https://doi.org/10.15407/ujpe63.7.645

P. KOSOBUTSKYY

National University "Lviv Polytechnic"
(12, S. Bandera Str., Lviv 79013, Ukraine; e-mail: petkosob@gmail.com)

OPTIMAL REGULARITIES OF THE NORMAL DISTRIBUTION FOR ESTIMATING THE SAMPLE STATISTICS OF THE RESULTS OF A PHYSICAL EXPERIMENT

Basic probabilistic principles for the formation of the normal distribution for random fluctuations of physical quantities under the action of independent random factors on the physical system have been formulated. The emphasis is made on the integrated approach to the probabilistic statistical analysis of a sample of experimental results.

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

1. Introduction

It is known that Laplace's determinism is inherent in such a complex system as an ideal gas of N particles in the state of thermal equilibrium. But, as was shown by Krylov, the collisions of particles with one another make the motion of a multiparticle system unstable by Lyapunov: the motion becomes chaotic, and a deterministic model is not suitable for its description.

A true role of random factors in the formation of the equilibrium state in a complex system was understood as long ago as by Maxwell, who substantiated the physical meaning of the probability function. The correctness of its essence was confirmed in the experiments carried out by Brown and Perrin, who established that, although the average value of the collective displacement parameter Δx for a walking particle equals zero, $\langle \Delta x \rangle = 0$, its average quadratic value differs from zero, $\left\langle (\Delta x)^2 \right\rangle \neq 0$. After the fundamental theoretical works by Maxwell and Boltzmann, the importance of the role of probabilistic approaches in physical modeling became ultimately clear.

The central limit theorem is a cornerstone of the probability theory and mathematical statistics. The theorem asserts that if a measurable physical quantity is subjected to the action of plenty of random and independent (or weakly dependent) factors, each of which does not give a dominant contribution to the common result, then the statistical distribution of the sample values tends to the normal (Gaussian) distribution characterized by the following parameters: the mathematical expectation (ME) m_X or $M[X]$, and the mean square deviation (MSD) σ_X. The variance $D_X = \sigma_X^2$ as a characteristic of the random variable (RV) spread with respect to the ME has a quadratic dimension, which is not very convenient for statistics. Therefore, the parameter σ_X was introduced [1].

The normal distribution $N(m_X, \sigma_X)$ is most often used to construct physical models in the framework of the continuous RV method, and this distribution is considered in a lot of works [2]. The corresponding theoretical model is the best one developed analytically. It serves as a basis for the theory of statistical error estimates [3], the Alentsev–Fock method of spectrum resolution into separate modes [4], and probabilistic approaches to the study of the dynamics of linear and nonlinear systems [5] and quantum-mechanical phenomena [6, 7]. On its basis, the powerful resources for computer simulations of the dynamics of atomic-molecular compounds [8] and for other purposes [9] were created. On the basis of RVs with the Gaussian distributions, a whole direction in statistical optics was developed [10]. The normal distribution was also proved to take place in many chaotic systems such as the Boltzmann billiards, the Brownian motion in a harmonic potential, and so on.

The normal distribution is stable and can reproduce itself. Therefore, it was successfully used for the statistical simulation in modern high technologies,

such as laser cooling of atoms [11] and stochastic phenomena in radio engineering systems [12,13]. Furthermore, the Gaussian distribution belongs to distributions with "heavy tails" [1]. This circumstance allows one to assume, on the basis of the central limit theorem, that the sum of a sufficiently large number of small RVs is distributed normally, and the Gaussian distribution is a limiting case for a rather large number of distributions, irrespective of the regularities available for separate errors, because those specific regularities mutually compensate one another in the limit of their large number [14]. That is why the class of normal functions is most often chosen as an approximate model for the general set of random data; in this case, it is enough to evaluate only two parameters, m_X and σ_X, for their sample. The mixtures of distributions on their basis, which possess the infinite differentiability, smoothness, and so forth are widely used at various simulations [15–19].

The aim of this work is to demonstrate that the statistical analysis of the samples of experimental results and the choice of a theoretical model [20, 21] is not always accompanied by the proper attention to the basic principles of the probability theory and mathematical statistics. This point was marked earlier by the author in work [22].

2. Basic Principles Used to Verify the Normal Distribution of the Values Measured for a Quantity in the Framework of a Given Physical Model

Physical quantities or conditions of their measurement always undergo random fluctuations. Therefore, it is more correct to talk about their statistically averaged values. For this purpose, various probabilistic methods for processing the measurement results are applied, including those developed on the basis of linear and nonlinear RV transformations. Fluctuations are random deviations of macroscopic variables from their average (e.g., thermodynamically equilibrium) values. In macroscopic systems, fluctuations are a result of a large number of degrees of freedom, which have a macroscopic origin. In microscopic systems, fluctuations are associated with the corpuscular (atomic) structure of matter and electricity (the smallest piece of electricity is equal to one electron charge).

Fluctuations in atomic systems are induced by thermal vibrations of microparticles – electrons, ions, and so on – which are associated with the Brownian motion of atoms and molecules, molecular light scattering in the medium, thermal noise in electrical circuits, thermal radiation emission, and the fractional effect in electronics (it is induced by electron flux fluctuations during transfer processes). A chaotic character is inherent in the process of magnetic domain reversal in ferromagnetic media.

Under real measurement conditions, the distribution of the growing number of measured physical data always tends to the normal one. Here are the most important arguments for why the Gaussian law should be applied to the physical modeling:

1) a distribution of the type $N(m_X; \sigma_X^2)$ is a good mathematical model to describe phenomena and processes, by using random numbers;

2) arbitrary linear combinations of normally distributed RVs are also distributed normally;

3) a stochastic Gaussian process can be completely modeled theoretically with the use of only the first and second moments, which is not enough for other distributions;

4) if a normally distributed random signal is transmitted through a linear system, the output signal is also distributed normally.

Therefore, in the absence of the theoretical and experimental information on the character of a RV distribution (e.g., the measurement noise), it is most often assumed that a continuous variable is distributed normally, and a similar discrete one is distributed according to the Poisson law, which tends to the Gaussian distribution with increasing the sample size.

However, other distributions related to the normal one are also relevant in statistics. In particular, these are the χ^2-distribution, Student's t-distribution, and the Fisher–Snedecor F-distribution. If each of n independent random variables $\xi_1, \xi_2, ..., \xi_n$ is distributed

[1] The essence of the distribution with a "heavy tail" consists in that events considered as not frequent – e.g., the impact of aircrafts – are not actually rare from the viewpoint of the secure system functioning, so that their probability cannot be neglected. In the Gaussian model, a hypothetical incident is absent. At the same time, in the power-law model, although it is rare, but possible: $P(X > x) \cong x^{-\alpha}$ at $x \to \infty$, where the parameter $\alpha > 0$ describes the rate of distribution "tail" fading. The asymptotic distribution model is hyperbolic. Therefore, it is slower than the exponential one, and most of the probabilistic measure can be concentrated in the distribution tail.

according to the standard law $N(0,1)$, then the RVs of the type $\eta = \xi_1^2 + \xi_2^2 + ... + \xi_n^2$ are already distributed accroding to the χ^2 law with n degrees of freedom. If each of $n+1$ independent random variables $\xi_0, \xi_1, \xi_2, ..., \xi_n$ is distributed according to the $N(0,\sigma^2)$ law, then the RVs of the type

$$\eta = \frac{\xi_0}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2}}$$

are distributed according to Student's law with n degrees of freedom. If each of $n+m$ independent random variables $\xi_1, \xi_2, ..., \xi_n, \xi_{n+1}, ..., \xi_{n+m}$ is distributed according to the $N(0,\sigma^2)$ law, the RVs of the ratio type

$$\eta = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2}{\frac{1}{m}\sum\limits_{i=n+1}^{n+m}\xi_i^2}$$

are distributed according to the Fisher–Snedecor law with n and m degrees of freedom. An RV can often be converted into a normally distributed one, by using simple mathematical transformations. For instance, if a continuous random variable ξ is distributed lognormally with the parameters m_X and σ_X^2, then the random variable $\eta = \ln\xi$ has a normal two-parameter lognormal distribution $LN(m_X, \sigma_X^2)$, the $LN2$-distribution. If the $LN2$-distribution is also not normal, then a limiting parameter Σ is introduced before taking the logarithm. This parameter transforms the distribution into a three-parameter lognormal $LN3$-distribution [23], for which the random variable $Y = \ln(X - \Sigma)$ has a normal distribution.

The analysis of experimental data in order to verify their normal distribution by evaluating the variance $\sigma_X^2 = D_X$ and the displacement m_X from a sample of measurement data $\Omega_n = \{x_1, x_2, ..., x_n\}$ is recommended to perform according to the following algorithm. Using the method of moments, the sample parameters

$$\overline{X} = m_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

$$S^2 = \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 \qquad (1)$$

should be calculated. The calculated arithmetic average $\overline{X}$ is a consistent, unshifted, and effective estimate of the ME. The sample variance is estimated, by using formula (1), which is shifted, since $M\left[S^2\right] = \left(1 - \frac{1}{n}\right)\sigma_X^2 \neq \sigma_X^2$. The estimate shift appears due to the Boltzmann factor $1 - \frac{1}{n}$, but its role is nullified for sample sizes $n > 30$. If $n < 30$, the non-shifted estimate should be calculated by the formula

$$D_X = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2.$$

Therefore, for a normally distributed sample, formula (1) is not efficient. However, if the mathematical expectation m_X is known in advance, estimate (1) is unshifted, consistent, and effective.

The quality of estimates obtained for the parameters is evaluated on the basis of the estimate accuracy measure. Among such measures, the most standard one is the mean-square error. This measure generates the corresponding criterion for the estimate optimum, the minimum of the mean-square error. For unshifted estimates, the variance is a measure of their accuracy, and the variance minimum is the criterion of optimum.

But the parameters describing the position m_X of the extremum on the distribution density curve $f_X(x)$ and the distribution spread σ_X do not determine a physical quantity, but they characterize the form of the probability density function $f_X(x)$ for its random values. These notions are convenient for the normal, exponential, trapezoidal, and other distributions that use the method of moments, which is valid only if the integrals in corresponding calculations do not diverge. More universal is the concept of a distribution center, which is defined as the center of gravity of the distribution or the 50% quantile. The center of gravity of the RV distribution is a mechanical analog of the ME, if the probabilities of the values are taken as the masses of points.

In physics on the basis of the center-of-gravity model, it is substantiated that any body in an indefinite state tends to an equilibrium state. In exactly the same way, an arbitrary RV tends to its equilibrium (in the sense of the mean value), provided a significant number of measurements. Such an approach demands that only the zero-order moment and a parameter characterizing the distribution width should exist. For a symmetric distribution, the center of gravity coincides with its mode. But unlike the mode, the concept of the center of gravity of a distribution is valid for all distributions. For instance,

there is no ME for the Cauchy distribution, whereas the definition of the center of gravity for its distribution curve is eligible. The mode is also absent for a uniform distribution.

Fuctuations in chaotic systems are characterized by the dispersion, which relates the first and second initial moments to each other and characterizes the fluctuation intensity, so that the case $D_X = 0$ has no physical sense. For example, in an electric circuit, the dispersion mainly corresponds to the average electric power released on an active electric resistance by the alternating component of the electric voltage or current. In an alternating electric circuit, the parameter σ_X corresponds to the voltmeter or ammeter readings, if the constant component of electric signal is eliminated by inserting a capacitor into the circuit.

From the viewpoint of a theoretical modeling of the dispersion in a physical system, it is important to establish whether the physical quantities, whose values change randomly, are statistically dependent or independent. In the infinite interval of the RV value spread, $X \in [-\infty; +\infty]$, the variance is calculated as the second central moment

$$D_X = \int\limits_{-\infty}^{\infty} \left(x - \overline{X}\right)^2 C_X f_X(x) dx = \\ = \int\limits_{-\infty}^{\infty} \left((x)^2 + \overline{X}^2 - 2x\overline{X}\right) C_X f_X(x) dx \qquad (2)$$

and characterizes the spread of values with respect to the origin of the abscissa axis [2]. For statistically independent RVs, integral (2) equals

$$D_X = C_X \left\{ \int\limits_{-\infty}^{\infty} (x)^2 f_X(x) dx + \overline{X}^2 \int\limits_{-\infty}^{\infty} f_X(x) dx - \right. \\ \left. -2\overline{X} \int\limits_{-\infty}^{\infty} x f_X(x) dx \right\} = \overline{X^2} + \overline{X}^2 \int\limits_{-\infty}^{\infty} C_X f_X(x) dx - 2\overline{X}^2. \qquad (3)$$

Since the normalization condition has also to be satisfied in this case, the dispersion equation has the following form for statistically independent RVs:

$$D_X = \overline{X^2} - \overline{X}^2. \qquad (4)$$

Here, the mathematical expectation m_X of the random variable X was taken for the mean $\overline{X}$ as an unshifted, substantiated, and consistent estimate. Furthermore, if the distribution has a spread, this parameter is also an asymptotically normal estimate.

The statistical means $\overline{X}$ and $\overline{X^2}$ also play an important role in physical systems. For example, in the case of electric circuit, the mean square $\overline{X^2}$ can be associated with the time-averaged square of the random voltage or current. Then the mean square will be proportional to the average electric power released on an active resistance, $\sqrt{\overline{x(t)^2}}$. In electrical engineering, this value is referred to as effective.

On the other hand, when using probabilistic models for the physical modeling, one should take into account that, in effect, the magnitudes of physical quantities are not negative, so that the spread interval of their random values is actually semiinfinite: $X \in (0, +\infty)$. As the experience testifies, the probability distribution density $f_X(x)$ decreases proportionally to the $\exp\left(-x^2\right)$ at large deviations x from the mean m_X, with an extremum at the point $x = 0$. Therefore, more suitable for practical applications is the probability density distribution function, in which the random variable X is centered by subtracting the mean value m_X from the variable:

$$g(X) = X - m_X. \qquad (5)$$

In this case, the extremum of the exponential function $f_X(x) \approx \exp\left(-(x - m_X)^2\right)$ becomes shifted to the coordinate $x = m_X$ on the scale of the centered variable.

The change from the scale of the absolute values of the physical quantity X to the scale of the centered values $X - m_X$ does not distort the Gaussian curve, so that the centered variable can be normalized:

$$g(X) = U = \frac{X - m_X}{\sqrt{2}\sigma_X} = \sqrt{p_X}\,(X - m_X). \qquad (6)$$

This procedure makes it possible to use a dimensionless exponent in calculations and simplify the normal $N(m_X, \sigma_X)$ distribution to the standard one, $N(0, 1)$, with the zero ME and the variance equal to one. It is

[2] In the probability theory, the variance is a measure of the spread from the mean value. In mathematical statistics, this parameter characterizes the spread of quantitative values in a statistical sample with respect to the mean value, i.e. the ME of the squared deviation of a RV from its ME.

symmetric; its point of symmetry is the mode coordinate, whose value coincides with the center of gravity, i.e. the ME, and the corresponding odd central moments equal zero, although the inverse statement is not always true. In particular, the equality of the third central moment to zero is only a necessary condition for the distribution to be symmetric.

The normalization operation (6) does not change the coordinate of the maximum of the exponential function

$$\exp\left(-\frac{(x-m_X)^2}{\sigma_X^2}\right).$$

Therefore, the density function $f_X(x)$ has the following analytic form:

$$f_X(x) = \frac{C_X}{\sqrt{2\pi\sigma_X^2}} e^{-\left(\frac{x-m_X}{\sqrt{2}\sigma_X}\right)^2} = \\ = C_X\sqrt{\frac{p_X}{\pi}}\exp\left(-p_X(x-m_X)^2\right). \quad (7)$$

The semiinfinite interval $X \in (0, +\infty)$ for the normally distributed random variable X corresponds to the interval $U \in (-\sqrt{p}m_X, +\infty)$ for the centered normalized random variable U, with

$$x = 0 : u|_{x=0} = -\sqrt{p}m_X, \quad x = +\infty : u|_{x=+\infty} = +\infty. \quad (8)$$

The normalization integral to determine the constant C_X looks like

$$C_{X\in(0,+\infty)}\int_0^\infty f_X(x)dx = C_{X\in(0,+\infty)} \times \\ \times \int_0^\infty \sqrt{\frac{p_X}{\pi}}\exp\left(-p_X(x-m_X)^2\right)dx = \\ = \frac{C_{X\in(0,+\infty)}}{\sqrt{\pi}}\int_{-\sqrt{p_X}m_X}^\infty e^{-u^2}du = 1. \quad (9)$$

From the viewpoint of the geometric interpretation of a definite integral as an area under the curve $f_X(x)$,

$$\int_0^\infty f_{X\in(0,+\infty)}(x)dx < \int_{-\infty}^\infty f_{X\in(-\infty,+\infty)}(x)dx,$$

because the semiinfinite interval $X \in [0, +\infty)$ is obtained from the infinite interval $(-\infty, +\infty)$, by truncating all negative argument values. Therefore, the truncated normal distribution is applied to model the reliability of physical and engineering systems [24–27], the physical processes of charge transfer in electronic devices [28, 29], and other purposes.

Note that the theory of random processes with a truncated spread interval for RV values was developed as long ago as in works by Einstein and Smoluchowski [30, 31]. It was they who proposed one of the first mathematical algorithms of data processing to evaluate the distribution function and the probability density of empirical dependences, by analyzing a sample of experimental data. It is convenient to illustrate the relationship between the Brownian motion and the Gaussian distribution, by using the well-known Fokker–Planck equation (or the Kramers equation) [32, 33]

$$\frac{\partial}{\partial t}f(t,x) = D\frac{\partial^2}{\partial x^2}f(t,x). \quad (10)$$

This equation describes the motion of Brownian particles with the distribution density

$$f(t,x) = \frac{\partial}{\partial t}P(x,t)$$

under the action of random forces. Here, $P(x,t)$ is the probability to find a particle at the point with the coordinate x at the time t, and D is the diffusion coefficient. The corresponding solution of Eq. (10) looks like

$$f(t,x) = \frac{C}{\sqrt{4\pi D}}\frac{1}{\sqrt{t}}\exp\left(-\frac{x^2}{4Dt}\right). \quad (11)$$

Formally, solution (11) resembles the Gaussian function, if the quantity $2Dt$ is interpreted as the variance: $\sigma^2 = 2Dt$.

The Gaussian integrals for calculating the statistical means of $\overline{X}$ and $\overline{X^2}$ over a truncated interval of the random variable X are not expressed analytically in terms of elementary functions. A special nonelementary tabulated function of errors, $\mathrm{erf}(x)$, was introduced for their calculation [34]:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-\frac{u^2}{2}\right)du. \quad (12)$$

If an RV sample is described by the normal distribution with the standard deviation σ_X, the value of $\operatorname{erf}\left(\frac{m}{\sqrt{2}\sigma}\right)$ is equal to the probability of that the RV value does not deviate from the mean by more than m_X. The error function is related to the standard distribution function $N(0,1)$ by means of the normalized Laplace function,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) du. \tag{13}$$

Making use of the function $\operatorname{erf}(x)$ and following the algorithm

$$\mathrm{P}\left(X \leqslant x_a\right) = \begin{cases} 0.5 - \operatorname{erf}\left(\frac{m_X - x_a}{\sigma_X}\right) & \text{for} \quad x_a \leqslant m_X \\ 0.5 + \operatorname{erf}\left(\frac{x_a - m_X}{\sigma_X}\right) & \text{for} \quad x_a \geqslant m_X \end{cases} \tag{14}$$

the so-called standard of 3σ confidence intervals has been elaborated, and the following rules have been formulated:

$$\begin{aligned} &\mathrm{P}\left(|X - m_X| \leqslant \sigma_X\right) = 2\operatorname{erf}(2) = 0.685, \\ &\mathrm{P}\left(|X - m_X| \leqslant \sigma_X\right) = 2\operatorname{erf}(3) = 0.991, \\ &\mathrm{P}\left(|X - m_X| \leqslant \sigma_X\right) = 2\operatorname{erf}(4) = 0.998, \end{aligned} \tag{15}$$

which allow the parameters of a physical model with random fluctuations to be simulated with a required confidence probability.

The importance of the Gaussian function in the statistical modeling of physical systems is confirmed by the Maxwellian probabilistic distribution known in physics. Despite that the Cartesian components of the velocity vector for the chaotic motion of particles in an equilibrium system are distributed normally, the distribution of the particle velocity magnitudes obeys the Maxwellian distribution

$$\begin{aligned} f(\vartheta_z) &= \sqrt{\frac{1}{\pi}\frac{m}{2k_{\mathrm{B}}T}} \exp\left(-\frac{m\vartheta_z^2}{2k_{\mathrm{B}}T}\right) = \\ &= \sqrt{\frac{p_{\vartheta_Z}}{\pi}} \exp\left(-p_{\vartheta_Z}\vartheta_z^2\right) \Rightarrow \quad p_{\vartheta_Z} = \frac{m}{2k_{\mathrm{B}}T}, \end{aligned} \tag{16}$$

where $p_{\vartheta_Z} = \frac{m}{2k_{\mathrm{B}}T}$. By equating function (16) and the Gaussian function, we obtain

$$\frac{1}{2\sigma_{\vartheta_Z}^2} = \frac{m}{2k_{\mathrm{B}}T},$$

so that

$$\sigma_{\vartheta_Z} = \frac{k_{\mathrm{B}}T}{m}.$$

Hence, the MSD of the velocity projection onto the corresponding direction increases proportionally to the absolute temperature and inversely to the particle mass.

The probability analogs of the center of mass

$$y_c = \frac{\sum y_k m_k}{\sum m_k}$$

and the moment of inertia with respect to the center of mass

$$I_{m_X} = \sum (y_k - a)^2 m_k$$

are the ME and the variance, respectively. Therefore, it is convenient to use an analogy between the optimization of dynamic regularities at the rotation of a system of material points with the masses m_k and arranged along a massless rod and the algorithm for minimizing the fluctuation processes (the dispersion) in a physical system that evolves toward its equilibrium state. According to this analogy, in the problem of optimizing the dynamics of rotational motion of the system of material points around a selected rotation axis, the optimum value of a parameter a describing its spatial arrangement is to be determined, for which the moment of inertia (and, therefore, the fluctuation dispersion) is minimum. In the rotating system, every point is located at the rod at the distance y_k reckoned from either of the rod ends taken at the reference point, with the rotation axis being perpendicular to the rod.

When rotating, the radius vector y_k describes a circle with the area πy_k^2. Let us calculate the distribution density of the circle area, $f_{S_k}(s_k)$, characterized by the parameters m_S and σ_S, if the y_k-values are scattered according to the normal law $N(m_Y, \sigma_Y)$:

$$y_k = \frac{1}{\sqrt{\pi}}\sqrt{s_k}, \quad \left|\frac{dy_k}{ds_k}\right| = \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{s_k}}. \tag{17}$$

The inverse function of $s_k = \pi y_k^2$ is two-valued. Therefore, from the $f_{Y_k}(y_k)dy_k = f_{S_k}(s_k)ds_k$, we obtain

$$f_{S_k}(s_k) = f_{Y_k}(y_k)\left|\frac{dy_k}{ds_k}\right| = f_{Y_k}(y_k)\frac{1}{2\sqrt{\pi\, s_k}}. \tag{18}$$

The absolute value of the derivative was taken to make allowance for the change of the derivative sign. Therefore,

$$f_{S_k}(s_k) = \frac{\sqrt{p_{S_k}}}{2\pi\sqrt{s_k}}\Big\{\exp\left(-p_Y(\sqrt{s_k} - \sqrt{\pi}m_Y)^2\right) + \\ + \exp\left(-p_Y(\sqrt{s_k} + \sqrt{\pi}m_Y)^2\right)\Big\}, \quad (19)$$

where

$$p_Y = \frac{1}{2\pi\sigma_Y^2}, \quad p_{S_k} = \frac{1}{2\sigma_{S_k}^2}.$$

The area has positive values, and the statistical mean m_{S_k} equals

$$m_{S_k} = \overline{S_k} = \\ = \frac{\int\limits_0^{+\infty} s_k \left\{e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2}\right\} ds_k}{\int\limits_0^{+\infty} \left\{e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2}\right\} ds_k} = \\ = \frac{\sqrt{2}}{\pi\sigma_Y}\left\{\frac{\sqrt{\pi}}{4p_Y\sqrt{p_Y}} + \pi\, m_Y^2 \frac{\sqrt{\pi}}{2\sqrt{p_Y}}\right\} = \pi\left(\sigma_Y^2 + m_Y^2\right). \quad (20)$$

Let us analogously calculate the statistical mean $\overline{S_k^2}$:

$$\overline{S_k^2} = \\ = \frac{\int\limits_0^{+\infty} s_k^2 \left\{e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2}\right\} ds_k}{\int\limits_0^{+\infty} \left\{e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2}\right\} ds_k} = \\ = \pi^2\left(3\sigma_Y^4 + m_Y^4 + 6\sigma_Y^2 m_Y^2\right). \quad (21)$$

Hence, the MSD equals

$$\sigma_{S_k} = \pi(3\sigma_Y^4 + m_Y^4 + 6\sigma_Y^2 m_Y^2 - m_Y^4 - \\ - \sigma_Y^4 - 2\sigma_Y^2 m_Y^2)^{1/2} = \sqrt{2}\pi\sigma_Y(\sigma_Y^2 + 2m_Y^2)^{1/2}. \quad (22)$$

Since the mean-square deviation σ_{S_k} and the mathematical expectation m_{S_k} are no more independent of each other, the both being dependent on the statistically independent m_Y and σ_Y, it seems that the state with minimum MSD can be optimized only through the optimization of the m_Y- and σ_Y-values of the Gaussian RVs.

Finally, let us point out the following. When modeling a physical system with random influences, there arises a challenging issue concerning the sample volume of experimental data that was taken for the statistical analysis. According to the Lyapunov theorem [35], the probabilistic calculations are inapplicable, if the sample size $n \leq 20 \div 30$. If the sample size is small ($n < 30$), the estimate of the distribution MSD is unreliable. In this case, Student's distribution is used. This distribution is used to calculate the probability of that the deviation of X from m_X does not exceed a certain value $\sigma_{\overline{X}}$:

$$p\left(-\sigma_{\overline{X}} < \overline{X} - m_X < +\sigma_{\overline{X}}\right) = \\ = \frac{1}{\sigma_{\overline{X}}}\sqrt{\frac{2}{\pi}} \int\limits_{m_X-\sigma_{\overline{X}}}^{m_X+\sigma_{\overline{X}}} \exp\left(-\frac{\left(\overline{X} - m_X\right)^2}{2\sigma_{\overline{X}}^2}\right) d\overline{X}. \quad (23)$$

Then a deviation of the arithmetic mean from the true value of the measured quantity does not exceed the following value:

$$\Delta p = t_p S_{\overline{X}} = t_p \frac{S_X}{\sqrt{n}}. \quad (24)$$

with the accuracy of the approximate equality $\overline{X} \approx \\ \approx m_X$ growing, as the n-value increases. However, this conclusion does not imply that, in order to increase the accuracy of final results, one must, first of all, try to increase the accuracy of separate measurements. Student's coefficients t_p are tabulated for various probability values P and degrees of freedom $k = n - 1$. The limiting cases of Student's distribution are the Cauchy distribution at $k = 1$ (there is no MSD and variance, because the integral diverges, and no efficient estimation of the distribution center can be made; therefore, the median is used to determine the center) and the Gaussian distribution at $k \to \infty$. The Cauchy distribution (or the Lorentz distribution, or the Wigner–Breit distribution) plays an important role at the physical modeling. In particular, in the well-known Ohm and Hooke (or other) laws, provided that the volume of independent measurements is sufficient, the relationship between the voltage drop U_R across an active resistance R and the current through the latter or between the deformation Δx of a body and the force F_X applied to the latter is linear, which provides the normal distribution of their fluctuations. At the same time, the distributions of random changes of the proportionality coefficients $R = \frac{U_R}{I_R}$ and $k = \frac{F_X}{\Delta x}$ obey the Cauchy distribution. The proper attention is not always paid to this circumstance.

3. Conclusions

From the viewpoint of creating a perfect statistical portrait of a physical model, the regularities discussed above can serve as a first step to establishing the distribution origin of the values of stochastic quantities obtained as a result of experimental measurements. But this step is mandatory.

The estimation of a distribution law to a given confidence probability requires that the spectrum of statistical probabilistic researches to be expanded, including the overlapping degree of distribution intervals [36, 37], the parameter of the distance between them [38], simulation with the help of generalized [39, 40] and mixed [41, 42] distribution methods, and others. It is the complex approach to processing the experimental data that made it possible to successfully imply such high-tech methods of complex object visualization as the atomic force and tunneling microscopies, the nuclear magnetic resonance spectroscopy [43, 44], and others.

However, by no means, one should come to a conclusion that the stochastic regularities of physical quantities have to satisfy the normal distribution. There are physical processes, in which the stochastic regularities confirm the applicability of other laws, e.g., the exponential law [3] for the radioactive decay [45], the power law in the phenomena of synergetics and fractal dynamics [46,47], the lognormal law in the optimal forecasting of mineral localization on the basis of geophysical observations [48], and so forth. But this is a topic of another work.

1. G. Bohm, G. Zech. *Introduction to Statistics and Data Analysis for Physicists* (Deutsches Electronen-Synchrotron, 2010).
2. J.K. Patel, C.B. Read. *Handbook of the Normal Distribution* (Dekker, 1982) [ISBN: 0-8247-1541-1].
3. D.S. Sivia, J. Skilling. *Data Analysis: A Bayesian Tutorial* (Oxford Univ. Press, 2006) [ISBN: 978-0-19-856831-5].
4. M.V. Fock. Resolution of complex spectra into separate bands using theAlentsev method. *Trudy Fiz. Inst. Akad. Nauk SSSR* **63**, 3 (1972) (in Russian).
5. A.S. Gusev. *Probabilistic Methods in the Mechanics of Machines and Structures* (Moscow State Technical University, 2009) (in Russian) [ISBN: 978-5-7038-3160-1].
6. E. Wichmann. *Quantum Physics. Berkeley Physics Course* (McGraw-Hill, 2010), Vol. 4 [ISBN: 0070048614].
7. F. Reif. *Fundamentals Statistics and Thermal Physics* (Waveland Press, 1965) [ISBN: 1-57766-612-7].
8. http://www.d.umn.edu/∼psiders/courses/chem5650/gaussviewtutorial /tutorial.html.
9. V.M. Zolotarev. *One-Dimensional Stable Distributions* (Amer. Math. Soc., 1986).
10. P.V. Korolenko, M.S. Maganova. *Fundamentals of Statistical Methods in Optics* (Universitetskaya Kniga, 2010) (in Russian) [ISBN: 978-5-91304-126-5]..
11. F. Bardou, J.-Ph. Bouchaud, A. Aspect, C. Cohen-Tannoudji. *Lévy Statistics and Laser Cooling. How Rare Events Bring Atoms to Rest* (Cambridge Univ. Press, 2001).
12. O.N. Maslov. *Stable Distributions and Their Application in Radio Engineering* (Radio i Svyaz, 1994) (in Russian) [ISBN: 5-256-01187-1].
13. G. Samorodnitsky, M.S. Taqqu. *Stable Non-Gaussian Random Processes* (Chapman and Hall, 1994).
14. S. Holmes. *Sums of Random Variables: Statistics 116* (Stanford University, 1998).
15. D. Titterington, A. Smith, U. Makov. *Statistical Analysis of Finite Mixture Distributions* (Wiley, 1985) [ISBN: 0-471-90763-4].
16. N.A. Carreira-Perpiñán. Mode-finding for mixture of Gaussian distributions. *IEEE Trans. Pattern Analys. Mach. Intell.* **22**, 1318 (2000).
17. Ya.A. Fomin, G.R. Tarlovskii. *Statistical Theory of. Pattern Recognition* (Radio i Svyaz, 1986) (in Russian).
18. J.A. Hartigan. Distribution Problems in Clustering. In *Classification and Clustering,* edited by J. Van Ryzin (Academic Press, (1977), p. 45.
19. N.N. Aprausheva, I.A. Gorlach, A.A. Zhelnin, S.V. Sorokin. An experiment on automated statistical recognition of clouds. *J. Comput. Math. Math. Phys.* **38**, 1715 (1998).
20. M. Podrygalo, O. Isakova, A. Korobko. A new way to evaluate the coincidence of the results of theoretical and experimental studies *Metrolog. Prylady* **5**, 48 (2017) (in Ukrainian).
21. D.Z. Shmatko, O.V. Kochneva. Calculation of the probability of defect determination in pieces by non-destructive control. *Matematych. Model.* No. 1, 51 (2017) (in Ukrainian).
22. P. Kosobutskyy. On the simulation of the mathematical expectation and variance of samples for Gaussian-distributed random variables. *Ukr. Fiz. Zh.* **61**, 827 (2017) (in Ukrainian).
23. J.R. Stedinger, R.M. Vogel, E. Foufoula-Georgiou. Frequency anlysis of extreme events. In *Handbook of Applied Hydrology, Part 3*, edited by D. Maidment (McGraw-Hill, 1993), ch. 18 [ISBN: 978-0-07-171177-7].
24. A. Hald. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Scand. Actuar. J.* No. 1, 119 (1949).
25. B. Gnedenko, I. Ushakov. *Probabilistic Reliability Engineering* (Wiley Interscience, 1995) [ISBN: 0-471-30502-2].

[3] However, if the experiment is repeated n times, and the decay constant is determined every time, its statistics tends to the Gaussian distribution at $n \to \infty$.

26. V.R. Matvievsky. *Reliability of Engineering Systems* (MGIEM, 2002) (in Russian) [ISBN: 5-230-22198-4].
27. A. O'Connor, M. Modarres, A. Mosleh. *Probability Distributions Used in Reliability Engineering* (CRE University of Maryland, 2016) [ISBN: 978-0-9966468].
28. J. Xinzhang, L. Tao. An empirical formula for yield estimation from singly truncated performance data of qualified semiconductor devices. *J. Semicond.* **33**, 125008 (2012).
29. K. Gu, X. Jia, H. You, T. Liang. The yield estimation of semiconductor products based on truncated samples. *Int. J. Metrol. Qual. Eng.* **4**, 215 (2013).
30. A. Einstein. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Ann. der Phys.* **17**, 549 (1905).
31. A. Einstein, M. Smoluchowski. *Brownian Motion. Collection of Papers* (ONTI-GROTL, 1936) (in Russian).
32. H. Risken, T. Frank. *The Fokker–Planck Equation: Methods of Solutions and Applications* (Springer, 1989) [ISBN: 0-387-50498-2].
33. I.I. Gikhman, A.V. Skorokhod. *The Theory of Stochastic Processes I* (Springer, 2004) [ISBN: 3-540-20284-6].
34. E. Ng, M. Geller. A table of integrals of the error functions. *J. Res. Natl. Bureau Stand. B* **73B**, 1 (1969).
35. Sh. Sulaberidze. *Methods for Analyzing and Processing Measured Quantity Values* (Baltian State Technical University, 2013) (in Russian) [ISBN: 978-5-85546-742-0].
36. E. Inman, E. Bradley. The owerlapping coefficient as a measure of the agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun. Stat. Theor. Methods* **18**, 3851 (1989).
37. Z. Karian, E. Dudewicz. *Handbook of Fitting Statistical Distributions with R* (Taylor and Francis, 2011) [ISBN: 978-1-58488-711-9].
38. D.C. Dowson, B.V. Landau. The Fréchet distance between multivariate normal distributions. *J. Multivar. Anal.* **12**, 450 (1982).
39. M. Varanasi, B. Aazhang. Parametric generalized Gaussian density estimation. *J. Acoust. Soc. Am.* **86**, 1404 (1989).
40. S. Nadarajah. A generalized normal distribution. *J. Appl. Stat.* **32**, 685 (2005).
41. J. Behboodian. On a mixture of normal distributions. *Biometrica* **57**, 215 (1970).
42. N.N. Aprausheva, S.V. Sorokin. *Notes on Gaussian Mixtures* (Computation Center of the Russian Academy of Sciences, 2015) (in Russian).
43. B. Voigtlander. *Scanning Probe Microscopy. Atomic Force Microscopy and Scanning Tunneling Microscopy* (Springer, 2015) [ISBN: 978-3-662-45239-4].
44. N.M. Sergeev. *NMR Spectroscopy* (Moscow State University, 1981) (in Russian).
45. E. Rutherford, F. Soddy. A comparative study of the radioactivity of radium and thorium. *Phil. Mag.* **5**, 445 (1903).
46. T. Akhromeeva, S. Kurdyumov, G. Malinetskii, A. Samarskii. *Structures and Chaos in Nonlinear Media* (Fizmatlit, 2007) (in Russian) [ISBN: 978-5-9221-0887-4].
47. M. Schroder. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (Freeman, 1991) [ISBN: 0-7167-2136-8].
48. A.A. Nikitin. *Statistical Methods for Distinguishing Geophysical Anomalies* (Nedra, 1979) (in Russian).

Received 14.03.18.
Translated from Ukrainian by O.I. Voitenko

П. Кособуцький

ОПТИМАЛЬНІ ЗАКОНОМІРНОСТІ
НОРМАЛЬНОГО РОЗПОДІЛУ З ТОЧКИ ЗОРУ
ОЦІНКИ СТАТИСТИКИ ВИБІРКИ РЕЗУЛЬТАТІВ
ФІЗИЧНОГО ЕКСПЕРИМЕНТУ

Р е з ю м е

На підставі аналізу літературних джерел синтезовані базові ймовірнісні і принципи формування нормального розподілу випадкових розсіянь значень фізичних величин в умовах незалежних випадкових дій на фізичну систему. Зроблений наголос на комплексному підході ймовірнісно-статистичного аналізу вибірки результатів експериментальних вимірювань.