INTRODUCTION TO THE Theory of Error

by

YARDLEY BEERS

COPY #3.

111 LABORATORY LIBRARY
298 LECONTE HALL

ADDISON-WESLEY PUBLISHING COMPANY

Introduction to the THEORY OF ERROR

by

YARDLEY BEERS

National Bureau of Standards
Radio Standards Division
Boulder, Colorado

ADDISON-WESLEY PUBLISHING COMPANY
READING, MASSACHUSETTS • MENLO PARK, CALIFORNIA
LONDON • AMSTERDAM • DON MILLS, ONTARIO • SYDNEY

Library of Congress Catalog Card No. 57–8708

Second edition, third printing — June 1962

ISBN 0-201-00470-4
FGHIJKLMN-AL-798

CONTENTS

PREFACE TO THE SECOND EDITION

This book contains a detailed discussion of the fundamental principles of the theory of error, and those applications which are commonly encountered by advanced undergraduate and graduate students in the physical sciences. The approach is intermediate between the didactic treatments contained in the introductory chapters of many laboratory manuals and the lengthy treatises which are of interest mainly to experts in the field.

Laboratory manuals, which seldom do more than state the results of the theory, rarely give the student any insight into the underlying concepts, while in the present book a plausible presentation of the concepts has been the principal objective. The cost of the longer treatises is not justified in the opinion of many students of the physical sciences because most of the material is of little interest to them; furthermore, these treatises tend to adopt an isolated point of view and to neglect certain aspects of the practical conditions under which the student must estimate error. In the physical sciences, a sample containing more than ten measurements is rare, and a sample of this size is so small that it can hardly be called a statistical sample. Furthermore, there are usually present indeterminant errors comparable in magnitude to the statistical errors, and these indeterminant errors are not amenable to the theory. Under these circumstances, extreme rigor or the complication of the theory of small samples is not entirely necessary, and much of the estimation of error must be done by common sense, or what perhaps more candidly should be called guesswork. This book contains some detailed examples in which the author has attempted to illustrate how to deal with such situations.

In the second edition an attempt has been made to correct a number of mistakes which were unfortunately included in the first edition. Most of these were in the details of the concepts rather than in the mathematics. The sections which have received major revisions are IV-B, IV-I, V-B, and VI-B.

The author wishes to thank those readers of the first edition who sent him their criticisms and suggestions. Special thanks are due to Dr. Ralph Hoyt Bacon, whose detailed comments resulted in the greater part of the revision. Most of his suggestions have been incorporated. There were a few important questions upon which the author did not agree with Dr. Bacon; nevertheless, even in these matters Dr. Bacon's comments were of great value because they forced the author to clarify his thinking.

The author wishes to take this opportunity to thank Mrs. Olga Crawford, Director of Publications of the Addison-Wesley Publishing Company, for extensive editing of the manuscript of both editions. Finally, he

wishes to express his gratitude to Dorothy Sands Beers, for her encouragement and for her help in reading proof. Some of the preparation of the manuscript for the second edition took place while the author was on sabbatical leave from New York University and held a Fulbright Research Scholarship at the Division of Electrotechnology, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia.

YARDLEY BEERS

New York City
January, 1957

I. INTRODUCTION

Many people call physics an "exact science." Actually, it is exact only in degree. For example, the accepted value of the velocity of light was recently quoted as

$$(2.997923 \pm 0.000008) \times 10^{10} \text{ cm sec}^{-1}.$$

The figure 2.997923×10^{10} cm sec^{-1} represents the best estimate of the velocity, while that of 0.000008×10^{10} cm sec^{-1} is an indication of the reliability of the result.

Many factors employing human judgment are involved in the determination of any "accepted" value:

(1) In the use of measuring instruments, the last significant figure must often be obtained by estimating a fraction of the smallest division on some measuring instrument such as a meter stick or an ammeter.

(2) In arriving at a specific value, a number of experiments are performed, and an "average" is taken. In some instances, certain data may be excluded as being unreliable in the opinion of the observer.

(3) Experiments of different types have differing uncertainties. In the computation of the final value, a weighted average of the separate values must be taken. The more accurate experiments are given greater importance, and the assignment of weight factors is again a matter of opinion.

(4) The error of human judgment enters into all of the items above. Although the observer is guided by rules, these are based upon arbitrary assumptions and are themselves, in the last analysis, matters of opinion. If more than one group of observers performs an experiment, it is probable that there will be a difference in the final values submitted.

The estimated uncertainty in the result of an experiment, such as the value 0.000008×10^{10} cm sec^{-1} quoted above, takes into account the uncertainty assigned to the experiment by the observers by analysis of the methods used and the inaccuracies of the data. For example, the velocity of light cannot be measured directly; instead, the time required for light to travel a given distance is measured. Thus *time* and *distance* are the measured quantities, whereas the velocity is a *computed* quantity. The uncertainty in the velocity is a composite effect of the uncertainties of the measurements of time and distance, an example of the *propagation of error*.

In evaluating these uncertainties, the individual processes of measurement of time and distance must be investigated in detail. Such analysis of

E. Determinate and indeterminate errors. Errors which may be evaluated by some logical procedure, either theoretical or experimental, are called *determinate*, while others are called *indeterminate*.

Random errors are determinate because they may be evaluated by application of a theory which will be developed later. In some cases random or systematic errors may be evaluated by subsidiary experiments. In other cases it may be inherently impossible to evaluate systematic errors, and their presence may be inferred only indirectly by comparison with other measurements of the same quantity employing radically different methods. Systematic errors may sometimes be evaluated by calibration of the instruments against standards, and in these cases whether the errors are determinate or indeterminate depends upon the availability of the standards.

F. Corrections. Determinate systematic errors and some determinate random errors may be removed by application of suitable corrections. For example, the measurements which are in error due to a kink in a steel tape may be eliminated by comparing the tape with a standard and subtracting the difference from all the measured values. Some of the random error of this tape may be due to expansion and contraction of the tape with fluctuations of temperature. By noting the temperature at the time of each measurement and ascertaining the coefficient of linear expansion of the tape, the individual values may be compensated for this effect.

G. Precision. If an experiment has small *random* errors, it is said to have high precision.

H. Accuracy. If an experiment has small *systematic* errors, it is said to have high accuracy.

I. Adjustment of data. This is the process of determining the "best" or what is generally called the *most probable value* from the data. If the length of a table is measured a number of times by the same method, by taking the average of the measurements we can obtain a value more precise than any of the individual ones. If some of the individual values are more precise than others, then a weighted average should be computed. These are examples of *adjustment of data* for directly measured quantities. For computed quantities the process may be specialized and complicated. Later we shall develop a method for determining the most probable value of the slope of a straight line representing the graph of linearly related measured quantities.

III. CLASSIFICATION OF ERRORS

A. Systematic errors.

(1) *Errors of calibration of instruments.*

(2) *Personal errors.* These are errors caused by habits of individual observers. For example, an observer may always introduce an error by consistently holding his head too far to the left while reading a needle and scale having parallax.

(3) *Experimental conditions.* If an instrument is used under constant experimental conditions (such as of pressure or temperature) different from those for which it was calibrated, and if no correction is made, a systematic error results.

(4) *Imperfect technique.* The measurement of viscosity by Poiseuille's law requires the measurement of the amount of liquid emerging from an apparatus in a given time. If a small amount of the liquid splashes out of the vessel which is used to catch it, a systematic error results.

B. Random errors.

(1) *Errors of judgment.* Most instruments require an estimate of the fraction of the smallest division, and the observer's estimate may vary from time to time for a variety of reasons.

(2) *Fluctuating conditions* (such as temperature, pressure, line voltage).

(3) *Small disturbances.* Examples of these are mechanical vibrations or, in electrical instruments, pickup of spurious signals from nearby rotating electrical machinery or other apparatus.

(4) *Definition.* Even if the measuring process were perfect, repeated measurements of the same quantity might still fail to agree because that quantity might not be precisely defined. For example, the "length" of a rectangular table is not an exact quantity. For a variety of reasons the edges are not smooth (at least if viewed under high magnification) nor are the edges accurately parallel. Thus even with a perfectly accurate device for measuring length, the value is found to vary depending upon just where on the cross section the "length" is measured. In nuclear physics errors of definition are commonly the largest source of error. (See Section VII.)

C. Illegitimate errors.

Most, or even all, of these types of error are always present, at least to a small degree, in the very best of experiments and they should be discussed in a written report. However, there are three types of avoidable errors which have no place in an experiment, and the trained reader of a report is justified in assuming that these are not present.

(1) *Blunders.* These are errors caused by outright mistakes in reading instruments, adjusting the conditions of the experiment, or performing calculations. These may be largely eliminated by care and by repetition of the experiments and calculations.

(2) *Errors of computation.* The mathematical machinery selected for calculating the results of an experiment (such as slide rules, logarithm tables, adding machines) should have errors small enough to be completely negligible in comparison with the natural errors of the experiment. Thus if the data are accurate to five significant figures, it is highly improper to use a slide rule capable of being read to only three figures, and then in the report to list "slide rule error" as a source of error. Such a slide rule should be used for calculating the results of an experiment having only three or preferably only two significant figures. On the other hand, if the experiment does give five significant figures, five- or six-place logarithm tables or some other more accurate means of calculation should be used.

(3) *Chaotic errors.* If the effects of disturbances become unreasonably large—that is, large compared with the natural random errors—they are called *chaotic errors.* In such situations the experiment should be discontinued until the source of the disturbance is removed.

IV. RANDOM ERROR OF A MEASURED QUANTITY

A. Proof that the sum of the squares of the deviations about the average value is a minimum. Previously it was stated that there exists a "most probable" value of a quantity. In the case of directly measured quantities this is generally, but arbitrarily, assumed to be the arithmetic average of the individual measurements.

Let $x_1, x_2, \ldots, x_k$ be measured values and let $\bar{x}$ be the most probable value. Then the differences between the respective measured values and $\bar{x}$ are called *deviations*. (They are also called *residuals*.)
For a given value x_n, the deviation is defined as

$$\delta x_n \equiv x_n - \bar{x}. \tag{1}$$

If $\bar{x}$ is the arithmetic average of the k measurements, by definition

$$\bar{x} \equiv \frac{x_1 + x_2 + x_3 + \cdots + x_n + \cdots + x_k}{k} = \frac{\sum_{n=1}^{k} x_n}{k}, \tag{1A}$$

and the arithmetic average of the deviations may be shown to be zero. To show this, we compute the sum of the deviations as evaluated by Eq. (1) and then divide by the total number k. The sum of the first terms on the right-hand sides of equations of the type of Eq. (1) is just the sum of the measured values $\sum_{n=1}^{k} x_n$, which by Eq. (1A) is $k\bar{x}$. This sum is equal but opposite in sign to the sum of the second terms. Therefore the arithmetic average of the deviations is zero.

However, the squares of the deviations are all positive, of course, and therefore the sum of these squares does not vanish. Next we shall prove that if we use the average of the measured values for $\bar{x}$, the sum of squares of the deviations is a minimum. From this result we may infer that the "most probable" value of a computed quantity is the one for which the sum of squares of the deviations, when properly defined, is a minimum. We shall define a quantity called the *standard deviation*, which is simply related to this minimum value of the sum of the squares of the deviations and which is one of the quantities used for specifying error quantitatively.

To carry out this proof, we first compute the squares of the deviations by use of Eq. (1).

$$
\begin{aligned}
(\delta x_1)^2 &= x_1^2 - 2x_1\bar{x} + \bar{x}^2,\\
(\delta x_2)^2 &= x_2^2 - 2x_2\bar{x} + \bar{x}^2,\\
&\vdots\\
(\delta x_n)^2 &= x_n^2 - 2x_n\bar{x} + \bar{x}^2,\\
&\vdots\\
(\delta x_k)^2 &= x_k^2 - 2x_k\bar{x} + \bar{x}^2
\end{aligned}
$$

$$
\begin{aligned}
\text{Sum} \equiv X = \sum_{n=1}^{k} (\delta x_n)^2 &= (x_1^2 + x_2^2 + \cdots + x_n^2 + \cdots + x_k^2)\\
&\quad - 2\bar{x}(x_1 + x_2 + \cdots + x_n + \cdots + x_k) + k\bar{x}^2\\
&= \sum_{n=1}^{k} x_n^2 - 2\bar{x} \sum_{n=1}^{k} x_n + k\bar{x}^2, \qquad (2)
\end{aligned}
$$

where k is the total number of measurements.

In order to find the value of $\bar{x}$ which makes X a minimum, we differentiate X with respect to $\bar{x}$. Since all the x_n's have been fixed by experiment, their differentials are zero. We place the derivative equal to zero and solve for $\bar{x}$:

$$\frac{dX}{d\bar{x}} = -2 \sum_{n=1}^{k} x_n + 2k\bar{x} = 0,$$

or

$$\bar{x} = \frac{\sum_{n=1}^{k} x_n}{k}, \qquad (3)$$

which is the arithmetic average.

B. Definition of standard deviation. The precision of an experiment can be indicated by a graphical plot of the distribution of the deviations, such as is shown in Fig. 1 (Section IV-E). Such a plot is obtained by dividing the range of observed deviations into equal intervals, noting how many of the observed deviations lie within each interval, and plotting these numbers against the average of the deviation in each respective interval. While such graphs are informative in many respects, they provide no direct basis for the evaluation of the error in a quantity which is to be computed from two or more measured quantities. Therefore, we wish to introduce one or more quantities, to be evaluated numerically, which can represent the gross properties of such a graph. In particular, we wish to find a quantity which is related to the width of such a graph, since we note that in an experiment of high precision the graph tends to have a sharp maximum, while in an experiment of low precision the graph is relatively broad. In fact, having defined an appropriate quantity, in many cases we may choose not to actually plot the graph but merely to retain it as a concept.

One quantity which we can employ is the *root mean square* (*rms*) *deviation* s', which is given by the following relation:

$$s' = \sqrt{\frac{(\delta x_1)^2 + (\delta x_2)^2 + \cdots + (\delta x_n)^2 + \cdots + (\delta x_k)^2}{k}}$$

$$= \sqrt{\frac{\sum_{n=1}^{k} (\delta x_n)^2}{k}}. \tag{4}$$

An alternative formula which frequently is more convenient for numerical computations may be obtained by substituting into Eq. (4) the expression for $\sum_{n=1}^{k} (\delta x_n)^2$ from Eq. (2), bearing in mind that Eq. (2) may be simplified by making use of the fact [see Eq. (1A)] that $\sum_{n=1}^{k} x_n = k\bar{x}$. When these substitutions have been made,

$$s' = \sqrt{\frac{\sum_{n=1}^{k} x_n^2 - k\bar{x}^2}{k}}. \tag{4A}$$

Unfortunately the quantity s' does not have the widest possible significance, because it indicates only how a particular set of k values deviate from their average. We do not know whether or not this quantity depends systematically upon the number of values k in the set. Furthermore, the errors we are now considering are completely random, so that a second set of k measurements generally does not yield an average value identical with the first, nor an identical set of deviations, because of what are called *statistical fluctuations.*

To establish a quantity which has greater significance, we employ the concept that two such sets of k measurements are two samples of the entire *universe* of measurements which might be made, the number of measurements in the universe being infinite. The quantity which is of interest to us is the *standard deviation* s which is the rms deviation of the individual measurements about the *universe average.* The square of this quantity frequently is called the *variance.* Of course, it is quite impractical to make all of the measurements in the universe and then determine s exactly from a direct calculation, but, as we shall show in Section V-B, we can obtain an estimate of it from either of the following expressions:

$$s = \sqrt{\frac{\sum_{n=1}^{k} (\delta x_n)^2}{k - 1}}, \tag{5}$$

$$s = \sqrt{\frac{\sum_{n=1}^{k} x_n^2 - k\bar{x}^2}{k - 1}}. \tag{5A}$$

The relation of Eq. (5) to Eq. (5A) will be recognized to be the same as that of Eq. (4) to Eq. (4A).

The distinction between s' and s is important conceptually. Numerically, the difference between them is generally quite trivial. They are given by nearly identical formulas, the factor k in the denominator of the expression for s' being replaced by $k - 1$ in the expression for s. When k becomes very large, the expressions approach equality. When k is as small as 5, the difference is only about 12 percent, and usually this is not large compared with the indeterminate errors present.

The fact that s is larger than s' is to be expected, because we know that the sum of squares of deviations about the sample average is a minimum. Since the universe average generally does not coincide with the sample average, the sum of squares of the deviations pertaining to the finite sample about the universe average is not a minimum.

It is interesting to note that the factor $k - 1$ which appears in the denominator of the expression for s is equal to the number of functionally independent deviations. When $k = 1$, the concept of deviation is meaningless. When $k = 2$, we may compute two deviations, but they are always equal in magnitude, although different in sign. We may infer then that although we may compute k different deviations, there is one relation between them, and therefore they are not completely independent. This relation is, as we have seen, that their algebraic sum is equal to zero, or, what we have shown is equivalent, that the sum of the squares of the deviations is a minimum.

Whether s' and s are calculated directly by Eq. (4) and Eq. (5) or indirectly by Eq. (4A) and Eq. (5A), respectively, is a matter of taste or of convenience. Many observers prefer the direct method because they like to see the individual deviations. The most laborious part of the computation is the calculation of the sum of the squares, and this is carried out more easily by the direct method, in which the numbers are the smallest possible. However, the discrepancies between the individual values of x are usually only in the last one or two significant figures, and the indirect method can be simplified by subtracting from the individual x's some convenient number containing all of the invariant significant figures. Even then the calculation of the sum of squares is somewhat more laborious than with the direct method, but, on the other hand, the subtraction process is somewhat simpler with some number chosen by convenience rather than with the arithmetic average. The great convenience of the indirect method comes in more advanced statistical analyses where not only the standard deviation but other quantities beyond the scope of the present treatment are calculated.* In Section IV-E we shall illustrate the direct method, while in Section VIII-B we shall illustrate the indirect method.

C. Average deviation. As indicated previously, the algebraic sum of the deviations is zero. Therefore, if we are to define an average deviation, we must add the deviations without regard to sign and then divide by the number of observations. Then the average deviation is

*For example, see R. H. Bacon, *Am. J. Phys.* **14,** 84 (1946).

$$a = \frac{\sum_{n=1}^{k} |\delta x_n|}{k}. \tag{6}$$

D. Relation of average deviation to standard deviation. Percentage deviation. Basically the average deviation is of less direct significance than the standard deviation. Nevertheless, the calculation of the average deviation involves less arithmetic than that of the standard deviation. Later we shall show, subject to some assumptions which are usually valid, that when the number of measurements is large, the ratio of the standard deviation to the average deviation approaches 1.25, so that the standard deviation can be estimated by computing the average deviation and multiplying by 1.25. Such an estimate is not as reliable as one computed from Eq. (5) or Eq. (5A) because with a finite sample the actual ratio is subject to statistical fluctuations and can range from 1 to approximately $\sqrt{k/2}$. However, in practice, such an estimate often may be adequate, especially if there are large indeterminate errors in the experiment.

It is to be noted that both s and a have the same dimensions as x. Thus if x is a length in centimeters, then both s and a are in centimeters. They may be converted to dimensionless quantities by dividing by the average $\bar{x}$. The *fractional standard deviation* is defined as

$$S = \frac{s}{\bar{x}}, \tag{7}$$

and the *fractional average deviation* as

$$A = \frac{a}{\bar{x}}. \tag{8}$$

Both S and A are frequently expressed in percent by multiplying the values given in Eqs. (7) and (8), respectively, by 100.

The fractional deviations S and A have significance only when the measurements are referred to some physically significant zero rather than an arbitrary one. When the zero has a physical significance, S and A are always independent of the units. Even when the measurements are referred to a proper zero, S and A have limited usefulness when the individual deviations become appreciable in comparison to $\bar{x}$.

Illustrations of measurements referred to arbitrary zeros are those of temperature in degrees centigrade and in degrees fahrenheit. The zeros of these scales were chosen purely for human convenience. In fact, since the zeros of the two scales do not coincide, a temperature when expressed in degrees fahrenheit has a different fractional standard deviation than the same temperature when expressed in degrees centigrade.

E. Example: Graphical representation of measurements as a distribution. Suppose that a certain length is measured 51 times. The results which might be obtained are shown in Table I. In the first column appear the various measured values and in the second column the number of times

TABLE I

x (cm)	No. of occurrences m	mx (cm)	$\delta x = x - \bar{x}$ (cm)	$m\lvert\delta x\rvert$ (cm)	$(\delta x)^2$ (cm^2)	$m(\delta x)^2$ (cm)^2
1.01	1	1.01	−0.04	0.04	16×10^{-4}	16×10^{-4}
1.02	3	3.06	−0.03	0.09	9	27
1.03	6	6.18	−0.02	0.12	4	24
1.04	8	8.32	−0.01	0.08	1	8
1.05	10	10.50	0.00	0.00	0	0
1.06	7	7.42	+0.01	0.07	1	7
1.07	8	8.56	+0.02	0.16	4	32
1.08	4	4.32	+0.03	0.12	9	36
1.09	3	3.27	+0.04	0.12	16	48
1.10	0	0.00	+0.05	0.00	25	0
1.11	1	1.11	+0.06	0.06	36	36
	$51 = k = \sum m$	$53.75 = \sum_{n=1}^{51} x_n$		$0.86 = \sum_{n=1}^{51} \lvert\delta x_n\rvert$		$234 \times 10^{-4} = \sum_{n=1}^{51} (\delta x_n)^2$

Average:

$$\bar{x} = \frac{\sum_{n=1}^{51} x_n}{k} = \frac{53.75}{51} = 1.054 \text{ cm.}$$

Average deviation:

$$a = \frac{\sum_{n=1}^{51} |\delta x_n|}{k} = \frac{0.86}{51} = 0.016_8 \text{ cm.}$$

Fractional average deviation:

$$A = \frac{a}{\bar{x}} = \frac{0.0168}{1.054} = 0.016_0 \text{ or } 1.6\%.$$

Standard deviation:

$$s = \sqrt{\frac{\sum_{n=1}^{51} (\delta x_n)^2}{k - 1}} = \sqrt{\frac{234 \times 10^{-4}}{50}} = 0.021_6 \text{ cm.}$$

Fractional standard deviation:

$$S = \frac{s}{\bar{x}} = \frac{0.0216}{1.054} = 0.020_4 \text{ or } 2.0\%.$$

each has occurred. For example, in the third row we note the value 1.03 cm, and to the right in the next column the number 6. This means that 6 of the 51 measurements yielded 1.03 cm.

It is to be noted that both $\Sigma x_n = 53.75$ cm and $\bar{x} = 1.054$ cm are given to four significant figures, although the individual measurements are given only to three. The justification is as follows. Since the individual values are significant to three figures, the last digits in the mx column are at least partially significant. Therefore the grand total will be at least partially significant to the nearest 0.01 cm. Such a procedure is plausible because the average value $\bar{x}$ is more precise than any of the individual values and therefore requires more significant figures than the individual values. (For computing the deviations in Table I, a value of $\bar{x} = 1.05$ cm was used.)

Note that from the directly computed values we obtain the ratio

$$\frac{s}{a} = \frac{0.0216}{0.0168} = 1.29,$$

instead of the 1.25 predicted by theory. This discrepancy is due in part to the fact that the individual deviations are known to only one significant figure, and therefore the precision to which s and a are known is limited. Also, only 51 measurements were employed in this calculation while the theory is based upon the assumption of a very large number, ideally infinite.

The data in Table I may be represented graphically as shown in Fig. 1. We imagine the range of values of x to be divided into equal intervals Δx, and plot the number of values of x lying in the interval versus the average value of x within that interval. Thus the six measurements of 1.03 might be thought of as lying in an 0.01-cm interval centered upon 1.03 cm; that is, between 1.025 and 1.035 cm. Thus we plot 6 on our vertical scale versus 1.03 cm on our horizontal scale. Since with a small number (such as 51) these points do not lie on a smooth curve, it is conventional to represent such a plot by a histogram consisting of a series of horizontal lines of length Δx centered upon the individual points, the ends of adjacent horizontal lines being connected by vertical lines of appropriate length.

If we were to make another 51 measurements and plot the corresponding graph, we would, in general, get a graph which does not coincide with the one given. In other words, in this second set of measurements we might not have obtained six measurements of 1.03 cm, but five, or seven, or even eight. Thus this distribution is subject to what are called *statistical fluctuations.* If we had repeated this process with 500 measurements, we would have found that the relative fluctuations became smaller, and with 5000 they would have been smaller still. At the same time, if we improved our measuring technique to get more significant figures we could use smaller values of the interval Δx. Thus we can conclude that as the number of measurements is indefinitely increased while the width of the intervals Δx is steadily decreased, the histogram approaches a smooth curve. In Section IV-G we shall develop a theory that gives the shape of the smooth curve

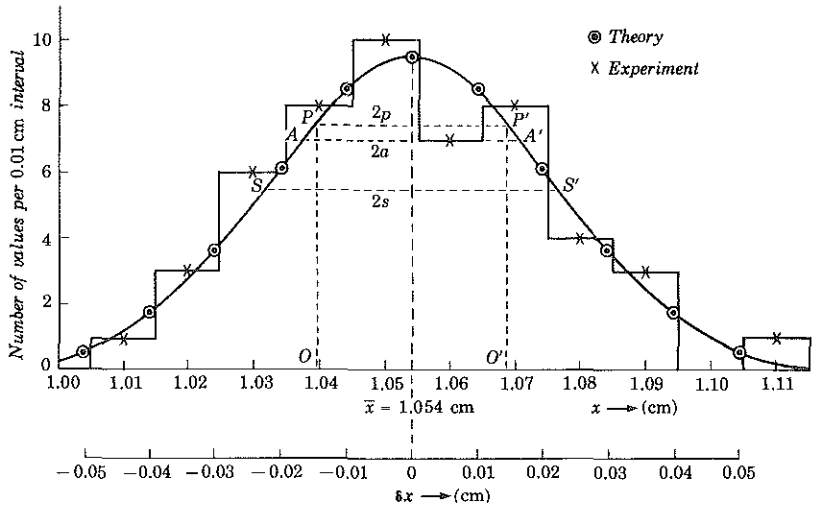

FIGURE 1.

which the histograms approach under these conditions. The smooth curve plotted in Fig. 1 has been calculated using this theory.

It is to be observed that this theoretical curve has a maximum at the average $\bar{x}$ and is symmetrical about this maximum. In fact, the fundamental assumptions used in the theory are just these: that the maximum should occur at the average and that the curve should be symmetrical. That is, values of x less than $\bar{x}$ and values greater than $\bar{x}$ are equally likely. Such assumptions are generally valid when the curve is relatively narrow; that is, when the fractional deviations S and A are not greater than a few percent, as is usual in most circumstances which the student is likely to encounter. This type of curve is called a *Gauss error curve* or a *normal error curve*. In nuclear physics, however, the student may meet some measurements which do not comply with these assumptions and therefore do not conform to a Gauss or normal error law but to another law called the *Poisson law*. (See Section VII.)

For dealing with the theoretical curve, it is more convenient to consider the distribution as plotted against δx rather than against x. Since $\delta x = x - \bar{x}$, this procedure merely consists in moving the origin from $x = 0$ to $x = \bar{x}$ and leaves the shape of the graph unchanged. Shown at the bottom of Fig. 1 is a second horizontal axis which is used in this case.

For reference there are shown on Fig. 1 the following points: A, where $x = \bar{x} - a$; A', where $x = \bar{x} + a$; S, where $x = \bar{x} - s$; S', where $x = \bar{x} + s$. Also are shown four points P, P', O, and O', whose significance will be discussed later.

In Fig. 1 we have partially illustrated the facts that the measurements of an experimental quantity may be represented graphically as a distribution and furthermore that this distribution approaches a definite theoretical

shape as the number of measurements becomes very large. Often we do not wish to take the trouble to actually plot such a distribution, or to find a numerical formula corresponding to the theoretical curve, but we retain these as important concepts. The relative width of the theoretical curve, whether actually drawn or merely imagined, is an indication of the precision of the measurements. After we have derived an analytical expression for the theoretical curve, we shall see that its width may be represented in terms of a single parameter called the *precision index*. Still later we shall show that the precision index is simply related to the standard and average deviations and that therefore these quantities acquire additional significance, since they may also be used indirectly to specify the sharpness of the maximum. Furthermore, as a by-product, we shall obtain the theoretical value of the ratio of the standard deviation to the average deviation. Finally, from the analytical expression for the theoretical curve we shall show that it is possible to set up a procedure for deciding which data should be rejected as being unreliable. The sections immediately to follow will be devoted to the derivation of this expression for the theoretical curve, and to its applications.

F. Probability. If, in our example, we should ask what value would be obtained if we should make another (52nd) determination, it is obvious that we cannot predict the exact value that will occur. All that we can say is that values near the average of 1.054 cm are much more likely to occur than any others. Whereas we cannot predict by theory the value another measurement will yield, we can say something about the "chance" that some particular value will be obtained. "Chance," when expressed quantitatively in mathematical language, is called *probability*, and has the following definition.

If the number of ways in which an event may happen or fail to happen can be analyzed into c successes and b failures, each equally likely to occur, the probability of success in a single trial is

$$W = \frac{c}{c + b}, \tag{9}$$

and the probability of failure is

$$Q = \frac{b}{c + b}. \tag{10}$$

It follows from this definition of probability that the magnitude of a probability (such as W or Q) can never exceed unity; that a probability of unity is to be interpreted as "certainty"; and that the sum of the probabilities of all possible events must be equal to unity. In a simple situation involving two types of event ("success" and "failure")

$$W + Q = 1. \tag{11}$$

EXAMPLE. A box contains 10 white balls and 15 black balls of identical size and weight. What would be the probability, in a dark room, that a white ball would be withdrawn from the box on the first try? There are 10 possible ways of success (withdrawing a white ball) and 15 possible ways of failure (withdrawing a black ball). Thus, $c = 10$, $b = 15$, and $W = 10/(10 + 15) = 0.4$ or 40%, and consequently $Q = 0.6$ or 60%.

Another interpretation of probability is to imagine that the situation leading to an event is repeated a very great number of times. Then W would be the fraction of the total number of times which turned out to be successes, and Q would be the fraction which turned out to be failures. Strictly speaking, this would be true only if the number of repetitions was very large indeed (approaching infinity). For a finite number of repetitions, the fraction of observed successes may differ from W because of statistical fluctuations. As the number of repetitions increases, the agreement between the observed and the theoretical results improves.

The probability that a combination of two events W occurs is equal to the product of the separate probabilities W_1 and W_2, provided that the two events are completely independent.

EXAMPLE. Suppose that in addition to our box with 10 white balls and 15 black balls, we have a second box with 20 white balls and 30 black balls. What would be the probability that we would obtain two white balls on withdrawing one ball from each box?

By Eq. (9), $W_1 = W_2 = 0.4$. Then $W = W_1W_2 = (0.4)^2 = 0.16$. In other words, in a large number of trials, 0.4 of the times we would be successful in drawing a white ball from the first box, and then only 0.4 of these preliminary successes would be followed by success at the second box. If we had withdrawn both balls from the box containing 10 white and 15 black balls, however, W would have had a different value, because W_2 is not 0.4. Once we have withdrawn one white ball, only 9 are left. Thus $W_2 = 9/(9 + 15) = 3/8 = 0.375$. Then $W = (0.4)(0.375) = 0.15$.

It can be inferred that the probability that more than two events take place in combination is the product of the probabilities of the separate events, provided that they are independent.

G. Derivation of the Gauss or "normal" error law. In the following derivation it will be convenient to employ a change of notation. We shall use z to denote a deviation in x (previously denoted by δx), since we shall have occasion to calculate derivatives with respect to this quantity. Thus, $z_1 = x_1 - \bar{x}$, $z_2 = x_2 - \bar{x}$, $z_3 = x_3 - \bar{x}$, and, in general, $z_n = x_n - \bar{x}$. Also, we note that a determination may yield a value in the interval between x and $x + \Delta x$, or a deviation between z and $z + \Delta z$. From the definition of z, an interval Δx wide in x corresponds to an interval Δz of equal width in z (i.e., $\Delta x = \Delta z$).

From Fig. 1, we find that the probability that one determination yields a value in the range of deviation from z to $z + \Delta z$ depends on two factors: (1) the width of the interval Δz, and (2) some unknown function of z, which we shall denote by $f(z)$. This function is assumed to have a maximum at

$z = 0$ ($x = \bar{x}$), and approaches zero when z becomes very large in either a positive or negative sense. The dependence upon the interval is obvious. In our example, we see that in a single interval of 0.01 cm, from 1.055 to 1.065 cm, we have 7 measurements, and in the interval of 0.01 cm between 1.065 and 1.075 cm we have 8 measurements. If we divide our data into intervals of 0.02 cm each, such as from 1.055 to 1.075 cm, we would have 15 measurements within the interval or, on the average, twice as many as in one of these 0.01 cm intervals.

The function $f(z)$, sometimes called the *probability amplitude*, is defined as the probability of obtaining a deviation lying in a unit interval centered at z. Since the probability is, in general, proportional to $f(z)$ and Δz, it is proportional to their product. The probability that a value lies in an interval of width Δz centered at z is

$$W = f(z)\,\Delta z.$$

If we make k measurements, neglecting statistical fluctuations, the number found to lie in this range is given by

$$kW = kf(z)\,\Delta z. \tag{12}$$

We shall determine the analytical form of $f(z)$ according to the method of Reddick and Miller.* We assume that a large number of values $x_1, x_2, \ldots, x_k$ have been obtained by experiment. These have deviations $z_1, z_2, \ldots, z_k$, respectively. If the range of values of z is divided into intervals Δz small enough so that not more than one value occurs in each interval, we may say that

$$\begin{aligned}
f(z_1)\,\Delta z &= \text{probability of the deviation } z_1,\\
f(z_2)\,\Delta z &= \text{probability of the deviation } z_2,\\
&\vdots\\
f(z_n)\,\Delta z &= \text{probability of the deviation } z_n,\\
&\vdots\\
f(z_k)\,\Delta z &= \text{probability of the deviation } z_k.
\end{aligned}$$

Therefore the probability of this combination of deviations is

$$W = f(z_1)f(z_2)\cdots f(z_k)(\Delta z)^k, \tag{13}$$

or, taking the logarithm of both sides,

$$\ln W = \ln f(z_1) + \ln f(z_2) + \cdots + \ln f(z_k) + k \ln (\Delta z). \tag{14}$$

**Op. cit.*, p. 357 (see list of references).

We choose the function $f(z)$ subject to the following restrictions:

(1) The total number of deviations over the entire range of possible values of z (from $-\infty$ to $+\infty$) must be equal to the total number of measurements k or, by use of Eq. (12),

$$k = k\int_{-\infty}^{+\infty} f(z)dz,$$

or

$$1 = \int_{-\infty}^{+\infty} f(z)dz. \tag{15}$$

Equation (15) is a mathematical expression of our previous statement that the sum of all possible probabilities must be unity (certainty).

(2) We assume that the most probable value $\bar{x}$ is the average of the measured values. From the definition of *average*, it follows that

$$z_1 + z_2 + \cdots + z_k = \sum_{n=1}^{k} z_n = 0. \tag{16}$$

(3) When $\bar{x}$ is the average of the measured values, W (and $\ln W$) becomes a maximum. Therefore

$$\frac{d(\ln W)}{d\bar{x}} = 0.$$

The probability W depends indirectly on $\bar{x}$ through the z's. By use of indirect differentiation, from Eq. (14),

$$\begin{aligned}\frac{d(\ln W)}{d\bar{x}} &= \frac{\partial(\ln W)}{\partial z_1}\frac{dz_1}{d\bar{x}} + \frac{\partial(\ln W)}{\partial z_2}\frac{dz_2}{d\bar{x}} + \cdots + \frac{\partial(\ln W)}{\partial z_k}\frac{dz_k}{d\bar{x}} \\ &= \frac{df(z_1)/dz_1}{f(z_1)}\frac{dz_1}{d\bar{x}} + \frac{df(z_2)/dz_2}{f(z_2)}\frac{dz_2}{d\bar{x}} + \cdots + \frac{df(z_k)/dz_k}{f(z_k)}\frac{dz_k}{d\bar{x}} \\ &= 0. \end{aligned} \tag{17}$$

The last term in Eq. (14) is a constant, and its derivative is zero. Since each term involves only one z, we may write Eq. (17) as

$$\sum_{n=1}^{k} \phi(z_n) = \phi(z_1) + \phi(z_2) + \cdots + \phi(z_k) = 0, \tag{18}$$

wherein a typical term is defined by

$$\phi(z_n) = \frac{df(z_n)/dz_n}{f(z_n)}. \tag{19}$$

In arriving at Eq. (18) from Eq. (17), we have employed the fact that from the definition of z [Eq. (1)],

$$\frac{dz_1}{d\bar{x}} = \frac{dz_2}{d\bar{x}} = \cdots = \frac{dz_k}{d\bar{x}} = -1.$$

Next we assume, in line with the common technique of solving differential equations, that $\phi(z)$ is given by a power series:

$$\phi(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \ldots, \tag{20}$$

where the a's are constants to be determined by substituting the expression for $\phi(z)$ from Eq. (20) into Eq. (18).

$$\begin{aligned}
\phi(z_1) &= a_0 + a_1 z_1 + a_2 z_1^2 + a_3 z_1^3 + \cdots, \\
\phi(z_2) &= a_0 + a_1 z_2 + a_2 z_2^2 + a_3 z_2^3 + \cdots, \\
&\vdots \\
\phi(z_k) &= a_0 + a_1 z_k + a_2 z_k^2 + a_3 z_k^3 + \cdots \\
\hline
\text{Sum} = \sum_{n=1}^{k} \phi(z_n) &= k a_0 + a_1 \sum_{n=1}^{k} z_n + a_2 \sum_{n=1}^{k} z_n^2 + a_3 \sum_{n=1}^{k} z_n^3 + \cdots \\
&= 0.
\end{aligned} \tag{21}$$

Equation (21) can be satisfied if each of the individual terms is equal to zero. This may be accomplished by having all of the a's except a_1 equal to zero. From Eq. (16), $\sum_{n=1}^{k} z_n = 0$, and thus a_1 need not be equal to zero. All the other a's must be equal to zero, however, in order to satisfy Eq. (21) even when the z's vary because of statistical fluctuations. By use of Eqs. (19) and (20),

$$\phi(z) = \frac{df(z)/dz}{f(z)} = a_1 z,$$

or

$$\frac{df(z)}{f(z)} = a_1 z\, dz.$$

Upon integration,

$$\ln f(z) = \frac{a_1 z^2}{2} + \ln C,$$

or

$$f(z) = C e^{a_1 z^2/2}, \tag{22}$$

where C is a constant of integration.

Since $f(z)$ is known to have a maximum at $z = 0$, and since we expect $f(z)$ to approach zero for z very large in either a positive or negative sense, a_1 must be negative. Thus, it is appropriate to replace it by a constant h defined by

$$h^2 = -\frac{a_1}{2},$$

yielding

$$f(z) = Ce^{-h^2z^2}. \tag{23}$$

The constant C is evaluated by substituting the expression for $f(z)$ from Eq. (23) into Eq. (15), and performing the indicated integration. This integral must be evaluated by special methods.* The result is that $C = h/\sqrt{\pi}$, and the probability that a value occurs with a deviation from z to $z + \Delta z$ is

$$f(z)\Delta z = \frac{h}{\sqrt{\pi}} e^{-h^2z^2}\Delta z. \tag{24}$$

The number of measurements with this deviation, on the average (neglecting statistical fluctuations), is

$$kf(z)\Delta z = \frac{kh}{\sqrt{\pi}} e^{-h^2z^2}\Delta z. \tag{25}$$

The graph of $f(z)$ vs. z is sketched in Fig. 2. From this figure, we see that a large value of h yields a high, narrow curve, which corresponds to an experiment of high precision. A small value of h yields a low, wide curve, which corresponds to an experiment of low precision. Consequently, h is called the *precision index*. The geometrical interpretation of Eq. (15) is that the area under any such curve must be unity, and thus the areas under all such curves must be the same.

The probability of obtaining a measurement with a deviation lying in the range from $-Z$ to $+Z$, where Z is some arbitrarily chosen limit, is found by integrating Eq. (24) from $-Z$ to $+Z$; because of the symmetry of the integrand about $z = 0$, this integral is twice the integral from zero to $+Z$. Except for the particular case of the definite integral from zero to infinity, this integral cannot be evaluated in closed form, but values can be obtained to any desired accuracy by numerical integration or by approximation methods. Such values may be found in almost all standard compilations of mathematical tables. For such a table to be independent of the particular value of h, it is necessary to make a transformation of the variable. Some

---

*Reddick and Miller, *op. cit.*, p. 359. See also A. G. Worthing and J. Geffner, *Treatment of Experimental Data*, John Wiley and Sons, Inc., New York (1943), p. 155. This integral is listed as No. 492 by B. O. Pierce, *A Short Table of Integrals*, Ginn and Co., Boston (1929).

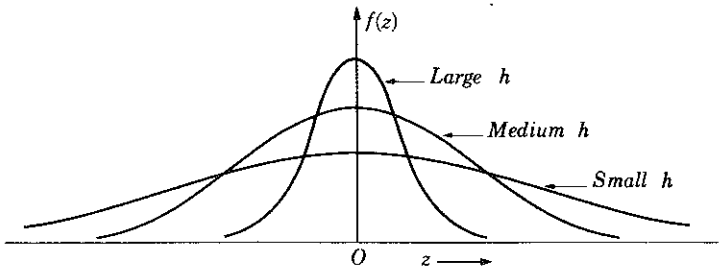

FIGURE 2.

compilations employ the variable $t = hz$, while others, utilizing a relation between h and the standard deviation s derived below as Eq. (29), employ the variable $t' = z/s$. Then the probability of obtaining a deviation having a magnitude of Z or smaller is given by

$$W_{-Z,Z} = 2\int_0^Z f(z)dz = \frac{2}{\sqrt{\pi}}\int_0^Z e^{-h^2z^2}(hdz)$$

$$= \frac{2}{\sqrt{\pi}}\int_0^{t=hZ} e^{-t^2}dt \qquad (26)$$

$$= \sqrt{\frac{2}{\pi}}\int_0^{t'=Z/s} e^{-t'^2/2}\,dt'. \qquad (26A)$$

Data concerning this integral may also be expressed in terms of the probability of obtaining a deviation of magnitude equal to or greater than Z, which, of course, is $1 - W_{-Z,Z}$. Some values of this quantity will be given later in Table II.

H. Probable error. Relation between the various quantities used to express precision.

(1) *Probable error.* The probable error p is that magnitude of deviation whose probability of being exceeded is one-half. In other words, p is that value of Z which, when substituted into Eq. (26), yields $W_{-Z,Z} = \frac{1}{2}$. This consideration gives

$$p = \frac{0.4769}{h}. \qquad (27)$$

The geometrical interpretation of the meaning of p is given in Fig. 3. Vertical lines drawn under the curve $f(z)$ vs. z at the points $z = p$ and $z = -p$ divide the area under the curve into three parts. The central part (crosshatched) is equal to the sum of the other two. In Fig. 1, the lines OP and $O'P'$ have the same meaning.

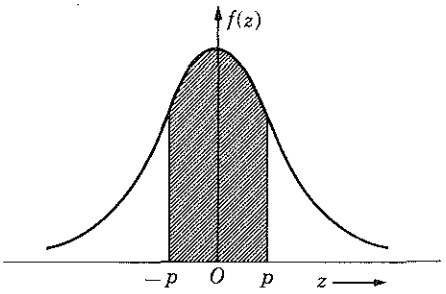

FIGURE 3.

(2) *Standard deviation.* The probable error p and the precision index h may be evaluated in terms of the standard deviation s or the average deviation a. The number of values between z and $z + \Delta z$ is given by Eq. (25). Each of these contributes an amount z^2 to the sum of the squares of the deviations. The derivation of the Gauss error law has been based upon the implication of an infinite number of measurements; hence the Gaussian distribution is a property of the universe and not of the sample. Therefore, in accord with the definition given in Section IV-B, we can determine the standard deviation by computing the sum of the squares of the deviations by integration and then dividing by the number of measurements k. This division by k cancels the factor of k contained in Eq. (25), and

$$s^2 = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} z^2 e^{-h^2 z^2}\, dz = -\frac{1}{h\sqrt{\pi}} \int_0^{\infty} z e^{-h^2 z^2} (-2h^2 z\, dz).$$

On integration by parts,

$$s^2 = -\frac{1}{h\sqrt{\pi}} \left(\left[z e^{-h^2 z^2} \right]_0^{\infty} - \int_0^{\infty} e^{-h^2 z^2}\, dz \right). \tag{28}$$

The first term on the right vanishes at both limits. The second may be evaluated by means of the tables, as it involves the same integral as in Eq. (26). The result is

$$s = \frac{1}{h\sqrt{2}}. \tag{29}$$

(3) *Average deviation.* By an analogous procedure, the sum of the deviations without regard to sign and the average deviation a may be calculated. The contribution of all the values with negative deviations is equal to that of all the values with positive deviations. By reference to Eqs. (6) and (25),

$$a = 2\int_0^\infty zf(z)dz = \frac{2h}{\sqrt{\pi}}\int_0^\infty ze^{-h^2z^2}dz = \frac{1}{h\sqrt{\pi}}. \tag{30}$$

Equations (27), (29), and (30) may be summarized by a multiple proportionality:*

$$p:a:s:1/h = 0.477:0.564:0.707:1.000. \tag{31}$$

Thus,

$$\frac{p}{s} = \frac{0.477}{0.707} = 0.6745, \tag{31A}$$

a result which is often quoted and used for determining p. Also,

$$\frac{s}{a} = \frac{0.707}{0.564} = 1.25, \tag{32}$$

as has been mentioned in Section IV-E.

The quantities s and a are obtained directly by analysis of the data in the manner illustrated, while h and p are calculated from s or a by means of Eq. (31).

In the example given in Table I:

$$p = 0.6745s = 0.6745 \times 0.021_6 = 0.014_6 \text{ cm} \quad \text{[by Eq. (31A)]}$$

and

$$h = \frac{0.707}{s} = 33 \text{ cm}^{-1}. \qquad \text{[by Eq. (31)]}$$

The equation for the curve in Fig. 1 is obtained by substituting the above value of h and the values $k = 51$ and $\Delta z = 0.01$ cm into Eq. (25), yielding

$$kf(z)\Delta z = \frac{51 \times 33}{\sqrt{\pi}} e^{-(33z)^2} \times 0.01 = 9.5\, e^{-(33z)^2}.$$

I. Rejection of data. When a measurement is repeated several times, quite frequently most of the resulting values are found to lie close together while one or two others differ very much from this group. The question immediately arises as to whether these apparently anomalous measurements should be rejected.

In cases where physical disturbances were known to be present when the questionable measurements were made, obviously they should be rejected.

---

*Worthing and Geffner, *op. cit.*, p. 158.

In fact, even if the data obtained during a disturbance appear to agree with others, they should be rejected.

In other situations where no disturbances were known to be present, it is highly controversial whether the data should be rejected. Some observers contend that all such data should be retained. Ideally, so many additional measurements should be made that retaining or rejecting the dubious measurements has little effect on the average value, but unfortunately such a procedure is often impractical.

Other observers are inclined to reject the questionable measurements on the basis that the probability of obtaining such large deviations is so negligible that they must have resulted from some abnormal cause. Perhaps there were physical disturbances that were not detected. In an electrical experiment, the line voltage may have dropped briefly, but the observer, writing in his notebook, may not have seen the meters flicker. Alternatively, the observer may have made outright blunders in reading the instruments.

While there is considerable justification for the latter point of view, it presents at least two difficulties. In the first place, there is no clear-cut basis for deciding just how small the probability must be to be considered negligible, and thus any specific criterion for rejecting data is a matter of debate. In the second place, if some definite criterion is adopted, and then applied without discretion, it may lead to an unreasonable situation. In applying it to the original set of data, the observer rejects some of the measurements. He may then suppose that some of the remaining discrepancies are also due to abnormal causes, and apply the criterion to the remaining data, and rejecting other measurements. This reasoning leads to successive applications, each resulting in rejections, until ultimately nearly all of the data have been rejected. Obviously, any such repetitive applica-

TABLE II

Probability of obtaining a deviation greater than some preassigned magnitude

Deviation relative to standard deviation $\lvert t' \rvert = \lvert Z \rvert / s$	Probability $1 - W_{-z,z}$
$0.6745 = p/s$	0.50
0.707	0.48
$0.80 = a/s$	0.43
1.0	0.32
$1.41 = 1/sh$	0.16
1.5	0.13
2.0	0.05
2.5	0.012
3.0	0.003

tion of a criterion is ridiculous. Most observers who employ such a procedure at all apply it only to the original data.

The Gauss error law provides a rational basis for setting up a criterion, if one is to be used. If the probability of obtaining deviations greater in magnitude than some limit $|z|$ is negligible compared to $1/k$, where k is the number of measurements, then all measurements with deviations greater in magnitude than $|z|$ are to be rejected. The probability of obtaining a deviation greater than $|z|$ is $1 - W_{-z,z}$, where $W_{-z,z}$ is given by Eq. (26). The quantitative definition of the word "negligible" is a matter of opinion which will be left to the reader. However, for his convenience some representative values of $1 - W_{-z,z}$ as a function of the deviation are given in Table II. Note that this probability decreases very rapidly as $|z|$ increases beyond s.

While the Gauss error law provides a simple method of rejecting data rationally, its use in this application is only approximately valid. More valid but more complicated methods exist. The Gaussian distribution is a property of the universe of measurements and only approximates the distribution of a particular sample, and with the usual case of small samples this approximation is not very good. A better method involves the use of the statistics of small samples, a subject which is beyond the scope of this book. This method is discussed in the paper by Thompson.*

*W. R. Thompson, *Annals of Mathematical Statistics* **6**, 214 (1935).

V. PROPAGATION OF ERROR

Let us suppose that a quantity V is to be computed from two measured quantities x and y by means of a theoretical formula. In other words, $V = V(x, y)$. Then V is in error by an amount dV as a result of the errors dx and dy in the measured quantities x and y. (The quantities dV, dx, and dy must be comparable. If one represents a probable error, the others must also. If one represents an average deviation, the others must be average deviations, etc.) The relation between dV and dx and dy may be determined from differential calculus and other methods we have previously employed. There are two limiting situations. Usually the error dx has no relation to the error dy, and they are said to be *independent*. For example, suppose that we determine the speed of an object by observing the time it takes to travel a measured distance. There is no reason to believe that if the time measurement is too large, the distance measurement is necessarily also too large. On the other hand, situations may exist where the errors are related, and these are called *nonindependent*. For example, suppose that an area is determined by measuring the length and width of a rectangle with a tape measure which has stretched since its original calibration. Obviously, the measurements of both length and width will then be too small. Sometimes both independent and nonindependent errors are present.

In the case of independent errors, there is clearly some possibility of compensation. When the error in x causes V to be too large, the error in y may cause it to be too small. Thus we should expect that, on the average, the total error of V will be algebraically less than the sum of the separate contributions of dx and dy. We shall soon see that a logical way of adding the separate contributions is to take the square root of the sum of their squares, and this method does have the expected compensating property. On the other hand, there is no such possibility of compensation for nonindependent errors, and therefore these contributions do add algebraically.

In addition to independent and nonindependent errors there is a third class known as *correlated* errors, in which the deviations of x are systematically related to those of y. The distinction between independent and correlated errors may be illustrated by the following two situations. On the one hand, we might determine the area of a rectangle rapidly, under constant conditions, by measuring the length x and the width y ten times each with the same tape. If we were to consider the sum of the products of the deviations, $\Sigma \delta x_n \delta y_n$, the only logical pairing of individual deviations of x with individual deviations of y would be a completely random one, and we would expect on the average that the sum would be zero. Then the errors associated with these deviations would be completely independent. In addition

there would be some nonindependent error; for example, if the temperature were not known very well, there would be error in the correction for the expansion of the tape, which would affect both measurements in the same way. The presence of this type of error cannot be detected by an analysis of the deviations.

On the other hand, suppose we were to make one measurement of the length and one of the width each day for ten days, and suppose that there were a considerable variation in termperature. If we had proper information, we could correct the individual measurements for the effect of the expansion to a very high precision; in this case the deviations would be independent, and on the average no larger than if all the measurements were made at the same time. However, if for some reason we should ignore the effect of variation of temperature, the deviations would be larger on the average. Furthermore, the situation suggests that we should pair a measurement of width with the one of length made on the same day. Then on the hotter days δx and δy would both tend to be negative and on the cooler ones they would both tend to be positive. Hence positive terms in the sum $\Sigma \delta x_n \delta y_n$ would be more likely to occur than negative ones, and the sum would not be zero. (In the limit where the expansion effects entirely predominate over the completely random ones, the sum would be approximately equal to $s_x s_y$.) Under these conditons the errors in x and y would be said to correlated. It should be noted that if the variations in temperature were small compared with the difference between the average temperature and the temperature at which the tape was calibrated, there would also be some purely nonindependent error associated with the error in correction of the average temperature.

In a sense, correlated errors are not an intrinsically new class of errors, but rather errors whose independent and nonindependent portions cannot be separated. The method for their addition is intermediate between the methods for independent and nonindependent errors: in dealing with correlated errors it is necessary to define the *correlation coefficient* ρ_{xy}, which is the average value of the sum $\Sigma \delta x_n \delta y_n$ for the universes of measurements of x and y, divided by $s_x s_y$. In the sum the deviations are paired in accordance with some known or suspected correlation. Advanced books on statistics show that the best estimate of ρ_{xy} from a finite sample of k pairs of measurements of x and y is given by

$$\rho_{xy} = \frac{1}{(k-1)s_x s_y} \sum_{n=1}^{k} (\delta x_n \delta y_n), \tag{33}$$

or

$$\rho_{xy} = \frac{1}{(k-1)s_x s_y} \left[\sum_{n=1}^{k} (x_n y_n) - k\bar{x}\,\bar{y} \right]. \tag{33A}$$

In the case of purely independent errors, of course, ρ_{xy} is equal to zero, while for completely correlated errors ρ_{xy} has the value of either $+1$ or -1.

Purely independent errors are much more common than correlated ones, and they may usually be recognized as being independent from the physical context. In such cases ρ_{xy} may be set equal to zero *a priori*. In cases where correlation is known or suspected to exist it is necessary to calculate ρ_{xy}.

A. General rule for combining independent and correlated errors. Let us suppose that we make a very large number k of measurements of both x and y. Ultimately we shall let k become infinite, so that our sample will include the entire universes of measurements of x and y. If we pair the measurements of x and y in accordance with some known or suspected correlation, we may compute k values of V:

$$V_1 = V(x_1, y_1),\ V_2 = V_2(x_2, y_2), \ldots, V_n = V(x_n y_n), \ldots,$$
$$V_k = V(x_k, y_k).$$

The measurements of x and of y may be averaged in the usual way to obtain $\bar{x}$ and $\bar{y}$. We shall make the fundamental assumption that the best possible (most probable) value of V is $V(\bar{x}, \bar{y})$. In general, each value V_n differs from $V(\bar{x}, \bar{y})$ by some amount

$$\delta V_n = V_n - V(\bar{x}, \bar{y}), \tag{34A}$$

which we shall call the *deviation* of V_n. From the differential calculus, to the first order of approximation,

$$\delta V_n = \frac{\partial V}{\partial x}\,\delta x_n + \frac{\partial V}{\partial y}\,\delta y_n, \tag{34B}$$

where $\delta x_n = x_n - \bar{x}$ and $\delta y_n = y_n - \bar{y}$, respectively, are the deviations of x_n and y_n. The average value of $\delta V = (\sum_{n=1}^{k} \delta V_n)/k = 0$, since $\sum_{n=1}^{k} \delta x_n = 0$ and $\sum_{n=1}^{k} \delta y_n = 0$. Therefore, to a first approximation, which in ordinary cases is valid,

$$\overline{V} = V(\bar{x}, \bar{y}). \tag{35}$$

In Eq. (34A) we have defined δV_n in a manner not strictly consistent with the way in which we defined δx_n and δy_n in Eq. (1). However, using Eq. (35), we see that these two possible definitions are equivalent in most circumstances.

Next let us compute the average value of the sum of the squares of the deviations by the use of Eq. (34B):

$$\frac{\sum_{n=1}^{k} (\delta V_n)^2}{k} = \left(\frac{\partial V}{\partial x}\right)^2 \frac{\sum_{n=1}^{k} (\delta x_n)^2}{k} + \left(\frac{\partial V}{\partial y}\right)^2 \frac{\sum_{n=1}^{k} (\delta y_n)^2}{k}$$
$$+ 2\left(\frac{\partial V}{\partial x}\right)\left(\frac{\partial V}{\partial y}\right) \frac{\sum_{n=1}^{k} (\delta x_n \delta y_n)}{k}. \tag{36}$$

If we take the square root of each side and take the limit as k becomes infinite, then the quantity on the left becomes, by definition, the standard deviation of V, which we denote by s_V. The various sums which appear on the right side are expressible in terms of the standard deviations s_x and s_y and the correlation coefficient ρ_{xy}, by use of their respective definitions. Hence

$$s_V = \sqrt{\left(\frac{\partial V}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial V}{\partial y}\right)^2 s_y^2 + 2\rho_{xy}\left(\frac{\partial V}{\partial x}\right)\left(\frac{\partial V}{\partial y}\right) s_x s_y}. \qquad (37)$$

With a finite number of measurements, we obtain the best estimate of s_V by substituting into Eq. (37) the best estimates of s_x and s_y from Eq. (5) or Eq. (5A), and the best estimate of ρ_{xy} from Eq. (33) or Eq. (33A). In the case of correlated errors we normally have the same number of measurements of x as of y. With measurements known *a priori* to be independent, ρ_{xy} is set equal to zero; in this case, in general, the number of measurements of x differs from the number of measurements of y.

Up to this point we have assumed that V is a function of only two measured quantities x and y. Of course, V may be a function of any number of measured quantities. The generalization of Eq. (37) to include the contributions of these quantities is obvious. It is merely necessary to include under the radical on the right additional terms of the same types as are already present but pertaining to the other quantities. For example, if a third quantity w is involved, the following terms would be added:

$$\left(\frac{\partial V}{\partial w}\right)^2 s_w^2 + 2\rho_{wx}\left(\frac{\partial V}{\partial w}\right)\left(\frac{\partial V}{\partial x}\right) s_w s_x + 2\rho_{wy}\left(\frac{\partial V}{\partial w}\right)\left(\frac{\partial V}{\partial y}\right) s_w s_y.$$

The justification for this process may be seen by reexamining the previous derivation with the inclusion of these quantities in mind. Any of the special rules for combining errors which are based upon Eq. (37) may be generalized to include other measured quantities in a similar way.

We expect the probable errors, average deviations, and reciprocals of precision indices of V, x, and y to be in the ratios to their respective standard deviations given by Eq. (31). Therefore the best estimate of the probable error of V is obtained by replacing s_V, s_x, and s_y in Eq. (37) by p_V, p_x, and p_y; and similarly for the average deviation a_V and the reciprocal of the precision index $(1/h_V)$.

B. The standard deviation of an average. Up to this point we have been concerned with estimating the standard deviation of a single measurement. In other words, in Section IV we have developed a procedure whereby from the analysis of k observations we estimate the error we would have if we had made only one observation. This, of course, is a peculiar bit of hindsight. Having the k measurements, we know that the best available value is the average $\bar{x}$, and therefore we are mainly interested in the estimate of the error of $\bar{x}$.

We could, of course, make several sets of k readings, compute their respective averages, and then apply the previously developed methods to these averages. We could also compute the average of these averages, which would be still more precise, and inquire into its error. This reasoning could go on indefinitely without coming to a logical end. Fortunately, by application of Eq. (37) we can deduce a procedure for estimating the error in the average of k measured values without having to repeat the set of measurements.

We make use of the fact that the average $\bar{x}$ is a quantity computed from the measured quantities $x_1, x_2, \ldots, x_k$. Thus, for this purpose, V is to be replaced by $\bar{x}$. By definition,

$$\bar{x} = \frac{1}{k}(x_1 + x_2 + \cdots + x_n + \cdots + x_k). \tag{38}$$

Therefore,

$$\frac{\partial \bar{x}}{\partial x_1} = \frac{\partial \bar{x}}{\partial x_2} = \cdots = \frac{\partial \bar{x}}{\partial x_n} = \cdots = \frac{\partial \bar{x}}{\partial x_k} = \frac{1}{k}. \tag{39}$$

Since we are considering only the random errors in the measurements, the x_n's may be considered independent, and their mutual correlation coefficients placed equal to zero. Then by Eq. (37), modified to include the effects of k quantities, the standard deviation of the average is

$$s_{\bar{x}} = \sqrt{\frac{1}{k^2}(s_{x_1}^2 + s_{x_2}^2 + \cdots + s_{x_n}^2 + \cdots s_{x_k}^2)} = \frac{s}{\sqrt{k}}, \tag{40}$$

since each $s_{x_n} = s$ and the number of them is k. Thus the standard deviation of the average is equal to the standard deviation of an individual measurement, divided by the square root of the number of the independent measurements. In other words, the precision improves in proportion to the square root of the number of measurements in the sample. This is a fundamental principle of statistics. In Eq. (85), which will follow, we shall show that the same principle applies when the number of observations is the random variable.

The standard deviation s, which appears in Eq. (40), can be evaluated by using Eq. (5) or Eq. (5A). We are now in a position to derive these equations, which were given without proof. We recall that s is the rms deviation about the universe average, which we shall denote by $\bar{x}_u$. This quantity may be computed from Eq. (4), after evaluating $\sum_{n=1}^{k} (\delta x_n)^2$ from Eq. (2) with $\bar{x}$ replaced by $\bar{x}_u$. If, to avoid confusion, we now denote the sample average by $\bar{x}_k$, then from Eq. (1) the quantity $\sum_{n=1}^{k} x_n$, which appears in the second term of Eq. (2), equals $k\bar{x}_k$. Making these substitutions and squaring, we obtain

$$s^2 = \frac{\sum_{n=1}^{k} x_n^2}{k} - 2\bar{x}_u\bar{x}_k + \bar{x}_u^2. \tag{41}$$

The square of the rms deviation about the sample average is, from Eq. (4A),

$$s'^2 = \frac{\sum_{n=1}^{k} x_n^2}{k} - \bar{x}_k^2. \tag{41A}$$

Subtracting Eq. (41A) from Eq. (41), we obtain

$$\begin{aligned} s^2 - s'^2 &= \bar{x}_k^2 - 2\bar{x}_u\bar{x}_k + \bar{x}_u^2 \\ &= (\bar{x}_k - \bar{x}_u)^2. \end{aligned} \tag{41B}$$

The quantity inside the parentheses of Eq. (41B) is the deviation of $\bar{x}_k$ about $\bar{x}_u$. We estimate this to be equal to $s_{\bar{x}}$, which may be expressed in terms of s by means of Eq. (40). We may then solve for s in terms of s', with the result that

$$s = s' \sqrt{\frac{k}{k-1}}. \tag{42}$$

When the expressions for s' from Eqs. (4) and (4A) are substituted into Eq. (42) we obtain, respectively, Eqs. (5) and (5A), which are the desired results.

Logically, we expect the averages $\bar{x}$ and $\bar{y}$ to have the same degree of correlation as do individual pairs of measurements. In fact, it is shown in advanced treatises that $\rho_{\bar{x}\bar{y}} = \rho_{xy}$. Therefore we expect the standard deviations of averages to combine in the same way as the standard deviations of individual measurements. (This statement is especially plausible in the case where we have k measurements of both x and y. Then $s_{\bar{x}}/s_x = s_{\bar{y}}/s_y = 1/\sqrt{k}$, and therefore we expect also that $s_{\bar{V}}/s_V = 1/\sqrt{k}$.) Accordingly we may write

$$s_{\bar{V}} = \sqrt{\left(\frac{\partial V}{\partial x}\right)^2 s_{\bar{x}}^2 + \left(\frac{\partial V}{\partial y}\right)^2 s_{\bar{y}}^2 + 2\rho_{xy}\left(\frac{\partial V}{\partial x}\right)\left(\frac{\partial V}{\partial y}\right) s_{\bar{x}} s_{\bar{y}}}. \tag{43}$$

C. Method of combining nonindependent errors. If the errors of x and y are due to some common cause, the error of V is given directly by the differential calculus:

$$dV = \frac{\partial V}{\partial x}\,dx + \frac{\partial V}{\partial y}\,dy. \tag{44}$$

Note that for the case of completely correlated errors ($\rho_{xy} = 1$), Eq. (37) reduces to the same mathematical form as Eq. (44). However, completely correlated and nonindependent errors are of very different origins. Completely correlated errors are associated with the individual measurements of x and y and could be detected by computing the correlation coefficient from

the deviations and finding that it equals unity. Nonindependent errors are due to external causes and cannot be detected by a study of the deviations. For example, in the first determination of the area of a rectangle discussed in the introductory paragraphs of this chapter, the nonindependent errors are those associated with the calibration of the tape and the correction for thermal expansion. These may random in nature, so that they can be evaluated by application of the previously developed theory to the calibration or correction process, or they may be of systematic nature, so that they can only be estimated, as will be discussed in further detail in Section V-F.

For these reasons, it is inappropriate to denote the nonindependent errors by the symbols s_V, s_x, and s_y. Instead we have used dV, dx, and dy in Eq. (44).

Note that nonindependent and completely correlated errors are combined by algebraic addition. On the other hand, completely independent errors ($\rho_{xy} = 0$) are combined by taking the square root of the sum of their squares, as we see from Eq. (37).

When both nonindependent and correlated errors are present, they can be computed from Eqs. (44) and (37), respectively. However, these groups of errors are completely independent of each other: the total error in V is found by calculating $\sqrt{dV^2 + s_V^2}$.

D. Graphical addition of errors. Equation (37) may be simply interpreted in terms of the addition of vectors (i.e. the law of cosines for the triangle). It can be seen that s_V is just the resultant of the vectors $(\partial V/\partial x)s_x$ and $(\partial V/\partial y)s_y$, and ρ_{xy} is equal to minus the cosine of the angle between them.

Two cases of special interest are illustrated in Fig. 4. In (a) is shown the situation of completely independent errors ($\rho_{xy} = 0$), in which the original vectors are at right angles, while in (b) is shown that of nonindependent (and also completely correlated) errors, in which the vectors are parallel.

It is well known that when one arm of a right triangle is short compared with the other, the hypotenuse is not much longer than the long arm, and variations of the short arm have little effect upon the hypotenuse. Therefore, when one independent error is appreciably smaller than the other, it may generally be neglected; it would be a waste of time to attempt to improve the measurement of the quantity to which this term belongs. These

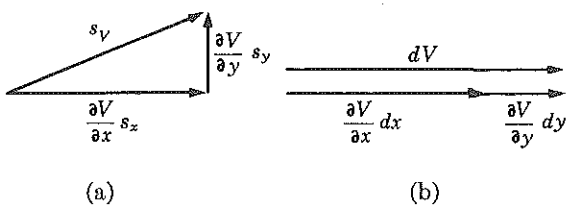

FIG. 4. (a) Graphical addition of completely independent errors. (b) Graphical addition of nonindependent errors.

TABLE III

Illustration of Eq. (37)

$$\left(\frac{\partial V}{\partial x}\right)s_x = 1.000$$

$\left(\frac{\partial V}{\partial y}\right)s_y$	s_V
1.000	1.414
0.500	1.225
0.250	1.118
0.100	1.049
0.010	1.005
0.001	1.0005

remarks are further illustrated by Table III, where s_V is computed for various values of the second term when the first is assumed equal to unity.

E. Special rules for combining errors. The rules expressed by Eqs. (37), (43), and (44) are valid regardless of the mathematical form of $V = V(x, y)$. A number of special rules for such commonly occurring functions as sums and products may be derived from these equations. The computed quantity is given in terms of two measured quantities; generalized forms for more than two variables may be obtained by the methods discussed at the end of Section V–A. These rules are given for independent errors, which are the ones usually encountered. The corresponding rules for nonindependent errors may be obtained by substituting ordinary algebraic addition for addition by taking the square root of the sum of the squares, and replacing the standard deviations by the estimated amounts of errors.

(1) *Sum and difference.* $V = x \pm y$.

$$\frac{\partial V}{\partial x} = 1 \text{ and } \frac{\partial V}{\partial y} = \pm 1.$$

Therefore, by Eq. (37),

$$s_V = \sqrt{s_x^2 + s_y^2}. \tag{45}$$

In other words, the *absolute* standard deviation of a quantity which is the sum or difference of measured quantities is equal to the square root of the sum of the squares of the *absolute* standard deviations of the measured quantities.

(2) *Product of factors raised to various powers.* Let

$$V = x^m y^q, \tag{46}$$

where m and q are known exactly from theory:

$$\frac{\partial V}{\partial x} = mx^{m-1}y^q \quad \text{and} \quad \frac{\partial V}{\partial y} = qx^m y^{q-1}.$$

Substituting these values into Eq. (37):

$$s_V = \sqrt{m^2x^{2(m-1)}y^{2q}s_x^2 + q^2x^{2m}y^{2(q-1)}s_y^2}. \tag{47}$$

Dividing Eq. (47) by Eq. (46):

$$\frac{s_V}{V} = \sqrt{m^2\left(\frac{s_x}{x}\right)^2 + q^2\left(\frac{s_y}{y}\right)^2}. \tag{48}$$

However, according to Eq. (7), the fractional standard deviations are, respectively,

$$S_V = \frac{s_V}{V}, \quad S_x = \frac{s_x}{x}, \quad S_y = \frac{s_y}{y}.$$

Therefore Eq. (48) becomes

$$S_V = \sqrt{m^2S_x^2 + q^2S_y^2}. \tag{49}$$

In other words, if a computed quantity is equal to the product of measured quantities raised to various powers, then the *fractional* standard deviation of the computed quantity is equal to the square root of the sum of terms consisting of the squares of the *fractional* deviations of the measured quantities multiplied by the squares of their respective powers. Thus it can be seen that particular care should be devoted to measurement of the quantity with the larger exponent.

(3) *Simple product or quotient.* In this case, $V = xy$ or $V = x/y$. Therefore, $m = 1$ and $q = \pm 1$ in treatment (2) above. Then Eq. (49) becomes

$$S_V = \sqrt{S_x^2 + S_y^2}. \tag{50}$$

The fractional standard deviation in a product or quotient is equal to the square root of the sum of the squares of the fractional standard deviations of the measured quantities.

(4) *The logarithm of a quantity.* If $V = B \ln x$, where B is a constant known exactly from theory, $dV/dx = B/x$, and $s_V = (B/x)s_x$. Therefore

$$s_V = BS_V. \tag{51}$$

Hence the absolute standard deviation of the natural logarithm of x (the case $B = 1$) is equal to the fractional standard deviation of x, while the absolute standard deviation of the logarithm to the base 10 (the case $B = 0.434$) is equal to 0.434 times the fractional standard deviation of x.

F. Estimation of systematic errors. As mentioned in Section II, systematic errors may be either determinate or indeterminate. When they are determinate, the original error may be removed by application of a suitable correction. However, this correction generally involves random errors, which must be combined with other random errors by the methods described.

A complete analysis of an experiment requires the listing of all possible sources of indeterminate systematic error. In many cases an observer can make an intelligent guess as to the magnitude of these errors. Since they are usually independent, systematic errors may be combined with random errors by the methods of combining independent errors previously described, that is, by taking the square root of the sum of their squares. For this purpose, many experimenters estimate the *limit of error*, the maximum amount by which the quantity may reasonably be supposed to be in error. Other observers believe such a procedure too conservative, since large errors are relatively improbable compared with small ones. Therefore, instead of using the full estimated value of limit of error, these observers reduce it, perhaps by one-third. Since these are matters of opinion, no firm rules can be given, and each experimenter must use his own judgment.

In some cases where it may be difficult or impossible to estimate the magnitude of a source of indeterminate systematic error, it may be possible to give the sign. For example, suppose an experiment involves trapping in a vessel the liquid emerging at high speed from a pipe, and later weighing the liquid that has been collected. In such a measurement a systematic error might result from loss of liquid by splashing. There is no possibility, however, that *extra* liquid might be collected. In such cases, the discussion of error of the complete experiment should not only list the source of error of the measured quantity, and its sign, but should also consider its influence in regard to sign on the error of any quantity computed from this measured one.

G. Selection of instruments. The preceding paragraphs dealt with the problem of calculating the error in a computed quantity V in terms of the errors in the directly measured quantities. A converse problem also exists: how to select instruments or techniques for making the measurements in such a way that the error in V is smaller than some preassigned value. This question may be anwered by reversing the previous reasoning.

First of all, suppose that V is to be determined from K independently measured quantities $x, y, \ldots$, and each is to make an equal contribution to s_V. The standard deviation s_V is given in terms of the standard deviations of the measured quantities by means of Eq. (37), generalized to include K quantities. Then if the contributions are to be equal, the terms under the

radical must be equal, and the required values of the standard deviations in the measured quantities are

$$s_x = \frac{s_V}{\sqrt{K}\,\dfrac{\partial V}{\partial x}}, \qquad s_y = \frac{s_V}{\sqrt{K}\,\dfrac{\partial V}{\partial y}},$$

and so on.

The assumption of equal contributions is usually unrealistic. In Section V–D it was shown that when the magnitudes of the contributions of two measured quantities are considerably different, the smaller one has very little effect on the result and often can be neglected. Therefore, although K quantities may be involved, only rarely do more than two of them say, x and y, have an appreciable effect. We can select x and y by inspection. If we suppose that the term involving y under the radical of Eq. (37) is W^2 times as large as the term involving x, it can easily be seen that the required values of s_x and s_y are

$$s_x = \frac{s_V}{\sqrt{1 + W^2}\,\dfrac{\partial V}{\partial x}} \quad \text{and} \quad s_y = \frac{W s_V}{\sqrt{1 + W^2}\,\dfrac{\partial V}{\partial y}}.$$

In rare situations where more than two unequal contributions must be considered, the previous reasoning may be extended to cover them.

VI. SPECIAL TOPICS IN THE ADJUSTMENT OF DATA

Adjustment of data concerns the selection of the "most probable value" of a quantity. In the case of a measured quantity this most probable value, as we have indicated in Section IV–A, is the arithmetic average of the measured values if these are assumed to have equal precision. In the case of a quantity whose value may be computed directly from measured quantities by some theoretical formula, we showed in Section V–A that the most probable value is calculated from the average values of the measured quantities, that is, $V(\bar{x}, \bar{y})$. Most situations normally encountered fall into one of these two categories. However, there are a few other common cases which should be considered.

A. Weighted averages. Suppose that we measure x by two methods, obtaining values x_1 and x_2, which are known to have standard deviations s_1 and s_2, respectively. A simple arithmetic average would not give the "best" value because it would make x_1 and x_2 equal in importance, while their errors are assumed to be different. A *weighted average* is then used as a means of determining the best value. The question is how to assign the correct value of the weight factor w_1 to x_1 and of w_2 to x_2 to give the minimum standard deviation in the average value $\bar{x}$. By definition,

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2} = \frac{x_1 + w x_2}{1 + w}, \tag{52}$$

where $w = w_2/w_1$.

We may compute $s_{\bar{x}}$, the standard deviation in $\bar{x}$, by application of Eq. (37), replacing V by $\bar{x}$, x by x_1, and y by x_2. From Eq. (52),

$$\frac{\partial \bar{x}}{\partial x_1} = \frac{1}{1 + w} \quad \text{and} \quad \frac{\partial \bar{x}}{\partial x_2} = \frac{w}{1 + w}.$$

Therefore,

$$s_{\bar{x}} = \sqrt{\frac{s_1^2 + w^2 s_2^2}{(1 + w)^2}}. \tag{53}$$

The optimum value of w is found by calculating $\partial s_{\bar{x}}/\partial w$ from Eq. (53) and making this equal to zero. Then

$$w = \frac{w_2}{w_1} = \left(\frac{s_1}{s_2}\right)^2. \tag{54}$$

Thus the optimum weight factors are inversely proportional to the squares

of the respective standard deviations (or average deviations, or probable errors). The generalization of this treatment to the situation where more than two values are to be averaged is obvious.

One consequence of this rule is that if one of the deviations is appreciably larger than the other, the measured value corresponding to it has little effect on the average value $\bar{x}$. In fact, if the ratio of the standard deviations is more than about 2, the less precise value can be neglected completely and the more precise directly measured value can be considered as the final result, $\bar{x}$.

B. Best fit of a straight line. Frequently, two quantities x and y, both directly measurable, are related by some theoretical formula $y = y(x)$ which involves some unknown constants to be evaluated from the observed data. The most familiar and simple example, of course, is when y is related to x by the equation of a straight line:

$$y = a + bx, \tag{55}$$

where a is the intercept on the y-axis and b is the slope. In a more complicated formula there may be a larger number of such constants. If there are m such constants, we could obtain m pairs of values of x and y, substitute these into the relation $y = y(x)$, obtain m equations relating the constants, and then solve these simultaneously for the constants. However, the precision is limited by this process, since it requires exactly m values, no more and no less. If we have available more than m values, we cannot utilize all of our information in this way to improve the precision in the calculated values of the constants, because they would be algebraically overdetermined. Hence it is necessary to develop an improved method of calculation which utilizes all the data. The general method uses the principles that we have employed many times previously in this work. However, in any one case the particular method depends on the exact nature of the relation $y = y(x)$. The treatment for the linear case, based upon certain simplifying assumptions, is well known and is given below as an example. The methods used in this treatment may be extended to apply to more complicated problems of this type. Examples of these applications may be found in various textbooks.* Recently Hudson** has developed a general method which is particularly useful with transcendental functions.

Now let us consider the formulation of the problem of fitting a straight line to some experimental data. If we were to obtain a pair of "true" measurements of x and y, these would be represented graphically as a point, and the straight line would be expected to pass through it. However, actually both x and y are subject to experimental errors, and thus the position of the point is not exactly determined. Therefore the ideal point might be considered as having expanded to an ellipse, the ratio of whose axes is

*See for example, Worthing and Geffner, *op. cit.*, Chapter XI.

G. E. Hudson, *Am. J. Phys.* **21, 362 (1953).

s_y/s_x. Because of the experimental errors, the centers of such ellipses cannot be expected to lie on the straight line, but they can be expected to be distributed equally on either side of it. (In this application the word "equally" has several possible meanings, three of which may be inferred from the following discussion.) In general, the sizes of these hypothetical ellipses are different because the errors are functions of the variables, and therefore the centers lie in a band of varying width. Hence the importance of each pair of measurements in determining the parameters of the line differs; this effect can be included in the calculations by the introduction of appropriate weight factors.

This discussion suggests that the best straight line is the one chosen in such a way that the sum of the squares of the distances from the centers of the ellipses to the line, measured along some appropriate direction, is a minimum. The appropriate direction depends upon the relative errors of x and y. In the case where x and y have the same physical dimensions and are plotted to the same scale, and $s_x = s_y$, the appropriate direction is the perpendicular. In most other cases, the appropriate direction is oblique. Often x and y have different dimensions and their graphical scales bear an arbitrary relationship; in this case there is usually no unique way of selecting the appropriate direction, but in a few instances the original variables can be transformed to others which do have the same dimensions. In any event, such a process is very complicated algebraically, and the labor required is rarely justified even in the situations where, in principle, the transformation can be carried out uniquely. Those readers who are interested in the details of the method may find them in a paper by Bacon.*

A much simpler procedure results if we assume that one of the quantities, let us say x, is measured exactly, while all of the errors are concentrated in the other variable y. Graphically this situation is represented by vertical lines centered upon the experimental points (x_n, y_n). The actual errors of x may be taken into account indirectly by increasing the lengths of the lines an appropriate amount beyond the lengths required to represent the actual errors of y. The deviation of y_n is defined by the relation

$$\delta y_n = y_n - (a + bx_n). \tag{56}$$

Graphically, δy_n is the vertical distance from the point (y_n, x_n) to the desired straight line of best fit. Then we choose the constants a and b in the equation of this line in such a way that the weighted sum of the squares of these deviations is a minimum. This choice is independent of the relative dimensions of x and y.

Alternatively, we could determine a and b by supposing that all of the errors have been concentrated in x, and then following a similar procedure. These values of a and b usually would not coincide with the first set, but the differences generally would not be of great practical importance. The values

*R. H. Bacon, *Am. J. Phys.* **21**, 428 (1953).

which would be obtained by having the error divided between x and y may be supposed to be intermediate.

We shall proceed with the assumption that all of the error is concentrated in y. For simplicity* we shall further assume that the absolute error of y is independent of x. Thus the plotted experimental values (x_n, y_n) surround the line of best fit in a band of constant width.

From Eq. (56),

$$(\delta y_n)^2 = y_n^2 + a^2 + b^2x_n^2 - 2ay_n + 2abx_n - 2bx_ny_n. \tag{57}$$

If k is the total number of pairs of values, the sum of the squares of the deviations of y is

$$\sum_{n=1}^{k} (\delta y_n)^2 = \sum_{n=1}^{k} y_n^2 + ka^2 + b^2 \sum_{n=1}^{k} x_n^2 - 2a \sum_{n=1}^{k} (y_n - bx_n) - 2b \sum_{n=1}^{k} (x_ny_n). \tag{58}$$

We select a and b in such a way as to make this sum a minimum. To do this, we differentiate partially with respect to a and b and place the two partial derivatives each equal to zero. (For simplicity we shall omit the indices on the summation signs.)

$$\frac{\partial[\sum(\delta y_n)^2]}{\partial a} = 2ka - 2\sum y_n + 2b\sum x_n = 0, \tag{59}$$

and

$$\frac{\partial[\sum(\delta y_n)^2]}{\partial b} = 2b\sum x_n^2 + 2a\sum x_n - 2\sum(x_ny_n) = 0. \tag{60}$$

The desired values of a and b are obtained by solving these equations simultaneously:

$$a = \frac{\sum x_n^2 \sum y_n - \sum x_n \sum (x_ny_n)}{k\sum x_n^2 - (\sum x_n)^2}, \tag{61}$$

*Actually the calculations are not greatly complicated by omitting this assumption, but the situations which are of most practical interest are those for which this assumption is valid. As an exercise the student might carry through the theory with weight factors included, or, alternatively, reference can be made to R. H. Bacon, *Am. J. Phys.* **21**, 428 (1953).

and

$$b = \frac{k\sum(x_n y_n) - \sum x_n \sum y_n}{k\sum x_n^2 - (\sum x_n)^2}. \tag{62}$$

It should be pointed out that Σx_n^2 and $(\Sigma x_n)^2$ are not the same: Σx_n^2 implies that each value of x is squared and then a sum is made of these squares, while $(\Sigma x_n)^2$ implies that all of the values of x are added together and then this sum is squared. In other words, the order of performing the operations of raising to the second power and of addition is interchanged. There is an analogous distinction between $\Sigma(x_n y_n)$ and $\Sigma x_n \Sigma y_n$.

The evaluation of the indicated sums is generally tedious and is not justified unless the calculations retain all of the significant figures yielded by the data. It is best done with the aid of a calculating machine. Note that the denominators in Eqs. (61) and (62) are identical, and therefore this quantity need be calculated only once. Once these results are obtained, they are useful for a number of reasons. For instance, the slope or intercept may be of direct physical interest. Also, these values when substituted into Eq. (55) are useful for interpolating or extrapolating (that is, for calculating a value y_0 corresponding to some arbitrarily chosen x_0).

Next we consider how to compute the standard deviations of the quantities a, b, and y_0. Our data consists of k pairs of measurements of x and y, which are samples of the universe of all such measurements. The standard deviations s_a, s_b, and s_0 are then the rms deviations about a_u, b_u, and y_{0u}, which are the quantities that would be computed from the entire universe were it possible to make all of the measurements contained in it. These standard deviations, as we shall show, may be expressed in terms of s_y, the standard deviation of y about the line defined by a_u and b_u. In analogy to Eq. (5), it is possible to estimate s_y as follows:

$$s_y = \sqrt{\frac{\sum(\delta y_n)^2}{k - 2}}. \tag{63}$$

We shall not attempt here to rigorously justify the factor $k - 2$ which appears in the denominator on the right in place of the factor $k - 1$ appearing in Eq. (5). Such a justification may be found in a paper by Bacon.* However, with reference to the discussion which followed the introduction of Eq. (5), it should be pointed out that $k - 2$ is the number of independent determinations of δy_n. While there are k different δy_n's which may be computed from the data, there are two relations between them. These are given by Eqs. (59) and (60). The number of constraints, two, is also equal to the number of the constants to be determined, namely a and b. (When we consider a directly measured quantity, as in Section IV–B, there is one relation between the deviations and one constant to be determined, namely, $\bar{x}$.)

*R. H. Bacon, *Am. J. Phys.* **21**, 428 (1953).

The quantity $\Sigma(\delta y_n)^2$ in Eq. (63) may be computed by first calculating all of the deviations directly, substituting into Eq. (56), in turn, the experimentally determined pairs of values of x and y. However, such a direct calculation is cumbersome and generally not as convenient as an indirect one. Indirectly, $\Sigma(\delta y_n)^2$ may be computed from Eq. (58) after expressions for a and b have been substituted into it from Eqs. (61) and (62). The result is

$$\sum(\delta y_n)^2 = \sum y_n^2 - \frac{(\sum y_n)^2 \sum x_n^2 - 2\sum(x_n y_n)\sum x_n \sum y_n + k[\sum(x_n y_n)]^2}{k\sum x_n^2 - (\sum x_n)^2}. \tag{63A}$$

While Eq. (63A) appears formidable, it should be noticed that except for Σy_n^2, all of the quantities represent numbers, or the squares of numbers, which have previously been obtained in the calculation of a and b.

The standard deviation s_a may be found in the usual way, that is, by application of Eq. (37). There are k contributions appearing under the radical, resulting from the errors due to the k separate y's. The contribution due to one of these, let us say y_j, may be found by differentiating Eq. (61) with respect to y_j and multiplying by the standard deviation of y_j, which is s_y. The variable y_j appears in Eq. (61) in one of the terms of the sum Σy_n and also one of the terms of $\Sigma x_n y_n$. Its contribution to the error is thus

$$s_y \frac{\partial a}{\partial y_j} = s_y \frac{\sum x_n^2 - (\sum x_n) x_j}{k\sum x_n^2 - (\sum x_n)^2}. \tag{64}$$

The standard deviation s_a is found by squaring the terms of the type given by Eq. (64), adding them (that is, summing the index j from 1 to k), and taking the square root of this sum. In this process it should be remembered that $\Sigma x_j = \Sigma x_n$ and $\Sigma x_j^2 = \Sigma x_n^2$. The result is

$$s_a = s_y \sqrt{\frac{\sum x_n^2}{k\sum x_n^2 - (\sum x_n)^2}}. \tag{65}$$

In a similar manner, the standard deviation of the slope s_b may be found:

$$s_b = s_y \sqrt{\frac{k}{k\sum x_n^2 - (\sum x_n)^2}}. \tag{66}$$

A computed value y_0 is found by substitution of the desired value of the independent variable x_0 into Eq. (55):

$$y_0 = a + bx_0. \tag{67}$$

The contribution to the standard deviation in y_0 as a result of the deviation of one of the measured values y_j is

$$\frac{\partial y_0}{\partial y_j}\,\delta y_j = \left[\frac{\partial a}{\partial y_j} + x_0\,\frac{\partial b}{\partial y_j}\right]\delta y_j. \tag{68}$$

The partial derivatives on the right side of Eq. (68) are evaluated and then the standard deviation s_0 of y_0 may also be computed in the manner outlined above. The result is

$$s_0 = s_y\sqrt{\frac{\sum x_n^2 - 2x_0\sum x_n + kx_0^2}{k\sum x_n^2 - (\sum x_n)^2}}. \tag{69}$$

By placing the derivative of s_0 with respect to x_0 equal to zero, it may be shown that s_0 has a minimum value

$$s_{0\ \min} = \frac{s_y}{\sqrt{k}} \tag{69A}$$

at $x_0 = \Sigma x_n/k$, that is, when x_0 is equal to the average of the measured values of x.

On the other hand, if the magnitude of x_0 becomes very large, the first two terms inside the parentheses of Eq. (69) may be neglected in comparison with the third term. By reference to Eq. (66), it may be seen that in this case Eq. (69) approaches the following:

$$s_0 = |x_0|s_b. \tag{69B}$$

Thus, for extreme extrapolations, the standard deviation of the computed value is proportional to the value of the independent variable. Also, when the magnitude of x_0 is large, a may be neglected in comparison with bx_0 in Eq. (67). Then it may be seen that if Eq. (69B) is divided by this approximate form of Eq. (67).

$$\frac{s_0}{y_0} = \frac{s_b}{b} \quad \text{or} \quad S_0 = S_b.$$

Under these conditions, the fractional deviation in y_0 approaches a constant value, the fractional standard deviation of the slope b.

C. Miscellaneous applications of the least squares fit of a straight line. A number of equations not originally in the linear form may be transformed into the form of Eq. (55), and the foregoing treatment may be applied. For example, consider the following relation, which is the mathe-

matical form of both Newton's law of cooling and the law of radioactive decay:

$$z = Ae^{-Bt}, \tag{70}$$

where A and B are constants to be determined, and z and t are measured quantities. Taking logarithms of both sides of Eq. (70),

$$\ln z = \ln A - Bt.$$

This may be put into the desired form by placing $\ln z = y$, $\ln A = a$, $B = -b$, and $t = x$.

In the application of Eq. (70) to radioactive decay, the errors in the time $t = x$ are usually of negligible effect in comparison with the statistical errors of the counting rate z, and the background counting rate is usually small compared with the counting rate obtained with the source. Under these conditions, the validity of the hypothesis that the standard deviation of y is independent of x can be ensured by the simple procedure of employing the same number of counts in each measurement, as may be seen by reference to Eq. (51) and to Eq. (85) below. Also, in this case, s_y may be computed more easily from Eq. (85) than from Eqs. (63) and (63A). However, if some instrumental defect or the presence of disturbances is suspected, it may be desirable to compute s_y by both methods as a check.

Another type of equation which may be transformed into linear form is

$$u = Cv^n + D. \tag{71}$$

If n is known exactly from theory C and D are to be determined, appropriate substitutions are (1) $x = v^n$, $y = u$, $a = D$, and $b = C$. If $D = 0$ by theory but if C and n are to be determined, possible substitutions are (2) $x = \ln v$, $y = \ln u$, $a = \ln C$, and $b = n$; or (3) $x = \ln u$, $y = \ln v$, $a = -(1/n)\ln C$, and $b = 1/n$.

If in principle $D = 0$ but if the actual location of the origin ($u = 0$, $v = 0$) is subject to error, and if n is known from theory, as in the example of the calibration of a Jolly balance illustrated in Section VIII–B, appropriate substitutions are (1) above or (4) $x = u^{1/n}$, $y = v$, and $b = C^{-1/n}$. In such a case it is appropriate to use [$x = 0$, $y = y(0)$] as one of the pairs of measurements. Then, as the line of best fit cannot be expected to go through the origin exactly, the computed value of a from Eq. (61) cannot be expected to be zero, although it should not be large compared with s_a from Eq. (65).

On the other hand, if the location of the origin is known exactly ($D = 0$), it is necessary to rework the basic theory with $a = 0$ in Eq. (56). Now Eq. (59) does not exist, and we obtain the desired value of b simply by setting $a = 0$ in Eq. (60):

$$b = \frac{\sum(x_n y_n)}{\sum x_n^2}. \tag{71A}$$

It is to be noted that the largest measured values of x and y have the greatest importance in determining b in Eq. (71A). Possible substitutions are (1) with $D = a = 0$, and (4) above.

This discussion may also apply in some situations in which C and D are not true constants. They may be functions of other measured quantities; if these are known with a relatively high precision, the effects of their errors may be neglected to a reasonable approximation. They may be slowly varying functions of u and v themselves; in this case we may employ the method of successive approximations, in which we use approximate values of u and v to correct C or D for this slow variation. In particular, if D is a true constant plus such a slowly varying function, it would be appropriate to employ substitutions of the type (1) above, with a equal to the truly constant portion of D and y equal to u minus the variable portion of D. If $D = 0$ but if C is a true constant plus a slowly varying function, we may utilize an analogous procedure with substitutions (2) or (3).

Where we have a choice of possible substitutions, according to our hypothesis we should choose for y whichever quantity has the larger errors. Also we should keep in mind the hypothesis that the absolute deviation of y is independent of x. Should there be a slight dependence it may be neglected, of course, as a reasonable approximation, but if the dependence is large, the theory must be modified to include appropriate weight factors.

VII. THE STATISTICAL ERRORS OF NUCLEAR PHYSICS

Particles are emitted by radioactive decay or by artificial disintegration at completely random intervals of time. Therefore, the time t necessary for the observation of any finite number of counts N is subject to statistical fluctuations giving rise to an error in the observed counting rate $n = N/t$. This is fundamental to the phenomenon of radiation and not a property of the instruments and is an error of definition. We shall prove that the standard deviation is approximately $\sqrt{N}$ by first showing that the number of counts observed in a given time obeys what is called a Poisson distribution, and then showing that when the average number $\overline{N}$ becomes very large it approaches a Gaussian distribution. This result was first obtained by Bateman.*

Let the probability that N particles are observed in time t be P_N. Suppose that the time t is divided into b equal intervals so small that the probability of the emission of two particles within an interval is negligible. The probability of the emission of one particle in a given interval is then $\overline{N}/b$. The probability of emission of N particles in the first N intervals and none in the remaining $b - N$ is (see Section IV–F):

$$(\overline{N}/b)^N(1 - \overline{N}/b)^{b-N}. \tag{72}$$

However, this is only one possible way of obtaining N particles in the total time t. The first particle could have been in any one of the b intervals, the second in any one of the remaining $b - 1$, the third in any of the remaining $b - 2$, and, finally, the Nth in any of the remaining $b + 1 - N$. The Nth particle has the choice between the interval it occupies and all the $b - N$ unoccupied intervals. Thus the number of ways of distributing the N particles in the b intervals is

$$b(b - 1)(b - 2) \cdots (b - N + 1).$$

However, not all of these ways are essentially different, since the particles may be interchanged without influencing the result. The number of essentially different ways is found by dividing the above by the number of ways of interchanging the particles. Any one of the N particles can be chosen as the first, any one of the remaining $N - 1$ as the second, any one

*H. Bateman, *Phil. Mag.* **20**, 704 (1910).

of the remaining $N - 2$ as the third, and so on. Thus the number of ways of interchanging the particles is $N!$. Hence the probability of obtaining N counts is

$$P_N = \frac{b(b-1)\cdots(b-N+1)}{N!}\left(\frac{\overline{N}}{b}\right)^N\left(1-\frac{\overline{N}}{b}\right)^{b-N}. \tag{73}$$

This is called the binomial distribution law. If b approaches infinity, the first factor on the right approaches $b^N/N!$, while the last factor approaches e^{-N}, yielding the formula for a Poisson distribution:

$$P_N = \frac{\overline{N}^N e^{-\overline{N}}}{N!}, \tag{74}$$

the factor b^N having canceled out.

Strictly speaking, Eq. (74) is valid only for integral values of N. However, a continuous curve of P_N vs. N plotted through the points specified by this formula yields a curve with an unsymmetrical maximum for small values of $\overline{N}$. As $\overline{N}$ becomes larger, the curve becomes more symmetrical about the maximum, while the relative height-to-width ratio of the maximum increases. Analytically, Eq. (74) can be shown to approach a Gauss error curve, as follows.

According to Stirling's approximation for the factorial of a large number, for which the error is less than 1 percent when the number is greater than 10,

$$N! = \sqrt{2\pi}\, N^{N+\frac{1}{2}} e^{-N}. \tag{75}$$

Therefore,

$$\overline{N}! = \sqrt{2\pi\overline{N}}\, \overline{N}^{\overline{N}} e^{-\overline{N}},$$

and

$$P_{\overline{N}} = \frac{1}{\sqrt{2\pi\overline{N}}}. \tag{76}$$

If Y represents the natural logarithm of P_N,

$$Y = -\overline{N} + N \ln \overline{N} - \ln\sqrt{2\pi} - (N + \tfrac{1}{2}) \ln N + N, \tag{77}$$

and

$$\frac{dY}{dN} = \ln \overline{N} - (N + \tfrac{1}{2})\left(\frac{1}{N}\right) - \ln N + 1. \tag{78}$$

This latter expression must be zero when P_N is a maximum, yielding

$$\overline{N} - Ne^{1/2N} = 0. \tag{79}$$

Since N is postulated as being a large number, $e^{1/2N}$ is essentially unity and therefore the maximum occurs very nearly at $N = \overline{N}$.

We now expand Y about $\overline{N}$ in a Taylor series, and retain only the first two nonvanishing terms. As the maximum occurs very close to $\overline{N}$, the first derivative of Y with respect to N is zero, and this expansion becomes

$$Y(N) = Y(\overline{N}) + \frac{1}{2}\frac{d^2Y}{dN^2}x^2$$

$$= -\ln\sqrt{2\pi\overline{N}} - \frac{1}{2}\left(\frac{1}{\overline{N}} - \frac{1}{2\overline{N}^2}\right)x^2, \tag{80}$$

where $x = N - \overline{N}$. Since $\overline{N}$ is a large number, the second term in the parentheses of Eq. (80) may be neglected in comparison with the first. Then if the antilogarithm is taken:

$$P_N = \frac{1}{\sqrt{2\pi\overline{N}}}\, e^{-(x^2/2\overline{N})}. \tag{81}$$

By reference to Eq. (24), we see that Eq. (81) is in the form of the Gauss error law with a precision index $h = 1/\sqrt{2\overline{N}}$. According to Eq. (29), the standard deviation $s = \sqrt{2}/h$. Thus, if on the average $\overline{N}$ counts are observed in a given time t, the standard deviation in the number of counts is

$$s = \sqrt{\overline{N}}. \tag{82}$$

From this point on we shall drop the bar above the N. We usually make only a single reading, which is assumed to be near enough to $\overline{N}$ to be adequate for the evaluation of s.

Generally, we are not interested in s, the standard deviation in the number of counts, but rather in the standard deviation of the counting rate. The counting rate is given by

$$n = \frac{N}{t}, \tag{83}$$

where t is the time of observation, which is assumed to be measured with such high precision that its error may be neglected.

By the use of Eqs. (82) and (83), we see that the standard deviation in the counting rate is

$$s_n = \frac{\sqrt{N}}{t} = \sqrt{\frac{n}{t}}, \tag{84}$$

and the fractional standard deviation is

$$S_n = \frac{1}{\sqrt{N}} \tag{85}$$

Thus the fractional standard deviation decreases in proportion to the square root of the number of counts, being 10% for $N = 100$, 1% for $N = 10{,}000$, and 0.1% for $N = 1{,}000{,}000$. High-precision measurements require observation of a large number of counts, which involves long observation times if the counting rates are low.

In practice, all counting instruments have a background counting rate β when no known source is present. When the source is present, the counting rate increases to n_2. The counting rate due to the source then is

$$n_1 = n_2 - \beta. \tag{86}$$

By reference to Eqs. (45) and (84), we find the standard deviation in n_1 to be

$$s_1 = \sqrt{\frac{n_2}{t_2} + \frac{\beta}{t_\beta}}, \tag{87}$$

where t_2 is the time of observation of source plus background and t_β is the time of observation of the background.

If the source is a weak one, the second term in Eq. (87) is important and considerable time must be spent measuring the background. We might inquire into the most efficient division of time between measurement of background counting rate and of counting rate due to the source. It can be shown that if the total time $t_2 + t_\beta$ is a constant, s_1 is a minimum when the times of observation are made proportional to the square roots of the counting rates. That is,

$$\frac{t_2}{t_\beta} = \sqrt{\frac{n_2}{n_\beta}}. \tag{88}$$

The reasoning leading up to Eq. (88) may be applied to develop the most effective procedure in experiments involving more than one source* and a variable background.**

The previous discussion has dealt with the purely statistical errors of counting instruments. These instruments are also subject to other errors and corrections which are, however, beyond the scope of this work. It should be mentioned that statistical errors are sometimes important in so-called integrating instruments, which do not detect individual particles but the average effects of many particles.†

---

*Y. Beers, *Rev. Sci. Inst.* **13**, 72 (1942).

R. L. Loevinger and M. Berman, *Nucleonics* **9, No. 1, 26 (1951).

†Y. Beers, *Phys. Rev.* **63**, 77 (1943). Also, John Strong *et al.*, *Procedures in Experimental Physics*, Prentice-Hall, Inc., New York (1938), Chapter VI.

VIII. EXAMPLES

Analyses of data from two standard experiments from undergraduate laboratory courses are presented in the following pages. The data were taken under conditions similar to those normally encountered in such courses rather than under research conditions where the observer might have had access to standards for calibration or to more than one instrument of a given type for comparison. Yet if these facilities had been available, the same principles would have been applied, only more extensively.

The role of human opinion was discussed in the Introduction, and in accord with those remarks, the analyses presented here involve in great measure selection of data and estimation of sources of error in circumstances not amenable to theory. It is not the purpose of this writer to attempt to make the reader agree with him quantitatively, nor does he claim authoritative judgment. His purpose is to point out the things which an experimenter must consider and which he must evaluate in a manner that seems reasonable.

The theory which has been developed is applied wherever possible. However, criticism might justly be made of the present use of this theory, which is based upon the assumption of many more measurements than are contained in these examples. For complete rigor, an elaborate theory based upon small samples should be used. However, it is to be remembered that errors evaluated by the theory are to be added to others obtained by estimate or, more candidly, by pure guesswork. When the procedure is viewed from the result, a rigorous theory is hardly in order. The number of measurements in these examples is typical not only of undergraduate experiments but also, unfortunately, of many research experiments. Yet, in spite of its lack of rigor, the theory serves a useful purpose in that it shows what the important errors are and what must be done to perform a better experiment.

A. Torsion pendulum experiment.

Description. A body of axial symmetry is fastened on its axis to the lower end of a vertical wire whose upper end is held by a rigid, fixed clamp. By twisting the body and releasing it, the system is set into oscillation with a period

$$T = 2\pi \sqrt{\frac{I}{k}}, \tag{89}$$

where I is the moment of inertia of the body and k is the torque constant of the wire, i.e., the torque per unit angle of twist.

In spite of its symmetry, the shape of the body does not allow its moment of inertia to be calculated easily. However, provision is made for slipping onto it a thin circular ring of mass m, inner diameter D_1, and outer diameter D_2. With its axis coinciding with the wire, the moment of inertia of the ring, I', may be calculated from the expression

$$I' = \frac{m}{8}(D_1^2 + D_2^2). \tag{90}$$

With this ring in place, a period T' is measured. From these data it is possible to compute both k and I by means of the following formulas:

$$k = \frac{4\pi^2 I'}{T'^2 - T^2}, \tag{91}$$

and

$$I = I' \frac{T^2}{T'^2 - T^2}. \tag{92}$$

The torsion modulus of the material of which the wire is composed is given by

$$M = \frac{32Lk}{\pi d^4}, \tag{93}$$

where L is the length of the wire and d its diameter.

The derivation of these formulas follows from methods to be found in standard texts.* It is to be noted, however, that here the formulas are expressed in terms of diameters rather than, as is conventional, in terms of radii, because the diameters are the directly measured quantities.

Data and calculations. For evaluation of error, the quantities to be determined must be expressed in terms of directly measured quantities or in terms of computed quantities that are statistically independent. The moment of inertia I' is, of course, a computed quantity whose value is involved in obtaining the desired end results, k, I, and M. It is statistically independent, however, because D_1 and D_2 are measured with a vernier caliper, while the other lengths L and d are measured with a meter stick and a micrometer, respectively, and because the measurement of mass is not otherwise involved. Therefore we may proceed step by step, first calculating I' and evaluating its error, and then performing the calculations indicated by Eqs. (91) and (92). Since k is then statistically independent of L and d, we take the computed value of k and calculate M from Eq. (93). If I' were not statistically independent, we would have to eliminate it between these equations and obtain k, I, and M in terms of measured quantities.

---

*See, for example, F. W. Sears, *Mechanics, Heat, and Sound*, Addison-Wesley Publishing Co., Inc., Reading, Mass. (2nd ed., 1952).

In the present example, we choose to compute average deviations from the data by use of Eq. (6), since this procedure involves less arithmetic than the calculation of standard deviations. However, an estimate of the standard deviations may be obtained from these values of the average deviations by use of Eq. (31). Since these computations have already been illustrated in Section IV–E, not all of the details will be given here.

Evaluation of I' and its error. The measurements of D_1 and D_2 were taken at different positions on the ring, and therefore the errors are partly errors of definition, since the surfaces cannot be expected to be perfectly circular in cross section. The data are given in Table IV.

TABLE IV

Data on D_1 and D_2.

D_1		D_2	
Value	*Deviation*	*Value*	*Deviation*
7.82 cm	+0.02 cm	8.79 cm	−0.04 cm
7.82	+0.02	8.82	−0.01
7.83	+0.03	8.85	+0.02
7.74	−0.06	8.87	+0.04
7.78	−0.02	8.84	+0.01
7.79	−0.01	8.82	−0.01
7.79_7 cm	0.02_7 cm	8.83_3 cm	0.01_5 cm
Average	Average deviation	Average	Average deviation

The average deviations in Table IV represent estimates of the errors of single measurements. The average deviations of the means are found by dividing by the square root of the number of observations (that is, 6), giving 0.01_1 cm for D_1 and 0.00_6 cm for D_2. However, these are estimates of only the random contributions to the errors of D_1 and D_2. There are also non-independent systematic errors due to the calibration of the caliper, which are indeterminate with the facilities available. It is reasonable to assume the presence of such errors amounting to one scale division, or 0.01 cm, which is comparable with the random errors. Therefore, both experimental and systematic errors must be included in the analysis. If one were very much smaller than the other, it could have been neglected, in accordance with the argument in Section V-C.

The largest observed deviations contained in Table IV are 0.06 cm for D_1 and 0.04 cm for D_2. These are, respectively, 2.2 and 2.7 times the average deviations, or 1.8 and 2.1 times the estimated standard deviations.

According to Table II the probabilities of obtaining deviations of these magnitudes or greater are roughly 0.11 and 0.04. While these probabilities are slightly smaller than $1/k = 0.17$, they do not seem to this author sufficiently small to consider rejection of the corresponding data.

The mass of the ring was measured with a platform balance. The ring was placed on the left-hand pan, and balance was obtained by using three different combinations of the standard masses. The procedure was repeated with the ring on the right-hand pan. The results are shown in Table V. The standard masses were tested for internal consistency by balancing various combinations against each other and against the slider. The reading of the slider was added to the mass in the right-hand pan. (See Table VI.)

Clearly, these data do not warrant application of the theory of error, and the error can be evaluated only by estimate. The author estimates that the error in the mass due to all causes may reasonably be taken as an average deviation of 0.2 gm. (Since both the ring and the standard masses were made of brass, the correction for the buoyant force of air is identically zero.)

According to Eq. (90),

$$I' = \frac{380.0}{8}\left[(7.79_7)^2 + (8.83_3)^2\right] = 6.59_3 \times 10^3 \text{ gm-cm}^2.$$

TABLE V

Mass of Ring

	Ring on left	*Ring on right*
	379.0 gm	380.3 gm
	379.8	380.2
	380.0	380.3
Average	379.6 gm	380.3 gm

Average of both sets = 380.0 gm

TABLE VI

Consistency of Masses

Mass on left	*Mass on right*
100 + 100 gm	200.1 gm
2	1.9
4	4.0
100 + 1	101.0

TABLE VIII

Data on Periods

T'				T				
Number of oscillations	*Time* (sec)	*Period* (sec)	*Deviation* (sec)	*Number of oscillations*	*Time* (sec)	*Period* (sec)	*Deviation* (sec) *from average of* 4.45_4 sec	*Deviation* (sec) *from average of* 4.46_7
20	140.2	7.01	−0.01	20	89.5	4.48	+0.03	+0.01
20	140.4	7.02	0.00	20	89.5	4.48	+0.03	+0.01
20	140.6	7.03	+0.01	21	93.0	4.43	−0.02	−0.03
20	140.1	7.01	−0.01	20	89.2	4.46	+0.01	−0.01
20	140.5	7.03	+0.01	20	89.4	4.47	+0.02	0.00
20	140.4	7.02	0.00	20	87.6	4.38	−0.07	——
				20	89.5	4.48	+0.03	+0.01
Mean 7.02_0 sec Average deviation: 0.00_7 Average deviation of mean: 0.00_3 sec Fractional average deviation of mean: 4×10^{-4}				Mean 4.45_4 sec Average deviation: 0.03_4 sec If the measurement $T = 4.38$ sec is rejected, the mean becomes 4.46_7 sec with an average deviation of 0.01_0 sec or an average deviation of 0.00_4 sec of the mean.				

values of the period, but the analysis of the deviations would have to take into account their weight factors in a manner which is beyond the scope of this book.

Upon completing this calculation we note that the average deviation of one measurement of T is five times that of one measurement of T'. Since these observations were made in the same way with the same equipment, the average deviations of their universes should be equal. A ratio as large as 5 seems unreasonable even for small samples of six or seven measurements. (According to advanced considerations, the probability of obtaining such a large ratio is less than 0.003). Therefore, we inquire whether one of the measurements of T should be rejected. The one of 4.38 sec is to be suspected because of its large deviation, 2.1 times the average. According to Table II, the probability of obtaining a deviation this large or greater is roughly 0.11. Since this is not very small in comparison to $1/k = 0.14$, this consideration does not seem to justify the rejection of this measurement. However, the estimate of the errors should be based upon all the data available, and all that data should be consistent. Because the measurements of T and T' were made in the same way, it is logical, although not

TABLE IX

Total Error of k by the General Method

Source	*Average deviation*	*Partial derivative*	*Contribution*	*Square of contribution*
I'	19 gm–cm^2	$\frac{k}{I'} = 1.32$	$19 \times 1.32 = 25$	625
T'	3×10^{-3} sec	$\frac{-2kT'}{T'^2 - T^2} = -4.27 \times 10^3$	$-3 \times 4.27 = -12.8$	164
T	4×10^{-3} sec	$\frac{2kT}{T'^2 - T^2} = 2.74 \times 10^3$	$4 \times 2.74 = 11$	121

Sum of squares of contributions 910

Absolute average deviation $= \sqrt{910} = 30$ dyne–cm per radian.

Fractional average deviation $= \frac{30}{8.87_3} \times 10^{-3} = 0.3_4\%$.

rigorously correct, for us to consider the six measurements of T and the seven of T' as a single sample of thirteen measurements. The average of all thirteen deviations is 0.01_9 sec, and according to Table II, the probability of obtaining a deviation of 0.07 sec with such an average deviation is about 0.003. Furthermore, if we reject the value 4.38 sec, we obtain an average of 4.46_7 sec and an average deviation of 0.01_0 sec for one measurement of T, and this is roughly equal to 0.00_7 sec for T'. Therefore, the author has concluded that some blunder has been made in the 4.38 sec measurement and has rejected it. In examining the surviving sample of six measurements of T', we note that the one of 4.43 sec has a probability of occurrence of 0.05, which is somewhat smaller than $1/k = 0.17$; however, particularly with such a small sample, this consideration does not seem to justify the rejection of this measurement, especially as the validity of repetitive application of a rejection procedure is highly dubious.

If we should inquire into a possible cause of a blunder in the rejected 4.38 sec measurement, we note that if $19\frac{1}{2}$ oscillations had taken place within the observed time of 87.6 sec, the period would have been 4.49 sec, and this is in reasonable agreement with the other measurements. Therefore we may suppose that $19\frac{1}{2}$ oscillations were miscounted as 20.

According to Eq. (91),

$$k = \frac{4\pi^2 \times 6.59_3 \times 10^3}{(7.02_0)^2 - (4.46_7)^2} = 8.87_3 \times 10^3 \text{ dyne-cm per radian.}$$

TABLE IX

Total Error of k by the General Method

Source	*Average deviation*	*Partial derivative*	*Contribution*	*Square of contribution*
I'	19 gm–cm²	$\frac{k}{I'} = 1.32$	$19 \times 1.32 = 25$	625
T'	3×10^{-3} sec	$\frac{-2kT'}{T'^2 - T^2} = -4.27 \times 10^3$	$-3 \times 4.27 = -12.8$	164
T	4×10^{-3} sec	$\frac{2kT}{T'^2 - T^2} = 2.74 \times 10^3$	$4 \times 2.74 = 11$	121

Sum of squares of contributions 910

Absolute average deviation $= \sqrt{910} = 30$ dyne-cm per radian.

Fractional average deviation $= \frac{30}{8.87_3} \times 10^{-3} = 0.3_4\%$.

rigorously correct, for us to consider the six measurements of T and the seven of T' as a single sample of thirteen measurements. The average of all thirteen deviations is 0.01_9 sec, and according to Table II, the probability of obtaining a deviation of 0.07 sec with such an average deviation is about 0.003. Furthermore, if we reject the value 4.38 sec, we obtain an average of 4.46_7 sec and an average deviation of 0.01_0 sec for one measurement of T, and this is roughly equal to 0.00_7 sec for T'. Therefore, the author has concluded that some blunder has been made in the 4.38 sec measurement and has rejected it. In examining the surviving sample of six measurements of T', we note that the one of 4.43 sec has a probability of occurrence of 0.05, which is somewhat smaller than $1/k = 0.17$; however, particularly with such a small sample, this consideration does not seem to justify the rejection of this measurement, especially as the validity of repetitive application of a rejection procedure is highly dubious.

If we should inquire into a possible cause of a blunder in the rejected 4.38 sec measurement, we note that if $19\frac{1}{2}$ oscillations had taken place within the observed time of 87.6 sec, the period would have been 4.49 sec, and this is in reasonable agreement with the other measurements. Therefore we may suppose that $19\frac{1}{2}$ oscillations were miscounted as 20.

According to Eq. (91),

$$k = \frac{4\pi^2 \times 6.59_3 \times 10^3}{(7.02_0)^2 - (4.46_7)^2} = 8.87_3 \times 10^3 \text{ dyne-cm per radian.}$$

TABLE X

Calculation of Error of M

Source	*Average fractional deviation*	*Power*	*Magnitude of contribution*	*Square of contribution*
L	0.1%	1	$0.1 \times 1 = 0.1\%$	0.01
k	0.34%	1	$0.34 \times 1 = 0.34\%$	0.12
d	0.13%	−4	$0.13 \times 4 = 0.52\%$	0.27
			Sum of squares of contributions	0.40
			Fractional average deviation of M is $\sqrt{0.40} = 0.6\%$	

Eq. (49) as generalized for three variables. The general method is considerably more tedious, and will not be used. Computation by the special method is given in Table X.

Conclusions. The wire used in this experiment was made of steel whose exact constitution was unknown. *The Handbook of Chemistry and Physics* (18th ed.) lists values ranging from 7.79 to 8.11 $\times$ 10^{11} dynes per cm^2 for the torsion modulus (or modulus of rigidity) of various types of steel. The discrepancy between the value obtained in the present experiment and the listed values is 4 percent or greater, which is more than 5 times the assigned average deviation. However, moduli for other common metals listed range from 2.37 $\times$ 10^{11} dynes per cm^2 for aluminum to 14.81 $\times$ 10^{11} dynes per cm^2 for tungsten. The relatively large discrepancy between the experimental and the listed values for steel does not necessarily indicate any large unknown error in this experiment, for two reasons. First, the exact constitution of the wire was unknown, and second, this particular sample had been in use in the laboratory for some time and undoubtedly had been subjected to some abuse.

In our analysis, we recall that the principal error of I' was due to the systematic error of calibration in the venier caliper used for measuring the diameters of the ring. This could have been eliminated by correction if calibration facilities had been available. Since the principal error of k as given in Table IX was due to I', the principal error of k may be traced indirectly to this same source. Thus further measurements on the periods T and T' would not have materially improved the value of k. The largest contribution to the error of M, as shown in Table X, was due to the measurement of the wire diameter d. Since d itself was measured with relatively high precision, its contribution predominated as a result of the fact that d appeared to the fourth power in Eq. (93), the theoretical formula for M. Thus any improvement in the value of M would have to be obtained mainly by improving the measurement of d. The contribution of k, which

was traced to the vernier caliper, is the second source of error and is not entirely negligible.

B. Calibration of a Jolly balance. A Jolly balance employs the extension of a vertical helical spring to measure an applied force. Since the spring is assumed to be used within its elastic limit,

$$F = Kz, \tag{94}$$

where F is the applied force, z is the extension of the spring, and K is a constant of proportionality called the force constant or spring constant. If the balance is to be used for measuring unknown forces in terms of their corresponding values of the extension, K must be determined. Calibration is effected by hanging various known weights on the spring and measuring the corresponding values of z. One method for finding K is to plot F vs. z graphically and obtain the slope of the straight line drawn through the experimental points, or as near to them as possible. This procedure gives no means of estimating the error, however, and in the present experiment it is not suitable because the numerical value of the slope would not have a precision corresponding to that of the original data. We shall use the method developed in Section VI–B, which overcomes these disadvantages, although at the expense of considerable labor.

In our instrument, the extension was produced by moving the support which held the upper end of the spring along a vertical scale until a horizontal cross bar, hung from the lower end of the spring above the weight pan, just barely failed to touch a fixed indicator.

The data and calculations based on Section VI–B are given in Table XI. The first column gives the applied forces, expressed for convenience in static units, grams weight. The second column lists the corresponding positions of the support of the upper end of the spring as measured from an arbitrary zero. The third column gives the extensions of the spring, obtained by subtracting 40.426 cm (the position for zero force) from the figures in the second column. Each of these positions was determined five times, and the respective standard deviations were computed by use of Eq. (5), by the method illustrated in Section IV–E. These standard deviations are given in the fourth column. While there may appear to be some variation in these figures, the agreement is excellent considering the small number of values upon which each is based. We conclude that the absolute random error of the extension is independent of the applied force, and therefore the application of the theory in the latter part of Section VI–B seems justified. (The 3-gm measurement was not completely independent, since this was obtained by combining the 1-gm and 2-gm weights used in the preceding values. However, this is neglected in the following calculations.)

In the application of the theory in Section VI–B, there are two possible ways of associating the quantities in Eq. (94) with those in Eq. (55):

$$(1)\ x = F,\ \ y = z,\ \ b = 1/K,$$

TABLE XI

Data and Calculations for Calibration of Jolly Balance

Applied force (gm wt) (x)	*Position of support* (cm)	*Extension* (cm) (y)	*Standard deviation (from experiment)* (cm)	x^2	xy	$\delta y = y - a - bx$ (cm)	y^2
0.000	40.426	0.000	9×10^{-3}	0.000	0.000	-5×10^{-3}	0.00000
1.000	41.206	0.780	7×10^{-3}	1.000	0.780	-1×10^{-3}	0.60840
2.000	42.002	1.576	9×10^{-3}	4.000	3.152	$+19 \times 10^{-3}$	2.48378
3.000	42.758	2.332	5×10^{-3}	9.000	6.996	0×10^{-3}	5.43822
4.000	43.508	3.082	5×10^{-3}	16.000	12.328	-26×10^{-3}	9.49872
5.000	44.324	3.898	7×10^{-3}	25.000	19.490	$+14 \times 10^{-3}$	15.19440
15.000 $=\Sigma x_n$		11.668 $=\Sigma y_n$	7×10^{-3} Average	55.000 $=\Sigma x_n^2$	42.746 $=\Sigma x_n y_n$		33.22352 $= \Sigma y_n^2$

and

$$(2)\quad x = z,\ y = F,\ b = K.$$

In both cases, a has the theoretical value of zero. By logic, there is no clear choice of method. In the calibration process, the force is the independent variable and logically should be associated with x, while in the use of the instrument after calibration the converse is true. The first system was chosen arbitrarily and is used in the remaining columns of Table XI.

If we take values of the quantities from the bottom of Table XI and also place the number of measurements $k = 6$, the values of a and b may be calculated from Eqs. (61) and (62), respectively:

$$a = \frac{(55.000)(11.668) - (15.000)(42.746)}{D'} = 0.005 \text{ cm}$$

and

$$b = \frac{(6)(42.746) - (15.000)(11.668)}{D'} = 0.775_8 \text{ cm per gm wt},$$

where

$$D' = k\Sigma x_n^2 - (\Sigma x_n)^2 = (6)(55.000) - (15.000)^2 = 105.00 \text{ (gm wt)}^2.$$

In this application, the interpretation to be given to a is that it is the deviation associated with the point (0.000, 0.000). Since this is an experi-

mental point subject to errors like other points, we cannot expect our line of best fit to go through it exactly.

Let us determine s_y by the indirect method, in which $\Sigma(\delta y_n)^2$ is found by the use of Eq. (63A):

$$\begin{aligned}\Sigma(\delta y_n)^2 &= 33.22352 \\ &\quad - \frac{(11.668)^2(55.000) - 2(42.746)(15.000)(11.668) + 6(42.746)^2}{D'} \\ &= 33.22352 - \frac{7{,}487.822 - 14{,}962.810 + 10{,}963.323}{D'} \\ &= 33.22352 - 33.22224 = 1.28 \times 10^{-3}\ \text{cm}^2.\end{aligned}$$

A calculation by the direct method gave 1.26×10^{-3} cm^2, which can be considered to be in good agreement.

The question arises why it was necessary to retain as many as eight digits in some of the numbers appearing in this calculation when the original numbers, whose products give these numbers, are given to only five digits and probably are significant to only four. If these numbers were statistically independent, four, or possibly five, digits would be all that would be justified. However, correlations exist between these numbers because they are computed from the same set of data. Therefore systematic differences persist in spite of much larger statistical variations.

By substitution of $\Sigma(\delta y_n)^2 = 1.28 \times 10^{-3}$ cm^2 and $k = 6$ into Eq. (63) we obtain

$$s_y = 18 \times 10^{-3}\ \text{cm}.$$

This is more than twice as great as the average of 7×10^{-3} cm obtained from analysis of the data on the individual readings, as given at the bottom of the fourth column of Table XI. This discrepancy is a real one. It cannot be attributed to statistical errors, since both values were computed from the same set of data and also because it is too large for this to be the probable source. The discrepancy can then be explained only by the presence of another source of error, due to the calibration of the weights. We denote the quantity at the bottom of the fourth column of Table XI by s_1 and the standard deviation in the calibration of the weights by s_2. Then, since these are independent,

$$s_y^2 = s_1^2 + s_2^2.$$

From this and from the numerical values given above, $s_2 = 17 \times 10^{-3}$ cm. It is more logical to express this in terms of units of force. To do this, we divide this value by the force constant K and obtain 22 mg wt. Thus our specification of the applied force in Table XI to four significant figures is not completely justified.

The preceding calculation has been based upon the assumption that all the errors are concentrated in y (displacement) and none are in x (force); it has also utilized the minimization of the sum of the squares of the deviations of y. However, it has led to a contradiction of these assumptions, as it has compelled us to conclude that there are large errors in the measurements of force as represented by s_2. Therefore, a more elaborate method of selecting the best line is required, in principle. As we have said previously, whatever such a method is, it should yield a straight line having parameters intermediate between those obtained by the extremes of assuming (a) that all the errors are concentrated in the displacement, as above, and (b) that all the errors are concentrated in the force. Therefore, to set bounds on these parameters, a similar calculation was carried out employing the other substitution of variables which was suggested earlier: $x = z$, $y = F$, $b = K$. The slope and the intercept on the displacement axis obtained in this second calculation turned out to be identical with the original ones, namely, 0.775_8 cm per gm wt and 0.005 cm. Therefore an elaborate procedure such as this would be justified only if the individual measurements had greater accuracy than the present ones.

The determination of the error s_2 from the present data would have been impossible without the application of the theory in Section VI–B. The large value of s_2 indicates that the only improvement which would have been obtained by use of a more sensitive spring would have been to increase the accuracy of relative measurements of force. It would not have improved the absolute values unless a better set of weights had been used for calibration.

The result $s_y = 18 \times 10^3$ cm may be used to calculate the standard deviations of a and b. From Eq. (65),

$$s_a = 18 \times 10^{-3} \sqrt{\frac{55.000}{D'}} = 13 \times 10^{-3} \text{ cm.}$$

The value of a is 5×10^{-3} cm, and the difference between this and the expected value, zero, is thus considerably less than the standard deviation.

In a similar manner, the standard deviation of b may be determined from Eq. (66):

$$s_b = 18 \times 10^{-3} \sqrt{\frac{6}{D'}} = 4.3 \times 10^{-3} \text{ cm per gm wt.}$$

The fractional standard deviation is $S_b = 0.55\%$.

Finally, we should estimate the error to be assigned to unknown forces to be measured by this instrument. From Eq. (69A), it is seen that s_0, the standard deviation in a computed value of y, is a minimum when x is equal to the average of the measured values. In the present experiment, this occurs at $x = 2.5$ gm wt, where the extension has a standard deviation, by Eq. (69A), of 7×10^{-3} cm. Since with the instrument used in this experi-

ment the extension is measured and the force is computed, it is more logical to express this as an equivalent standard deviation in the force. This may readily be effected by dividing by the value of the slope $b = K$, giving 9 mg wt.

The computation of forces between the extremes used in the calibration is, of course, in the range of "interpolation." At both of these limits, $x_0 = 0$ and $x_0 = 5$ gm wt, s_0 equals 13×10^{-3} cm, or the equivalent of 17 mg wt, according to Eq. (69). If the force is greater than 5 gm wt (that is, lies in the range of "extrapolation"), s_0 increases further. The maximum value which could be accommodated with the present scale and spring would be with the upper end of the spring at the end of the scale at 100 cm. This would correspond to an extension of the spring of about 60 cm or an applied force of approximately 45 gm wt. For such a force the standard deviation, according to Eq. (69), is the equivalent of 0.22 gm wt or 0.48%. This is very nearly equal to the limiting value for the fractional standard deviation $S_b = 0.54\%$.

REFERENCES

The following references have been useful to the author in the preparation of this book, and should prove of value to the reader desiring more extensive information.

GENERAL

1. R. H. Bacon, *American Journal of Physics* **14**, 84 and 198 (1946).
2. R. H. Bacon, *American Journal of Physics* **21**, 428 (1953).
3. R. T. Birge, *American Physics Teacher* **7**, 351 (1939).
4. R. T. Birge, *Physical Review* **40**, 207 (1932).
5. R. S. Burington and D. C. May, *Handbook of Probability and Statistics.* Sandusky, Ohio: Handbook Publishers, Inc., 1953.
6. W. E. Deming, *Statistical Adjustment of Data,* 2nd ed. New York: John Wiley & Sons, Inc., 1943.
7. W. E. Deming and R. T. Birge, *Reviews of Modern Physics* **6**, 119 (1934).
8. T. C. Fry, *Probability and Its Engineering Uses.* New York: D. Van Nostrand Co., Inc., 1928.
9. G. E. Hudson, *American Journal of Physics* **21**, 362 (1953).
10. F. Kohlrausch, *Lehrbuch der Praktischen Physik,* 16th ed. Leipzig: B. G. Teubner, 1930.
11. H. W. Reddick and F. H. Miller, *Advanced Mathematics for Engineers,* Chapter IX. New York: John Wiley & Sons, Inc., 1947.
12. J. B. Scarborough, *Numerical Mathematical Analysis.* Baltimore: The Johns Hopkins Press, 3rd ed., 1955.
13. A. G. Worthing and J. Geffner, *Treatment of Experimental Data.* New York: John Wiley & Sons, Inc., 1943.

STATISTICAL ERRORS OF NUCLEAR PHYSICS

1. W. C. Elmore, *Nucleonics* **6**, No. 1, 26 (1950).
2. R. A. Fisher, *Statistical Methods for Research Workers.* London: Oliver and Boyd, 1946.
3. D. Halliday, *Introductory Nuclear Physics,* Section 20. New York: John Wiley & Sons, Inc., 1950.
4. S. A. Korff, *Electron and Nuclear Counters,* 2nd ed. New York: D. Van Nostrand Co., Inc., 1955.
5. F. Proschan, *American Journal of Physics* **21**, 520 (1953).
6. L. J. Rainwater and C. S. Wu, *Nucleonics* **1**, No. 2, 60 (1947); **2**, No. 1, 42 (1948).
7. E. Rutherford, J. Chadwick, and C. D. Ellis, *Radiations from Radioactive Substances,* Chapter VII. New York: Cambridge University Press, 1930.
8. John Strong *et al., Procedures in Experimental Physics,* Chapters VI and VII. New York: Prentice-Hall, Inc., 1938.