ISSN 0278-6419, Moscow University Computational Mathematics and Cybernetics, 2017, Vol. 41, No. 2, pp. 76–80. © Allerton Press, Inc., 2017.
Original Russian Text © V.G. Ushakov, N.G. Ushakov, 2017, published in Vestnik Moskovskogo Universiteta, Seriya 15: Vychislitel'naya Matematika i Kibernetika, 2017, No. 2, pp. 26–29.

Boundaries of the Precision of Restoring Information Lost after Rounding the Results from Observations

V. G. Ushakov[1*] and N. G. Ushakov[2]**

[1] *Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119991 Russia; Institute of Informatics, Russian Academy of Sciences*

[2] *Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

Received November 11, 2016

Abstract—Lower and upper estimates are obtained for deviations of the limit of a selected mean from estimated mathematical expectations when rounded data are processed. Different cases of error distribution are considered: normal, Simpson (triangle), and Laplace (double exponential)distributions.

Keywords: rounded data, extract average, normal distribution, Laplace distribution, Simpson distribution.

DOI: 10.3103/S0278641917020091

1. INTRODUCTION

Interest in the problem of processing rounded data [1–4] has recently grown for several reasons, including the rapid growth of computer technologies that normally produce large volumes of data. As was shown in [5], the measuring error can be used in the statistical processing of rounded data to reduce the effect of the rounding error. In many cases, it is worth artificially increasing the measuring error in order to improve the precision of the final results. In this work, we obtain the upper and lower boundaries of precision for estimating the mathematical expectation of an observable random quantity if the distribution of the measuring error satisfies one of three types of distribution, with distribution density $f(x)$ and characteristic function $\varphi(t)$:

1) a normal distribution with $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ and $\varphi(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$;

2) a Laplace distribution with $f(x) = \dfrac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x-\mu|}$ and $\varphi(t) = \dfrac{2e^{i\mu t}}{\sigma^2 t^2 + 2}$;

3) a Simpson distribution with $f(x) = \dfrac{1}{\sqrt{6}\sigma} - \dfrac{1}{6\sigma^2}|x-\mu|$ and $\varphi(t) = \dfrac{2\sin^2\left(\sqrt{\frac{3}{2}}\sigma t\right)}{3\sigma^2 t^2} e^{i\mu t}$.

In all three cases, μ is the mathematical expectation and σ^2 is the dispersion.

Below, we assume (without loss of generality) that the discretization step is 1. We introduce the following notation: The integral and fractional part of real number x is denoted by $[x]$ and $\{x\}$, respectively. We denote the rounded value by x^* up to the closest integer. Thus, $x^* = [x]$ if $\{x\} < 1/2$, and $x^* = [x] + 1$ if $\{x\} \geqslant 1/2$. Note that $x^* = [x + 1/2]$. The almost certain limit of a sequence of random quantities $X_1, X_2, \ldots$ is denoted by $\lim\limits_{n\to\infty}(\text{a. c.})X_n$.

To find the upper and lower estimates, we use the following:

*E-mail: vgushakov@mail.ru.
**E-mail: ushakov@math.ntnu.no.

Lemma. *Let Y be an absolutely continuous random quantity with distribution density $f(y)$ and characteristic function $\varphi(t)$. If $\varphi(t)$ is absolutely integrable, we then have the equality*

$$\mathbf{E}\{Y\} = \int_0^1 y \sum_{n=-\infty}^{\infty} f(y+n)dy = \frac{1}{2} - \sum_{n=1}^{\infty} \frac{\operatorname{Im}\varphi(2\pi n)}{\pi n}.$$

Proof. Using the Poisson summation formula (see, e.g., [6])

$$\sum_{n=-\infty}^{\infty} f(y+n) = \sum_{n=-\infty}^{\infty} \varphi(2\pi n) e^{-iy2\pi n},$$

we obtain the required equality

$$\mathbf{E}\{Y\} = \sum_{n=-\infty}^{\infty} \varphi(2\pi n) \int_0^1 y e^{-iy2\pi n} dy = \frac{1}{2} - \sum_{n\neq 0} \varphi(2\pi n)\frac{i}{2\pi n} = \frac{1}{2} - \sum_{n=1}^{\infty} \frac{\operatorname{Im}\varphi(2\pi n)}{\pi n}.$$

2. UPPER ESTIMATES

Let $X_1, X_2, \ldots$ be a sequence of independent equally distributed random quantities with unknown mathematical expectation μ, and $\varepsilon_1, \varepsilon_2, \ldots$ be a sequence with one of the distributions described above with mathematical expectation 0 and dispersion σ^2. Consider the sequence of rounded data $(X_1+\varepsilon_1)^*, (X_1+\varepsilon_2)^*, \ldots$. We write

$$\Delta(\mu,\sigma) = \left| \lim_{n\to\infty} (\text{a. c.}) \frac{1}{n} \sum_{i=1}^{n} (X_i + \varepsilon_i)^* - \mu \right|.$$

Theorem 1. *For any μ, depending on the distribution of $\varepsilon_n, n = 1, 2, \ldots$, one of the following inequalities holds*:

1) $\Delta(\mu,\sigma) < \dfrac{1}{\pi}\left(1 + \dfrac{1}{4\pi^2\sigma^2}\right) e^{-2\pi^2\sigma^2}$ *with the normal distribution*;

2) $\Delta(\mu,\sigma) < \dfrac{1}{\pi\left(2\pi^2\sigma^2+1\right)} + \dfrac{1}{4\pi^3\sigma^2}$ *with the Laplace distribution*;

3) $\Delta(\mu,\sigma) \leqslant \dfrac{\left|\sin\left(\sqrt{6}\pi\sigma\right)\right|}{36\pi\sigma^2}$ *with the Simpson distribution.*

Proof. Let $\psi(t)$ and $\varphi(t)$ be the characteristic functions of random quantities $X_1, X_2, \ldots$ and $\varepsilon_1, \varepsilon_2, \ldots$, respectively. Since

$$\Delta(\mu,\sigma) = \left| \mathbf{E}\left[X_i + \varepsilon_i + \frac{1}{2}\right] - \mu \right| = \left| \frac{1}{2} - \mathbf{E}\left\{X_i + \varepsilon_i + \frac{1}{2}\right\} \right|,$$

according to the Lemma we have

$$\Delta(\mu,\sigma) = \sum_{n=1}^{\infty} \frac{\operatorname{Im}\left(\varphi(2\pi n)\psi(2\pi n)e^{i\pi n}\right)}{\pi n} = \sum_{n=1}^{\infty} \frac{(-1)^n \operatorname{Im}\left(\psi(2\pi n)\right)\varphi(2\pi n)}{\pi n}.$$

This implies

$$\Delta(\mu,\sigma) \leqslant \sum_{n=1}^{\infty} \frac{\varphi(2\pi n)}{\pi n}.$$

Let us consider each type of distribution of the random quantities $\varepsilon_n, n = 1, 2, \ldots$, separately.

1. For the normal distribution, we obtain:

$$\sum_{n=1}^{\infty} \frac{\varphi(2\pi n)}{\pi n} = \sum_{n=1}^{\infty} \frac{e^{-2\pi^2\sigma^2 n^2}}{\pi n} < \frac{1}{\pi}\left(e^{-2\pi^2\sigma^2} + \int_1^{\infty} \frac{e^{-2\pi^2\sigma^2 x^2}}{x} dx\right) < \frac{4\pi^2\sigma^2+1}{4\pi^3\sigma^2} e^{-2\pi^2\sigma^2}.$$

2. For the Laplace distribution, we obtain:

$$\sum_{n=1}^{\infty}\frac{\varphi(2\pi n)}{\pi n}=\sum_{n=1}^{\infty}\frac{1}{\pi n}\frac{1}{2\pi^2\sigma^2n^2+1}<\frac{1}{\pi\left(2\pi^2\sigma^2+1\right)}+\frac{1}{\pi}\int_1^{\infty}\frac{dx}{x\left(2\pi^2\sigma^2x^2+1\right)}$$
$$=\frac{1}{\pi\left(2\pi^2\sigma^2+1\right)}+\frac{1}{2\pi}\ln\left(1+\frac{1}{2\pi^2\sigma^2}\right)<\frac{1}{\pi\left(2\pi^2\sigma^2+1\right)}+\frac{1}{4\pi^3\sigma^2}.$$

3. For the Simpson distribution, we obtain:

$$\sum_{n=1}^{\infty}\frac{\varphi(2\pi n)}{\pi n}=\sum_{n=1}^{\infty}\frac{\sin^2\left(\sqrt{6}\pi\sigma n\right)}{6\pi^3\sigma^2n^3}.$$

Since $|\sin(n\alpha)|\leqslant n\sin\alpha$, we have

$$\Delta(\mu,\sigma)\leqslant\frac{\left|\sin\left(\sqrt{6}\pi\sigma\right)\right|}{6\pi^3\sigma^2}\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\left|\sin\left(\sqrt{6}\pi\sigma\right)\right|}{36\pi\sigma^2}.$$

The theorem is proved.

3. LOWER ESTIMATES

When finding upper estimates for the deviation of the limit of a selected average from the estimated mathematical expectation in the previous section, we considered only the additional random measuring error given by the sequence $\varepsilon_1,\varepsilon_2,\ldots$. To find nontrivial lower estimates, we must consider all random measuring errors. We therefore do not not distinguish initial random errors $X_1,X_2,\ldots$ and additional errors $\varepsilon_1,\varepsilon_2,\ldots$ in observations.

Let $X_1,X_2,\ldots$ be a sequence of independent equally distributed random quantities having one of the three distributions described above, with mathematical expectation μ and dispersion σ^2. We then consider the sequence of rounded values $X_1^*,X_1^*,\ldots$. We are interested in the lower boundaries for the deviation of limit of selected average of rounded observations $\frac{1}{n}\sum_{i=1}^{n}X_i^*$ from μ for the worst value of μ and the dependence of these boundaries on σ. We write

$$\Delta(\sigma)=\sup_{\mu}\Delta(\mu,\sigma)=\sup_{\mu}\left|\lim_{n\to\infty}(\text{a. c.})\frac{1}{n}\sum_{i=1}^{n}X_i^*-\mu\right|.$$

Theorem 2. *Depending on the distribution of $X_1,X_2,\ldots$, one of the following inequalities holds*:

1) $\Delta(\sigma)>\dfrac{e^{-2\pi^2\sigma^2}}{\pi}-\dfrac{e^{-18\pi^2\sigma^2}}{3\pi}\left(1+\dfrac{1}{6\pi^2\sigma^2}\right)$ *in the case of the normal distribution*;

2) $\Delta(\sigma)>\dfrac{1}{\pi\left(2\pi^2\sigma^2+1\right)}-\dfrac{1}{16\pi^3\sigma^2}$ *in the case of the Laplace distribution*;

3) $\Delta(\sigma)\geqslant\dfrac{\left|\sin\left(\sqrt{6}\pi\sigma\right)\right|}{6\pi^3\sigma^2}\left(\left|\sin\left(\sqrt{6}\pi\sigma\right)\right|-0.4\right)$ *in teh case of the Simpson distribution.*

Proof. Note that according to the strengthened law of large numbers, we have

$$\Delta(\sigma)=\sup_{\mu}|\mathbf{E}X_i^*-\mu|=\sup_{\mu}\left|\mathbf{E}\left[X_i+\frac{1}{2}\right]-\mu\right|=$$
$$=\sup_{\mu}\left|\frac{1}{2}-\mathbf{E}\left\{X_i+\frac{1}{2}\right\}\right|=\sup_{\mu}\left|\frac{1}{2}-\mathbf{E}\left\{X_i\right\}\right|.$$

According to the Lemma, we have

$$\mathbf{E}\{X_i\}=\frac{1}{2}-\sum_{k=1}^{\infty}\frac{\operatorname{Im}\varphi(2\pi k)}{\pi k}.$$

Let us consider each type of distribution $X_1, X_2, \dots$ separately.

1. With a normal distribution, we obtain

$$\mathbf{E}\{X_i\} = \frac{1}{2} - \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{k}e^{-2\pi^2\sigma^2k^2}\sin(\mu 2\pi k)$$

and therefore

$$\sup_{\mu}\left|\frac{1}{2} - \mathbf{E}\{X_i\}\right| = \frac{1}{\pi}\sup_{\mu}\sum_{k=1}^{\infty}\frac{1}{k}e^{-2\pi^2\sigma^2k^2}\sin(\mu 2\pi k).$$

We obviously have the inequality

$$\sup_{\mu}\sum_{k=1}^{\infty}\frac{1}{k}e^{-2\pi^2\sigma^2k^2}\sin(\mu 2\pi k) \geqslant e^{-2\pi^2\sigma^2} - \sum_{k=3}^{\infty}\frac{1}{k}e^{-2\pi^2\sigma^2k^2}.$$

Let us find an upper estimate for the sum in the right hand side of the latter inequality. We have

$$\sum_{k=3}^{\infty}\frac{1}{k}e^{-2\pi^2\sigma^2k^2} < \frac{1}{3}e^{-18\pi^2\sigma^2} + \int_3^{\infty}\frac{1}{x}e^{-2\pi^2\sigma^2x^2}dx.$$

In addition, we obtain

$$\int_3^{\infty}\frac{1}{x}e^{-2\pi^2\sigma^2x^2}dx < \frac{1}{3}\int_3^{\infty}e^{-6\pi^2\sigma^2x}dx = \frac{1}{18\pi^2\sigma^2}e^{-18\pi^2\sigma^2}.$$

This implies

$$\sum_{k=3}^{\infty}\frac{1}{k}e^{-2\pi^2\sigma^2k^2} < \frac{1}{3}e^{-18\pi^2\sigma^2}\left(1 + \frac{1}{6\pi^2\sigma^2}\right)$$

and therefore

$$\sup_{\mu}\left|\frac{1}{2} - \mathbf{E}\{X_i\}\right| > \frac{1}{\pi}\left(e^{-2\pi^2\sigma^2} - \frac{1}{3}e^{-18\pi^2\sigma^2}\left(1 + \frac{1}{6\pi^2\sigma^2}\right)\right).$$

2. With the Laplace distribution, we obtain

$$\sup_{\mu}\left|\frac{1}{2} - \mathbf{E}\{X_i\}\right| = \frac{1}{\pi}\sup_{\mu}\sum_{k=1}^{\infty}\frac{1}{k}\frac{\sin(\mu 2\pi k)}{1 + 2\pi^2\sigma^2k^2}.$$

We therefore have

$$\sup_{\mu}\left|\frac{1}{2} - \mathbf{E}\{X_i\}\right| > \frac{1}{\pi}\frac{1}{1 + 2\pi^2\sigma^2} - \frac{1}{\pi}\sum_{k=3}^{\infty}\frac{1}{k}\frac{1}{1 + 2\pi^2\sigma^2k^2}$$

$$> \frac{1}{\pi}\frac{1}{1 + 2\pi^2\sigma^2} - \int_2^{\infty}\frac{dx}{x(1 + 2\pi^2\sigma^2x^2)} > \frac{1}{\pi}\frac{1}{1 + 2\pi^2\sigma^2} - \frac{1}{16\pi^3\sigma^2}.$$

3. With the Simpson distribution, we obtain

$$\sup_{\mu}\left|\frac{1}{2} - \mathbf{E}\{X_i\}\right| \geqslant \frac{\sin^2(\sqrt{6}\sigma\pi)}{6\pi^3\sigma^2} - \sum_{k=3}^{\infty}\frac{\sin^2(\sqrt{6}\sigma\pi k)}{6\pi^3\sigma^2k^3}$$

$$\geqslant \frac{\sin^2(\sqrt{6}\sigma\pi)}{6\pi^3\sigma^2} - \frac{\left|\sin\left(\sqrt{6}\pi\sigma\right)\right|}{6\pi^3\sigma^2}\left(\frac{\pi^2}{6} - \frac{5}{4}\right) \geqslant \frac{\left|\sin\left(\sqrt{6}\pi\sigma\right)\right|}{6\pi^3\sigma^2}\left(\left|\sin\left(\sqrt{6}\pi\sigma\right)\right| - 0.4\right).$$

The theorem is proved.

Comparison of upper and lower estimates shows that they have equal orders of decline as $\sigma \to \infty$. Estimates become useless for very small values of σ,: lower estimates become negative, and upper estimates become very large. However, this reflects the problem: If the measuring error is tens of times smaller than the rounding error (the discretization step), it has virtually no effect on the precision of the final result.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Science Foundation, project no. 14-11-00364.

REFERENCES

1. W. M. Li and Z. D. Bai, "Rounded data analysis based on multi-layer ranked set sampling," Acta. Math. Sin. **27**, 2507–2518 (2011).
2. H. Schneeweiss, J. Komlos, and A. S. Ahmad, "Symmetric and asymmetric rounding: a review and some new results," AStA Adv. Stat. Anal. **94**, 247–271 (2010).
3. B. Wang and W. Wertelecki, "Density estimation for data with rounding errors," Comput. Stat. Data Anal. **65**, 4–12 (2013).
4. N. Zhao and Z. Bai, "Analysis of rounded data in mixture normal model," Stat. Papers **53**, 895–914 (2012).
5. V. G. Ushakov and N. G. Ushakov, "On averaging of rounded data ," Inform. Primen. **9** (4), 116–119 (2015).
6. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1957), Vol. 2.

Translated by A. Stoyanovsky