A TEXT-BOOK

ON THE

METHOD OF LEAST SQUARES.

BY

MANSFIELD MERRIMAN,

MEMBER OF AMERICAN MATHEMATICAL SOCIETY.

EIGHTH EDITION, REVISED

TOTAL ISSUE, EIGHT THOUSAND.

NEW YORK

JOHN WILEY & SONS, INC.

LONDON: CHAPMAN & HALL, LIMITED

1915

THE SCIENTIFIC PRESS
ROBERT DRUMMOND AND COMPANY
BROOKLYN, N. Y.

CONTENTS.

CHAPTER I.

INTRODUCTION.

CHAPTER II.

LAW OF PROBABILITY OF ERROR.

CHAPTER III.

THE ADJUSTMENT OF OBSERVATIONS.

CHAPTER IV.

THE PRECISION OF OBSERVATIONS

CHAPTER V.

DIRECT OBSERVATIONS ON A SINGLE QUANTITY.

CHAPTER VI.

FUNCTIONS OF OBSERVED QUANTITIES.

CHAPTER VII.

INDEPENDENT OBSERVATIONS ON SEVERAL QUANTITIES.

CHAPTER VIII.

CONDITIONED OBSERVATIONS.

CHAPTER IX.

THE DISCUSSION OF OBSERVATIONS.

CHAPTER X.

SOLUTION OF NORMAL EQUATIONS.

CHAPTER XI.

APPENDIX AND TABLES.

TABLES.

CHAPTER II.

LAW OF PROBABILITY OF ERROR.

17. The probability of an assigned accidental error in a set of measurements is the ratio of the number of errors of that magnitude to the total number of errors. It is proposed, in this chapter, to investigate the relation between the magnitude of an error and its probability.

Axioms derived from Experience.

18. An analogy often referred to in the Method of Least Squares is that between bullet-marks on a target and errors of observations. The marksman answers to an observer; the position of a bullet-mark, to an observation; and its distance from the centre, to an error. If the marksman be skilled, and all constant errors, like the effect of gravitation, be eliminated in the sighting of the rifle, it is recognized that the deviations of the bullet-marks, or errors, are quite regular and symmetrical. First, it is observed that small errors are more frequent than large ones; secondly, that errors on one side are about as frequent as on the other; and, thirdly, that very large errors do not occur. Further: it is recognized, that, the greater the skill of the marksman, the nearer are the marks to his point of aim.

For instance, in the Report of the Chief of Ordnance for 1878, Appendix S′, Plate VI, is a record of one thousand shots fired deliberately (that is, with precision) from a battery-gun, at a target two hundred yards distant. The target was fifty-two feet long by eleven feet high, and the point of aim was its central horizontal line. All of the shots struck the target; there being few, however, near the upper and lower edges, and nearly the same number above the central horizontal line as below it. On the record, horizontal lines are drawn, dividing the target into eleven equal divisions; and a count of the number of shots in each of these divisions gives the following results:

In top division	1 shot
In second division	4 shots
In third division	10 shots
In fourth division	89 shots
In fifth division	190 shots
In middle division	212 shots
In seventh division	204 shots
In eighth division	193 shots
In ninth division	79 shots
In tenth division	16 shots
In bottom division	2 shots
Total	1,000 shots

On Fig. 3 is shown, by means of ordinates, the distribution of these shots; A being the top division, B the middle, and C the bottom division. It will be observed that there is a slight preponderance of shots below the centre, and there is reason to believe that this is due to a constant error of gravitation not entirely eliminated in the sighting of the gun.

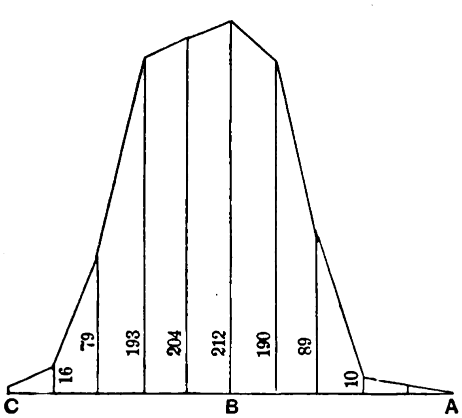

Fig. 3.

19. The distribution of the errors or residuals in the case of direct observations is similar to that of the deviations just

discussed. For instance, in the United States Coast Survey Report for 1854, p. *91, are given a hundred measurements of angles of the primary triangulation in Massachusetts. The residual errors (Art. 8) found by subtracting each measurement from the most probable values are distributed as follows:

Between +6″.0 and +5″.0	1 error
Between +5.0 and +4.0	2 errors
Between +4.0 and +3.0	2 errors
Between +3.0 and +2.0	3 errors
Between +2.0 and +1.0	13 errors
Between +1.0 and 0.0	26 errors
Between 0.0 and −1.0	26 errors
Between −1.0 and −2.0	17 errors
Between −2.0 and −3.0	8 errors
Between −3.0 and −4.0	2 errors
Total	100 errors

Here also it is recognized that small errors are more frequent than large ones, that positive and negative errors are nearly equal in number, and that very large errors do not occur. In this case the largest residual error was 5″.2; but, with a less precise method of observation, the limits of error would evidently be wider.

20. The axioms derived from experience are, hence, the following:

Small errors are more frequent than large ones.
Positive and negative errors are equally frequent.
Very large errors do not occur.

These axioms are the foundation of all the subsequent reasoning.

The Probability Curve.

21. In precise observations, then, the probability of a small error is greater than that of a large one, positive and negative errors are equally probable, and the probability of a very large error is zero. The words "very large" may seem somewhat vague when used in general, although in any particular case the meaning is clear; thus, with a theodolite reading to seconds, 20″ would be very large, and with a transit reading to minutes, 5′ would be very large. Really, in every class of measurements there is a limit, l, such that all the positive errors are included between 0 and $+l$, and all the negative ones between 0 and $-l$.

22. Hence the probability of an error is a function of that error; so that, calling x any error and y its probability, the law of probability of error is represented by an equation

$$y = f(x),$$

and will be determined, if the form of $f(x)$ can be found. If, then, y be taken as an ordinate, and x as an abscissa, this may be regarded as the equation of a curve which must be of a form to agree with the three fundamental axioms; namely, its maximum ordinate OA must correspond to the error zero; it must be symmetrical with respect to the axis of Y, since positive and negative errors of equal magnitude are equally probable; as x increases numerically, the value of y must decrease, and, when x becomes very large, y must be zero. Fig. 4 represents such a curve, OP and OM being errors, and PB and MC their respective probabilities. Further: since different measurements have different degrees of accuracy, each class of observations will have a distinct curve of its own.

The curve represented in Fig. 4 is called the probability curve. In order to determine its equation, it is necessary to consider y as a continuous function of x. This is evidently perfectly allowable; since, as the precision of observations is increased, the successive values of x are separated by smaller and smaller intervals. The requirement of the third axiom, that y must be zero for all values of x greater than the limit $\pm l$, is apparently an embarrassing one, as it is impossible to determine a continuous function of x which shall become zero for $x = \pm l$ and also be zero for all values of x from $\pm l$ to $\pm \infty$. But, since this limit l can never be accurately assigned, it will be best to extend the limits to $\pm \infty$, and determine the curve in such a way that

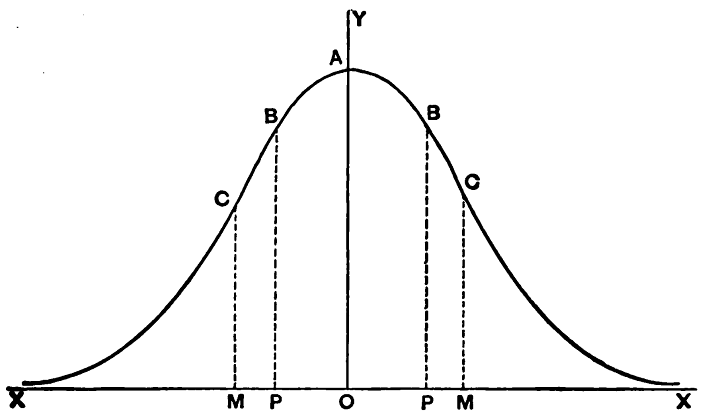

Fig. 4.

the value of y, although not zero for large values of x, will be so very small as to be practically inappreciable. The equation of the probability curve will be the mathematical expression of the law of probability of errors of observation. Two deductions of this law will be given; the first that of Hagen, and the second that of Gauss.

First Deduction of the Law of Error.

23. Hagen's demonstration rests on the following hypothesis or axiom, derived from experience:

An error is the algebraic sum of an indefinitely great number of small elementary errors which are all equal, and each of which is equally likely to be positive or negative.

To illustrate: suppose that, by several observations with a levelling instrument and rod, the difference in elevation between two points has been determined. This value is greater or less than the true difference of level by a small error, x. This error x is the result of numerous causes acting at every observation: the instrument is not perfectly level, the wind shakes it, the sun's heat expands one side of it, the level-bubbles are not accurately made, the glass gives an indistinct definition, the tripod is not firm, the eye of the observer is not in perfect order, there is irregular refraction of the atmosphere, the man at the rod does not hold it vertical, the turning-points are not always good ones, the graduation of the rod is poor, the target is not properly clamped, the rod-man errs in taking the reading, and many others. Again: each of these causes may be subdivided into others; for instance, the error in reading the rod may be due, perhaps, to the accumulated result of hundreds of little causes. The total error, x, may hence be fairly regarded as resulting from the combination of an indefinitely great number of small elementary errors; and no reason can be assigned why one of these should be more likely to be positive than negative, or negative than positive.

24. Now, it is evident that it is more probable that the number of positive elementary errors should be approximately equal to the number of negative ones than that either should be markedly in excess, and that the probability of the elementary errors being either all positive or all negative is exceedingly small. In the first case the actual error is small, and in the second large; and so the probabilities of small errors are the greatest, and the probability of a very large error is practically zero. These correspond to the properties which the probability curve must possess.

Let Δx represent the magnitude of an elementary error, and m the number of those errors. The probability that any Δx will be positive is $\frac{1}{2}$, and that it will be negative is also $\frac{1}{2}$. The probability that all of the m elementary errors will be positive is hence $(\frac{1}{2})^m$; the probability that $m - 1$ will be positive and 1 negative is $m(\frac{1}{2})^{m-1}(\frac{1}{2})^1$; and the probabilities of all the respective cases will be given by the corresponding terms of the binomial formula (Art. 14). When all of the m elementary errors are positive, the resulting error of observation is $+ m.\Delta x$; when $m - 1$ are positive and 1 negative, the resulting error is $+ (m - 1)\Delta x - \Delta x$, or $+ (m - 2)\Delta x$. If $m - n$ elementary errors are positive and the remaining n are negative, the resulting error is $+ (m - n)\Delta x - n.\Delta x$, or $+ (m - 2n)\Delta x$, and the probability of this particular combination is given by the $n + 1^{\text{th}}$ term of the expansion of the binomial $(\frac{1}{2} + \frac{1}{2})^m$. It is easy then to write the following table:

Elementary Errors Δx.	Resulting Error x.	Its Probability y.
If m are + and 0 are −	$m\Delta x$	$\left(\frac{1}{2}\right)^m$
If $m-1$ are + and 1 is −	$(m-2)\Delta x$	$m\left(\frac{1}{2}\right)^m$
If $m-2$ are + and 2 are −	$(m-4)\Delta x$	$\frac{m(m-1)}{1.2}\left(\frac{1}{2}\right)^m$
If $m-3$ are + and 3 are −	$(m-6)\Delta x$	$\frac{m(m-1)(m-2)}{1.2.3}\left(\frac{1}{2}\right)^m$
. . . .	. . . . . .	. . . .
If $m-n$ are + and n are −	$(m-2n)\Delta x$	$\frac{m(m-1)(m-2)\ldots(m-n+1)}{1.2.3\ldots n}\left(\frac{1}{2}\right)^m$
If $m-n-1$ are + and $n+1$ are −	$(m-2n-2)\Delta x$	$\frac{m(m-1)(m-2)\ldots(m-n)}{1.2.3\ldots n+1}\left(\frac{1}{2}\right)^m$
. . . .	. . . . . .	. . . .

25. In the curve $y = f(x)$ let OM be any error x, and MC its probability y; also let OP be an error x' less in magnitude, and PB its corresponding probability y'. Then, from the figure,

$$\text{limit}\ \frac{BD}{CD} = \text{limit}\ \frac{y - y'}{x - x'} = \frac{dy}{dx}$$

is the differential equation of the curve. To deduce, then, the law of probability of error, it is only necessary to find $\frac{y-y'}{x-x'}$ in terms of y and x, pass to the limit, place it equal to $\frac{dy}{dx}$, and perform the integration.

If x' be taken as the error next less in magnitude to x, the difference $x - x'$ equals $2\Delta x$, and the value of $\frac{y-y'}{x-x'}$ is the limit $\frac{dy}{dx}$ if the curve is to be continuous.

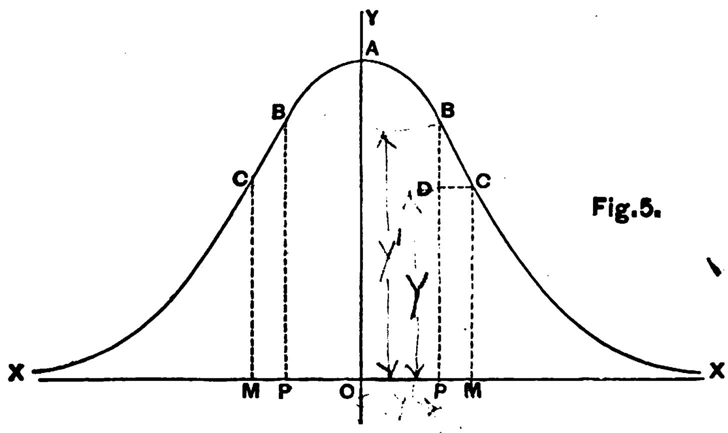

Fig.5.

26. For the two consecutive errors x and x' take (from Art. 24) the two general values

$$x = (m - 2n)\Delta x, \quad \text{and} \quad x' = (m - 2n - 2)\Delta x.$$

The ratio of the probabilities of these errors is

$$\frac{y'}{y} = \frac{m - n}{n + 1},$$

which, after inserting for n its value in terms of x, m, and Δx, may be put into the form

$$y - y' = y\frac{2(\Delta x - x)}{(m + 2)\Delta x - x} = y\frac{-2x}{m\Delta x}.$$

Here Δx in the numerator vanishes in comparison with x. In the denominator, 2 vanishes compared with m, and $m\Delta x$ is the maximum positive error, which is so large that x vanishes in comparison with it. The differential equation, then, is

$$\frac{dy}{dx} = \frac{y - y'}{2\Delta x} = -\frac{yx}{m(\Delta x)^2},$$

or

$$\frac{dy}{dx} = -2h^2yx,$$

in which $2h^2$ has been written to represent the quantity $\frac{1}{m(\Delta x)^2}$. The integration of this equation gives

$$\log y = -h^2x^2 + k',$$

in which k' is the constant of integration, and the logarithm is in the Napierian system. By passing from logarithms to numbers

$$y = e^{-h^2x^2 + k'} = e^{-h^2x^2}\,e^{k'},$$

in which e is the base of the Napierian system. Since $e^{k'}$ is a constant, this may be written

$$(1) \qquad y = ke^{-h^2x^2},$$

and this is the equation of the probability curve, or the equation expressing the law of probability of errors of observation.

This equation satisfies the conditions imposed in Art. 22, for y is a maximum when x is 0; it is symmetrical with respect to the axis of Y, since equal positive and negative values of x give equal values of y, and when x becomes very large, y is very small. The constants k and h will be particularly considered hereafter.

Second Deduction of the Law of Error.

27. Gauss's demonstration is based on the following hypothesis or axiom, established by experience:

The most probable value of a quantity which is observed directly several times, with equal care, is the arithmetical mean of the measurements.

The average or arithmetical mean has always been accepted and used as the best rule for combining direct observations of equal precision upon one and the same quantity. This universal acceptance may be regarded as sufficient to justify the axiom that it gives the most probable value, the words "most probable" being used in the sense of Art. 13; for after all, as Laplace has said, the theory of probability is nothing but common sense reduced to calculation. If the measurements be but two in number, the arithmetical mean is undoubtedly the most probable value; and, for a greater number, mankind, from the remotest antiquity, has been accustomed to regard it as such.

It is a characteristic of the arithmetical mean that it renders the algebraic sum of the residual errors zero. To show this, let $M_1, M_2 \ldots M_n$, be n measurements of a quantity; then the arithmetical mean of these is,

$$z = \frac{M_1 + M_2 + M_3 + \ldots + M_n}{n}.$$

This equation may be written

$$nz = M_1 + M_2 + M_3 + \ldots + M_n,$$

which by transposition becomes

$$(z - M_1) + (z - M_2) + (z - M_3) + \ldots + (z - M_n) = 0;$$

that is to say, the arithmetical mean requires that the algebraic sum of the residual errors shall be zero. To take a numerical illustration, let 730.4, 730.5, and 730.9 be three measurements of the length of a line. The arithmetical mean is 730.6, giving the residuals +0.2, +0.1, and −0.3, whose algebraic sum is 0.

28. Consider the general case of indirect observations, in which it is required to find the most probable values of quantities by measurements on functions of those quantities. For simplicity, only two quantities, z_1 and z_2, will be considered; although the reasoning is general, and applies to any number. Let n observations be made on functions of z_1 and z_2, from which it is required to find the most probable values of z_1 and z_2. The differences between the observations and the corresponding true values of the functions are errors $x_1, x_2 \ldots x_n$, each of which is also a function of z_1 and z_2. The probabilities of these errors are

$$y_1 = f(x_1), \quad y_2 = f(x_2) \ldots y_n = f(x_n).$$

And by Art. 12 the probability of committing the given system of errors is

$$P = y_1 y_2 y_3 \ldots y_n = f(x_1)\,f(x_2) \ldots f(x_n).$$

Applying logarithms to this expression, it becomes

$$\log P = \log f(x_1) + \log f(x_2) + \ldots + \log f(x_n).$$

Now, the most probable values of the unknown quantities z_1 and z_2 are those which render P a maximum (Art. 13), and hence the derivative of P with respect to each of these variables must be equal to zero. Indicating the differentiation, the following equations result:

$$\frac{dP}{Pdz_1} = \frac{df(x_1)}{f(x_1)\,dz_1} + \frac{df(x_2)}{f(x_2)\,dz_1} + \ldots + \frac{df(x_n)}{f(x_n)\,dz_1} = 0,$$

$$\frac{dP}{Pdz_2} = \frac{df(x_1)}{f(x_1)\,dz_2} + \frac{df(x_2)}{f(x_2)\,dz_2} + \ldots + \frac{df(x_n)}{f(x_n)\,dz_2} = 0.$$

Since in general $df(x) = \phi(x)f(x)dx$, these may be written

$$\phi(x_1)\frac{dx_1}{dz_1} + \phi(x_2)\frac{dx_2}{dz_1} + \ldots + \phi(x_n)\frac{dx_n}{dz_1} = 0,$$

$$\phi(x_1)\frac{dx_1}{dz_2} + \phi(x_2)\frac{dx_2}{dz_2} + \ldots + \phi(x_n)\frac{dx_n}{dz_2} = 0,$$

and, being as many in number as there are unknown quantities, they will determine the values of those unknown quantities as soon as the form of the function ϕ is known.

Since these equations are general, and applicable to any number of unknown quantities, the form of the function ϕ may be determined from any special but known case. Such is that in which there is but one unknown quantity, and the observations are taken directly upon that quantity. Thus, if there be only the quantity z, and the measurements give for it the values M_1, $M_2 \ldots M_n$, the errors are,

$$x_1 = z - M_1,\ x_2 = z - M_2 \ldots x_n = z - M_n,$$

from which

$$\frac{dx_1}{dz} = \frac{dx_2}{dz} = \ldots = \frac{dx_n}{dz} = 1,$$

and the first equation above becomes

$$\phi(x_1) + \phi(x_2) + \phi(x_3) + \ldots + \phi(x_n) = 0.$$

In this case, also, the arithmetical mean is the most probable value, and the algebraic sum of the residuals will be zero, or, if v denote any residual in general,

$$v_1 + v_2 + v_3 + \ldots + v_n = 0.$$

Now, if the number of observations, n, is very large, the residuals v will coincide with the errors x (Art. 8), and

$$x_1 + x_2 + x_3 + \ldots + x_n = 0.$$

This equation can only agree with that above when ϕ signifies multiplication by a constant, or when

$$\phi(x_1) + \phi(x_2) + \ldots + \phi(x_n) = cx_1 + cx_2 + \ldots + cx_n.$$

Replacing in this the values of $\phi(x_1)$, $\phi(x_2)$, etc., it becomes

$$\frac{df(x_1)}{f(x_1)dx_1} + \frac{df(x_2)}{f(x_2)dx_2} + \text{etc.} = cx_1 + cx_2 + \text{etc.};$$

and, since this is true whatever be the number of observations, the corresponding terms in the two members are equal. Hence, if x be any error, and $y = f(x)$,

$$\frac{df(x)}{f(x)dx} = \frac{dy}{ydx} = cx.$$

Multiplying both members by dx, and integrating,

$$\log y = \frac{cx^2}{2} + k',$$

Passing from logarithms to numbers,

$$y = e^{\frac{1}{2}cx^2} e^{k'}.$$

Here the constant c must be essentially negative, since the probability y should decrease as x increases numerically; replacing it, then, by $-2h^2$, and also putting $e^{k'} = k$, there results

$$(1) \qquad y = ke^{-h^2x^2},$$

which is the equation of the probability curve, or the equation expressing the law of probability of errors of observation.

Discussion of the Curve $y = ke^{-h^2x^2}$.

29. Since positive and negative values of x numerically equal give equal values of y, the curve is symmetrical with respect to the axis of Y. The maximum value of y is for $x = 0$, when

$y = k$; k is, hence, the probability of the error 0. As x increases numerically, y decreases; and when $x = \infty$, y becomes 0. The value of the first derivative is

$$\frac{dy}{dx} = -2kh^2e^{-h^2x^2}x,$$

which becomes zero when $x = 0$ and when $x = \pm\infty$, indicating that the curve is horizontal over the origin, and that the axis of x is an asymptote. The value of the second derivative is

$$\frac{d^2y}{dx^2} = -2kh^2e^{-h^2x^2}(-2h^2x^2 + 1),$$

which becomes 0 when $-2h^2x^2 + 1 = 0$, indicating that the curve has an inflection-point when $x = \pm\frac{1}{h\sqrt{2}}$.

To show further the form of the curve, the following values have been computed, taking k and h each as unity:

$$y = e^{-x^2} = \frac{1}{e^{x^2}}.$$

x	y	x	y
.0	1.0000	±1.8	0.0392
±0.2	0.9608	±2.0	0.0183
±0.4	0.8521	±2.2	0.0079
±0.6	0.6977	±2.4	0.0032
±0.8	0.5273	±2.6	0.0012
±1.0	0.3679	±2.8	0.0004
±1.2	0.2370	±3.0	0.0001
±1.4	0.1409		
±1.6	0.0773	±∞	0.0000

The curve in Fig. 4 is constructed from these values, the vertical scale being double the horizontal. C is the inflection-point, whose abscissa OM is 0.707.

30. The constant h is a quantity of the same kind as $\frac{1}{x}$, since the exponent h^2x^2 must be an abstract number. Methods will be hereafter explained by which its value may be determined for given observations. The probability of an assigned error x' decreases as h increases; and hence, the more precise the observations, the greater is h. For this reason h may be called "the measure of precision."

The constant k is an abstract number; and, since it is the probability of the error 0, it is larger for good observations than for poor ones. The more precise the measurements, the larger is k.

The Probability Integral.

31. To determine the value of the constant k, and also to investigate the probability of an error falling between assigned limits, the following reasoning may be employed:

Let $x', x_1, x_2 \ldots x$ be a series of errors, x' being the smallest, x_1 the next following, and x the last; the differences between the successive values being equal, and x' being any error. Then, by Art. 11, the probability of committing one of these errors, that is, the probability of committing an error lying between x' and x, is the sum of the separate probabilities $ke^{-h^2x'^2}$, $ke^{-h^2x_1^2}$, etc.; or, if P denote this sum,

$$P = k(e^{-h^2x'^2} + e^{-h^2x_1^2} + e^{-h^2x_2^2} + \ldots + e^{-h^2x^2}),$$

which may be written

$$P = k\Sigma_{x'}^{x} e^{-h^2x^2},$$

the notation $\Sigma_{x'}^{x}$ denoting summation from x' to x inclusive.

To replace the sign of summation by that of integration, dx must be the interval between the successive values of the errors, and then the probability that an error will lie between any two limits x' and x is

$$P = \frac{k}{dx}\int_{x'}^{x} e^{-h^2x^2}dx.$$

Now, it is certain that the error will lie between $-\infty$ and $+\infty$, and, as unity is the symbol for certainty,

$$1 = \frac{k}{dx}\int_{-\infty}^{+\infty} e^{-h^2x^2}dx.$$

The value of the definite integral in this expression is $\frac{\sqrt{\pi}}{h}$.*
Hence

$$1 = \frac{k\sqrt{\pi}}{hdx},$$

---

* The following method of determining this integral is nearly that presented by Sturm in his Cours d'Analyse, Paris, 1857, vol. ii. p. 16.

The integral $\int e^{-h^2x^2}dx$ expresses the area between the probability curve and the axis of X, and, since the curve is symmetrical to the axis of Y, that integral between the limits $-\infty$ and $+\infty$ will be equal to double the integral between the limits 0 and $+\infty$. Placing also $hx = t$,

$$\int_{-\infty}^{+\infty} e^{-h^2x^2}dx = \frac{2}{h}\int_0^{\infty} e^{-t^2}dt,$$

and the integral in the second member is to be determined.

Take three co-ordinate rectangular axes OT, OU, and OV, and change t into u, then

$$A = \int_0^{\infty} e^{-t^2}dt = \text{area between curve } VtT \text{ and axes,}$$

$$A = \int_0^{\infty} e^{-u^2}du = \text{area between curve } VuU \text{ and axes,}$$

and $$A^2 = \int_0^{\infty}\int_0^{\infty} e^{-t^2-u^2}dtdu.$$

from which the value of k is

$$k = \frac{hdx}{\sqrt{\pi}}.$$

The equation of the probability curve now becomes

(2) $$y = hdx\pi^{-\frac{1}{2}}e^{-h^2x^2},$$

and the probability that an error lies between any two given limits x' and x becomes

(3) $$P = \frac{h}{\sqrt{\pi}}\int_{x'}^{x} e^{-h^2x^2}dx.$$

Equations (1), (2), and (3) are the fundamental ones in the theory of accidental errors of observation.

32. The probability that an error lies between the limits $-x$ and $+x$ is double the probability that it lies between the limits 0 and $+x$, on account of the symmetry of the curve. Hence

(4) $$P = \frac{2h}{\sqrt{\pi}}\int_{0}^{x} e^{-h^2x^2}dx$$

Now $v = e^{-t^2}$ is the equation of the curve VtT, and $v = e^{-u^2}$ is the equation of VuU, and, if either of these curves revolves about the axis of V, it generates a surface whose equation is $v = e^{-t^2-u^2}$. Hence the double integral A^2 is one-fourth of the volume included between that surface and the horizontal plane. If a series of cylinders concentric with the axis V form the volume, the area of the ring included between two whose radii are r and $r + dr$ is $2\pi rdr$, and the corresponding height is $v = e^{-t^2-u^2} = e^{-r^2}$. Hence one-fourth of the volume is

$$A^2 = \frac{1}{4}\int_0^{\infty} e^{-r^2} 2\pi rdr,$$

which, since $\int e^{-r^2}2rdr = e^{-r^2}$, is equal to $\frac{\pi}{4}$. Therefore

$$A = \int_0^{\infty} e^{-t^2}dt = \frac{\sqrt{\pi}}{2},$$

and hence, finally,

$$\int_{-\infty}^{+\infty} e^{-h^2x^2}dx = \frac{2}{h}\int_0^{\infty} e^{-t^2}dt = \frac{\sqrt{\pi}}{h}.$$

expresses the probability that an error is numerically less than x. This may be written in the form

$$P=\frac{2}{\sqrt{\pi}}\int_0^{hx} e^{-h^2x^2}d(hx), \tag{4}$$

and is called the probability integral.

As the number of errors of the magnitude x is proportional to the probability y, and as P in equation (4) is merely the summation of the probabilities of all errors between $-x$ and $+x$, the number of errors between these limits is also proportional to P. Now, P is the area of the probability curve between the limits $-x$ and $+x$, the whole area being unity. Hence the number of errors between two assigned limits ought to bear the same ratio to the whole number of errors as the value of P between these limits does to unity.

By the usual methods of the integral calculus the value of the probability integral corresponding to successive numerical values of hx may be computed.* A table of these values is given at the end of this volume (Table I.).

---

* First put $hx = t$, then $\frac{2}{\sqrt{\pi}}\int_0^t e^{-t^2}dt$ is the integral to be evaluated. By developing e^{-t^2} into a series by Maclaurin's formula, the following results:

$$P=\frac{2}{\sqrt{\pi}}\left(t-\frac{t^3}{3}+\frac{1}{1.2}\cdot\frac{t^5}{5}-\frac{1}{1.2.3}\cdot\frac{t^7}{7}+\text{etc.}\right),$$

which is convenient for small values of t. For large values integrate by parts, thus

$$\int e^{-t^2}dt=-\frac{1}{2t}e^{-t^2}-\frac{1}{2}\int\frac{e^{-t^2}}{t^2}dt$$

$$=-\frac{1}{2t}e^{-t^2}+\frac{1}{2^2t^3}e^{-t^2}+\frac{3}{2^2}\int\frac{e^{-t^2}}{t^4}dt.$$

To illustrate the use of this table, consider the case of $hx = 1.24$, for which $P = 0.9205$. Here 0.9205 is the probability that an error will be numerically less than $\frac{1.24}{h}$; or, in other words, if there be 10,000 observations, it is to be expected that in 9,205 of them the errors would lie between $-\frac{1.24}{h}$ and $+\frac{1.24}{h}$, and in the remaining 795 outside of these limits.

Comparison of Theory and Experience.

33. By means of Table I the theory employed in the deductions of equations (1), (2), (3), and (4) may be tested. To use the table it is necessary to know the value of the constant h. Granting for the present that it may be determined, the following examples will exemplify the accordance of theory and experience.

For the one hundred residual errors discussed in Art. 19, the value of h may be determined to be $\frac{1}{2''.236}$.

And since $\int_0^\infty e^{-t^2}dt = \frac{\sqrt{\pi}}{2}$, as shown in the preceding footnote,

$$\int_0^t e^{-t^2}dt = \frac{\sqrt{\pi}}{2} - \int_t^\infty e^{-t^2}dt,$$

from which $P = 1 - \frac{e^{-t^2}}{t\sqrt{\pi}}\left[1 - \frac{1}{2t^2} + \frac{1.3}{(2t^2)^2} - \frac{1.3.5}{(2t^2)^3} + \text{etc.}\right]$

From these two series the values of P can be found to any required degree of accuracy for all values of t or hx.

Then from the table the following values of P are taken:

for $x = 1''.0$ with $hx = 0.447$ the area $P = 0.473$
for $x = 2.0$ with $hx = 0.894$ the area $P = 0.794$
for $x = 3.0$ with $hx = 1.341$ the area $P = 0.942$
for $x = 4.0$ with $hx = 1.788$ the area $P = 0.989$
for $x = 5.0$ with $hx = 2.235$ the area $P = 0.998$
for $x = \infty$ with $hx = \infty$ the area $P = 1.000$

Now, these probabilities or areas P are proportional to the number of errors less than the corresponding values of x. Hence multiplying them by 100, the total number of errors, and subtracting each from that following, the number of theoretical errors between the successive values of x is found. The following is a comparison of the number of actual and theoretical errors:

Limits	Actual Errors.	Theoretical Errors.	Differences.
0″.0 and 1″.0	52	47	+5
1.0 and 2.0	30	32	−2
2.0 and 3.0	11	15	−4
3.0 and 4.0	4	5	−1
4.0 and 5.0	2	1	+1
5.0 and 6.0	1	0	+1
6.0 and ∞	0	0	0

The agreement between theory and experience, though not exact, is very satisfactory when the small number of observations is considered.

34. Numerous comparisons like the above have been made by different authors, and substantial agreement has always been found between the actual distribution of errors and the

theoretical distribution required by equations (2) and (4). The following is a comparison by Bessel of the errors of three hundred observations of the right ascensions of stars:

Limits.	Actual Errors.	Theoretical Errors.	Differences.
$0^s.0$ and $0^s.1$	114	107	+7
0.1 and 0.2	84	87	−3
0.2 and 0.3	53	57	−4
0.3 and 0.4	24	30	−6
0.4 and 0.5	14	13	+1
0.5 and 0.6	6	5	+1
0.6 and 0.7	3	1	+2
0.7 and 0.8	1	0	+1
0.8 and 0.9	1	0	+1
0.9 and ∞	0	0	0

The differences are here relatively smaller than in the previous case. And in general it is observed that the agreement between theory and experience is closer, the greater the number of errors or residuals considered in the comparison.

Whatever may be thought of the theoretical deductions of the law of probability of error, there can be no doubt but that its practical demonstration by experience is entirely satisfactory.

Remarks on the Fundamental Formulas.

35. The two equations of the probability curve,

(1) $$y = ke^{-h^2x^2},$$

(2) $$y = h.dx.\pi^{-\frac{1}{2}}e^{-h^2x^2},$$

are identical, and the former has already been discussed at length. In the latter, dx for any special case is the interval between successive values of x. For instance, if observations of an angle be carried to tenths of seconds, dx is $0''.1$; if to hundredths of seconds, dx is $0''.01$; and if a continuous curve is considered, dx is the differential of x. As y is an abstract number, $h.dx$ must likewise be abstract, and hence h must be a quantity of the same kind as $\frac{1}{dx}$. The probability of the error 0 is $\frac{hdx}{\sqrt{\pi}}$; thus in measuring angles to hundredths of seconds, the probability that an error is $0''.00$ is $\frac{0''.01h}{\sqrt{\pi}}$. As this increases with h, the value of h may be regarded as a measure of the precision of the observations. Methods of determining h are given in Chap. IV.

36. The two probability integrals,

$$P = \frac{h}{\sqrt{\pi}}\int_{x'}^{x} e^{-h^2x^2}dx, \tag{3}$$

$$P = \frac{2}{\sqrt{\pi}}\int_{0}^{hx} e^{-h^2x^2}dhx, \tag{4}$$

are identical, except in their limits. The first gives the probability that an error will lie between any two limits x' and x; and the second, the probability that it lies between the limits $-x$ and $+x$, or that it is numerically less than x. The second is then a particular case of the first. Table I refers only to (4); and from it by simple addition or subtraction the probability can be found for any two assigned limits. For example, the probability that an error lies between $-2''.0$ and $+4''.0$ is the sum of the probabilities for the limits $0''.0$ to $2''.0$ and $0''.0$ to $4''.0$; and the probability that an error is between $+2''.0$ and $+4''.0$ is the difference of the probabilities of those limits.

The integral P is simply the summation of the values of y

between the assigned limits, or $P = \Sigma y$, as required by the principle of Art. 11 to express the probability of an error lying between those limits.

37. *Problems and Queries.*

1. Can cases be imagined where positive and negative errors are not equally probable?

2. An angle is measured to tenths of seconds by two observers, and the value of h for the first observer is double that for the second. Draw the two curves of probability of error.

3. Show that the arithmetical mean of two measurements is the only value that can be logically chosen to represent the quantity.

4. The reciprocal of h for the bullet-marks in Art. 18 is 2.33 feet. Compare the actual distribution of errors with the theoretical.

5. Draw a curve for each of the equations $y = ke^{-x^2}$ and $y = ke^{-4x^2}$, assuming a convenient value for k. Show that the value of k should have been taken different in the two equations.

6. Explain how the value of π might be determined by experiments with the help of equation (2).