It is extremely difficult to carry out such investigations by means of analytical methods only. The results given have been obtained by the method of statistical simulation and computer analysis being developed [1]. The method has given a good account of itself; it enables to obtain fundamental knowledge about statistical regularities with fewer difficulties.

[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)

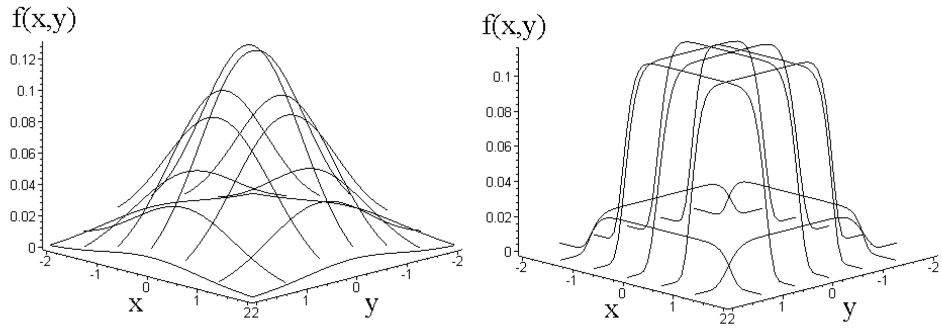

Figure 1: Simulated two-dimensional distributions, constructed for the form parameter values $\lambda = 2$ (on the left) and $\lambda = 10$ (on the right)

2 Simulation of multidimensional random variables

The key point of simulating statistic distributions of correlation analysis is the necessity of multidimensional pseudorandom vectors simulation. For the simulation of pseudorandom normal vectors the algorithm described in [2] in detail has been used. This algorithm has proved to be a good one. In [3] it was proposed to use this procedure for simulation of multidimensional non-normal distributions with the mathematical expectation and covariance matrix given. The main modification made is to use the manifold of one-dimensional variables $\{Z_i\}, i = \overline{1,m}$ [2] distributed by a certain non-normal law instead of the standard normal law.

For simulation of different systems $\{Z_i\}, i = \overline{1,m}$ it is convenient to use the exponential distribution family with the density function

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

where λ is the form parameter. The special cases of this law are the Laplace distribution (with $\lambda = 1$), the normal distribution ($\lambda = 2$); the limiting cases are the Cauchy distribution ($\lambda \to 0$) and the uniform distribution ($\lambda \to +\infty$). By changing the form parameter we can give a continuous "moving" of a simulated multidimensional law away from the normal distribution and make it more flat-topped than the normal law for $\lambda > 2$ or more peaked for $0 < \lambda < 2$. Pseudorandom $\bar{X}$ vectors from the normal distribution are formed with $\lambda = 2$.

As an output we obtain some non-normal multidimensional law. Such procedure doesn't enable to simulate multidimensional variables with a given distribution function and given mathematical expectation and covariance matrix. However it enables to generate pseudorandom vectors which have the distribution differing from the normal (according to the modeling process) with given mathematical expectation and covariance matrix.

In figure 1 the form of density functions obtained by simulation of two-dimensional

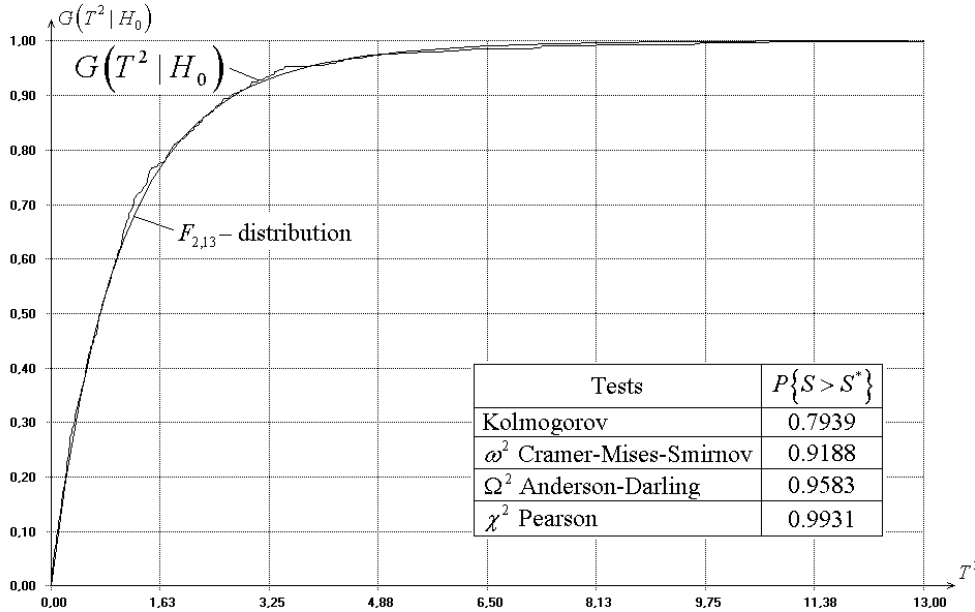

Figure 2: T^2 statistic distribution for $m = 2$, $n = 15$, $\lambda = 5$

vectors for $\lambda = 2$ (the normal distribution density, on the left) and $\lambda = 10$ (on the right). As it is seen, in the second case the obtained flat-topped distribution considerably differs from the normal law.

3 The investigation of test statistic distributions, used in the correlation analysis

The investigation of statistic distributions in case of normal observations. At the first stage of investigations we have investigated correlation analysis statistic distributions when observations submit the multinormal distribution law by statistical modeling technique with the purpose to test the reliability of the method used. The investigated criteria for testing hypotheses, the formulars of statistics and their limiting distributions are given in [4]. Simulation and investigation of the empirical distributions of the classical correlation analysis statistics have shown their good agreement with the corresponding theoretical limiting distributions.

The investigation of statistic distribution in case of non-normal observations. Correlation analysis statistic distributions have been investigated for multidimensional laws, simulated with the proposed procedure [3] with different values of the parameter $\lambda > 1$ of the distribution family (1).

In fig. 2 the empirical distribution for investigated statistic are given with the corresponding classical statistic limiting distribution. There are also the achieved significance levels $P\{S > S^*\}$, obtained after testing goodness-of-fit of empirical statistic distributions to the limiting distributions of classical statistics by the χ^2 Pearson, Kolmogorov, ω^2 Cramer-Mises-Smirnov, Ω^2 Anderson-Darling tests.

In fig. 2 it is shown the empirical distribution of the statistic $T^2 = \frac{n(n-m)}{m(n-1)}(\hat{M} - M_0)^T \hat{\Sigma}^{-1}(\hat{M} - M_0)$, (where n is sample size, m is vector dimension), which is used in testing hypothesis $H_0 : \bar{M} = \bar{M}_0$ about the mathematical expectation vector with unknown covariance matrix Σ [4], obtained for $\lambda = 5$, and the limiting $F_{m,n-m}$-distribution. High significance levels achieved by all goodness-of-fit tests enable to assert that there is no significant changes in the limiting statistic distribution.

4 Conclusions

Correlation analysis statistics, used for testing hypotheses on mathematical expectations, on pair, partial and partial correlation coefficient in case of multidimensional distributions, have been discussed. The limiting distributions of these statistics have been shown to change insignificantly if an observed law differs from the normal distribution in a rather wide range (more peaked or more flat-topped distributions). The empirical statistic distributions has proved to be in good agreement with the corresponding limiting distributions obtained in the classical correlation analysis under the assumption of observed data normality. With the results obtained being used, methods of classical correlation analysis can be valid in a wider area of applications.

This conclusion doesn't take into account the criteria for testing hypotheses on covariance matrixes for multidimensional law (hypothesis of the type $H_0 : \Sigma = \Sigma_0$ with known or unknown mathematical expectation vector). The limiting distributions of the statistics, used for testing of such hypotheses, essentially depend on the observed multidimensional law.

References

[1] Lemeshko B.Yu. (2000). Computer methods of investigations of statistical regularities. *In the book "Modeling, automation and optimization of high technologies"*, pp. 18-19. (in Russian)

[2] Ermakov S.M., Michailov G.A. (1982). *Statistical modeling*. Nauka, Moscow. (in Russian)

[3] Lemeshko B.Yu., Pomadin S.S. (2002). An approach to the simulation of pseudo-random vectors with "given" numerical characteristics by non-normal distribution laws. *Proceedings of the ISTC "Information Science and Telecommunication Problems"*. pp. 121-122. (in Russian)

[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)