USSR ACADEMY OF SCIENCES,
B.P. KONSTANTINOV NUCLEAR PHYSICS INSTITUTE, LENINGRAD

Report 45

ON THE STATISTICAL MEANING OF EXPERIMENTAL DATA

A.P. Bukhvostov

Leningrad 1973

Translated at CERN by B. Hodge
(Original: Russian)
Revised by N. Mouravieff

(CERN Trans. Int. 74-5)

Geneva
November 1974

Communicated to
"Yadernaya Fisika"

Abstract

On the basis of "Review of Particle Data" tables, an analysis is made of the probability distribution for the values of a quantity a, using its measured value A and experimental error Δ. Contrary to the widespread belief, this distribution is quite different from the normal one and exhibits a significantly slower fall-off with growing deviation. The divergence with the normal law is several orders of magnitude. A good approximation of the empirical distribution can be given by a simple exponential. Arguments are put forward to the effect that this law is natural; it would appear to be convenient to use it for estimating the confidence level of experimental data.

*****

When experimenters measure a certain quantity a, they usually express the result obtained in the form

$$\Delta = A \pm \Delta, \tag{1}$$

i.e. they indicate two numbers: A - the most probable value of the measured quantity, and Δ - the standard deviation. It is usually considered that the probability of finding the true value of quantity a in the interval between a and $a+da$ is provided by the normal distribution formula:

$$dw(a) = \frac{1}{\sqrt{2\pi}\,\Delta} e^{-\frac{(a-A)^2}{2\Delta^2}} \tag{2}$$

It is true that the grounds for using this formula are not very sound, and this is stressed fairly frequently; nevertheless, it is widely used for evaluating the confidence level of the results. In particular, it is generally considered that the probability of a divergence of more than three standard deviations between the measured and true values is about 0.3%, and that of more than five standard deviations about 10^{-6} etc. It frequently happens, however, that more accurate measurements at a later date provide a value which diverges very considerably (by several standard deviations) from that previously indicated. Such divergences are usually due to various systematic errors. This provides grounds for a certain lack of confidence in experimental results. Many physicists hold the opinion that a reasonable evaluation of the reliability of the results may be obtained by increasing the experimental error quoted by about three times. The present work is devoted to an analysis of this phenomenon and constitutes an attempt to provide a quantitative description of it.

As will be seen from what follows, the real distribution of probabilities decreases much more slowly than in accordance with formula (2), so that the probability of a divergence exceeding, for example, five standard deviations, is about 10^{-2} instead of 10^{-6}. To explain this, it is natural to use as a basis the fact that the standard deviation (dispersion) Δ is determined rather imprecisely in practice, so that the inaccuracy in determining it is comparable with its actual value.

A straightforward analysis of this kind was carried out on the basis of the tables by Rosenfeld et al., (1964)/1/, from which were taken all of the results of the individual experimental papers relating to particles which are stable with regard to strong interactions. Only those results were discarded which were not of the form (1), i.e. either did not contain an error value at all, or were given in the form of the inequality $a < A$. The total number of different data was 235. The "accurate" values of the measured quantities were taken from the table of 1972 /2/. As in reality they too are approximate, and for some of them the error has over the past years changed only slightly, it was necessary to make a certain selection among the 235 experimental values. The criterion for the selection was the ratio between the old error Δ and the new one δ. The examination included those data for which the Δ/δ ratio exceeded a certain figure *). In order to be able to estimate the influence of arbitrary selection of this number, two values were selected for it; 2.5 and 4. The number of data fell to 209 in the case $\Delta/\delta > 2.5$, and to 181 in the case $\Delta/\delta > 4$. Their distribution over the various ranges of the value $x = \frac{|a - A|}{\Delta}$ is shown in Table I.

Even a cursory glance at this table shows the very wide range of the distribution being studied: its tail extends up to $x > 6$. The probability of this deviation, computed in accordance with formula (2), ought to have been of the order of 10^{-9}. This can be seen even more clearly from figure 1, where the integral distribution of the x data from Table I is shown in the form of a semi-logarithmic graph. The vertical axis represents the relative number of measurements for which x exceeds the given value (broken lines). The continuous line 1 relates to data of the group $\Delta/\delta > 4$, and the chain-dotted line 2 relates to the data where $\Delta/\delta > 2.5$. Over almost all their length, these two lines practically coincide; this is natural since the number of measurements in the range $4 > \Delta/\delta > 2.5$ is comparatively small.

---

*) It is stressed that the new value of δ is the result of averaging a large number of measurements which are at present known, whereas Δ relates to individual experimental papers.

The curve 3, which is given for comparison, illustrates the distribution to which formula (2) leads, i.e. the function

$$\sqrt{\frac{2}{\pi}} \int_{x}^{\infty} e^{-\frac{x^2}{2}} dx = 1 - \Phi\left(\frac{x}{\sqrt{2}}\right). \qquad (6)$$

It will be seen that the "experimental" curves are consistently higher than the "theoretical" curve, the divergence attaining many orders of magnitude at the end of the curve. It is important to note that even at small values of x the divergence is quite significant (when $x = 3$ it is more than one order of magnitude). Thus, the predictions of formula (2) are not confirmed experimentally.

In order to improve the agreement with formula (2), we may assume that the experimenters are underrating their errors, and we can increase accordingly the measurement error Δ contained in (2) by a factor of 2, for example. (There are no grounds for applying a factor of 3; this would give a distribution which decreases too slowly). This will, however, lead to a marked overestimate of probability in the region of intermediate x values between 1 and 4 (the chain-dotted line 4 in figure 1). It may be concluded that the Gaussian-type formulae are not suitable for evaluating the confidence level of the experimental results.

A much better approximation of the experimental values of probability is given by the simple exponential dependence;

$$p(a, A, \Delta) \sim \frac{1}{\Delta} e^{-k \frac{|a-A|}{\Delta}}, \qquad (7)$$

where the coefficient k in the exponent is, with a good degree of accuracy, equal to unity (the straight line 5 in figure 1). The last-mentioned situation is probably accidental but, as far as the more general expression (7) is concerned, arguments can be put forward for its occurence instead of the generally accepted normal law (2).

Use of formula (2) assumes that the value of the rms error (distribution dispersion) is precisely known. There are, however, grounds for believing that the measurement error indicated by the experimenters do not give absolutely precise dispersion values. Furthermore, the error in their determination may quite easily be of the order of the value itself. In order to determine the total error of a complex present-day experiment, account must be taken of a very large number of different factors whose influence in many cases is estimated more or less roughly. If the total error were only statistical, it would be determined fairly well. Unfortunately, the real situation is quite different. As a rule, the measurement results are affected by systematic errors, and it may prove a very complex matter to take these into account. Furthermore, there almost always remains some uncertainty as to whether account has been taken of all the substantial factors which may lead to systematic errors.

In view of the above remarks, the following formal approach may be adopted. We shall consider that the experimental error quoted Δ provides, not an accurate, but only the most probable dispersion value.

Instead of knowing the precise dispersion value, we shall assume a certain distribution $\tau(\Delta', \Delta)$, which gives the probability that the true dispersion value falls between Δ' and $\Delta' + d\Delta'$:

$$d w\, (\Delta', \Delta) = \tau(\Delta', \Delta)\, d\Delta', \tag{8}$$

where Δ is the experimental error, which we shall assume to be equal to the most probable value of Δ' . Assuming that for each given dispersion value Δ' the distribution (2) is correct, we obtain

$$\rho(a; A, \Delta) = \frac{1}{\sqrt{2\pi}} \int \frac{d\Delta'}{\Delta'}\, \tau(\Delta', \Delta)\, e^{-\frac{(a-A)^2}{2\Delta'^2}} \tag{9}$$

This formula is reduced to (2) only when the function $\tau(\Delta', \Delta)$ in the significant range of Δ' changes very rapidly in comparison with $\frac{1}{\Delta'}\, e^{-\frac{(a-A)^2}{2\Delta'^2}}$ and may thus be replaced by $\delta(\Delta' - \Delta)$ (since the integral from it is equal to 1). If, however, this condition is not fulfilled, formula (9) gives a very different result.

This will be seen if we examine the behaviour of the function ρ at large deviations $|a - A|$. In this case, the main contribution to the integral (9) is given by the region of large Δ', and consequently the asymptotic form of the function τ at $\Delta' \to \infty$ is substantial. We shall assume that this asymptotic form, in accordance with the normal Δ' distribution, has a Gaussian form

$$\tau(\Delta', \Delta) \sim e^{-\nu\left(\frac{\Delta'}{\Delta}\right)^2}, \tag{10}$$

where ν is a certain number. Then, in integral (9) the region of the saddle point defined by the condition

$$\frac{\partial}{\partial \Delta'}\left[\nu\left(\frac{\Delta'}{\Delta}\right)^2 + \frac{(a-A)^2}{2\Delta'^2}\right] = 0,$$

or

$$\left(\frac{\Delta'}{\Delta}\right)^2 = \frac{x}{\sqrt{2\nu}}, \qquad x = \frac{|a-A|}{\Delta}.$$

is substantial.

Hence, we obtain for ρ at large values of x

$$\rho \sim e^{-x\sqrt{2\nu}}, \tag{11}$$

which coincides with (7) at $K = \sqrt{2\nu}$.

Let us note that a distribution of the form

$$\tau(\Delta', \Delta) = \frac{\Delta'}{\Delta^2} e^{-\frac{1}{2}\left(\frac{\Delta'}{\Delta}\right)^2} \tag{12}$$

gives exactly formula (7) with $K = 1$, which in this case is correct for any x (not only asymptotically). Generally speaking, the value of k, in accordance with formulae (11) and (12) is determined by the rate at which the distribution of $\tau(\Delta', \Delta)$ declines and obviously must depend on the type of experiments, their complexity, the number and degree of confidence in the assumptions made etc. - in other words, on how reliably one can determine the rms error of the experiment. We may consider $k = 1$ as the empirical average value at which formula (7) gives a satisfactory evaluation of the probability of deviation from a given experimental result.

Let us sum up. It follows from what has been stated that the use of the normal distribution formula (2) for evaluating the degree of confidence in experimental results leads to completely wrong conclusions. In order to eliminate the possible deviation with a probability in excess of, for example, 99% it is necessary to have a much wider range than that normally used (3 standard deviations which is based on formula (2). In order to obtain a clear picture, we indicate in Table II the comparison of the probability of a divergence exceeding n standard deviations, determined from formulae (2), (7) and from empirical data. From this we can see that a confidence level of 99% is attained only when there is a margin of 5 standard deviations. If we adhere to this principle, the appearance from time to time of results which are seriously divergent will not in any way seem extraordinary.

The author expresse his gratitude to V.M. Shekhter for the interest he has constantly shown in this work and for useful discussions.

*****

Table I

Distribution of data for $x = \frac{|a - A|}{\Delta}$

Range of x	Number of data $\frac{\Delta}{\sigma} > 4$	$4 > \frac{\Delta}{\sigma} > 2.5$
0 - 0.5	55	6
0.5 - 1	54	7
1 - 1.5	29	7
1.5 - 2	21	3
2 - 2.5	10	1
2.5 - 3	4	3
3 - 3.5	2	-
3.5 - 4	1	-
4 - 4.5	1	-
4.5 - 5	1	-
5 - 5.5	2	-
5.5 - 6	-	-
6 - 6.5	1	1
Total:	181	28

Table II

Probability that the precise value departs from the measurement result by more than n errors

Number of standard deviations n	Normal distribution (formula 2)	Exponent (formula (7) for $k = 1$)	Empirical result
1	1/3.2	1/2.7	1/2.5
2	1/22	1/7.3	1/8
3	1/370	1/20	1/23
4	1/16000	1/55	1/40
5	1/1700000	1/150	~ 1/60

References

1. A.H. Rosenfeld, A. Barbaro-Galtieri, W.H. Barkas, P.L. Bastien, J. Kirz, M. Roos. Rev.Mod.Phys. 36, 977 (1964).

2. P. Soding, J. Bartels, A. Barbaro-Galtieri, J.E. Enstrom, T.A. Lasinski, A. Rittenberg, A.H. Rosenfeld, T.G. Trippe, N. Barash-Schmidt, C. Bricman, V. Chaloupka, M. Roos. Phys.Lett., 39B, 1 (1972).

Paper received by Publishing Department 17 April 1973.

*****

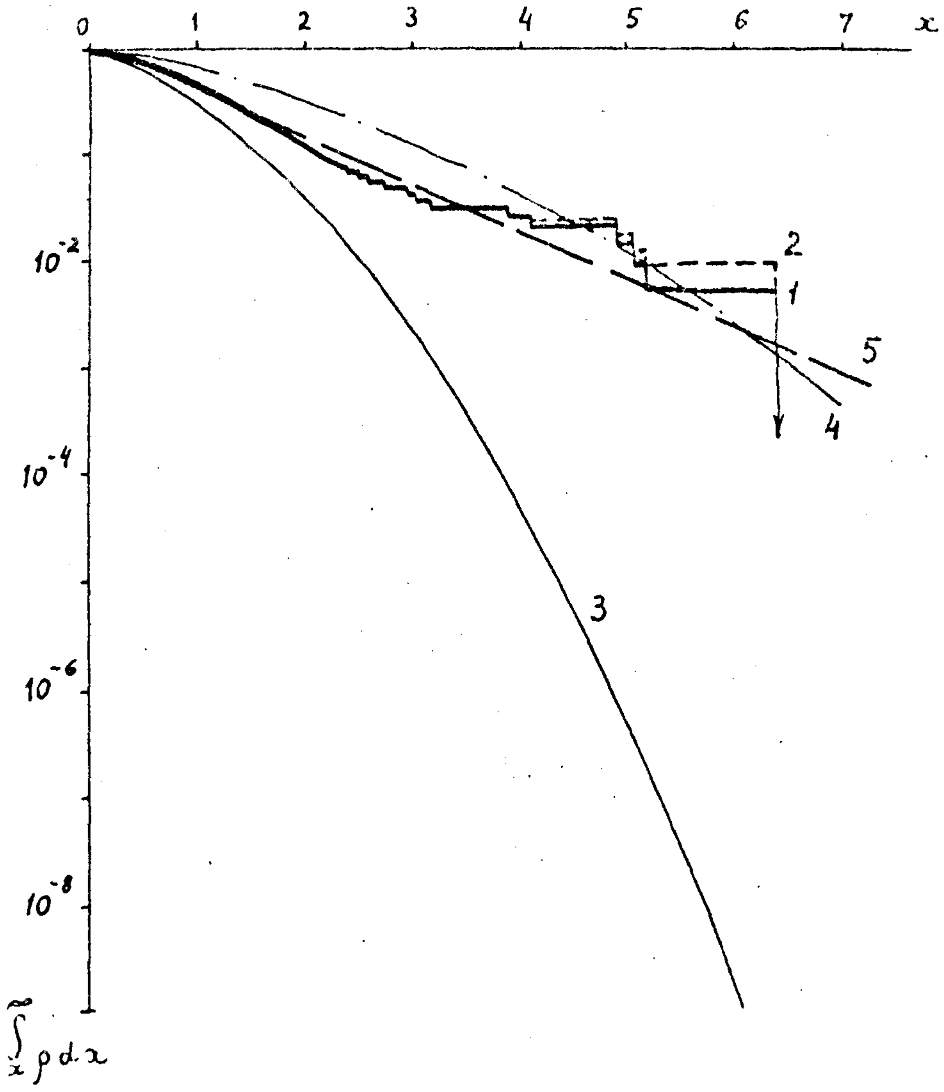

Fig. 1. Relative number of measurements the results of which depart from the precise values by more than x standard deviations :

Curve 1 : empirical data, $\frac{\Delta}{\sigma} > 4$;

Curve 2 : empirical data, $\frac{\Delta}{\sigma} > 2{,}5$;

Curve 3 : normal law, formula (6) ;

Curve 4 : normal law with twice the error;

Curve 5 : exponential law, formula (7), for $K = 1$.