DOI 10.1007/s10958-020-04760-x
*Journal of Mathematical Sciences, Vol. 246, No. 4, April, 2020*

# AN ESTIMATE OF STABILITY OF RECONSTRUCTION OF THE NORMAL DISTRIBUTION TYPE

**A. P. Ushakova[1] and N. G. Ushakov[2]**

One of the characterization problems of statistics is reconstruction of types when observations can have different location and/or scale parameters. In these cases, invariant statistics are used, and one of the main problems is uniqueness and stability of the reconstruction. There are a number of works devoted to this problem. In this article, we obtain a new estimate of stability of the normal type. It improves previously obtained estimates.

## 1. Introduction

The following problem often arises in applications. Suppose that there are a number of small independent samples such that in each small sample observations are independent and identically distributed while from sample to sample they have different values of location and scale parameters. The problem is to identify the distribution of the entire sample. First works in this direction include [4, 7], where, in particular, Zinger [7] solved a problem of characterization of the normal family (called also normal type) posed by A. N. Kolmogorov.

For such problems, it is necessary to use statistics which do not depend on the location and scale parameters. Reconstruction of the type of initial distribution from distribution of such a statistic is an actual problem, in particular, for goodness of fit testing. Prokhorov [5] solved the problem of uniqueness of the reconstruction and its qualitative stability in quite a general case. The quantitative stability of the reconstruction has been studied by a number of authors; see, for example, [1, 2]. In [6], it was proved that the upper bound of stability has the order $\varepsilon^{1/3}L(\varepsilon)$, where $\varepsilon$ is the distance between distributions of invariant statistics, and $L(\varepsilon)$ is a slowly varying function. In the present work, this estimate is improved.

In what follows we suppose (without loss of generality) that the small subsamples have size 3, i.e., the minimal necessary size. An essential feature of this work is that instead of the usually used two-dimensional statistic $(X_1-X_3, X_2-X_3)$ we will use a family of one-dimensional statistics that are linear combinations of $X$s.

## 2. Estimate of stability

In what follows we denote by $\Phi(z)$ the standard normal distribution function. Let $X_1, X_2, X_3$ be independent identically distributed random variables with distribution function $F(x-\theta)$ and unit variance, and let $Z_1, Z_2, Z_3$ be independent random variables having the standard normal distribution. Denote characteristic functions of $X_i$ and $Z_i$ by $f(t)$ and $g(t)$. Consider the set of three-dimensional vectors $(\alpha_1, \alpha_2, \alpha_3)$ satisfying the following conditions:

$$\alpha_1+\alpha_2+\alpha_3=0,\ \alpha_1^2+\alpha_2^2+\alpha_3^2=2. \tag{1}$$

Consider statistics $X_\alpha=\alpha_1X_1+\alpha_2X_2+\alpha_3X_3$ and $Z_\alpha=\alpha_1Z_1+\alpha_2Z_2+\alpha_3Z_3$. Denote their distribution functions by $F_{X,\alpha}(x)$ and $\Phi_{Z,\alpha}(x)$ and characteristic functions by $\psi_\alpha(t)$ and $\phi_\alpha(t)$.

[1] Institute of Microelectronics Technology, Russian Academy of Sciences, Chernogolovka, Russia, e-mail: al.ushakova@gmail.com

[2] Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway, e-mail: nikolai.ushakov@ntnu.no

Proceedings of the XXXV International Seminar on Stability Problems for Stochastic Models, Perm, Russia, September 24–28, 2018. Part I.

 

**Theorem 1.** *Suppose that the following inequalities hold*:

$$\sup_{\alpha} \int_{-\infty}^{\infty} d|(F_{X,\alpha} - \Phi_{Z,\alpha})| \leqslant \varepsilon < \frac{1}{2}, \tag{2}$$

$$\sup_{\alpha} \int_{0}^{\infty} x d|(F_{X,\alpha}(x) - (1 - F_{X,\alpha}(-x)))| \leqslant \varepsilon\sqrt{\varepsilon}. \tag{3}$$

*Then there exists* $\bar{\theta}$ *such that*

$$|F(x-\bar{\theta}) - \Phi(x)| \leqslant \sqrt{\varepsilon} L(\varepsilon), \tag{4}$$

*where* $L(\varepsilon)$ *is a slowly varying function.*

**Proof.** Let (2) and (3) be satisfied. Since $\alpha_3 = -(\alpha_1 + \alpha_2)$, we obtain that

$$\alpha_1^2 + \alpha_2^2 + \alpha_1\alpha_2 = 1. \tag{5}$$

Characteristic functions of $X_\alpha$ and $Z_\alpha$ can be written as follows:

$$\psi_\alpha(t) = f(\alpha_1 t) f(\alpha_2 t) \overline{f((\alpha_1 + \alpha_2)t)},$$

$$\phi_\alpha(t) = e^{-t^2} = g(\alpha_1 t) g(\alpha_2 t) g((\alpha_1 + \alpha_2)t).$$

Using (2), we obtain

$$\sup_{\alpha,t} |\psi_\alpha(t) - \phi_\alpha(t)| \leqslant \sup_{\alpha} \int_{-\infty}^{\infty} d|(F_{X,\alpha} - \Phi_{Z,\alpha})| \leqslant \varepsilon;$$

therefore

$$|f(\alpha_1 t) f(\alpha_2 t) \overline{f((\alpha_1 + \alpha_2)t)} - g(\alpha_1 t) g(\alpha_2 t) g((\alpha_1 + \alpha_2)t)| \leqslant \sup_{\alpha,t} |\psi_\alpha(t) - \phi_\alpha(t)| \leqslant \varepsilon. \tag{6}$$

Let $\alpha_2 = 0$. Then (6) implies that

$$||f(\alpha_1 t)|^2 - |g(\alpha_1 t)|^2| \leqslant \varepsilon$$

for any $t$, i.e.,

$$||f(t)|^2 - |g(t)|^2| \leqslant \varepsilon$$

or

$$||f(t)| - |g(t)|| \leqslant \sqrt{\varepsilon} \tag{7}$$

for all $t$.

Set $T_0 = \sqrt{-\ln(2\varepsilon)}$. If $|t| \geqslant T_0$, then, due to (7),

$$|f(t) - g(t)| \leqslant |f(t)| + |g(t)| \leqslant 2|g(t)| + \sqrt{\varepsilon} \leqslant (2\sqrt{2} + 1)\sqrt{\varepsilon}. \tag{8}$$

If $|t| \leqslant T_0$, then

$$||\psi_\alpha(t)| - |\phi_\alpha(t)|| \leqslant |\psi_\alpha(t) - \phi_\alpha(t)| \leqslant \varepsilon,$$

hence

$$|\psi_\alpha(t)| \geqslant \phi_\alpha(T_0) - \varepsilon \geqslant \varepsilon > 0$$

and

$$|f(t)| \geqslant |g(t)| - \sqrt{\varepsilon} \geqslant g(T_0) - \sqrt{\varepsilon} = \sqrt{2\varepsilon} - \sqrt{\varepsilon} > 0.$$

Thus, for $|t| \leqslant T_0$, the functions $\psi_\alpha(t)$ and $f(t)$ have no zeros, and, therefore, there exist principal values of arguments of these functions, and they are continuous. Let us estimate these principal values (for $|t| \leqslant T_0$). For the imaginary part of $\psi_\alpha(t)$, using simple algebra, we obtain

$$|\Im\psi_\alpha(t)| \leqslant \int_0^\infty |tx| d|(F_{X,\alpha}(x) - (1 - F_{X,\alpha}(-x)))| \leqslant |t|\varepsilon\sqrt{\varepsilon}$$

(the last inequality follows from (3)). Thus for $|t| \leqslant T_0$,

$$|\Im\psi_\alpha(t)| \leqslant \varepsilon\sqrt{\varepsilon}|t|, \ |\psi_\alpha(t)| \geqslant \varepsilon. \tag{9}$$

Show that $\Re\psi_\alpha(t) > 0$ for $|t| \leqslant T_0$. Assume the contrary. Then there exists $u$ such that $|u| \leqslant T_0$, and $\Re\psi_\alpha(u) = 0$. Then $|\psi_\alpha(u)| = |\Im\psi_\alpha(u)|$. Using (9), we get

$$\varepsilon \leqslant |\psi_\alpha(u)| = |\Im\psi_\alpha(u)| \leqslant \varepsilon\sqrt{\varepsilon}|u| \leqslant \varepsilon\sqrt{\varepsilon}T_0 = \varepsilon\sqrt{\varepsilon}\sqrt{-\ln 2\varepsilon},$$

which cannot be when $\varepsilon < 1/2$.

Denote the principal value of the argument of the function $f(t)$ by $\mathrm{Arg} f(t)$. Using (9) and (15), we obtain that

$$|\mathrm{Arg}\psi_\alpha(t)| \leqslant \frac{\pi}{2}\sqrt{\varepsilon}t$$

for $0 \leqslant t \leqslant T_0$. But then for these $t$s

$$|\mathrm{Arg} f(\alpha_1 t) + \mathrm{Arg} f(\alpha_2 t) - \mathrm{Arg} f(\alpha_1 t + \alpha_2 t)| = |\mathrm{Arg}\psi_\alpha(t)| \leqslant \frac{\pi}{2}\sqrt{\varepsilon}t. \tag{10}$$

Denote $a(t) = \mathrm{Arg} f(t)$. We have

$$|a(\alpha_1 t) + a(\alpha_2 t) - a(\alpha_1 t + \alpha_2 t)| \leqslant \frac{\pi}{2}\sqrt{\varepsilon}t, \tag{11}$$

where $a(x)$ is a continuous function such that $a(0) = 0$, $a(-x) = -a(x)$, $0 \leqslant t \leqslant T_0$, and (5) holds. In order to use Lemma 1, we prove that (11) implies

$$|a(t_1) + a(t_2) - a(t_1 + t_2)| \leqslant \frac{\pi}{2}\sqrt{\varepsilon}(t_1 + t_2)$$

for any

$$0 \leqslant t_1 \leqslant T_0, \ 0 \leqslant t_2 \leqslant T_0, \ 0 \leqslant t_1 + t_2 \leqslant T_0. \tag{12}$$

Consider arbitrary $t_1$, $t_2$ satisfying (12) and set $t = \sqrt{t_1^2 + t_2^2 + t_1 t_2}$, $\alpha_1 = t_1/t$, $\alpha_2 = t_2/t$. Then

$$\alpha_1^2 + \alpha_2^2 + \alpha_1\alpha_2 = \frac{1}{t^2}(t_1^2 + t_2^2 + t_1 t_2) = 1.$$

Evidently these $\alpha$s satisfy the conditions $0 \leqslant \alpha_1 \leqslant 1$, $0 \leqslant \alpha_2 \leqslant 1$. Substitute $\alpha_1 t = t_1$ and $\alpha_2 t = t_2$ in (11). Then

$$|a(t_1) + a(t_2) - a(t_1 + t_2)| \leqslant \frac{\pi}{2}\sqrt{\varepsilon}(t_1 + t_2)$$

for all $t_1, t_2$ satisfying (12). Choose now the location parameter $\theta_0$ for the function $F(x - \theta)$ so that $\mathrm{Arg} f(T_0) = 0$ and use Lemma 1. Then we get the inequality

$$|\mathrm{Arg} f(t)| \leqslant \frac{\pi}{2}\sqrt{\varepsilon}(T_0 + 5) \tag{13}$$

for $|t| \leqslant T_0$ (for negative values of $t$, the inequality holds because the argument is an odd function). Using inequality (16), we obtain from (7) and (13)

$$|f(t) - g(t)| \leqslant \left(\frac{\pi}{2}(T_0 + 5) + 1\right)\sqrt{\varepsilon} = \sqrt{\varepsilon}\left(\frac{\pi}{2}(\sqrt{-\ln 2\varepsilon} + 5) + 1\right) \tag{14}$$

for $|t| \leqslant T_0$. For $|t| > T_0$ inequality (14) holds due to (8).

Use of Lemma 2 completes the proof. It is sufficient to set in Lemma 2

$$\epsilon = \sqrt{\varepsilon}\left(\frac{\pi}{2}(\sqrt{-\ln 2\varepsilon} + 5) + 1\right), \ T = \sqrt{\varepsilon}, \ L = 2T.$$

## 3. Auxiliary results

This section contains auxiliary results that are used in the proof of Theorem 1.

**Lemma 1.** *Let $a(t)$ be a continuous function satisfying the following conditions*: $a(0) = 0$, $a(T_0) = 0$,

$$|a(t_1) + a(t_2) - a(t_1 + t_2)| \leqslant \varepsilon(t_1 + t_2)$$

*for any $t_1 \geqslant 0$, $t_2 \geqslant 0$, $t_1 + t_2 \leqslant T_0$, where $\varepsilon > 0$, $T_0 \geqslant 1$. Then*

$$|a(t)| \leqslant \varepsilon(T_0 + 5).$$

The lemma follows from Lemma 1 of [6].

**Lemma 2.** *Let $F(x)$ be a nondecreasing function, $G(x)$ be a function of bounded variation, and $F(-\infty) = G(-\infty)$. Denote the corresponding Fourier–Stieltjes transforms by $f(t)$ and $g(t)$. Let $G(x)$ be differentiable, and*

$$\sup_x |G'(x)| \leqslant c.$$

*If $|f(t) - g(t)| < \epsilon$ for $|t| < T$, then for any $L > 2/T$, the following inequality holds*:

$$\sup_x |F(x) - G(x)| < A\left(\epsilon \log(LT) + \frac{c}{T} + \gamma(L)\right),$$

*where*

$$\gamma(L) = \operatorname*{Var}_{-\infty<x<\infty} G(x) - \sup_x \operatorname*{Var}_{x\leqslant y<x+L} G(y)$$

(Var *is the total variation*).

This lemma was obtained in [3].

We point out also the following elementary inequalities. Let $|z|$ be a complex number $z = x + iy$. If $x > 0$ and $|y|/|x| \leqslant \eta < 1$, then

$$|\arg z| \leqslant \frac{\pi}{2}\eta. \tag{15}$$

Let $z$ and $t$ be two complex numbers such that $|z| \leqslant 1$, $|t| \leqslant 1$. Then

$$|z - t| \leqslant ||z| - |t|| + |\arg z - \arg t|. \tag{16}$$

## REFERENCES

1. A. M. Kagan and L. B. Klebanov, "Estimation of the stability in the problem of reconstructing the additive type of a distribution," *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI)*, **61**, No. 136, 68–74 (1976).

2. L. B. Klebanov, "More on estimating the stability in the problem of reconstructing the additive type of a distribution," *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI)*, **87**, 74–78 (1979).
3. L. D. Meshalkin and B. A. Rogozin, "An estimate of the distance between distribution functions from the closeness of their characteristic functions and its application to the central limit theorem," in: *Limit Theorems of Probability Theory*, Tashkent (1963), pp. 49–55.
4. A. A. Petrov, "Verification of statistical hypotheses on the type of a distribution based on small samples," *Theor. Prob. Appl.* **1**, No. 2, 248–271 (1956).
5. Yu. V. Prokhorov, "A characterization of a class of probability distributions by the distributions of certain statistics," *Theor. Prob. Appl.*, **10**, No. 4, 479–487 (1965).
6. A. P. Ushakova, "Estimates of stability of characterization of additive types of distributions," *J. Math. Sci.*, **89**, No. 5, 1582–1589 (1998).
7. A. A. Zinger, "On a problem of A.N. Kolmogorov," *Vestn. Leningrad. Univ.*, **11**, No. 1, 53–56 (1956).