Probability, computers, and the processing of experimental data

V. N. Tutubalin

M. V. Lomonosov State University, Moscow
(Submitted 15 February 1993)
Usp. Fiz. Nauk **163**, 93–109 (July 1993)

A discussion is presented in which it is considered how properly to evaluate the results of a statistical analysis of the data of physical experiments and how, in particular, to teach statistical methods to students. The case is made that hopes traditionally placed on statistical procedures and methods are often betrayed. Ultimately, the reason is that usual probability model of errors of observation are, as a rule, invalid if considered as physical models. However, if the results of data analysis are evaluated in the frameword of a statistical paradigm, then it is entirely possible that they will be useful for physical applications.

1. INTRODUCTION

The occasion for writing this paper was my reading of the very interesting paper by Yu. I. Alimov and Yu. A. Kravtsov "Is probability a 'normal' physical quantity?"[1] On the whole, I am in basic agreement with the critical views regarding the application of probability methods as expressed in that paper. There are, however, certain differences in the general theoretical premises that lead to considerable differences in the practical conclusions. Briefly, these differences involve the following.

It is sinful, of course, to doubt that in fundamental physics probability is a "normal" physical quantity, i.e., a physical quantity is true and/or noble to the extent that physical values exist. There exists in addition an entire world of other physical phenomena different from the world of fundamental physics, in which there naturally arise statistical ensembles of various processes and averaging over an ensemble. With the aid of correlations, calculated by averaging over an ensemble, a new approach to randomness and determinateness is possible in this world (see the last part of Ref. 1 and also Refs. 2 and 3). Of course, in this world, too, probability is a normal physical quantity.

However, there is yet a third world—a world of errors in measurements and observations, which was basically the subject of Ref. 1. The concepts and methods of the theory of probability have been applied to this world for 200 years. It appears to me that historical experience shows quite definitely that in this world probability *is not* a normal physical quantity, since it does not have either truth or nobility in the required measure. If we assess the probability-statistical methods of analyzing experimental results from the point of view of a physical paradigm, then we are obliged to discard them all as being insufficiently reliable. At the same time, I am completely convinced that these methods are extremely helpful in the area in which they are applicable, and will try to demonstrate this by example in this paper. In other words, renouncing the application of many concepts and methods of mathematical statistics (which were put forth in Ref. 1), I propose to replace them by an approach in the framework of another paradigm, which I will attempt to describe in this paper.

I have already written a great deal about the applications of the theory of probability and recently the publishing house of Moscow State University has granted me what is a rare luxury in these times—that of publishing my book.[4] However, in the same recent years I have been given—I do not even know how to say it: the good fortune or luxury?—to work on statistical methods in a new field for me: clinical physiology. This field has proved to be, as is not strange in our times, a completely virgin field in the sense of understanding and applying statistical methods. Of course, each experimental mean is accompanied by its standard error, but these errors have been interpreted either incorrectly or not at all; but actually they contain valuable information that allows one to plan further investigations. As a result, although my opinion of the truth and nobility of probability methods still stands as before, my opinion of their usefulness has been greatly reinforced. This will be the import of a new publication.

The core of the discussion will be to pose the question of what part of the theory of probability and mathematical statistics should be included in the university curriculum. I am more than ever convinced that under the name "theory of errors" or "theory of analyzing experimental results" the student may be presented with a fantastic mixture of statements, some of which are incorrect mathematically, others are meaningless, and yet others are not supported by the experience of the historical development of science. It appears to me that the root cause of this situation is that the probability theory of errors was in fact conceived of by Gauss and Laplace as a physical theory, but as such it did not stand the test of time. Nonetheless, it was not discarded because it demonstrated its usefulness over and over again. Therefore, the result was some disorder and confusion in the concepts, to say nothing about the unavoidable human limitations, a topic I plan to discuss in some detail.

2. THE CONCEPT OF SCIENTIFIC UTILITY

According to the dictionary of Dal', a wise man is one that combines truth with utility. As with any human, I would of course like to be a wise man, but I know that I will not succeed. Specifically, I wish to combine the bitter

 1063-7869/93/070628-14$10.00

truth that probability models are not adequate in the analysis of observations with the conviction that they are very useful. What is meant by utility for science?

I found it difficult to choose the proper term: philosophy or religion. If we use the term "philosophy" in its original meaning "love of wisdom", then the term will be philosophy. On the other hand, the term "philosophy of science" is used in a number of scientific disciplines, each of which has its own particular questions, but, however, not the question of what is noble and useful. If, on the other hand, I choose the word "religion" in the sense of L. N. Tolstoy, as a study of that which is noble, then the proper term will be religion. However, L. N. Tolstoy regarded science as a trifle not worthy of attention, and would be offended by the use of the term "religion of science". In any case, I see with interest that I cannot avoid tracings of the Christian dogma, and this is yet another proof that in its depths human thought is the same in all times.

What, then, is the analog of original sin for the person engaged in science? The answer, of course, is that knowledge, ability, experience, and the practical skill of each human being comprise only a negligible part of the entire volume of scientific knowledge; in short, original sin is the sin of stupidity. Our stupidity is revealed inescapably and constantly; one need only stray slightly beyond the limits of his competence. An individual, of course, is different. For example, I have many times been convinced that A. N. Kolmogorov understood things an order of magnitude faster than the mathematicians around him, although they themselves were not lacking in intellect. But some of the things that A. N. Kolmogorov did, it is better to forget as quickly and firmly as possible.

And what then is the blessed condition for the scientist? Well, of course it is the historical development of science, during which that which is true was maintained and that which is false was rejected and forgotten. We shall not dispute the fact that present-day physical theories are more profound than those that existed previously, the facts are more extensive, and from the modern viewpoint we can quite correctly judge the past. (We shall at once apply this principle to the theory of probability). But this state of grace is a present-day miracle, since it directly contradicts our everyday experience. Indeed, no matter how the sin of stupidity is overcome and is atoned for in science by collective efforts, from everyday experience we now know very well that when a discussion arises in any collective body, from kindergarten to the Academic Council, and from the Academic Council to the Supreme Council, it is carried on the at the level of the most ignorant (i.e., the least competent as regards the issue) of its members?!

Silence, i.e., nonparticipation in the discussion, is frequently not only golden, but also virtuous. The renowned P. L. Chebyshev did not often participate in the sessions of the Academy of Sciences (see Ref. 5, p. 275), except, of course, when he was asked to give a scientific opinion. However, not long before his death it seems his intellect changed somewhat, and he participated in a discussion of the following question (Ref. 5, p. 288). In Russia of that day the law required that a certain number of copies of a dissertation be published at the expense of the candidate. The candidates requested that when the essential part of the dissertation was a paper published in the journals of the Academy of Sciences then the author receive not the usual 50 reprints, but 100, so that they could be counted as publications of the dissertation. Since the type-setting and all the preparations for publication would be all the same, to demand additional reprints would be downright ridiculous, and in the tone of the minutes published in Ref. 5, one might assume that a positive solution was arrived at. However, P. L. Chebyshev said: "...Is it consistent with the dignity of the Academy..." that it generally considers such commercial nonsense as the defense of a dissertation? The academicians of the physical-mathematical division were surprised at this and carried over the question to the general session. As a result, they resolved "not to enter into an examination of what the authors wish to do" with their publications. They were the object of shame in that the simple and reasonable request was inconsistent with the dignity of the Academy.

In brief, we cannot love ourselves in science for we are sinful to abhorrence (as in the Christian dogma), and we can only love the mysterious and miraculous blessing of the historical development, thanks to which each subsequent generation of scholars chooses and maintains the truth arising from all the complex discussions of the preceding generation without being burdened with all the complications of these discussions. The role of probability methods of analyzing information in this process involves the fact that with the help of these methods one can analyze and turn to scientific use a relatively large part of the observational data (we shall see that two or three tens of numbers is already a large quantity of data), which otherwise would be impossible to comprehend. It is true that here there is a relatively high risk of drawing incorrect conclusions, but if we recall and set out hopes on the blessing of the historical development, this risk may perhaps not be so important.

3. THE BLESSING OF THE HISTORICAL DEVELOPMENT AND THE THEORY OF PROBABILITY

Perhaps it makes sense to recall the structure of the principal mathematical concepts of the theory of probability in general and the probability theory of errors in particular. Laplace imagined a random event in the form of drawing a ball from an urn. The random quantity is a function of the ball (i.e., each ball in the urn is put into correspondence with a number). As far as I can judge, Laplace did not have the notion of a distribution density of a random quantity that can take any real value.

In the hundred years after Laplace, his urn was considerably improved by the mathematicians: now we can place in the urn all real numbers, as well as vectors, matrices, quaternions (and, anything at all), and even functions of a real or otherwise variable. The concept of equal probability of selection from such a set of objects loses its meaning (by the way, Laplace never did insist on equal probability), but in order to think in a mathematically correct way about a random choice we must turn to the mathematical measure theory. The mathematical kitchen of measure theory and measurable functions appears to many as being too exotic, and in fact it is so, at least with respect to measure in the space of functions, i.e., the theory of random processes (for more detail see Ref. 4). But I personally think that this kitchen is unavoidable not only in the education of professional mathematicians, but also in the education of physicists and in general in any broad mathematical education. Actually, we shall forget about the theory of probability and think about what is a self-adjoint operator. We must define it (in the simplest case), we postulate, in the space of functions that are quadratically summable. But in this space we must include all the measurable functions, otherwise it will not be complete, and this is extremely disappointing. Thus if we wish to instruct a student of quantum mechanics then he must know what a measurable function is. But then, even in the exposition of the theory of probability it is natural to use this concept, and we arrive at one or another variant of the axiomatics of Kolmogorov. Let us see how the probability model of measurement errors looks like from that vantage point.

(Here it is necessary to make a stipulation. At the present time many authors study theoretically and practically models of errors in the form of dependent random quantities, i.e., as random processes (see e.g., Refs. 6 and 7). I believe, however, that the progress that has been made is not so much that one can consider introducing the results into student education. Very frequently in the analysis of factual data we do not have sufficient information for the model of random processes to be useful for errors. Therefore in this paper we consider only the model of independent random errors.

The idea that the measurements of any particular quantity must be repeated several times to extract information on the accuracy of each particular measurement is an old one. In the time of Laplace, in particular, with astronomic and geodesic measurements as an example, it was well known that the results of repeated highly accurate measurements could not be repeated precisely and usually oscillated chaotically, being reminiscent of drawing numbered balls from some mysterious urn. Thus each observation has the form

$$x_i=a+\delta_i, \quad i=1,\dots,n, \tag{1}$$

where a is the unknown true value of the measured quantity (in the simplest case, which we shall quickly abandon, a is the same for all n observations) and δ_i is the error of the i-th observation. The errors vary chaotically. We shall say provisionally that Laplace and Gauss made the decisive step by declaring the errors of the observations to be random quantities (it would be more correct to say that in the era of Laplace and Gauss this viewpoint was applied in science). The modern version is conceived in the following way. There are some spatially elementary events Ω with a probability measure on them. By an element $\omega\in\Omega$, i.e., a single elementary event, many statisticians mean the direct result of an observation (or value of an error), but in my opinion this is unorthodox, and the correct way of understanding the meaning of ω is as the set of all conceivable parameters that influence the error of observation, and which therefore are functions of ω, i.e., random quantities. The question arises, of course, why should we think about all these parameters, of which we have only the dimmest knowledge? However, the issue is that the second important element of the concept of Laplace and Gauss is the assumption of statistical independence of the errors of the individual observations. If the error δ_1 of the first observation is a function of the set of interfering parameters ω_1 and the error δ_2 of the second observation is a function of the set ω_2, then we in fact assume that between the first and second observation so much time has elapsed that the sets ω_1 and ω_2 are statistically independent. Thus if we think of Ω as proposed here, then we should certainly not be surprised that the classical theory of errors is inapplicable in radar observations, which can be made very frequently (Ref. 6). We must think the set of all n observations as a function on the n-fold direct product of probability space with itself (for greater detail, see e.g., Ref. 4). Finally the third and essentially unalterable element of the classical concept is the assumption of the absence of systematic errors: it is assumed that the mathematical expectation is $\mathbf{M}\delta_i=0$. It is even frequently assumed that the normal probability distribution applies to errors, but in general little is changed if it is assumed that all the errors have the same probability distribution even though it is not the normal distribution.

What follows from all these assumptions if they are correct? It follows, if not directly on a level of mathematical rigor, then on a level close to mathematically rigorous (and on a level exceeding the requirements of a physical paradigm), that there are a number of very remarkable things. For example, as an estimate of the unknown value a we take, naturally, $\bar{x}=\Sigma x_i/n$. Do you wish to know how large the error is? Very well, let us calculate

$$s^2=\frac{\sum_i (x_i-\bar{x})^2}{n-1}.$$

Then with a probability of 0.95 we have

$$|\bar{x}-a|\leqslant\frac{1.96s}{n^{1/2}}, \tag{2}$$

and if a reliability higher than 0.95 is desired, 0.99, let us say, then it is necessary simply to replace 1.96 with 2.57. If expression (2) is written in the form

$$\bar{x}-\frac{1.96s}{n^{1/2}}\leqslant a\leqslant\bar{x}+\frac{1.96s}{n^{1/2}},$$

then we obtain the so-called *confidence interval*, i.e., the interval with random ends (depending on the results of the observation), which catches the nonrandom but unknown to us value of a with a probability of 0.95. The confidence intervals in this (modern for us) meaning was already widely used by Laplace. In 1845, in his master's dissertation, Chebyshev calculated from the data of Cavendish the confidence interval for the mean density of the Earth with a reliability of 0.9924794 (Ref. 5, p. 85). Therefore, the authors of Ref. 1 were not entirely correct in calling the confidence intervals the Fisher intervals. A student and Fisher at the beginning of this century only introduced into the theory and practice of calculating confidence intervals some small refinements which are important only for a small number of observations (less than ten) and for a normal distribution of errors.

Before turning to a more complicated situation, we should emphasize the following fundamental feature of the classical theory of errors as well as of other methods of the statistical processing of information: to apply them it is quite unnecessary to know what method was used to take the measurements, or indeed, in general what it is that was measured. It is sufficient to know only the numbers $x_1,...,x_n$ themselves. Herein lies the particular power (in some situations) and the particular weakness (in other situations) of statistical methods.

It would be incorrect to reduce the classical theory of errors (as is sometimes thoughtlessly done) to the analysis of measurements of one and the same quantity. The principal creation of Laplace and Gauss in the area of analyzing experimental results is the method of least squares. This, as it appears to me, is also the minimum required statistical education of students of physics and other specializations with a broad mathematical training. Even at the beginning of the last century (and more so now) the principal interest lies not in the so-called "direct" observations, where in each experiment one measures the value of the quantity of interest, but in indirect observations, where the quantity of interest appears as a parameter in the directly measured function. For example, in astronomy, one is interested in the parameters of the orbit of some object of the solar system, and he measures the angles that characterize its positions on the celestial sphere at a series of instants of time. The geodesist is interested in the length an arc of the Earth's meridian, and measures the sides and angles of a triangle. Later in this paper I will give an example where the author was interested in the electromotive force and the internal resistance of a current source, and he measured the current-voltage characteristic. The main mark of indirect measurements is that they have some laws of nature, i.e., equations, which the measurements must satisfy if they were perfectly accurate, and the main idea of Gauss and Laplace was that according to the "discrepancy" or "residue", i.e., according to how much the equation is violated for the real measurements, one can:

1) Estimate the magnitude of the error of the individual measurements.

2) Select more-correct values of the parameters that enter into the dependences studied, having estimated the accuracy of their measurements.

3) Understand whether the laws of nature that are studied are indeed satisfied with the required accuracy, or if they are in need of correction.

Of course it is necessary that the number of observations be larger than the number of parameters to be evaluated. The main calculational idea is that at first a small number of observations (by hypothesis observations of sufficient accuracy) are used for an approximate estimate of the values of the parameters, and then the equations that express the laws of nature are linearized in the neighborhood of these estimates and in this way the arbitrary equations are replaced by linear ones. As a result we obtain the following mathematical model of the method of least squares

$$x_i=a_i+\delta_i, \quad i=1,...,n, \tag{3}$$

where a_i is the exact value of the quantity measured in the i-th experiment and δ_i is the error of this measurement. In a manner that is different in each case but is always quite natural and straightforward, the laws of nature that are used go over to the proposition that the vector $a=(a_1,...a_n)$ belongs to a known submanifold of n-dimensional space. For example, for the model (1) it is, so to speak, the "bisector" that is expressed by the equations $a_1=a_2=...=a_n$. The errors δ_i are assumed to be independent random quantities with $\mathbf{M}\delta_i=0$. For their dispersions $D\delta_i$, the following very intelligent model was constructed by the classicists:

$$\mathbf{D}\delta_i=\frac{\sigma^2}{w_i},$$

where σ^2 is a parameter that is to be determined from the discrepancy (there is only a single unknown parameter for the dispersion) and w_i are numbers that are *known* in some way and are called the weights of the observations. In the simplest case (for example, for the model (1)) equal weights are used (where they can be taken equal to unity) but in general the weights are determined from some assumptions or statistical data regarding errors prior to the measurements and the conditions of linearization of the laws of nature.

From the point of view of the actual calculations, the mathematics associated with the method of least squares can be very complicated and tedious. Fortunately, we can now use the facilities of computers for carrying out the calculations, while the principal part may be expounded simply. The first such exposition in the Russian language was given by A. N. Kolmogorov[8] (originally published in 1946). It is possible to go slightly further and almost entirely free the exposition from mathematical formulas, using only simple facts from linear algebra (see, e.g., Ref. 4). However, in 1946 Kolmogorov in his book,[8] which was addressed to a wide audience, was able to expound the theory of the method of least squares without mentioning whether the model based on the theory was valid or not. Now, however, after almost fifty years new facts have been accumulated (see, e.g., Refs. 6 and 7), and to tacitly assume that everything is in order in this respect is simply impossible. I plan to examine the following questions:

1) Is the classical probability model of the theory of errors correct for real errors?

2) Is this model comprehensible even in present-day practice in teaching and in scientific investigations?

3) Wherein, nevertheless, might lie its utility?

4. IS THE CLASSICAL MODEL CORRECT?

It is not necessary, of course to imagine that the classical authors thought that the theory of errors was applicable to any measurements. Little can be done to improve the quality of measurements by merely mathematical operations: above all it is necessary to improve the measurement methods: a project that physics, chemistry, and materials science can participate in. At a certain stage metrology is involved, and of course there is nothing to prevent good instruction to the observers. We can speak of the probability model when it is applied to those errors that yet remain after all means have been taken to improve the apparatus and to train the observers. The classical approach assumes that with good observations random errors will begin to play the major role. However, if we adopt the point of view from which physics judges the validity of a particular theoretical model, then we must inevitably acknowledge that the classical model, which was undoubtedly conceived of as a physical model, has not stood the test of time. Brief arguments are as follows.

From a general theoretical point of view the physicist cannot recognize any subjective or other kind of probability and prefers to speak of probability in those cases where there is some ensemble of experiments having a certain statistical stability (see, e.g., Ref. 1). In the 200 years that have passed since the time of Laplace and Gauss science has made no progress in the fundamental issue—that of when statistical stability arises. To learn whether it does or not can only be done from experiment, and the authors of Ref. 1 have rightly insisted on a careful examination of the experimental record. It might appear strange that a statistician begins his work with an examination of the experimental record without looking into the experimental method, but if one takes into account the sin of stupidity, then it becomes clear that this classical course is conceived quite to the point. In fact, there is a transfer of experimental material from one specialist (the experimental physicist), who has been taught to make measurements on a particular device but cannot subject his data to analysis, to another specialist (in mathematical statistics) who knows the data analysis methods but would certainly burn out an electrical multimeter as soon as he handled it, when by absent-mindedness he connects it incorrectly. An analysis of the experimental record of the measurements is not unlike a give-and-take procedure.

But there is every reason to suppose that in the majority of cases this procedure terminates with unfortunate results: statistical stability may be repudiated. Let us take an example of a geodesic nature: let us assume that the measurements consisted of measuring, from the site of one landmark, the angle between the directions to two other landmarks. Malicious tongues may say that there exists the phenomenon of horizontal, or lateral refraction, in which, because of the horizontal gradient of the refractive index of the air the beam of light will deviate in the horizontal direction by some amount that depends on the weather conditions. In this case two records of the measurements given for different weather conditions will give rather sharply different results, and the statistical stability will be repudiated. The issue here is not the errors in the device and/or of the observer, but in the fact that the investigated object (the angle between two rays of observation) does not exist with the degree of accuracy with which we wish to measure it. Recognizing the problem, we define the true angle as an average of the values of the angle in all weather conditions. This average may be determined from observations if they are made over several seasons, many times each season; but how do we know whether the Earth's crust has moved somewhat during this time? Now, we can attempt to average over the motion of the Earth's crust, but this surely requires several centuries. Perhaps we can articulate the fundamental practical divergence between the opinion of the authors of Ref. 1 and mine: Let us assume that for any geodesic grid it happens that the situation is the one described here—there is no statistical nonuniformity. Does it make sense to smooth out the observations by the method of least squares? As I understand it, the authors of Ref. 1 would answer that question definitely in the negative. I, however, believe that to smooth out is proper—not for the sake of generating confidence intervals for the angles of the grid (in which there is really not much sense), but to determine with confidence for which of the triangles the lateral refraction is the most important.

The model of the theory of errors, as is the case for any other model, can also be verified according to the conclusions that come from it. The confidence intervals of the form (2) are sometimes brilliantly verified by subsequent observations of higher accuracy. For example, the confidence interval of Chebyshev for the mean density of the Earth (5.48 ± 0.1 g/cm^3) actually contains (and with room to spare) the modern value of 5.52 g/cm^3. But frequently the confidence intervals obtained in more and more accurate measurements of a particular quantity present a delightful picture: they indeed all become smaller, but the later ones do not contain any of the earlier ones, but lie in a somewhat different place on the numerical axis.

In addition to the prerequisite of randomness, i.e., statistical uniformity, the classical model contains other prerequisites, those of the independence of the errors and the definite relations between their variances (in the simplest case the variances must be equal), and also the prerequisite that there be no systematic errors. Let us present some factual information from the book of Novitskii *et al.*[7] from which it follows that for electrical measuring devices such things are impossible.

For electrical measuring devices there is a metrological service that periodically checks the devices (by comparison with more accurate devices) and adjusts them to maintain the error within specified limits (which are normalized to the per cent of the smallest division of the scale, and not to the measured value!). According to Ref. 7, the errors that arise in this method in the first approximation are linear functions of the measured quantity, the parameters of which vary with time (zero adjustments, the possibility of which has been foreseen by the design of the device, are more often harmful than helpful). When, during the periodic calibration, it is observed that the largest error of the device (which usually is usually obtained somewhere near the upper end of the scale) is close to being greater than the specified per cent, the device is adjusted with a large margin to spare so that it will operate for another several years before another adjustment. The benefit of a computer (if the data can equally well be fed into a computer) might be that the adjustment of the device would be replaced with a table of errors in the computer. In any case the true values of the errors depend both on the value of the measured quantity and on the time that has elapsed since the last calibration (as well as other troublesome factors that show up in the use of the device outside the calibrating laboratory).

In short, from the point of view of a physical paradigm we can scarcely attribute any status to the probability model of observational errors other than the status of a myth. One hundred years ago the concepts of "science" and "myth" were diametrically opposite: science was considered a noble thing and myth an ignoble thing. Now, however, the idea of the deep unity of all manifestations of human thought is more and more prevalent, so that one can scarcely take offence if, for example, modern cosmology is called a version of the myth of the creation of the world that corresponds to the level of science (and in general the culture) of the twentieth century. According to my observations, for the student youth today such a philosophy is entirely applicable, and I am speaking of my lectures on the theory of errors as an example of the foregoing.

A person cannot think of any objects and/or work with them without creating for himself some ideal model of their essence. For the anthropologists this is called animism, i.e., attributing to them a soul. For example, "the soul of the hunter runs on the soul of the skis over the soul of the snow in pursuit of the soul of the elk". In this sense these mathematical concepts such as random quantities, independence, mathematical expectation, etc., are a soul that we attribute to the specific results of the observations in order to be able to work with them. In the twentieth century, however, we wish to obtain from the analysis of observations scientific results, that is, results that are comparatively reliable and at least worthy of attention. Therefore our problem of the study the theory of errors does not include how to calculate any particular formula nor recommendations on the analysis of the observations (possibly our computer has already done this for us), nor does it include instructions on teaching how to understand which of the results of the data analysis done by the computer are reliable and what is the degree of the reliability.

5. IS THE MODEL OF THE THEORY OF ERRORS UNDERSTOOD IN THE MODERN PRACTICE OF INSTRUCTION AND SCIENTIFIC RESEARCH?

No, it is not understood. For example, we have seen how in the practical work of physics the instructor has tended to create a universal theory of errors in which any error, for instance:

1) The inaccuracy in the value of a particular quantity taken with an insufficient number of decimal points from the handbook.

2) Systematic errors in the device.

3) Errors committed by the student while reading a length or the time,—all are characterized uniquely, by its "sigma" which are added up according to the rule of the "square root of the sum of the squares". In any large ensemble of observations, for example, of all the students and all the problems of the course of experimental study, something similar might be, perhaps although of little interest, the question of what might be the error of carrying out the problems of the experimental laboratory course, averaged over all the observations, all the students, and all the problems the experimental laboratory course. However, in the framework of an individual student and an individual problem I do not see much sense in this approach.

As another example, let us consider the teaching of the classical science of geodesy, for whose needs the theory of errors was primarily created (I obtained this information from a textbook[9]). In geodesy, there are many mathematical problems, in particular, the calculational problems of the application of the methods of least squares; yet there is no other method that we have. There is no mention of what is the mathematical meaning of a random quantity or of independence—concepts which alone provide a framework within which the smoothing itself has meaning: the smoothing procedure is simply carried out to the end with no idea of how the weights of the observations are obtained. By the way, in our time it is entirely possible with the aid of radar to measure a very large number of distances between the nodes of a given triangulation grid. Each new measurement of the distance adds not just one superfluous relation to the equation, but many, because the known segments can be connected via very different chains, consisting of triangles of the grid. I have been given the impression that it is recommended that one should measure somewhat fewer lengths in order not to complicate the equation. It is also not clear how in the model of the observational weights the weights of the observations of the angle and of the measurement of the distance are put together: in the method of least squares it is assumed that there is only a single unknown parameter for the variance. In general, as it seems to me, in geodesy one should strive not so much for the comfort of the soul that comes when the smoothed values of the lengths and angles in the end do not contradict the theorems of geometry, but for a more useful application of the model of errors, for example for the selection of those observations whose errors are apparently large so as to repeat them if possible.

If we now speak of the general-science fate of the theory of errors, of course, after 200 years it has become clear that the probability model may be incorrect. However, the generally accepted alternative model is rather wretched: it leads to the result that the errors of observation δ_i can have a nonzero mathematical expectation: $\mathbf{M}\delta_i = c_i$, where the c_i are also called systematic errors. In the calculation of the systematic errors the identically distributed random quantities (random errors) must again remain independent. Here it is not possible to make the c_i arbitrarily dependent on i, since this model will be useless (anything at all can be explained by systematic errors). It is simplest to make them all equal: $c_1=...=c_n=c$, and then c is called in English the "inaccuracy" and in Russian the systematic error (in a single number). The quantity $(\mathbf{D}\delta_1)^{1/2}=...=(\mathbf{D}\delta_n)^{1/2}=\sigma$ in English is called the "imprecision" and in Russian it is called "sigma" or the standard deviation. A paradigm has been constructed according to which any observational data must be properly associated with their inaccuracies and imprecisions. But then, after all that has been said (on the basis of Ref. 7) regarding what in reality may be the observational errors, it is perfectly clear that systematic and random errors in actual fact do not exist. However, this form of the probability myth also can be very useful for analyzing large quantities of information.

If I propose to attribute to the probability model the status of a myth regarding the soul of the observational results, then I must be consistent and attribute to the statistical procedure of analyzing information the status of augury. Randomness and augury, as we know very well, are closely connected in deep psychology: one can augur by the results of any random (in the sense of unpredictable) "experiment": by the cards or the dice, by the flight of the birds, the entrails of animals, etc. It would not hurt to mention that here too there have always been specialists: some make their predictions according to the flight of the birds and others according to the entrails—no single person can master everything at once. In our time, in the analysis of experimental data one augurs with numbers that are the results of observation, and this is an art, distinct from the art of measurement. In the twentieth century, however, we wish to augur at a scientific level and it is time to tell how this can be possible and useful.

6. THE UTILITY OF PROBABILITY-STATISTICAL METHODS OF DATA ANALYSIS

6.1. Philosophical introduction: the paradigm of applied statistics

During a period of about twenty years at Moscow University there was a curious entity that in the end was called the Inter-Departmental Laboratory of Statistical Methods. It was planned (probably on the model of the Indian Statistics Institute) and directed by A. N. Kolmogorov, and the assistant director was V. V. Nalimov. The laboratory in the end was disbanded: the statistics institute did not turn out successfully for some deep reasons that I am unable to analyze, but it is interesting what happened to it. What happened is that V. V. Nalimov when he was invited to join the laboratory was engaged in applied statistics, in particular the planning of experiments, which was based precisely on the classical model of errors of observation, but then his interests started to change. As long as this interest was science metrology, i.e., the analysis of the quantity of scientific publications and the number of their citations, this was quite tolerable for a statistics laboratory. But when he went into Indian philosophy, transcendental psychology, the probability model of speech,... well, to make a long story short, a number of young coworkers became disturbed by these inappropriate interests and launched an attack on this sort of business. The attack, by the way, ended only when A. N. Kolmogorov decided that V. V. Malimov, as a prominent scientist, had the right to occupy himself in any way he saw fit. But now after twenty and more years I see quite clearly that applied activity in the area of mathematical statistics can be successful only if the mathematics of Kolmogorov has combined in it the philosophy of Nalimov.

Philosophy here is understood in the sense of what we expect from the use of a particular method and what we consider good and bad. The philosophy of applied statistics of A. N. Kolmogorov reduces, in general, to a physical paradigm. He also provided remarkable examples of work that satisfy these requirements, the best known of which are work in the statistical theory of turbulence. However, A. N. Komogorov simply rejected a great deal of his work as defective for the reason that it "did not turn out", that is to say, probability did not prove to be a "normal" physical quantity. I felt it a pity that among these investigations, which were never published, was included the work on sun spots, because the preliminary analysis of the problem and the formulation of the problem, carried out in the seminar of A. N. Kolmogorov, made an indelible impression on me. In general, the philosophy of A. N. Kolmogorov, which in my youth I naturally followed blindly, now seems to me excessively rigid.

I do not plan to set out the philosophy of A. N. Kolmogorov in its entirety, but only in a comparatively trivial part that bears upon the philosophy of science, where it reduces to the fact that in relation to the object studied, scientific work often may be only a metaphor, in that it is similar to the object in some respects and not at all similar to it in other respects. Thus the probability model of the theory of errors, so to speak, at the microlevel of the errors of single measurements, is clearly not similar to the object. However, at the macrolevel—that is in the framework of a rather large collection of experimental information—it may give some average description of the object in the form of standard random errors and/or systematic errors. We can learn whether this description is sufficiently reliable, i.e., scientific, by dividing judiciously the available information into several parts and seeing whether the value of the calculated characteristics are stable in going from one part to another, or (if possible) by analyzing new information by the same means: one should get about the same characteristics as before. The utility of a statistical investigation, however, depends in general on the feedback to the experiment: if the results of the statistical analysis prompt the experimenter to some new ideas for an improved execution of the experiment, then this is very useful. Any combinations of reliability—i.e., stability of results—of a statistical description with its utility is possible. For example, a description may be unreliable but useful. Let us say, if by the method of Ref. 1 a statistical instability was observed in the record of an experiment, then this may be very helpful for improving the way the experiment is set up. A statistical analysis may also help to reveal (or make more convincing) any new effect; that is, it can afford an advantage equivalent to the acquisition of a more accurate apparatus.

TABLE I. Results of measurements of the current-voltage characteristic.

Resistance, R, Ω	Current, I, mA	Voltage for the various experiments, V				
		$V1$	$V2$	$V3$	$V4$	$V7$
	0	32	32	32,6	—	—
25000	1,2	30,1	30,1	29,4	—	—
20000	1,45	29,4	29,4	29,1	—	—
15000	1,85	28,5	28,5	28,5	—	—
10300	2,58	26,7	26,7	27	—	—
7500	3,3	25,4	25,4	25,6	—	—
5000	4,5	23,4	23,4	23,8	—	—
4000	5,4	22,2	22,2	22,6	—	—
3000	7	21,3	21,3	21,1	21,1	21,1
2000	10,5	21	21	20,8	20,8	20,8
1500	13,8	20,7	20,7	20,7	20,7	20,7
1200	17,4	20,5	20,5	20,3	20,3	20,3
1000	18,6	19,9	19,9	29,2	20,2	20,2
800	22,8	19,8	19,8	20	20,1	19,975
600	33	19,2	19,2	19,3	19,1	19,125
500	34	18,6	18,6	19,2	19,1	19,05
400	45	18	18	18,4	18,1	18,175
350	51	17,8	17,9	18,1	17,6	17,775
300	60	17	17,1	17,4	17	17,1
250	67	16,5	16,6	16,9	16,3	16,45
200	81	15,9	16	15,9	15,4	15,475
150	99	14,6	14,7	14,7	14	14,05
100	126	12,5	12,7	12,8	11,8	12,05
80	138	11,4	11,6	11,8	10,7	10,925
60	160	10	10,3	10,1	8,9	9,2
50	175	8,9	9,2	8,7	7,5	7,95
40	189	7,7	8,1	7,7	6,4	6,8
30	204	6,1	6,5	6,6	5	5,55
20	227	4,4	5	4,8	3,2	3,7
10	247	2,5	3,2	2,9	1,6	1,975
0	276	0	0,8	0,05	0,24*)	0,23*)

$V1$ is measured with a crude instrument; $V2$ lists the data of $V1$ with correction for the internal resistance of the device; $V3$ are the measurements by an accurate set of instruments; $V4$ are the measurements with the same set of instruments but in the reverse order of the experiment (from large to small values of I); $V7$ is an average over the four experiments.

*) $I = 266$ mA

Thanks to the introduction of computers it is now unnecessary to create a calculation bureau to carry out the calculations. However, a correct interpretation of the results of a statistical analysis (if only in the application of the method of least squares) can only be made by by someone that understands the probability model from which stem the algorithms of the analysis: a correct interpretation can never simply be purchased like a package of statistical software. Thus the expense associated with the statistical analysis consists not of acquiring a computer (it most likely is already at hand) but in hiring a new person.

6.2. An example of the interaction of an experiment and statistical analysis at the level a course in experimental physics

At one time I had need of a dc power supply for some everyday purpose. I put it together out of what was at hand: a transformer that stepped down 220/24 V, which was huge for my purposes, with a power of 250 VA; I also found diodes for a diode bridge, a choke (I measured a dc resistance of 70 Ω, and an inductance of 4.2 H) and a capacitor that fitted in perfectly, having a capacitance of 22 000 μF. Altogether the device weighed almost 8 kg, and I stopped to think what the properties of this wonder might be, and I started to take its current-voltage characteristic. Then I considered that, although perhaps not necessary for everyday use, it would not be such a bad idea to see what the statistical method would give for analyzing such material; moreover, it seemed to me that the results might have some pedagogical worth.

The results of all the measurements are listed in Table I. The first three columns R, I, and $V1$ were obtained in the following manner. With the aid of a TL-4M radio amateur multimeter the resistance of the rheostat simulating the load (and connected in parallel to the filter capacitor) was set to some value (column R). Then the multimeter used as an ammeter was connected in series with the load, and the current was measured (column I). Finally, the multimeter was connected as a voltmeter in parallel with the rheostat and the voltage drop across the load was measured (column $V1$). At first I assumed that the measuring device would not introduce any distortions. The data of

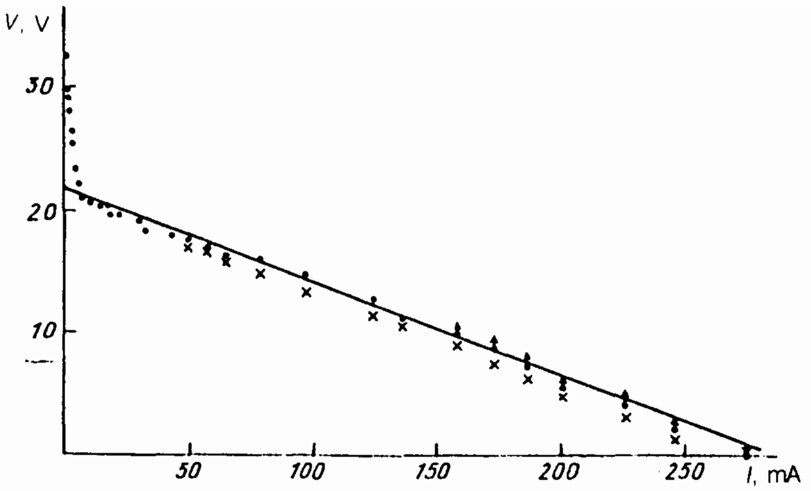

FIG. 1. Results of measurements of the current-voltage characteristic. The dots show the results of measurements by a crude instrument as the current I was raised. The straight line is the result of smoothing by eye the points in the linear region ($I \geqslant 7$ mA). The line practically coincides with the results of a linear regression analysis. The nature of the deviations from the straight line (insufficiently random) suggest that there is a nonlinearity. However, when the order of carrying out the experiment is changed (measuring while lowering the current) substantially different values of the voltage are obtained (crosses). Thus the current-voltage characteristic is not determined by the conditions of the experiment with such accuracy that one could attempt to study the characteristic having only the observations shown by the dots.

columns I and $V1$ are shown as points in Fig. 1.

Having obtained such a current-voltage characteristic, I was surprised (expecting to see a straight line) and stopped to think what my possibilities were for a theoretical analysis of the data. At one time I was taught that from a sinusoidal ac voltage a diode bridge puts out the modulus of the sinusoid, and then it is necessary with the aid of Kirchhoff's laws to set up a system of linear differential equations with constant coefficients describing the RCL circuit, with the modulus of the sinusoidal voltage on the right-hand side of one of the equations. But I was not measuring under transient conditions (because of the large capacitance of the capacitor in my circuit the transients last for a long time after turning on-up to minutes): in the measurements I waited for a steady state to be established. To study the steady-state regime, I knew, was simple: you expand the right-hand side in a Fourier series, and each harmonic passing through the filter is studied individually. However, my experimental apparatus under dc conditions was not at all sensitive to any harmonics, as can be seen by connecting it directly into a network as a dc voltmeter. Therefore I can assume that at the input to my filter there is only a dc voltage E equal to the integrated average of the absolute value of the sinusoidal signal, and nothing more. But then the law of nature that I am studying is obliged to have the form

$$V=E-rI, \tag{4}$$

where V and I are the voltage and current in the external circuit, and r is the dc resistance of the choke. Moreover, Ohm's law must be satisfied for the external circuit

$$V=RI, \tag{5}$$

where R is the external resistance.

According to the data of Fig. 1, the law of nature (4) is clearly violated, as is evident without any statistical analysis. It turns out that I discovered a new phenomenon, but, of course, it is new only in terms of my ignorance, while in radio engineering it is surely well known. (But if one makes a physics course of such material, then, of course one must explain to the students what it is all about). However, the law (4) is a physical one, and unlike a mathematical theorem, might not be absolutely or always true, but only serves as a good approximation. As judged from Fig. 1, linearity is satisfied over a very wide range of current $I \geqslant 7$ mA, which is the region of interest for practical purposes. Thus we obtain directly from experiment entirely reliable information on the region of applicability and accuracy of a physical law (4), and with respect to this information the term "augury" is entirely inappropriate.

Let us ask: could we obtain from Fig. 1 or any other place such clear and reliable information on the applicability of the probability model of the theory of errors to the possible measurement errors as may be provided by deviations of the data of Fig. 1 from a straight line (in the region $I \geqslant 7$ mA)? Of course not; and in this example the difference between physical measurements and the probability "soul" of the observations (and also between physical measurements and the statistical "augury") must, it seems to me, be perfectly clear. But now we shall see whether by some fortune telling it is possible to help the measurement.

The concept of least squares requires, generally speaking, complete smoothing of the measurements in accord with the laws of nature (4) and (5) (henceforth we shall refer only to the region of linearity $I \geqslant 7$ mA). But when the actual values of the first three columns of Table I are substituted into (5) very large discrepancies are observed, which are attributed to the crudeness of the resistance measurements by the multimeter in the ohmmeter mode. I did not find any reasonable means of carrying out a smoothing (what should be taken for the weight of the observations?) and preferred to set up the experiment in the following way: first (presumably) by means of the rheostat the current as shown in column I was set up in the external circuit and then the voltage drop shown in column $V1$ was measured. (Below, data of column R are used in a different way. The current and the voltage could be interchanged by solving Eq. (4) for I, but in that case the coefficients of the equation would have slightly less direct physical meaning than the coefficients of Eq. (4)).

TABLE II. Results of a regression analysis of the data of Table I.

Version of the Experiment	Degree of Polynomial	Constant	Coefficient of			Correlation Coefficient	Estimate of the Error of Measurement
			I	I^2	I^3		
$V1$	1	21,73	−0,0762			0,9987	0,34
	2	21,64	−0,0662	$-39 \cdot 10^{-6}$		0,9993	0,25
	3	21,65	−0,0792	$84 \cdot 10^{-6}$	$-3 \cdot 10^{-7}$	0,9999	0,21
$V2$	1	21,64	−0,0735			0,9989	0,30
	2	21,4	−0,0659	$-30 \cdot 10^{-6}$		0,9993	0,25
	3	21,63	−0,0781	$86 \cdot 10^{-6}$	$-2,8 \cdot 10^{-7}$	0,9995	0,22
$V3$	1	21,85	−0,0757			0,9989	0,30
	2	21,45	−0,0634	$-49 \cdot 10^{-6}$		0,9999	0,11
	3	21,56	−0,069	$4,9 \cdot 10^{-6}$	$-1,3 \cdot 10^{-7}$	0,9999	0,091
$V7$	1	21,8	−0,0797			0,9979	0,14
	2	21.61	−0,0738	$-23 \cdot 10^{-6}$		0,9999	0,057

The standard smoothing of the data of the polynomials was carried out by the method of least squares with equal weights of the observations. For a polynomial of degree 1, the constant has the physical meaning of an emf. The coefficient of I taken with the opposite sign is the internal resistance, and since V was measured in volts and I in mA, the internal resistance is in kΩ. The last column gives the estimate of the error of a single observation (standard deviation), measured in V.

Thus we are dealing with a linear regression on the values of column $V1$ on the values of column I We can carry out a linear regression over the points of Fig. 1 by eye (the straight line plotted in Fig. 1) or use the computer (the first column in Table II): the results are indistinguishable. The correlation coefficient is 0.9987 (I am pleased that in my amateur experiment I obtained such a good correlation coefficient, but here the issue is the relative simplicity of the technique).

I assume that most physicists not experienced in statistical augury, would, after obtaining this correlation coefficient, stop their measurements and their analysis: what more can you look for than the linear dependence? It is true that the values of E and r, $E=21.73$ V and $r=76.2\ \Omega$, evaluated from the regression equation, are somewhat different from the expected values. Actually, the ac output of the transformer was measured to be 25.4 V, which corresponds to a peak value of 36 V and an average value of 22.9 V. The dc resistance of the choke was measured to be 70 Ω. However, the reason for these discrepancies must probably be sought in the same place as the reason why with the external circuit disconnected the capacitor is charged not to the average value of the voltage, but almost to its peak value.

But someone that knows what statistical augury is is not so simply satisfied. Even the soothsayers that accompanied the emperor Julian on his ill-fated campaign, in order to see that all was in order, looked to see whether the sacred hens happily pecked their grain. The statistician will examine Fig. 1 not as other people will—by holding the sheet of paper in front of the eyes—but obliquely: placing the eyes almost in the plane of the figure along the extension of the line of regression. Then he will see that the deviations of the experimental points from a straight line do not behave as postulated, as independent random (chaotic) quantities: the points meander around the straight line like a sinusoid. This is terrible: the probability model is not satisfied, the sacred hens are not happy.

We must find the reason. For a start I was interested in the internal resistance of the multimeter in the voltmeter and ammeter modes, and I found that in the voltmeter mode it was quite high, and in the ammeter mode it was quite low. Using the value of R in the first column of Table I, let us calculate the correction (here is where excess information comes in handy) and show them by arrows in Fig. 1. Are the sacred hens happy now? The points on the right-hand side of the graph and lying farther than the rest from the line of regression were actually shifted almost onto the straight line, but the other points were also shifted. It can be seen that the line of regression must be drawn in a somewhat different way, and the nature of the deviations from the line after this remains as before. No, the sacred hens are not happy. Moreover, to the experienced eye it can be seen that they will not become happy even if instead of a linear regression we use a second or third degree polynomial. But let us carry out the appropriate computer calculations so that this fact is made clear also for any student of the art of augury.

The computer outputs all sorts of (in principle, useful) information (specifically, the STATGRAPHICS software package was used). But since the probability model is always the point in question, most of this information is not needed (for example, the significance level of the coefficient of regression or their standard errors; incidentally, it is doubtful if any nonspecialist knows the difference between the usual significance level and the limiting level, or the p-value, which the computer also gives). But there is one very important number—the mean square of the residues; taking the square root of this number we obtain the standard error of the observations (under the assumption that the regression model is correct). In the software package that was used there is also a very important graph of the residues, i.e., the differences between the regression line and the observations. For the not very skillful augur it replaces the operation of observing at an oblique angle. Let us look at Fig. 2, which shows a graph of the residues for the column $V1$ for a polynomial regression of degrees 1, 2, and 3. It can be seen that the residues change with increas-

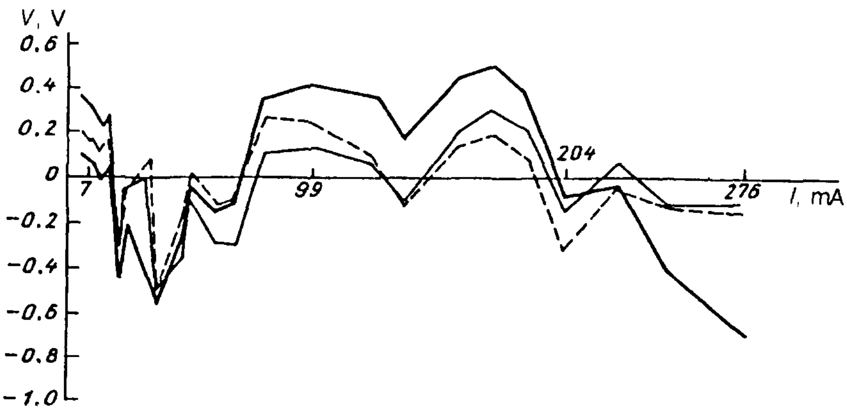

FIG. 2. Graphs of the residues obtained by smoothing the observations by regression polynomials (the crude measurements of column $V1$, Table I). The dark solid line represents the residuals for smoothing a first degree polynomial, the light solid line represents the residues for a second degree polynomial, and the dashed line is for a third degree polynomial. When the degree of the polynomial is increased the residues fall off somewhat and the graphs show mutual correlations. This means that the nonlinearity (as shown by the insufficiently random nature of the dark line) is not accounted for by polynomials of second or third degree.

ing degree of the polynomial, but in a somehow indecisive way, and the graphs of the residuals remain correlated. Also in Table II the estimates of the observation error fall off, but not sharply. This means that increasing the degree of the polynomial does not reveal the truth, but instead only the precision of the approximation increases because of the increase in the number of adjustable parameters. This is terrible: the true shape of the deviations from linearity is not made any clearer with the aid of a polynomial. The sacred hens refuse to take their food in the form of an increase in the degree of the regression. The pattern for the column $V2$ is almost the same, so it is not presented here (except for the regression in Table II). The corrections are ineffective.

A number of hypotheses arise. Can it be that the systematic errors of the multimeter while measuring the current and the voltage somehow combine together in such a way as to obtain a deviation from linearity of a complicated form? Or can it be that the reason stems from oscillations in the voltage in the network? This Gordian knot of doubts can be cut only with the aid of new measurements. Now the supply voltage is monitored by the voltmeter itself and is stabilized as well as possible with the use of a manual autotransformer. An electronic digital voltmeter and a pointer-type ammeter of class 0.5 are connected simultaneously in the measuring circuit. The rheostat is used to establish the previous values of the current, and the voltmeter records the value of the voltage (column $V3$). The graph of the residues is shown in Fig. 3, and Table II shows the results of a regression analysis.

Well, now it would appear that the sacred hens are happy. When the degree of the polynomial is increased to 2 the residues are reduced quite sharply, while with an increase to 3 they are essentially unchanged. The observation error turns out to be 0.11, which is quite reasonable, since it was noted that the tenth-volt readings on the electronic voltmeter were unstable (which, probably is explained by short-term variations in the voltage in the network, with a frequency of the order of 1 Hz). The third-degree regression polynomial, obviously, is not required (when the term in I^3 is introduced the coefficient of I^2 is also treated by the computer as insignificant). It would appear quite clear; a nonlinear effect has been found and is reliably describable by a second degree polynomial.

There is only one small cloud on the horizon. All the values in the columns $V2$ and $V3$ are very similar, respectively, except the last one. An attempt was made to repeat the last observation of column $V2$, and it was found... that the value was not repeated. It was found that if the experiment was carried out while increasing the current I, then the same values of the voltage were obtained (the columns $V1$, $V2$, and $V3$), and if, on the other hand, the current was steadily reduced by means of the rheostat, then other values of the voltage were obtained (the column $V4$ and

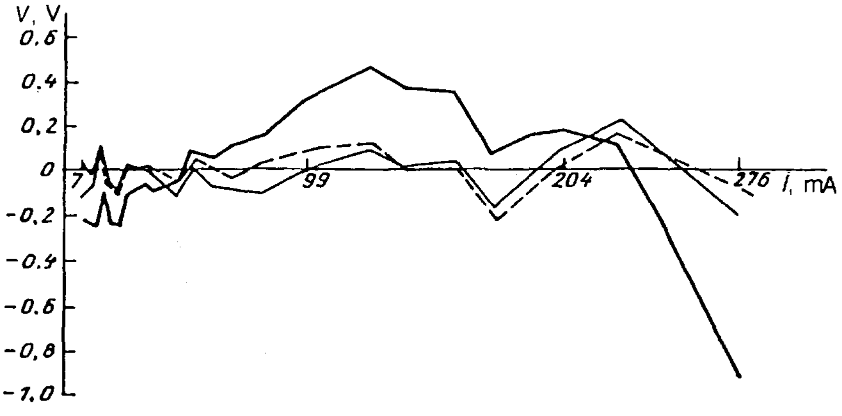

FIG. 3. Same as Fig. 2, but for the observations with instruments of greater accuracy (column $V3$, Table I). The nonlinearity is accounted for by a second degree polynomial.

the crosses in Fig. 1), which could differ from the other values by an amount the order of 1 V. This phenomenon (the current-voltage characteristic) is not determined by the conditions of the experiment with an accuracy that would allow me to study it. The reason for this is not clear, but in order of magnitude it might be entirely explained by a slight heating of the choke windings by the current. Thus the description of the device with the use of the electromotive force and the internal resistance is too crude even for practical purposes, because it turns out that the device might have some dynamic properties that are undesirable for a power supply.

Then which values of E and r are to be used? For an answer to this question I carried out two more experiments with increasing and decreasing current, and averaged the results with columns $V3$ and $V4$. The averages thus obtained are given in column $V7$, and the results of a regression analysis are given in the last two columns of Table II. If the data of column $V7$ are plotted on a large sheet of millimeter paper ($A3$ format), then it is easy to see that they are not fit by a straight line at all well, but are fit excellently by an upward-convex second-degree parabola. The data of Table II are in agreement with this conclusion: the coefficient of I^2 is negative, and the estimate of the error drops sharply in going from a straight line to a parabola. It should be noted that the error of 0.057 is smaller by almost a factor of two than the error of 0.11 obtained by smoothing the parabola of column $V3$. It may be that in the averaging of the data of four experiments an average parabola is obtained, and the random errors, being averaged become, as conjectured, smaller by a factor equal to the square root of the number of experiments.

Let us review what has happened. After the first series of measurements (column $V1$) a region of linearity of the current-voltage characteristic was discovered. From the data of the smoothing of the observations in this region the emf and the internal resistance of the current source were determined, but the question arose as to the accuracy of the determination, i.e., of the reliability intervals. But the reliability intervals obtained by computer are worthless if the probability model is violated, and this model is indeed violated because the deviations of the measurements from a straight line do not behave in a sufficiently random manner. The physically interesting question arises as to whether the current-voltage characteristic in the (supposedly) linear region is in fact linear or whether this effect should be attributed to errors in the measurement. The measurements, to the extent possible, are corrected (column $V2$) but this does not lead to any radical changes. The nonlinearity is rather strange: it is not caught by a polynomial of the second or third degrees. This consideration together with the idea of the possible variations of the voltage in the network accounts for the observations (in the example of the appearance of a nonlinearity). Fortunately, I was able to take more accurate measuring devices from the shelf (it was not for nothing that at one time I traded my vodka coupons for a decent second-hand laboratory ammeter). Measurements of higher accuracy gave a nonlinearity that was caught by a second degree parabola, but this improvement is evidently not accounted for by the fact that the device is more accurate, but by the fact that when the two devices are used instead of one the electrical circuit is not interrupted during the series of measurements, and therefore the temperature dynamics is more smooth (in time). The statistical example of the averaging of four series of measurements gives a generally excellent result that does not leave any doubt about the reality of the nonlinearity. Practically, one should use any values of the emf and internal resistance, from the first series of measurements ($V1$) or from the last ($V7$), it doesn't matter which, but now it is clear that this description is rather crude, and from the practical point of view my wonderful device, weighing 8 kg and constructed on the principles rooted in antiquity, is not entirely perfect: It would not have hurt to put a regulator in it.

All of this interaction of the measurements and their analysis went on without any use of the computer, which was not available when it was needed (unlike the ammeter): I simply viewed the sheet of millimeter paper. But the use of the computer makes the entire process simple, clear, and therefore accessible to all: one has the graphs of the residues and does not have to look at the graph at an oblique angle. If the situation concerns some multidimensional dependences, then without the computer the statistician would be entirely helpless.

7. CONCLUSION: THE BASIC PRINCIPLES OF TEACHING THE THEORY OF ERRORS

The probability theory of experimental errors is a hard nut to crack for the student, the more so because (as we have seen) things are not always entirely clear for the teacher. It was designed by the classical authors and can be somewhat logically presented only as an application of the theory of probability, which mathematically is quite complicated (since even in the simplest case of independence it in fact has to do with the direct product of a large number of probability spaces). But the theoretical sciences are good in that they can be understood relatively easily and quickly (of course, with the appropriate talent and training). It is well known (see, e.g., Ref. 4) how the theory of probability together with the fundamentals of the method of least squares can be presented (for adequately trained students) in a semester course, that is, in about 15 lectures. This amounts to 30 academic hours, or 22.5 astronomical hours, or three working days. Meanwhile, by the same accelerated method an accountant would be prepared in about a month.

However, having started from a purely mathematical science, a student to his surprise suddenly finds himself not in the field of physical science (as he would naturally expect and people have tried to believe for about 150 years), but in the field of statistical science, which in style of thought approaches such phenomena of our culture as economic statistics or (in the best case) demographic statistics. For the student of the physical, physical-technical, or natural sciences, and indeed in general almost any specialist, the statistical paradigm is alien and must be explained.

First, the reason why the student does not understand at once, is because there is no physical theory of errors. For example if science after 200 years still talks about the standard error σ of a device, then why has no one ever learned how to define it (with some precision, as a physical quantity) and list it in the instruction manual for the device as are the values of the main and additional errors? There the student should be taught that the instruction manual provides very incomplete values, those that are more or less guaranteed by the manufacturer for all the devices of that particular construction. Usually the real errors are considerably smaller. But, of course, to the scientist it is equally undesirable to underestimate as to overestimate the magnitude of the error: it may happen that he reads some new effect into the errors of observation and becomes famous without justification. However, when we begin to study real errors, then we find ourselves in a dense forest: the errors also depend on the individual device of that particular construction, and on the time elapsed since its last calibration, and possibly on the skill of the user; but the main phenomenon, the one that is being studied, might not reproduced entirely uniquely under the conditions of our experiment, and it is far from being perfectly clear how to take into account such errors, and it is completely impossible to list them in the certificate of the instrument. It is simply fortuitous that the classical authors invented the adjustable parameter σ, which at least in some average sense characterizes the accuracy of the reproduction and observation of some dependence. The parameter, in addition, is useful for soothing our souls, since it usually allows any experimental data to agree with the probability model of errors without any particularly conspicuous inconsistencies.

I believe the notion that, if the inconsistencies with the statistical model nonetheless turn out to exist, then, the analysis of their causes can undoubtedly be of use to physical investigations; this is a perfectly clear notion and indeed a trivial one for the normal student. Here we find the constantly acting reason why the probability analysis of experimental errors is not forgotten and not discarded by science, but quite the contrary, packages of computer software are written for such an analysis.

Of course, the averaged characteristic represented by the parameter σ, which expresses the correspondence of our measurements to the known laws of nature—this essence is possible not in the framework of a physical paradigm, but as statistical paradigm. To contrast this paradigm to a physical paradigm, employing the words such as "myth", or "soothsaying", or "sacred hens" in Section 6.2, in my opinion is very useful: it helps the student to understand the situation in not 150 years, but in 15 minutes. It is a good idea to bring up examples of model concepts from other disciplines. I have used some experiments from work in clinical physiology to cite the following example.

There exists a science—the physiology of respiration—which, in particular, serves as a necessary and very useful theoretical basis for developing a wide variety of breathing devices: from the aqualung to the breathing of pure oxygen at a reduced pressure in space, and from breathing in space to the use of artificial blood circulation and artificial oxygenation of the blood during surgical operations. This science recommends that the degree of inefficiency in the working of the lungs of a patient can be estimated by the index of venous mixing, or a shunt. In the textbook on physiology this index is introduced in the following way. It is temporarily forgotten that the lung has 300 million alveoli, and a model of the lung is considered consisting of a single alveolus and two capillaries. The first capillary is situated in the walls of the alveolus so that the blood passing through it (supposedly) is completely oxidized. The second capillary is located on the side, and the blood passing through it remains venous. A simple algebraic computation follows, and the fraction of the blood that passes through the second capillary (this is the shunt) turns out to be expressed in terms of the concentration of oxygen in the arterial and venous blood, and this latter quantity can easily be measured by modern techniques at the bedside of the patient. Actually, the efficiency of oxidation of the blood in the lungs depends on the relations between the ventilation and the blood flow in all of the 300 million alveoli combined (a discussion of this is found on another page of the same physiology textbook), so that in fact everything does not occur as in the model: there is neither complete discharge of the venous blood nor complete oxidation. However the point of the shunt index is that it accords with existing possibilities of measurement. It is known that in some situations the dogmatic use of the shunt leads to erroneous conclusions. This index reminds me of the parameter σ in the method of least squares, which is also good in that it is determined from the measurements.

A physical experiment can sometimes be constructed with a particular orientation to the probability model of errors. For example, one can create, as was done in Ref. 1, sets of repeated observations of some quantity and test their statistical stability. It makes sense to do this when the statistical stability or instability has an important physical meaning. Another example is the so-called "planning of an experiment", when we are not particularly interested in the exact form of some function of the experimental conditions (for example the yield of the useful product), but would rather maximize the value of this function. It is possible to plan the experiment for the quickest determination of the parameters of some theoretical function, being oriented towards a particular probability model of errors, but this to me personally appears as an unreliable business because of the manifestly nonphysical nature of the model of errors. Moreover, in this way one can miss the most interesting point: for example, in the experiment of Sec. 6.2, if one believes religiously in the linear model, then the experiments need only be done at $R=\infty$ and $R=0$, but in fact this is very unreasonable. It seems to me that in most cases the experiment must be planned from considerations of the physical investigation, but it must be planned very well if the results of the observations then are to be analyzed statistically, for example, by the method of least squares. It rarely happens that a statistical investigation does not yield anything of interest.

What is the price that must be paid for introducing statistical analysis where it was not previously used? The practical issue is that there must be a new person who will with the aid of a computer and a standard package of statistical software undertake a statistical analysis of the data, which would in any case be entered into a computer even if this person were not present. This may be a mathematician or a physicist by training, but this person must understand the mathematical theory of probability, since the algorithms of the statistical analysis always are based on some probability model for the actual data, and without understanding of to what extent this model may be incorrect one cannot interpret the results given by the computer. What role in science can this person claim to play? Of course, the answer is clear if we are speaking of a mathematician who in our hard times simply makes a living that way because nobody wants to hire him to do scientific work in the field of mathematics.

But what kind of work can be given to the physicist? For an answer to this question let us see first what is the lifetime of physical concepts and theories in science. In physics there once existed, for example the concept of phlogiston, whose lifetime must have been short because the scholastic tradition demanded that it be abused and humiliated in every way. But let us turn to the equation of heat conductivity, which in the textbooks is derived from Fourier's law (the amount of energy transported through an area is proportional to the temperature gradient). I do not see how the quantity of thermal energy is in principle any better than the quantity of phlogiston; consequently a physical concept can easily last three hundred years (of course, in different guises). However, the active scientific work of an individual scientist can be perhaps estimated to be of the order of thirty years. Consequently by taking up physics one can practically immortalize himself, leaving a trace in science for a period of time an order of magnitude longer than the duration of the work. For how much time might the results of a statistical analysis be of interest to science?

Millikan determined the charge of the electron in 1909, and as one might suppose, gave a confidence interval: the accuracy of the determination of the electron charge was estimated at 0.1% of the true value. This interval (as well as many others) later was not confirmed: the error was about 0.6% of the true value (it is taken into account that Millikan used a not entirely correct value of the viscosity of water). But in the book by Bronshtein,[10] the first edition of which came out in 1935, the author cites Millikan's estimate of the accuracy. Thus in this example the lifetime of a statistical result is estimated to be thirty years, i.e., the lifetime of one generation of scientists. This is quite understandable because statistical estimates of accuracy can retain their significance until the appearance of essentially improved methods and means of measurement. Thus if physics is compared with a bricklayer, whose work remains for a century in the edifice of science, then statistics can be compared with a plumber or electrician (nothing hinders replacing the water pipes or the electrical wiring without rebuilding the edifice). The role is historically more modest than that of the bricklayer but it is very important and necessary from the point of view of the concept of the utility of science, with which this paper was begun.

[1] Yu. I. Alimov and Yu. A. Kravtsov, "Is probability a 'natural' physical quantity?" Usp. Fiz. Nauk **162**, 149 (1992) [Sov. Phys. Usp. **35**, 606 (1992)].

[2] Yu. A. Kravtsov, "Randomness, determinateness, and predictability" Usp. Fiz. Nauk **158**, 92 (1989) [Sov. Phys. Usp. **32**, 434 (1989)].

[3] Yu. A. Kravtsov, "Randomness and predictability of dynamic chaos" in *Nonlinear Waves*, Vol. 2, *Dynamics and Evolution* [in Russian], Nauka, M. (1981).

[4] V. N. Tutubalin, *Theory of Probability and Random Processes. Fundamentals of the Mathematical Techniques and Applied Aspects* [in Russian], MGU, Moscow (1992).

[5] P. L. Chebyshev, *Collected Works*, Vol. 5 [in Russian], Academy of Sciences of the USSR, M., L. (1951).

[6] P. E. El'yasberg *Measurement Information: How Much Is Needed? How Should It Be Processed?* [in Russian], Nauka, M. (1983).

[7] P. V. Novitskii, I. A. Zograf, and V. S. Labunets, *Dynamics of Errors in the Means of Measurement* [in Russian], Energoatomizdat, L. (1990).

[8] A. N. Kolmogorov, "The basis of the method of least squares" in *Collection of the Works of A. N. Kolmogorov "Theory of Probability and Mathematical Statistics"* [in Russian], Nauka, M. (1986).

[9] Z. S. Khaimov, *The Principles of Advanced Geodesy. Textbook for Universities* [in Russian], Nedra, Moscow (1984).

[10] M. P. Bronshtein, *Atoms and Electrons*, Nauka, M. (1980).

Translated by J. R. Anderson