519.281

New Apparatus and Methods of Measurement

EXPERIMENTAL ERROR AND RELIABILITY OF SIMPLEST EXPERIMENTS

M. I. KORNFEL'D

Usp. Fiz. Nauk 85, 533-542 (March, 1965)

THE concept of experimental error does not have a clear meaning until we have indicated its reliability, i.e., the probability of finding the desired quantity *within the limits of error*.

The present note is devoted to an estimate of the reliability of the simplest experiments. In more complicated cases the reader should refer to the mathematical theory of the reduction of observations (cf. [1]).

1. COMPARISON OF OBJECTS

To distinguish one physical object from another means to indicate which of them A or B possesses a definite attribute.

However, because of unavoidable accidental disturbances the distinctive attribute may be ascribed not only to the object which in fact possesses this attribute, but also to the other object. In other words, there exists a probability W_A that the attribute will be ascribed to object A, and the probability $W_B = 1 - W_A$ that the attribute will be ascribed to object B.

In accordance with this the degree of distinguishability of the objects can be conveniently characterized by the quantity

$$U = W_A - W_B, \tag{1.1}$$

which is equal to zero for indistinguishable objects, and is equal to plus or minus unity for maximally distinguishable objects.

We assume that in the experimental determination of this quantity (i.e., in comparing the objects) n experiments (tests) have been carried out. If in m_A cases the attribute was ascribed to object A and in $m_B = n - m_A$ cases the attribute was ascribed to object B, then the degree of distinguishability is approximately equal to

$$\widetilde{U} = \frac{m_A}{n} - \frac{m_B}{n}. \tag{1.2}$$

The problem now consists of making a judgment of the true degree of distinguishability U on the basis of the approximate value of the degree of distinguishability $\widetilde{U}$.

This problem is equivalent to the statistical problem of finding the probability of an event on the basis of a number of independent experiments (cf., for example, [2], page 136).

Its solution shows that for $n \to \infty$, and in practice already for $n > 100$, the inequalities

Table I

p	0	0.7	0.9	0.99	0.999
t	0	1.04	1.64	2.58	3.29

$$\frac{\widetilde{U} - \beta\sqrt{1 - \widetilde{U}^2 + \beta^2}}{1+\beta^2} < U < \frac{\widetilde{U} + \beta\sqrt{1 - \widetilde{U}^2 + \beta^2}}{1+\beta^2}, \tag{1.3}$$

hold with the probability p, where

$$\beta = \frac{t}{\sqrt{n}},$$

while the dependence of t on p is given in Table I.

For $\beta \ll 1$ in place of the inequalities (1.3) we can write

$$U = \widetilde{U} \pm \Delta U,$$

where

$$\Delta U = \beta\sqrt{1 - \widetilde{U}^2 + \beta^2}. \tag{1.4}$$

In future we shall call ΔU the experimental error, and p the reliability in determining the degree of distinguishability.

Thus, the formulas above enable us to assert with a reliability p that the difference $U - \widetilde{U}$ does not lie outside the limits of error $\pm\Delta U$.

In scientific questions one can hardly be satisfied with the reliability of any assertion smaller than 0.9. In accepting this number, i.e., having reconciled oneself to the fact that in one case out of ten our assertion will turn out to be incorrect, it can be easily seen from (1.4) that in order to achieve a tolerably acceptable experimental error a very large number of experiments is required.

For $\widetilde{U}$ close to zero an experimental error of 0.1 corresponds to 250 experiments, an experimental error of 0.01 corresponds to 25,000 experiments, and an experimental error of 0.001 corresponds to two and a half million experiments!

For $\widetilde{U} \cong 1$ one needs a considerably smaller number of experiments. However, one must have in mind that in this case the approximate nature of (1.3) begins to be felt and the estimate of the number of experiments becomes inexact.

In order to illustrate the considerations outlined above we shall outline the results of an investigation of a procedure of comparison in which the distinguishability of objects was determined by the unaided eye (M. I. Kornfel'd [3]).

A circle drawn in India ink on white paper was

Table II

Δ, mm	0.10	0.15	0.20	0.25	0.30	0.40	1.00
$\tilde{U}$	0.09	0.38	0.57	0.62	0.75	0.78	0.98

photographed and prints were made with different magnifications from the negative so obtained. Seven pairs were formed from these prints. On one of the prints of each pair the circle had a diameter of 12 mm, and in the other one a diameter differing by Δ = 0.10; 0.15; 0.20; 0.25; 0.30; 0.40; 1.00 mm.

The experimenter by "shuffling" a given pair of prints by means of a special device was required each time to indicate which one of them contained the larger circle. The number of correct (and erroneous) conclusions was recorded automatically and from this by means of formula (1.2) the approximate degree of distinguishability of the circles $\tilde{U}$ was determined.

Twenty eight persons participated in this work. For each pair of prints one thousand experiments were carried out by different persons (100 experiments in one sitting). The results are shown in Table II and in Fig. 1 where the vertical lines correspond to the error evaluated with the aid of formula (1.4) for a reliability of 0.9 and for the number of experiments equal to 1000.

The curve shows first of all the existence of an interval of insensitivity situated in the diagram to the left of the vertical dotted line. The differences between circles lying within the limits of this interval are not detected by the eye even after a very large number of experiments, which apparently corresponds to the discrete nature of the receptors in the eye.

The shape of the curve beyond the interval of insensitivity can be explained by a superposition of random disturbances on the actual difference between the circles. The greater is this difference the smaller the role is played by the disturbances and, consequently, the closer to unity is the degree of distinguishability.

The data obtained show that the resolving power of the eye in the usual sense of this term depends on the degree of reliability with which we desire to distinguish one object from another, i.e., what fraction of erroneous conclusions we regard as being admissible. For example, for 30% erroneous conclusions the resolving power is equal to 0.15 mm, for 10% it is equal to 0.4 mm, for 1% it is equal to 1 mm etc.

2. MEASUREMENT OF MAGNITUDES

As a result of unavoidable random disturbances a repetition of the procedure of measuring a magnitude leads, as a rule, every time to a different result.

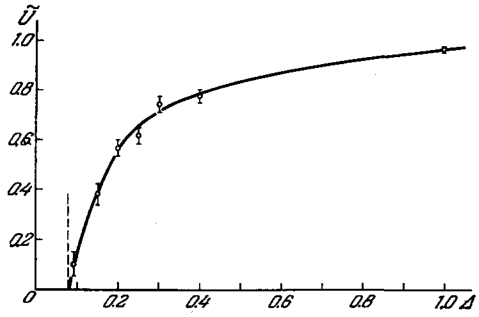

FIG. 1.

The problem consists of arriving at an estimate of the true magnitude on the basis of several numbers obtained in repeated measurements.

This problem can be solved in the following manner (cf. [4]).

In accordance with the definitions adopted in the preceding section, in comparing two indistinguishable objects *it is equally probable to ascribe a distinguishing* attribute to either one of them. In terms of the measuring operation this means that in measuring the magnitude of an object it is equally probable to obtain a number which is greater than or less than the true magnitude.

We denote the true magnitude by x and arrange the numbers obtained in the n measurements in increasing order: $x_1, x_2, x_3, \ldots, x_n$. Then, in accordance with what has been said the probability of the inequality $x < x_1$, as well as the probability of the inequality $x > x_n$, will be equal to $(1/2)^n$.

From this we get for the probability that neither one of these inequalities is satisfied, i.e., for the probability of the inequality $x_1 < x < x_n$,

$$p = 1 - 2(1/2)^n. \tag{2.1}$$

Thus, with a degree of reliability p one can assert that x lies between the limits of x_1 and x_n. In future instead of giving the limits of this interval we shall write

$$x = \tilde{x} \pm \Delta x,$$

where

$$\tilde{x} = \frac{1}{2}(x_1 + x_n), \tag{2.2a}$$

$$\Delta x = \frac{1}{2}(x_n - x_1). \tag{2.2b}$$

We shall call the first of these numbers the approximate value of the quantity, and the second number the experimental error of the measurement. When the number of measurements is great and the reliability calculated by means of (2.1) appears to us to be unnecessarily high, we can discard at the beginning and at the end of the series $x_1, x_2, x_3, \ldots, x_n$ respectively q and l numbers, and by this reduce the experimental error.

In order to calculate the reliability of the result obtained by this method one can utilize the expression

Table III

n	s		
	$p \cong 0.9$	$p \cong 0.99$	$p \cong 0.999$
10	2	—	—
20	6	4	2
30	10	8	6
40	14	12	10
50	19	16	13
60	23	20	17

$$p = 1 - \frac{1}{2^n}\left(\sum_{k=0}^{q} \frac{n!}{k!\,(n-k)!} + \sum_{k=0}^{l} \frac{n!}{k!\,(n-k)!}\right), \quad (2.3)$$

which represents a generalization of (2.1) to the case of the inequality $x_{q+1} < x < x_{n-l}$.

Setting for the sake of simplicity $q = l = s$, which corresponds to a decrease of the experimental error from $\frac{1}{2}(x_n - x_1)$ to $\frac{1}{2}(x_{n-s} - x_{s+1})$, and selecting a definite reliability, one can with the aid of (2.3) find the number s corresponding to this degree of reliability. In Table III and in Fig. 2 we have shown the dependence of s on n for $p \cong 0.9$, 0.99 and 0.999.

Let us also consider the case when one of the values of the quantity being measured sharply differs from all the rest. Usually such a value is simply discarded. However, by doing this we lower not only the experimental error but also the reliability of the result. Therefore, in each specific case one should first of all consider as to whether it is better to reduce the experimental error or the degree of reliability, and in accordance with this either to retain or to reject the "inconsistent" measurement.

On setting in formula (2.3) $q = 0$ and $l = 1$ we obtain

$$p = [1 - 2(1/2)^n] - \frac{n}{2^n}. \quad (2.4)$$

From this we can see that rejection of the "inconsistent" value is permissible only for sufficiently large values of n.

To illustrate the method of estimating the true value outlined above we shall utilize the 58 measurements of the charge of the electron by R. Millikan (in multiples of 10^{-10} cgs esu) (Table IV).

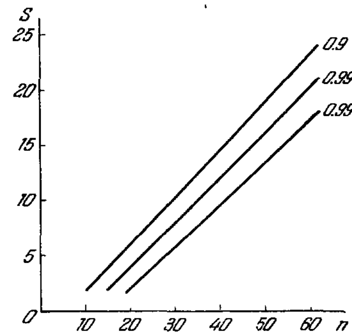

FIG. 2.

We arrange these numbers in increasing order:

4.740	4.747	4.749	4.758	4.761	4.764
4.764	4.764	4.765	4.767	4.768	4.769
4.769	4.771	4.771	4.772	4.772	4.772
4.774	4.775	4.775	4.776	4.777	4.777
4.778	4.779	4.779	4.779	4.781	4.781
4.782	4.783	4.783	4.785	4.785	4.785
4.788	4.788	4.789	4.789	4.790	4.790
4.790	4.791	4.791	4.791	4.792	4.792
4.795	4.797	4.799	4.799	4.801	4.805
4.806	4.808	4.809	4.810		

We select the degree of reliability $p = 0.9$. From Table III it can be seen that a reliability 0.9 for $n = 58$ corresponds to $s = 23$. Discarding 23 numbers from each end of the series of numbers given above we obtain the outer numbers of the remaining part: 4.777 and 4.785.

Consequently, the desired true value lies within the limits between 4.777 and 4.785, i.e., the charge of the electron with a reliability of 0.9 is equal to

$$(4.781 \pm 0.004) \cdot 10^{-10}.$$

We shall further investigate how does the estimate of the true value of the charge of the electron look for a small number of measurements, equal, for example, to five.

Utilizing formula (2.2) we find directly from Table IV for each of the eleven groups denoted by Roman numerals the approximate value and the experimental error. These results are shown in

Table IV

I	II	III	IV	V	VI
4.781	4.764	4.777	4.809	4.761	4.769
4.795	4.776	4.765	4.790	4.792	4.806
4.769	4.771	4.785	4.779	4.758	4.779
4.792	4.789	4.805	4.788	4.764	4.785
4.779	4.772	4.768	4.772	4.810	4.790
VII	VIII	IX	X	XI	XII
4.775	4.789	4.801	4.791	4.799	4.777
4.772	4.764	4.785	4.788	4.799	4.749
4.791	4.774	4.783	4.783	4.797	4.781
4.782	4.778	4.808	4.740	4.790	—
4.767	4.791	4.771	4.775	4.747	—

Table V

Group	Result	Group	Result
I	4.782±0.013	VII	4.779±0.012
II	4.777±0.012	VIII	4.777±0.014
III	4.775±0.020	IX	4.789±0.019
IV	4.779±0.019	X	4.765±0.025
V	4.784±0.026	XI	4.773±0.026
VI	4.787±0.019		

Table V; they correspond in accordance with formula (2.5) to a degree of reliability p = 0.94.

As can be seen from Table V a decrease in the number of measurements leads to an increase in the experimental error.

3. INVESTIGATION OF RELATIONS

The result of an experimental investigation of the relation between two quantities is given by the table:

$$\frac{x \mid x_1\ x_2\ x_3 \ldots\ x_N}{y \mid y_1\ y_2\ y_3 \ldots\ y_N}, \tag{3.1}$$

where

$$x_1 < x_2 < x_3 < \ldots < x_N.$$

On the basis of this table we have to make an estimate of the true function y = F(x).

Assuming, for simplicity, that only the y's are subject to experimental error and that these errors do not depend on x, we shall represent the data of the table on a graph representing each pair of values (x_i, y_i) by a corresponding point (Fig. 3).

Further we shall draw two smooth curves one above all these points and the other below all these points in such a way that the distance between the curves (along the y axis) would be a minimum and constant along the whole extent of the curve.

It can be easily seen that the distance between these limiting curves is equal to twice the experimental error which we would have obtained if we had measured y N times for a certain fixed value of x_i.

From this, in accordance with (2.1), the probability of finding the true value of y, $F(x_i)$, between the curves is equal to $1-2\,(1/2)^N$, while the probability that the true function at all points $x_1, x_2, x_3, \ldots, x_N$

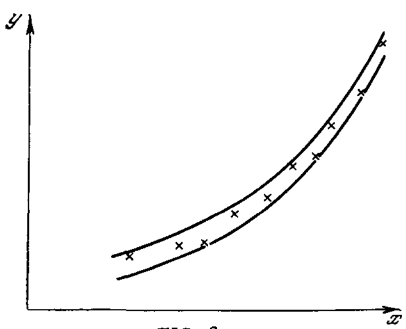

FIG. 3.

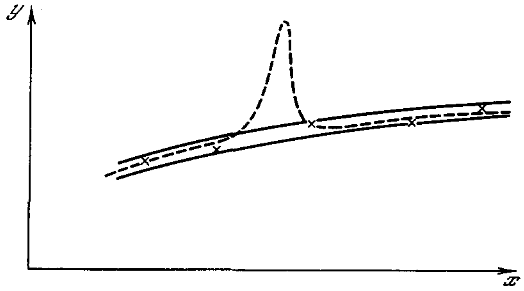

FIG. 4.

does not lie outside the limiting region will be given by

$$p = [1 - 2\,(1/2)^N]^N. \tag{3.2a}$$

We also note here that if at each point we have carried out several measurements of y, then

$$p = \left[1 - 2\left(\frac{1}{2}\right)^{\mathfrak{N}}\right]^N, \tag{3.2b}$$

where $\mathfrak{N}$ is the total number of measurements.

However, in this case the result can also be represented in a different manner. Having calculated for each x_i the approximate value $\tilde{y}_i$ and the experimental error Δy_i, we plot on the graph instead of the point (x_i, y_i) a vertical bar whose length is equal $2\Delta y_i$, and whose center coincides with $\tilde{y}_i$. Drawing the limiting curves respectively to join the top and the bottom ends of the bars we have

$$p = p_1 p_2 p_3 \ldots p_N, \tag{3.2c}$$

where $p_1, p_2, p_3, \ldots$ are the degrees of reliability corresponding to the approximate values $\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \ldots$

It is understood, of course, that the limitations imposed by the preceding formulas do not refer to the intervals between the neighboring values of x. For p arbitrarily close to unity the true function can have, for example, the shape shown in Fig. 4.

But even when we have reasons for assuming that in the "gaps" between the measurements the function does not go outside the boundaries of the limiting region, it cannot be determined completely accurately, since within the region there are contained a large number of different functions and it is not known which of them is the true one.

Therefore, we can call the curve $y = \tilde{F}(x)$ lying in the middle between the two boundary curves the approximate function, half of the interval between the bounding curves $\Delta F(x)$ the experimental error, and the probability p the reliability of the investigation of the relation (Fig. 5).

The data of the table (3.1) can also be represented in analytic form. With this aim in view utilizing the graph of the approximate function we obtain for equidistant values of x: $x_0, x_0 + \omega, x_0 + 2\omega, \ldots$ the values of y corresponding to them: $y_0, y_1, y_2, \ldots$ Further, we obtain the

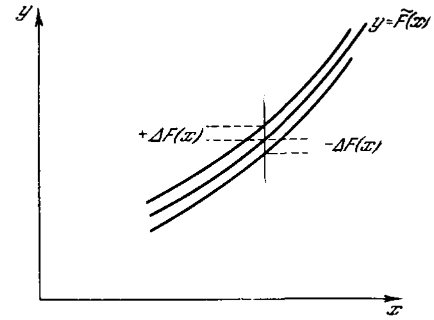

FIG. 5.

first differences: $\Delta_0^{(1)} = y_1 - y_0,\ \Delta_1^{(1)} = y_2 - y_1,$
$\Delta^{(1)} = y_3 - y_2, \ldots,$
the second differences: $\Delta_0^{(2)} = \Delta_1^{(1)} - \Delta_0^{(1)},\ \Delta_1^{(2)}$
$= \Delta_2^{(1)} - \Delta_1^{(1)}, \ldots,$
the third differences: $\Delta_0^{(3)} = \Delta_1^{(2)} - \Delta_0^{(2)}, \ldots$

$$\begin{array}{cccccc}
x_0 & y_0 & & & \\
 & & \Delta_0^{(1)} & & \\
x_1 & y_1 & & \Delta_0^{(2)} & \\
 & & \Delta_1^{(1)} & & \Delta_0^{(3)} \\
x_2 & y_2 & & \Delta_1^{(2)} & \\
 & & \Delta_2^{(1)} & & \Delta_1^{(3)} \\
x_3 & y_3 & & \Delta_2^{(2)} & \\
 & & \Delta_3^{(1)} & & \Delta_2^{(3)} \\
x_4 & y_4 & & \Delta_3^{(2)} & \\
 & & \Delta_4^{(1)} & & \\
x_5 & y_5 & & & \\
\vdots & \vdots & & &
\end{array}$$

If the m-th differences are constant, while the $(m+1)$-th differences are equal to zero, then the function under consideration can be represented by the polynomial $y = \tilde{a}_0 + \tilde{a}_1 x + \tilde{a}_2 x^2 + \ldots + \tilde{a}_m x^m$. Having obtained in this fashion the degree of the polynomial we can easily evaluate its coefficients:

$$m = 1 \left\{ \begin{array}{l} \tilde{a}_1 = \dfrac{\Delta^{(1)}}{\omega}, \\ \tilde{a}_0 = y - x\tilde{a}_1; \end{array} \right. \tag{3.3a}$$

$$m = 2 \left\{ \begin{array}{l} \tilde{a}_2 = \dfrac{\Delta^{(2)}}{2\omega^2}, \\ \tilde{a}_1 = \dfrac{\Delta^{(1)}}{\omega}, \\ \tilde{a}_0 = y - x\tilde{a}_1 - x^2\tilde{a}_2 \end{array} \right. \tag{3.3b}$$

etc., where y, x, $x^2 \ldots$, $\Delta^{(1)}$, $\Delta^{(2)}, \ldots$ correspond to values averaged over all the lines of the table of differences.

Table VI

t° C	y, sec.	t° C	y, sec.
—10.0	2.60	9.4	0.94
—5.4	2.01	14.8	1.06
1.0	1.34	19.4	1.25
4.6	1.08		

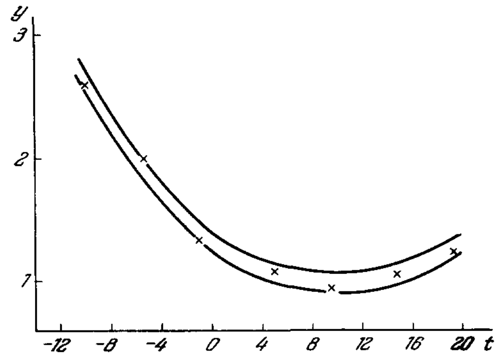

FIG. 6.

It now remains for us to find from the graph the experimental error $\Delta F(x)$ and to write the result so obtained in the form

$$y = \tilde{a}_0 + \tilde{a}_1 x + \tilde{a}_2 x^2 + \ldots + \tilde{a}_m x^m \pm \Delta F(x). \tag{3.4}$$

In order to illustrate the conclusions of this section we consider the following example.

The normal operating temperature of a chronometer is considered to be $T = 15$°C. P. G. Shtul'kerts, in an investigation of the effect of the deviations from this temperature $t = (T - 15)$ on the daily deviation of the chronometer, has obtained the results shown in Table VI.

It is required to make an estimate of the true function $y = F(t)$.

We plot on millimeter graph paper the data of the table and draw the limiting curves (Fig. 6). In accordance with formula (3.2a) for $N = 7$ $p \cong 0.9$. As a result of this one can assert with this degree of reliability that the true function lies between the limiting curves.

We now represent this result in analytic form. For the median curve (not shown in the diagram) we obtain the values of y corresponding to equidistant values of t, and construct Table VII.

Table VII

t	y	$\Delta^{(1)}$	$\Delta^{(2)}$	t^2
—10	2.62	—0.32		100
—8	2.30	—0.30	0.02	64
—6	2.00	—0.26	0.04	36
—4	1.74	—0.24	0.02	16
—2	1.50	—0.18	0.06	4
0	1.32	—0.12	0.06	0
2	1.20	—0.10	0.02	4
4	1.10	—0.06	0.04	16
6	1.04	—0.04	0.02	36
8	1.00	—0.02	0.02	64
10	0.98	+0.02	0.00	100
12	1.00	+0.04	0.02	144
14	1.04	+0.06	0.02	196
16	1.10	+0.10	0.04	256
18	1.20	+0.12	0.02	324
20	1.32			400
5.00	1.403	—0.0866	0.0285	110

Since, as can be seen from the table, the second differences are approximately constant, the approximate function can be approximated by the polynomial

$$y = \tilde{a}_0 + \tilde{a}_1 t + \tilde{a}_2 t^2.$$

Turning to (3.3b) we evaluate the coefficients:

$$\tilde{a}_2 = \frac{0{,}0285}{2 \cdot 4} = 0.0036,$$

$$\tilde{a}_1 = -\frac{0.0866}{2} - 2 \cdot 5.00 \cdot 0.0036 = -0.079,$$

$$\tilde{a}_0 = 1.403 + 5.00 \cdot 0.079 - 110 \cdot 0.0036 = 1.40.$$

From the graph of Fig. 6 we obtain the experimental error $\Delta F(x) = 0.08$. Thus, we obtain finally

$$y = 1.40 - 0.079t + 0.0036t^2 \pm 0.08.$$

[1] Yu. V. Linnik, Metod naimen'shikh kvadratov i osnovy teorii obrabotki nablyudeniĭ (The Method of Least Squares and the Foundations of the Theory of the Reduction of Observations), Moscow, Fizmatgiz, 1962.

[2] N. V. Smirnov and I. V. Dunin-Barkovskiĭ, Kratkiĭ kurs matematicheskoĭ statistiki (A Short Course on Mathematical Statistics), Moscow Fizmatgiz, 1959.

[3] M. I. Kornfel'd, ZhTF 24, 448 (1956).

[4] M. I. Kornfel'd, DAN SSSR 103, 213 (1955).

Translated by G. Volkoff