P. KOSOBUTSKYY

National University "Lviv Polytechnic"
(12, S. Bandera Str., Lviv 79012, Ukraine; e-mail: petkosob@gmail.com)

ON THE SIMULATION OF THE MATHEMATICAL EXPECTATION AND VARIANCE OF SAMPLES FOR GAUSSIAN-DISTRIBUTED RANDOM VARIABLES

*The derivation of propagation rules for the mean and the variance of physical quantities functionally connected by the transformations X^2, $\cos X$, $\sqrt{X}$, and $\arccos X$, which were proposed in Ukr. J. Phys. **61**, 345 (2016) and Ukr. J. Phys. **62**, 184 (2017), has been analyzed. It is shown that the substantiation of the "error propagation rules" was not based on the fundamentals of probability theory and mathematical statistics. Moreover, the proposed reduction of indices, $X \to \sqrt{X}$ and $X^2 \to X$, in the roots of the square equations forming a basis for the propagation formulas restricts the values of the normal distribution parameters m_X and σ_X.*

K e y w o r d s: normal distribution, expectation, variance, random variables, statistical averaging rules.

1. Introduction

Two works published recently in cite1a,1b were devoted to the derivation of the so-called rules of "error propagation" for the direct transformations $Y = g(X) = (X^2, \cos X)$ and the inverse to them $Z = g^{-1}(X) = (\sqrt{X}, \arccos X)$ of the random variable (RV) X obeying the normal distribution law $N(m_X, \sigma_X)$ [1]. Here, the parameters m_X and σ_X do not determine the physical quantity itself, but the distribution of the probability density for its random values. The error propagation rules were derived, by using elementary arithmetic transformations of quadratic equations for the variances of random quantities X, X^2, $\cos X$, and $\arccos X$. The required explicit forms of mean functions were calculated by applying tabulated integrals [3].

However, the substantiation of the obtained analytic expressions – the propagation rules – was not based on the fundamentals of probability theory and mathematical statistics [4–6]. Moreover, the obtained

relationships were tested on experimental samples that did not satisfy the criteria of normal distribution. In view of the challenging character of this approach to the statistical averaging of the results of physical researches, which always contain a fluctuating component, this work is devoted to the analysis of the substantiation of the error propagation formulas proposed in works [1, 2].

2. Theoretical Analysis

When analyzing the experimental data, the Gaussian statistical distribution $N(m_X, \sigma_X)$ is widely used. In this distribution, the probability density of obtaining some value for the measured quantity X depends on the variance $\sigma_X^2 = D_X$ and the shift $m_X = E_X$. The Gaussian distribution forms a basis for the theory of statistical estimation of errors [7]. The latter is based on the fundamental regularities of the curve describ-

---

[1] An experiment is associated with a random numerical value X. The function of X is real-valued and should satisfy all probability axioms.

ing the probability density for the standard $N(0, \sigma_X)$ distributions: symmetry, a single-mode character, and a maximum with the height $\frac{1}{\sqrt{2\pi\sigma_X^2}}$ located at the point with the coordinate E_X. At the height $\approx \frac{0.607}{\sqrt{2\pi\sigma_X^2}}$, the contour width eqials $2\sigma_X$, and the co-ordinate values at those points equal $E_X \mp \sigma_X$.

For the Gaussian curve, the mean coincides with the median and the mode, and the probability of large deviations Δx from m_X falls down proportion-ally to the exponential function $\exp\left(-\frac{\Delta x^2}{2\sigma_X^2}\right)$. Those properties allow the correspondence of an empirical distribution to the normal one to be rapidly esti-mated. Proper attention is often not given to this fact [1, 2], thereby violating the requirements concerning the verification of statistical models with respect to their Gaussian character.

The aim of works [1, 2] was to propose analytic rela-tionships that would make it possible to evaluate the parameters $E_{\sqrt{X}}$, $D_{\sqrt{X}}$, $E_{\arccos X}$, and $D_{\arccos X}$ of RVs subjected to the $Y = g^{-1}(X) = (\sqrt{X}, \arccos X)$ transformations – inverse to the direct $Y = g(X) = (X^2, \cos X)$ ones – in terms of the given pa-rameters E_X and D_X with the random values of the normally $N(m_X, \sigma_X)$-distributed random vari-able X. The cited author obtained them in the form

$$E_{\sqrt{X}}^4 = E_X^2 - D_X/2, \tag{1a}$$

$$D_{\sqrt{X}} = E_X - \sqrt{E_X^2 - D_X/2}. \tag{1b}$$

In order to analyze the statistical validity of the scheme of index reduction, which was applied in works [1, 2] while deriving relations (1), let us analyze the algorithm of their substantiation in a somewhat different statement.

In probability theory and mathematical statistics, a fundamental concept of RV is the variance D, as a measure of fluctuation intensity. It is equal to the mean of the square minus the square of the mean, and the inequality

$$D > 0 \tag{2}$$

is strictly obeyed for it.

If the initial random variable X is transformed quadratically, $Y = X^2$, then, provided that RVs are statistically independent, their covariance equals

zero, and the variance equations look like

$$D_X = E_{X^2} - E_X^2, \tag{3a}$$

$$D_{X^2} = E_{X^4} - E_{X^2}^2. \tag{3b}$$

They are satisfied for arbitrary RVs, irrespective of their distribution laws. Therefore, by analogy with Eqs. (3), the following system of equations can be composed for RV transformations according to the laws $\left(\arccos X, \sqrt{X}\right) \Leftarrow X \Rightarrow \left(X^2, \cos X\right)$:

$$D_{\sqrt{X}} + E_{\sqrt{X}}^2 = E_X, \tag{4a}$$

$$D_X + E_X^2 = E_{X^2}, \tag{4b}$$

$$D_{X^2} + E_{X^2}^2 = E_{X^4}, \tag{4c}$$

$$D_{\arccos X} + E_{\arccos X}^2 = \overline{(\arccos X)^2}, \tag{4d}$$

$$D_X + E_X^2 = \overline{X^2}, \tag{4f}$$

$$D_{\cos X} + E_{\cos X}^2 = \overline{(\cos X)^2}. \tag{4e}$$

Equations (4) are basic for the substantiation of ana-lytic relations used to estimate the variance and the-mathematical expectation of RV. They are valid for independent RVs[2] and do not depend on the type of probability distribution.

There is an opinion in the literature that the prob-lem can be solved, if the integrals of $E_{\sqrt{X}}$, E_X, E_{X^2}, E_{X^4}, $\overline{(\arccos X)^2}$, and $E_{\cos X}$ as the right-hand sides of Eqs. (4) can be expressed in terms of elementary functions. The author of works [1, 2] calculated the in-tegrated means $\overline{X^2}$, $\overline{X^4}$, and $\overline{\cos X}$ for the normally distributed, $N(m_X, \sigma_X)$, initial random variable X, for which random parameter values change within the limits

$$-\infty < m_X < +\infty, \quad \sigma > 0. \tag{5}$$

The corresponding calculations were made, by using tabulated integrals (3.462.2) and (3.896.2) taken from [3] (similar integration can be made by parts [8]). As a result, the following expressions were obtained:

$$E_X = m,$$

$$E_{X^2} = \sigma_X^2 + m_X^2,$$

$$E_{X^4} = 3\sigma_X^4 + 6\sigma_X^2 m_X^2 + m_X^4 = \tag{6}$$

$$= E_{X^2}^2 + 2\sigma_X^2(E_{X^2} + m_X^2),$$

$$E_{\cos X} = \exp\left(-\sigma_X^2/2\right)\cos m_X.$$

---

[2] For independent RVs, their covariance equals zero.

Then the system of equations (4) takes the form

$$D_{\sqrt{X}} = E_X - E_{\sqrt{X}}^2 \overset{Y=\sqrt{X}}{\Longleftrightarrow}$$
$$\overset{Y=\sqrt{X}}{\Longleftrightarrow} D_X = E_{X^2} - E_X^2 \overset{Y=X^2}{\Longleftrightarrow}$$
$$\overset{Y=X^2}{\Longleftrightarrow} D_{X^2} = 2D_X(E_{X^2} + E_X^2),$$
$$D_{\arccos X} = E_{\cos X} - E_{\arccos X}^2 \overset{Y=\arccos X}{\Longleftrightarrow} \qquad (7)$$
$$\overset{Y=\arccos X}{\Longleftrightarrow} D_X = E_{X^2} - E_X^2 \overset{Y=\cos X}{\Longleftrightarrow}$$
$$\overset{Y=\cos X}{\Longleftrightarrow} D_{\cos X} = \frac{1}{2}[1 - \exp(-D_X)] \times$$
$$\times \{1 - \exp(-D_X)\cos E_X\}.$$

This allows the following equations to be written down for the direct transformations $Y = X^2$ and $Y = \cos X$ of the random variable X:

$$D_X = E_{X^2} - E_X^2, \qquad (8a)$$
$$D_{X^2} = 2D_X(E_{X^2} + E_{X^2}^2), \qquad (8b)$$
$$D_X = E_{X^2} - E_X^2, \qquad (8c)$$
$$2D_{\cos X} - 1 = \left(\frac{E_{\cos X}}{E_X}\right)^4 \cos 2E_X - 2E_{\cos X}^2, \qquad (8d)$$

respectively. The corresponding solutions are [1, 2]: for Eqs. (8a) and (8b),

$$D_X = E_{X^2} - \sqrt{E_{X^2}^2 - D_{X^2}/2},$$
$$E_X^4 = E_{X^2}^2 - D_{X^2}/2; \qquad (9)$$

and, for Eqs. (8c) and (8d),

$$D_X = \ln\left(\frac{\cos^2 E_X}{E_{\cos X}^2}\right),$$
$$E_{\cos X} = E_X \exp(-D_X/2). \qquad (10)$$

The problem of finding the means

$$\overline{\sqrt{X}} = C \int_{-\infty}^{+\infty} \sqrt{x} \exp\left(-\frac{z^2}{2}\right) dx$$

and

$$\overline{\arccos x} = C \int_{-\infty}^{+\infty} \arccos x \cdot \exp\left(-\frac{z^2}{2}\right) dx, \qquad (11)$$

where $C = \frac{1}{\sqrt{2\pi\sigma_X^2}}$ and $z = \frac{x-m_X}{\sigma_X}$, for the normally $N(m_X, \sigma_X)$-distributed initial random variable X was not solved in works [1, 2]. Instead, in order

to determine $E_{\sqrt{X}}$, $D_{\sqrt{X}}$, $E_{\arccos X}$, and $D_{\arccos X}$ in Eqs. (7) (or their solutions (10) and (11)), the cited author proposed a scheme for reducing the indices following the algorithm

$$X \to \sqrt{X}, \quad X^2 \to X. \qquad (12)$$

Then Eqs. (9) can be used to compose the following system of equations:

$$D_{\sqrt{X}} = E_X - E_{\sqrt{X}}^2, \qquad (13a)$$
$$D_X = 2D_{\sqrt{X}}(E_X + E_{\sqrt{X}}^2), \qquad (13b)$$

with radicals (1) being its solutions. No statistical substantiation of algorithm (12) was given at that.

The plot of the solutions of Eqs. (13b) are parabolas with a maximum with the coordinates $(E_X^2, 0)$ for Eq. (1a) and $(E_X^2, D_{\sqrt{X}} = E_X)$ for Eq. (1b). This means that the restriction

$$2E_X^2 \geqslant D_X \qquad (14)$$

is imposed on the values of the quantities E_X and D_X, so that it is impossible to use a normally $N(m_X, \sigma_X)$-distributed initial random variable X with the mathematical expectation in the interval

$$m_X \leqslant \frac{\sigma_X}{\sqrt{2}}. \qquad (15)$$

There exists a similar restriction for the variance $D_{\arccos X}$ [1, 2]:

$$\sqrt{2}\sigma \leqslant 1 - E_y \leqslant 1 - E_y^2 \to$$
$$\to \sqrt{2}\sigma \leqslant 1 \leqslant 1 + E_y \to \sigma \leqslant 1/\sqrt{2}. \qquad (16)$$

Restrictions (14)–(16) contradict the initial condition of the problem about the random choice of the normally $N(m_X, \sigma_X)$-distributed initial random variable X within intervals (5) of the parameters E_X and D_X.

Let us present some probabilistic statistical arguments testifying to the incorrectness of the index reduction scheme. Nonlinear transformations, including those of type (12), change the form of distribution functions for the probability densities of transformed RVs [4–6]. The author of works [1, 2] carried out a standard calculation of means of the type E_X, D_X, E_{X^2}, D_{X^2}, and E_{X^4} for the random variable $X \in N(m_X; \sigma_X^2)$. On their basis, he composed an equation of type (4), making no analysis of the statistical regularities of transformed RVs.

Actually, in order to substantiate the equation for the variance of a transformed RV,

$$D_Y = \int\limits_{-\infty}^{\infty} \left(y - \overline{Y}\right)^2 f_Y(y)dy =$$

$$= \int\limits_{-\infty}^{\infty} \left(y^2 + \overline{Y}^2 - 2y\overline{Y}\right) f_Y(y)dy =$$

$$= \int\limits_{-\infty}^{\infty} y^2 f_Y(y)dy + \overline{Y}^2 \int\limits_{-\infty}^{\infty} f_Y(y)dy - 2\overline{Y} \int\limits_{-\infty}^{\infty} y f_Y(y)dy =$$

$$= \overline{Y^2} + \overline{Y}^2 \int\limits_{-\infty}^{\infty} f_Y(y)dy - \left(2\overline{Y}\right)^2, \qquad (17)$$

as well as its possible solution, the set of roots of the transformation function should be analyzed with respect to their existence. In addition, the regions, where this function is monotonic, should be determined, and, if needed, the normalization condition for the function $f_Y(y)$ describing the distribution of probability density for the transformed RV should be corrected by the normalizing factor in the term $\overline{Y}^2 \int_{-\infty}^{\infty} f_Y(y)dy$ of Eq. (17). As an example, the nonlinear transformation function $y = x^2$ is two-valued in the interval $(-\infty, +\infty)$ and has two roots: $X_1 = -\sqrt{Y}$ in the $x < 0$ range of values and $X_2 = +\sqrt{Y}$ in the $x > 0$ one. At the same time, the function of the radical type, $y = \sqrt{x}$, has only one root in the $x > 0$ range of values, and so forth.

To summarize, it is worth emphasizing the following point. Relations (1) agree with the approximate method, which is known in statistics for estimating transformed random variables Xs and is based on the theorem (see Theorem 2.18 in work [10])

$$E_{\sqrt{X}} \cong \sqrt{E_X}, \quad D_Y \cong D_X \left(\frac{dy}{dx}\right)^2 \bigg|_{X=E_X}, \qquad (18)$$

only in the limiting case $E_Y \gg D_Y$. So, if $Y = \sqrt{X}$, then $E_{\sqrt{X}} \cong \sqrt{m_X}$ and $D_{\sqrt{X}} \cong \frac{\sigma_X^2}{4m_X}$. Similarly, for other transformations,

$$Y = X^2 - E_{X^2} \cong m_X^2,$$
$$D_{X^2} \cong 4E_X^2 D_X = 4m_X^2 \sigma_X^2,$$

and

$$Y = \cos X - E_{\cos X} \cong \cos(E_X) = \cos(m_X),$$
$$D_{\cos X} = \sin^2 E_X D_X = \sin^2 m_X \sigma_X^2.$$

The case $E_Y \gg D_Y$ has a low practical importance. In the course of RV processing, the distribution curve $f_Y(y)$ is mainly centered in a vicinity of the sample mean. This procedure makes it possible, by optimizing statistically independent distribution parameters (for a normal two-parameter distribution, these are E_Y and D_Y), to apply standard algorithms for the development of a statistical model for a researched physical system.

3. Conclusions

1. The scheme of index reduction was proposed in works [1, 2] without a statistical substantiation. The obtained relations for the estimation of the mean of a normally distributed random variable agree with the approximate method known in the literature only in the limiting case $E_Y \gg D_Y$.

2. The problem formulated in works [1, 2] is challenging and has a practical value. Since it has not been solved properly, it should be studied further.

1. G.G. Rode. Propagation of measurement errors and measured means of a physical quantity for the elementary functions $\cos x$ and $\arccos x$. *Ukr. J. Phys.* **61**, 345 (2016).

2. G.G. Rode. Propagation of measurement errors and measured means of physical quantities for the elementary functions x^2 and $\sqrt{x}$. *Ukr. J. Phys.* **62**, 184 (2017).

3. I.S. Gradshtein, I M. Ryzhik. *Table of Integrals, Series, and Products* (Academic Press, 1980).

4. J. Mande. *The Statistical Analysis of Experimental Data* (Dover, 1964).

5. E. Suhir. *Applied Probability for Engineers and Scientistics* (McGraw-Hill, 1997).

6. T. Koski. *Lecture Notes: Probabiity and Random Processes at KTN for sf2940 Probability Theory* (KTN Royal Institute of Technology, 2017).

7. M.M. Dorozhovets. *Processing of Measurement Results* (Nat. Univ. "L'viv. Politekhnika", 2007) [ISBN: 978-966-553-640-6] (in Ukrainian).

8. L.M. Malyarets, A.V. Ignachkova, L.D. Shyrokorad. *Probability Theory and Mathematical Statistics in Exercises, Examples, and Problems* (Kharkiv Nat. Econom. Univ., 2010).

9. A.P. Prudnikov, Yu.A. Brychkov, O.I. Marichev. *Integrals and Series, Vol. 1: Elementary Functions* (Gordon and Breach, 1986).

10. Z. Pawlowski. *Wstęp do Statystyki Matematycznej* (Panstwowe Wydawnictwo Naukowe, 1965) (in Polish).

П. Кособуцький

СТОСОВНО МОДЕЛЮВАННЯ
МАТЕМАТИЧНОГО СПОДІВАННЯ
І ДИСПЕРСІЇ ВИБІРОК ГАУСОВО
РОЗПОДІЛЕНИХ ВИПАДКОВИХ ВЕЛИЧИН

Р е з ю м е

Проведений аналіз виведення запропонованих у роботах *УФЖ* **6**, №4, 355–362 (2017) та *УФЖ* **62**, №2, 184–190 (2017) правил переносу середнього і дисперсії фізичних величин, функціонально пов'язаних перетвореннями $X^2, \cos X, \sqrt{X}, \arccos X$. Показано, що обґрунтування "правил переносу похибок" не ґрунтуються на базових положеннях теорії ймовірностей і математичної статистики, а запропоноване пониження індексів $X \to \sqrt{X}; X^2 \to X$ в коренях квадратних рівнянь, покладеного в основу запису формул переносу, обмежує значення параметрів нормального розподілу m_X, σ_X.