
A Generalized Theory of the Combination of Observations so as to Obtain the Best Result

Author(s): Simon Newcomb

Source: *American Journal of Mathematics*, Vol. 8, No. 4 (Aug., 1886), pp. 343-366

Published by: The Johns Hopkins University Press

Stable URL: <http://www.jstor.org/stable/2369392>

Accessed: 04-10-2016 02:39 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://about.jstor.org/terms>



The Johns Hopkins University Press is collaborating with JSTOR to digitize, preserve and extend access to *American Journal of Mathematics*

A Generalized Theory of the Combination of Observations so as to Obtain the Best Result.

BY SIMON NEWCOMB.

§ 1. Introductory Remarks.

The accepted practice of combining observations rests upon the hypothesis that the frequency of errors follows a certain well-known law which may be expressed as follows: Let Δ be the amount by which the result of an observation may differ from the value obtained by taking the mean result of an infinity of similar observations. Δ will then be the *error* of the observation. The infinitesimal probability that an error will be contained between the limits Δ and $\Delta + d\Delta$ is supposed to be given by the equation

$$dp = \frac{h}{\sqrt{\pi}} e^{-h^2 \Delta^2} d\Delta,$$

in which h is the “modulus of precision” depending upon the accuracy of the observations, and e is the Naperian base.

The modulus h is commonly replaced by a probable error r , which term signifies such a magnitude that the number of errors less than r in absolute value is equal to the number which exceed r . The value of r is given in terms of the modulus h by the equation

$$r = \pm \frac{0.4769}{h}.$$

When the errors really follow the law in question, they diminish with extreme rapidity as Δ increases. For example, only one per cent. of the errors should fall without the limits $\pm 4r$.

As a matter of fact, however, the cases are quite exceptional in which the errors are found to really follow the law. The general rule is that much more than one per cent. of the errors exceed four times the probable error. In other

words, it is nearly always found that some of the outstanding errors seem abnormally large. The method of dealing with these abnormal errors has always been one of the most difficult questions in the treatment of observations. The common practice has been to consider the observations affected by them as abnormal, and to reject them in obtaining the final result. But we here meet with the difficulty that no positive criterion for determining whether an observation should or should not be treated as abnormal is possible. Several attempts have indeed been made to formulate such a criterion, the best known of which is that of Peirce.*

Peirce's criterion has always seemed to me subject to two serious objections. One is that it takes no account of any probable error of the observations under consideration which may be known beforehand, but proceeds as if the value of the probable error could be deduced from the comparison of the observations *inter se*. An immediate general consequence of this is that if all the errors of a system are multiplied by the same factor, the same observations are rejected as before, how small or great soever the factor may be.†

The second objection is that it takes no account of the fact that the *a priori* probability that an observer should make an abnormal observation varies with the observer, and places all observers on a level by regarding that probability as determined by a general mathematical principle applicable to all cases.

It is, however, well known that some observers make very few abnormal observations, while others are extremely liable to them. It is evident that if we are dealing with an observation whose error is so large that we doubt whether it should or should not be considered abnormal, our judgment must depend very largely upon any knowledge we may have of the carefulness of the observer.

The fact is, however, that any system of rejecting supposed abnormal observations is subject to the objection of leading to a result which is a discontinuous function of the separate errors of observation, and hence to results

* Gould's *Astronomical Journal*, Vol. II, p. 161.

† Certain results of Peirce's criterion in special cases, when applied to sets of three or four observations, do not seem to have been hitherto noticed. The following are cases in point:

Of a set of three observations none are ever rejected by it, no matter how much one may deviate from the mean of the other two.

In a set of four observations, if three agree exactly, the fourth will always be rejected if it differs ever so little from the others. More generally, if no one of the three results which agree best among themselves differs from the mean of the three by more than ϵ , then a fourth, which differs from that mean by more than 4ϵ , will be rejected. For example, if the results of four observations with a meridian circle were $0''.3$; $0''.4$; $0''.5$; $0''.8$, the last would be rejected.

which are sometimes indeterminate. Suppose, for example, that we are dealing with the mean of three observations, two of which are closely accordant, while the third differs from the mean of the other two by the quantity x . Let us represent the mean of the two accordant ones by the symbol m' ; then, if we include the discordant observation, the general expression for the mean result in terms of x will be

$$m = m' + \frac{1}{3}x.$$

In ordinary astronomical practice we retain this value of m so long as x does not exceed the limit which we consider that of a normal error. But, as soon as this limit is reached, we drop x entirely and take m' for the value of m . In other words, if we consider x to increase from zero, the adopted value of m will increase one-third as fast until the assigned limit is reached, and will then suddenly spring back from $m' + \frac{1}{3}x$ to m' . If the critical point at which x is to be rejected could be satisfactorily defined, this course would be less objectionable. But, as a matter of fact, it is to be determined by the judgment of the investigator, with the result that between certain wide limits the investigator must himself be doubtful whether he should take m' or $m' + \frac{1}{3}x$ as his result. Of course different investigators would reach different conclusions in special cases, and thus the most probable result is frequently indeterminate.

There are classes of important observations in which the proportion of large errors is so great that no separation into normal and abnormal observations is possible. This is the case in observations of transits of Venus and Mercury over the sun. A noteworthy instance has been given by the writer in his discussion of transits of Mercury.* By a comparison of 684 observations it was found that the errors of one-half of them were contained between the limits $\pm 6''.8$. If the errors followed the commonly assumed law, then only 5 of them should have exceeded ± 27 seconds. As a matter of fact, however, it was found that 49 exceeded these limits. Yet these 49 observations cannot be considered as wholly worthless, because their results are not scattered entirely at random, and are mostly included between comparatively narrow limits. They differ from the other observations only in having a larger probable error.

The case may be made clearer by reflecting that the law in question presupposes that the observations under consideration are all of the same general quality as regards liability to error; in other words, that they are all liable to

* *Astronomical Papers of the American Ephemeris*, Vol. I, pp. 379-383.

the same errors, and differ only in the accidental circumstances which give rise to the errors. If, however, this is not true; if, for example, we are furnished with a system of observations of which one portion have a small probable error, another a larger probable error, a third a yet larger one, and so on, then the errors of the whole system will not follow the law in question, but we shall find that large errors are disproportionately frequent. Now, this must be the case in nearly all astronomical and physical work.

From this another conclusion follows. In such a mixed system of observations the most probable result will be, not the arithmetical mean, but a mean obtained by giving less weight to the more discordant observations. This will be evident on reflecting that in such a case the more discordant results will probably belong to the observations having a larger probable error and therefore the less weight.

§ 2. *Modified Curves of Probability.*

The preceding considerations lead us to the further conclusion that the commonly received theory which presupposes that there must always be some one "most probable value" of a quantity determined by observations, lacks generality. The fact is that, in special cases, owing to a possibility of abnormal observations, the curve of probability may have a great variety of forms. As one example, let it be supposed that two mean declinations of a star, determined with a good meridian circle the micrometer-head of which is numbered at intervals of $5''$, differ from each other by a quantity approximating to $5''$. We then may make three hypotheses: that the observations are both normal, or that one or the other of them is in error by $5''$ through a mistake in recording.

According to the probability of the first hypothesis, and of either of the other two, we may have different curves. Assuming the instrument and observer to be so accurate that a difference of $5''$ between two normal observations is nearly out of the question, we shall have a curve of the form A. As the probability of the first hypothesis increases, the curve may assume the form B. If the observer is one never known to make mistakes in reading, the curve will approximate to its usual form.

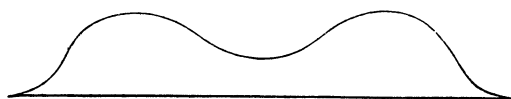


FIG. A.

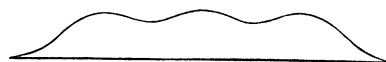


FIG. B.

Now, it is evident that in such a case as that indicated by curve A, we can assign no one most probable value of the observed quantity. The only complete statement we could make would be embodied in a table showing for each separate possible value of the required quantity the probability that the quantity had a value not differing from it by more than a small arbitrary amount. Assuming the intervals of the table to be taken to equal double this amount, the sum of all these probabilities would be unity.

Looking further into the matter, we see that this is a general method of expressing the conclusion to be derived in all cases from an observation or a series of observations. No matter how definite the primary value given by an observation may be, the actual conclusion to be drawn is always a series of separate probabilities that the quantity observed has some one of an infinite series of values. If the law of probability is that commonly assumed, then the probability of each assignable value is completely determined when the most probable value and the probable error are given. But such is not the case when the law of probability deviates from this form.

§ 3. *Evil and Worth of Erroneous Results.*

The question now arises whether, when we consider the most general case, in which there may be several maxima of probability, and when, therefore, no one most probable value can be assigned, it is possible to formulate any general principle by which a single value shall be preferred for acceptance above all others. Taking as an example such a case as A just given, it is clear that no such principle is possible without some antecedent hypothesis determining a law of choice between errors of different magnitudes. If, to fix the ideas, we suppose that in case A the results of the two separate observations were $0''.0$ and $5''.0$, then the three hypotheses will give us $0''.0$; $2''.5$; $5''.0$, as three values between which we are to choose. If we choose either the first or third, we shall have a probability of slightly less than one-half of being very near the truth, and an equal probability of being $5''$ in error, together with a very small probability of being about $2''.5$ in error. If, on the other hand, we take $2''.5$ as our result, we shall be almost sure of being between $2''$ and $3''$ in error, and no more. Our choice, then, must depend on whether a certainty of being $2''.5$ in error, or an even chance of each of the errors $0''$ and $5''$, is preferable. This again turns upon the question whether an error of $5''$ involves more or less than twice the evil of an error of $2''.5$.

The ordinary requirements of practical life are in favor of the view that the evil increases in a higher ratio than the simple magnitude of the error. As examples, if it is a case of an error in the position of a ship arising from an error of the Nautical Almanac, we readily see that the probability of the error leading to a shipwreck increases in a higher ratio than that of the simple error itself. Again, in the case in which, by the labor of continually increasing observations of a single quantity, we lessen the probable error, we know that it requires fourfold labor to reduce the probable error to one-half. It would seem, therefore, that the best hypothesis that we can adopt is that the evil of an error is proportional to the square of its magnitude.

A determination has more or less value according as it is less or more liable to errors. The simplest definition of the value of an observation that we can adopt is that it is inversely as the sum total of the evils to which it is subjected, each evil being multiplied by its probability. This also is in strict accord with the ordinary law of probability of a number of observations, since it involves the hypothesis that the value of a result is proportional to the number of observations on which it depends. As, however, the word "value," if used to express this conception, would be ambiguous in consequence of being applied to the simple amount of a quantity, we shall use the term *worth* to express the economic value of an assigned value as just described. We therefore have the definitions:

The *evil* of any value assigned to a quantity is equal to the sum of all the products obtained by multiplying the square of each possible error of that assigned value by the probability of its occurrence.*

The *worth* of any such value is inversely proportioned to its evil.

The value to which we are to give the preference is that whose worth is a maximum, or, which amounts to the same thing, whose evil is a minimum. This value, and the magnitude of the evil with which it is affected, will be two elements corresponding to "most probable value" and "probable error" in the usual theory.

§ 4. *Algebraic Expressions for the Evil.*

The general expression for these elements is readily obtained. Let us represent all possible values which the required quantity can have by the series

$$x_1, x_2, x_3, \dots x_n,$$

and let

$$p_1, p_2, p_3, \dots p_n$$

* This idea of an evil attached to each error, and proportional to the square of its magnitude, is due to Gauss (*Theoria Combinationis Observationum*, etc., Pars prior, §6), who applies the term *jactura* to the conception here called evil.

be the several probabilities of these values ; we necessarily have in this case

$$\sum_1^n p_i = 1.$$

Putting x for any value of the quantity in question, the evil of this value will be, by definition,

$$\begin{aligned} e &= p_1(x - x_1)^2 + p_2(x - x_2)^2 + \dots + p_n(x - x_n)^2 \\ &= x^2 - 2Ax + B, \end{aligned}$$

where

$$\begin{aligned} A &= \sum_1^n p_i x_i, \\ B &= \sum_1^n p_i x_i^2. \end{aligned}$$

This quantity is a minimum when

$$x = A,$$

which is, therefore, the value we are to prefer. The evil of this value is

$$e_0 = B - A^2.$$

The respective values of A^2 and B may be written, since $\Sigma p_i = 1$,

$$\begin{aligned} A^2 &= \sum_1^n \sum_{i,j}^2 p_i p_j x_i x_j, \\ B &= \sum_1^n \sum_{i,j}^2 p_i p_j x_i^2. \end{aligned}$$

The value of the minimum evil then becomes

$$e_0 = \frac{1}{2} \sum_1^n \sum_{i,j}^2 p_i p_j (x_i - x_j)^2.$$

It therefore appears that the inverse of this expression is the worth of the best value of the required quantity, which best value is given by the equation

$$x_0 = \sum_1^n p_i x_i. \quad (1)$$

In what precedes, the form of our equations is based upon the supposition that x_1, x_2, \dots, x_n are a finite number of discrete values which x may have. In the usual case, however, the unknown quantity may have all values between

certain wide limits, and the probability that it is contained between the limits x and $x + dx$ is given by an equation of the form

$$dp = \phi(x)dx,$$

dp being an infinitesimal probability. Since this is a pure number, it follows from this equation that, whatever physical quantity may be represented by x , $\phi(x)$ must be of the dimension -1 in this quantity.

Reducing the formula (1) to the present case, we find that the preferable value of x is given by the equation

$$x_0 = \int_{-\infty}^{+\infty} x\phi(x)dx. \quad (2)$$

The evil of this value is

$$\begin{aligned} e_0 &= \frac{1}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(y) \cdot \phi(z) (y - z)^2 dy dz \\ &= \int_{-\infty}^{+\infty} y^2 \phi y dy - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yz \phi(y) \phi(z) dy dz. \end{aligned} \quad (3)$$

From the definition of evil, it is a quantity of the dimensions $+2$ in those of the physical quantity in question. Its square root is therefore a definite quantity of the magnitude to be determined, which we may regard as an error.

An example of the results of the present theory will be found by applying it to the case of the usually assumed law of error, namely:

$$\phi(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}.$$

We then have $x_0 = 0$ as the preferable value of x , in accordance with the usual theory. In the expression for the evil of this value we have

$$\begin{aligned} A &= x_0 = 0, \\ B &= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{+\infty} y^2 e^{-h^2 y^2} dy = \frac{1}{2h^3}. \end{aligned} \quad (4)$$

This value of B is identical with the square of what is commonly called the *mean error*, which again is equal to $1.5 \times$ probable error nearly. Hence, in this special case, the evil is identical with the square of the mean error.

If, instead of taking zero as the value of x , we wish to express the evil of any other assumed value, we have the expression

$$e = x^2 + B = \varepsilon^2 + x^2,$$

ε being the mean error. If, instead of ε , we use r , the probable error, the expression will be

$$e = x^2 + 2.198r^2.$$

We readily find that if, instead of using the most probable value of a quantity, we adopt a value differing from it by its probable error, the evil will be increased by a little less than half its whole amount.

It appears, therefore, that, whatever the law of error, we may always find two quantities corresponding to "most probable value," and "probable error" of that value in the usual theory. One of these quantities will naturally be the best value of the required magnitude, or A itself. The other may be either the evil of the value A , or the change in the value of the magnitude required to increase its evil in a definite ratio, these last quantities being functions of the same quantity, and therefore of each other. If we present the result in the form

$$x = A \pm \sqrt{B - A^2},$$

the last term will be the "mean error" in the usual theory, and the change in x which would double its evil in the generalized theory. If we wish to express the quantity corresponding to the "probable error," we write

$$x = A \pm 0.6745\sqrt{B - A^2}.$$

§ 5. *Modified Law of Probability.*

The whole problem now before us is reduced to finding a curve of probability in the case of a number of observations of the same quantity. This problem naturally involves that of the law of error of the separate observations, and leads us to inquire what modification should be made in the usually assumed law in order that it may be applicable to all cases whatever.

The defect of the commonly assumed law, as represented by the equation

$$\phi(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2},$$

is that, in practice, large errors are more frequent than this equation would indicate them to be. This defect might be remedied by substituting some other function of x than $h^2 x^2$ as the exponent of e . The requirements of this function would be :

1. That it should be an even function of x , or of the form $f(x^2)$.
2. That it should become infinite when x did.
3. That it should increase less rapidly than $h^2 x^2$, or, more exactly, that the second derivative $\frac{\partial^2 f}{\partial x^2}$, instead of being a constant, should diminish with an increase

of x . Such a function can be formed by writing, instead of h^2 , an expression of the form

$$\frac{h^2(1+h^2x^2)}{1+h^2x^3},$$

so that we should have, for the exponent $f(x^2)$, an expression of the form

$$-\frac{h^3(x^2+h^2x^4)}{1+h^2x^3}.$$

The management of such an exponent might, however, prove inconvenient, and I shall adopt a law of error founded on the very probable hypothesis that we are dealing with a mixture of observations having various measures of precision.

Let us put

$$h_1, h_2 \dots h_n,$$

the possible separate values of the different measures of precision ;

$$p_1, p_2 \dots p_n,$$

the corresponding probabilities that an observation selected at random has any one of these several measures of precision.

Then, for an observation selected at random, the law of error will be

$$\phi(x) = \frac{1}{\sqrt{\pi}} \{ p_1 h_1 e^{-h_1^2 x^2} + p_2 h_2 e^{-h_2^2 x^2} + \dots + p_n h_n e^{-h_n^2 x^2} \}. \quad (5)$$

§ 6. *Deduction of Best Result.*

If we have m observations of the same quantity giving $x_1, x_2 \dots x_m$ as the observed values, then, assuming the law of error expressed in (5), the probability that the quantity is contained between the limits η and $\eta + d\eta$ is given by the equation

$$dp = \alpha \psi(\eta) d\eta,$$

in which $\psi(\eta) = \phi(x_1 - \eta) \phi(x_2 - \eta) \dots \phi(x_m - \eta),$ (6)

while α is a constant so chosen as to make

$$\alpha \int_{-\infty}^{+\infty} \psi(\eta) d\eta = 1.$$

The formula (2) then gives for the best value of x the expression

$$x = \int \eta dp = \alpha \int_{-\infty}^{+\infty} \eta \psi(\eta) d\eta = \frac{\int_{-\infty}^{+\infty} \eta \psi(\eta) d\eta}{\int_{-\infty}^{+\infty} \psi(\eta) d\eta}. \quad (7)$$

If $\psi(\eta)$ be multiplied by any constant factor, it will disappear from this expression ; we may therefore disregard all such factors in forming $\psi(\eta)$. We may

each mean being obtained by making an hypothesis respecting the distribution of the measures of precision $h_1, h_2 \dots h_n$ among the m separate observations. Since each observation may, independently of all the others, have any one of the n measures of precision, there will be n^m such hypotheses, each leading to a different mean, η . The final value of x is again a mean by weights of the results of the different hypotheses, the weight of each result being proportional to the probability of the hypothesis on which it depends, which is represented by w . This probability is a product of two factors, of which one, $\frac{P}{k}$, is proportional to the probability of the combination, while the other, $e^{\frac{b^2}{k^2} - c}$, is the probability of the combination of outstanding errors to which the hypothesis leads.

§ 7. *Application to an Example.*

Before showing how the preceding method may be simplified in practice, it may be of interest to give a simple numerical example of its rigorous application. Let it be granted that we have three observations of a class for which there is a probability of $\frac{2}{3}$ that an observation is good, and of $\frac{1}{3}$ that it is poor. Let the measure of precision of a good observation be 4, and of a poor one 1. Let the results of the three observations be

$$\text{I, 0; II, 0; III, 1.}$$

Since we have, *a priori*, no reason to distinguish between these results, the usual method of treatment would lead either to $\frac{1}{3}$ as the best result, or to the rejection of the third observation, and hence to the result 0.

From the point of view of the present paper, the agreement of the first two observations and the discordance of the third give color to the hypothesis that the first two observations are good and the third poor. On this hypothesis the best result would be $\frac{1}{3}$, the weights of the results being 16, 16 and 1. But, since every other hypothesis we can make would lead to a larger result, the best result must be greater than this.

The rigorous treatment of the problem gives

$$\phi(x) = \frac{1}{3\sqrt{\pi}} \left\{ 8e^{-16x^2} + e^{-x^2} \right\}.$$

Hence, when $x_1 = x_2 = 0$ and $x_3 = 1$,

$$\begin{aligned} \psi(\eta) = & 512e^{-48\eta^2 + 32\eta - 16} + 64e^{-33\eta^2 + 2\eta - 1} \\ & + 128e^{-33\eta^2 + 32\eta - 16} + 16e^{-18\eta^2 + 2\eta - 1} \\ & + 8e^{-18\eta^2 + 32\eta - 16} + e^{-3\eta^2 + 2\eta - 1}. \end{aligned}$$

These six terms correspond respectively to the six essentially different hypotheses which can be made respecting the distribution of the measures of precision among the different observations, it being observed that among the eight hypotheses there are two pairs such that the members of each pair lead to identical results. The results are tabulated as follows :

Hypothesis.	Results.	Probability.	Product.
	$\frac{b}{k^2}$	$\frac{P}{k} e^{\frac{b^2}{k^2} - c}$	
I, II and III all good,	$\frac{1}{3} = 0.3333$	0.003	0.001
I $\begin{cases} \text{good} \\ \text{bad} \end{cases}$; II $\begin{cases} \text{bad} \\ \text{good} \end{cases}$; III good,	$\frac{16}{33} = 0.4848$	0.006	0.003
I and II bad ; III good,	$\frac{8}{9} = 0.8889$	0.318	0.283
I and II good ; III bad,	$\frac{1}{33} = 0.0303$	4.224	0.128
I $\begin{cases} \text{good} \\ \text{bad} \end{cases}$; II $\begin{cases} \text{bad} \\ \text{good} \end{cases}$; III bad,	$\frac{1}{18} = 0.0555$	1.467	0.082
I, II and III all bad,	$\frac{1}{3} = 0.3333$	0.296	0.099
		$\Sigma = 6.314$	0.596

Hence, for the value of maximum worth, we have

$$x = 0.0944.$$

§ 8. *Modification of the Method when the Observations are Numerous.*

In order to apply the preceding method, it is necessary to know the respective probabilities that the measure of precision of any one observation has the several values h_1, h_2, \dots, h_n . These probabilities are determined from the actual distribution of the residuals with respect to magnitude as found by the study of large masses of observations. If it were found that in any class of observations the magnitudes of the residuals followed the commonly assumed law, we should have but one value of h . If, as will commonly be the case, we find a larger number of large residuals than would be given by the common theory, we assume one, two, three or more additional values of h , and determine how many observations we must assign to each class in order that the distribution may be represented by an equation of the form (5).

To carry out the rigorous process of finding the best mean value of x , we should form, by the equations (8) and (11), n^m different values of P, k, b, c, η and w , and thence, from (12), the required value of x .

To effect this, let us attach, or suppose to be attached, to each of the quantities P, k , etc., a system of m indices, i, j, k , etc., each index taking the values $1, 2 \dots n$. The system of indices

$$i, j, k \dots q$$

attached to a quantity will then indicate the special value of that quantity which results from assigning

$$\begin{array}{lll} \text{to } x_1 & \text{the precision} & h_i, \\ \text{to } x_2 & \text{"} & h_j, \\ \text{to } x_3 & \text{"} & h_k, \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ \text{to } x_m & \text{"} & h_q. \end{array}$$

Moreover, we shall, for brevity, represent the combination of indices

$$(i, j, k \dots q)$$

by the single symbol t .

Any one value of w in (11) may then be written in the form

$$w_t = \frac{P_t}{k_t} e^{-\frac{\Delta_t}{k_t^2}},$$

where we put, for brevity,

$$\Delta = k^2 c - b^2.$$

If we here substitute for k^2, b and c their values from (8), we find that this expression reduces to

$$\Delta = \sum_{i,j}^2 h_i^2 h_j^2 (x_i - x_j)^2, \\ (i = 1, 2, 3 \dots m-1; j = i+1, i+2 \dots m),$$

where h_i for the moment indicates the special value of h which, in any combination, is assigned to x_i . We may equally represent Δ in the form

$$\frac{1}{2} \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} h_i^2 h_j^2 (x_i - x_j)^2.$$

The first form will consist of $\frac{m(m-1)}{2}$ terms; the second of m^2 terms, if we count those which vanish through $i=j$.

Since Δ depends only on the differences between the x 's, it will remain unchanged when we subtract any constant from all of them. Let us then subtract η from all of them, putting

$$\xi_i = x_i - \eta;$$

would all be equal. But, in the case for which the preceding theory is constructed, each result x_i may have any one of the weights h_1^2, h_2^2 , etc.; and the equation (14) determines a certain mean among these weights which we are to assign to x_i . The coefficients W are functions of ξ , and admit of being tabulated as such in any special case.

In what precedes we have presupposed no difference of weights among the results $x_1 \dots x_m$ to be known *a priori*. But, since each observation may have any one of the weights $h_1 \dots h_n$, a certain mean weight w of each is determined *a posteriori*, as a function of its deviation from the general mean, by the equations (14). This mean weight can be tabulated as a function of ξ , and thus taken out from a table with a single argument.

If, however, we have some knowledge of an observation which leads us to assign it one precision rather than another, we may utilize this knowledge so as to modify the values of $w_j^{(i)}$. If h_1, h_2 , etc., are taken in the descending order of magnitude, then h_1^2 will be the weight of each observation of the best class. The theoretically best mode of dealing with such cases will, however, depend upon the circumstances of that case. Simplicity is so important an advantage that it will probably be found well to adopt the rule of replacing W by its product by the weight fixed from *a priori* considerations.

§ 9. *Application to Transits of Mercury over the Sun's Disc.*

I now propose to apply the preceding theory to the case of observed contacts of the inferior planets, Mercury and Venus, with the limb of the Sun. A peculiarity of the observations of these phenomena is the great number of them which investigators have had to reject on account of discordance of individual observations from the general mean. By suitable rejection very different final results may be obtained, and it is impossible to draw any line between those observations which should be rejected and those which should be retained.

In my discussion of observations of transits of Mercury,* I have shown that the residuals of 684 observations of the interior contact of the limbs of the Sun and Mercury are distributed as follows, the value of each residual being considered only to the nearest round 5 seconds:

*Astronomical Papers of the American Ephemeris, Vol. I.

Below — 27 seconds were 20 residuals.

of — 27, — 26, — 25, — 24, — 23	were	11	“
“ — 22, — 21, — 20, — 19, — 18	“	15	“
“ — 17, — 16, — 15, — 14, — 13	“	44	“
“ — 12, — 11, — 10, — 9, — 8	“	77	“
“ — 7, — 6, — 5, — 4, — 3	“	132	“
“ — 2, — 1, — 0, + 1, + 2	“	147	“
“ + 3, + 4, + 5, + 6, + 7	“	89	“
“ + 8, + 9, + 10, + 11, + 12	“	52	“
“ + 13, + 14, + 15, + 16, + 17	“	33	“
“ + 18, + 19, + 20, + 21, + 22	“	23	“
“ + 23, + 24, + 25, + 26, + 27	“	12	“

Exceeding + 27 “ 29 “

Collecting those of equal absolute value and classifying them according to the middle value of each group, we have the following comparison of actual and probable numbers, the latter being obtained on the usual theory, assuming a probable error of ± 6.67 , or a value of $\frac{1}{h} = 14.0$.

Mean Values of Residual.	Actual Number.	Probable Number.	A — P.
0	147	137	+ 10
5	221	240	— 19
10	129	166	— 37
15	77	88	— 11
20	38	36	+ 2
25	23	12	+ 11
> 27	49	5	+ 44.

There is, therefore, a large excess of both small and large residuals, which would have been yet more pronounced had the mean error been determined from the sum of the squares of all the residuals.

I find, by several trials, that the residuals which do not exceed 27 can be well represented by the following distribution of precisions and probable errors :

110 observations of $1:h = 6$	or probable error = 2.9,
100 “ “ “ 10 “ “ “	4.8,
400 “ “ “ 18 “ “ “	8.6,
50 “ “ “ 36 “ “ “	17.2.

The comparison of actual and probable numbers of residuals will then be as follows :

s.	Actual.	Probable.	$A - P.$
0	147	143	+ 4
5	221	220	+ 1
10	129	128	+ 1
15	77	76	+ 1
20	38	44	— 6
25	23	23	0
> 27	49	26	+ 23.

It must be understood that these four values of $1:h$ and of the consequent probable errors are not four entirely determinate quantities. Really we should consider that the precision has all values between the extreme limits; but it is not at all necessary to consider it as what it really is, a continuously varying quantity. All we have to do is to form an expression which shall represent the relation between the number and magnitude of the residuals; and this we do most conveniently by assuming three, four or more separate values of h , and then finding how many observations we have to assign to each class in order to represent the observed relation. From the above table we may infer that about one-third the observations of transits of Mercury belong to classes which might be called good or very good, the probable error ranging from $2\frac{1}{2}$ to 6 ; that more than half belong to an average class, of which the probable error may range from 6 to 12 seconds, and that about one-twelfth are made under such unfavorable circumstances that their probable error averages 17 . Even with this large probable error, we see that there is an excess of 23 residuals exceeding 27 seconds, so that we should have increased the number of this imperfect class. But I suspect that many of these arose from errors of a minute in the record, or from other pure blunders.

I am also inclined to think that the comparative excess of very small residuals, indicating that one-fifth of all the observations had a probable error as small as 3 , may be partly due to the fact that many of the residuals are deviations from the mean of a small number of observations, and that no comparison of the separate observations with the final theory founded on the whole mass of observations was made. On the whole, we may suppose that of the actual observations,

0.30 have a precision $h_1 = 1:10$,
 0.60 “ “ “ $h_2 = 1:18$,
 0.10 “ “ “ $h_3 = 1:36$.

This will give for the law of probability (*Eq. 5*):

$$\phi(x) = \frac{1}{\sqrt{\pi}} \left\{ .030e^{-\left(\frac{x}{10}\right)^2} + .0333 \dots e^{-\left(\frac{x}{18}\right)^2} + .00277 \dots e^{-\left(\frac{x}{36}\right)^2} \right\}.$$

We have, for the values of w_1 , w_2 and w_3 , as given by (13),

$$w_1 = .03000e^{-\left(\frac{\xi}{10}\right)^2}; h_1^2 = 1: 100,$$

$$w_2 = .03333e^{-\left(\frac{\xi}{18}\right)^2}; h_2^2 = 1: 324,$$

$$w_3 = .00277e^{-\left(\frac{\xi}{36}\right)^2}; h_3^2 = 1: 1296.$$

In the following table the quantities required are tabulated as a function of the residual ξ of an observation. The w 's are multiplied by 1,000, and the W 's by 10,000, so as to express them in convenient units:

ξ s.	w_1	w_2	w_3	W
0	30.0	33.3	2.8	6.1
2	28.8	32.9	2.8	6.1
4	25.6	31.7	2.7	5.9
6	20.9	29.8	2.7	5.7
8	15.8	27.4	2.6	5.3
10	11.0	24.5	2.6	4.9
12	7.1	21.4	2.5	4.5
14	4.2	18.2	2.4	4.0
16	2.3	15.1	2.3	3.6
18	1.2	12.3	2.2	3.3
20	0.55	9.70	2.04	3.0
22	0.23	7.49	1.91	2.8
24	0.09	5.63	1.78	2.6
26	0.03	4.14	1.65	2.5
28	0.01	2.96	1.52	2.3
30	0.00	2.07	1.39	2.2

If we could be sure that any one observation belonged to the best class, its weight on the above scale would be 10; were we sure it belonged to the intermediate class, its weight would be 3, and if it belonged to the worst class, it would be 0.77. The value of W for $\xi = 0$, namely, 6.1, falls below 10 in consequence of the probability that an observation of residual zero may belong to one of the poorer classes, and the value of W for an observation of residual

^{s.}
30 is above 0.77 on account of the possibility that such an observation may belong to the intermediate class.

§ 10. *Approximation Expression for the Evil.*

It remains to find an expression for the evil of the best result, as obtained by the preceding method. As already defined and shown, if the probability that the value of the observed quantity is contained between the limits x and $x + dx$ be

$$\theta x dx,$$

then the evil of any assumed value x_0 of the required quantity is given by the equation

$$E = \int_{-\infty}^{+\infty} (x - x_0)^2 \theta x dx.$$

In the case of m observed values of x , x_1, x_2, \dots, x_m , following the general law of error, we have

$$\theta x = \alpha \phi(x_1 - x) \phi(x_2 - x) \dots \phi(x_m - x),$$

the coefficient α being a constant, determined by the condition

$$\int_{-\infty}^{+\infty} \theta x dx = 1.$$

If we take for x_0 the best value of x , namely, the value which satisfies the condition

$$x_0 = \int_{-\infty}^{+\infty} x \theta x dx,$$

we have for its evil

$$E = \int_{-\infty}^{+\infty} x^2 \theta x dx - x_0^2.$$

It will be noticed that the function θx differs from $\psi(x)$ in (6) only in containing the factor α ; that is, we have

$$\theta x = \alpha \psi(x) = \alpha \Sigma P e^{-k^2 x^2 + 2bx - c}.$$

In consequence of α we may omit from ψx any constant factor, as $\sqrt{\pi}$. Then from

$$(9) \quad \psi(\eta) = \Sigma P e^{-k^2 \eta^2 + 2b\eta - c}$$

we have

$$\int_{-\infty}^{+\infty} x^2 \psi(x) dx = \Sigma P \frac{k^2 + 2b^2}{2k^5} e^{\frac{b^2}{k^2} - c}$$

α is determined by the condition

$$\alpha \int_{-\infty}^{+\infty} \psi(x) dx = 1 = \alpha \Sigma w.$$

Comparing with the equations (9) and (11), we find

$$\int_{-\infty}^{+\infty} x^2 \theta x dx = \frac{\Sigma w \eta^2}{\Sigma w} + \frac{\frac{1}{2} \Sigma \frac{w}{k^2}}{\Sigma w},$$

the sign Σ extending to the n^m distributions of the measures of precision. We thus have, for the minimum evil,

$$E = \frac{\Sigma w \eta^2}{\Sigma w} - \frac{(\Sigma w \eta)^2}{(\Sigma w)^2} + \frac{\frac{1}{2} \Sigma \frac{w}{k^2}}{\Sigma w} = \frac{\Sigma_{ij}^2 w_i w_j (\eta_i - \eta_j)^2}{(\Sigma w)^2} + \frac{\Sigma \frac{w}{2k^2}}{\Sigma w} \\ (i = 1, 2 \dots n^m - 1; j = i + 1, \dots n^m) \quad (15)$$

The second term of the right-hand member of this equation is a certain mean among the various values of $\frac{1}{2k^2}$, and coincides with the square of the "mean error" of the usual theory, which, as already shown, coincides with the "evil," as that term is here defined. If there is but one distribution of the k 's among the x 's, then there will be but one value of η , and the first term of the evil will vanish, so that we shall have no evil left except the "mean error." But when, as here supposed, the weights of the observations are themselves uncertain, then the last equation expresses the logical conclusion that, in order to obtain the total evil, we must add to the result of the mean uncertainty of the observations a quantity depending upon the uncertainty in the weights we should individually or collectively assign to them.

The first term of (15) comprises $\frac{\mu(\mu-1)}{2}$ terms ($\mu = n^m$); its actual computation is therefore out of the question. We may, however, remark that it admits of being expressed in the form

$$\Sigma \alpha_{i,j} (x_i - x_j)^2,$$

$\alpha_{i,j}$ being numerical coefficients. This form contains only $\frac{m(m-1)}{2}$ terms.

To show this, we remark that each value of η may be written in the form

$$\eta = p_1 x_1 + p_2 x_2 + \dots + p_m x_m,$$

where

$$p_1 + p_2 + p_3 + \dots + p_m = 1.$$

The difference of any two values of η multiplied by any factor, such as $\frac{\sqrt{w_i w_j}}{(\Sigma w)}$,

will therefore be of the form

$$\begin{aligned} & \mu_1 x_1 + \mu_2 x_2 + \dots + \mu_m x_m, \\ \text{where} \quad & \mu_1 + \mu_2 + \mu_3 + \dots + \mu_m = 0, \end{aligned} \quad (16)$$

and where each of the coefficients μ_1, μ_2 , etc., takes $\frac{n^m(n^m-1)}{2}$ values, corresponding to the number of differences of the η 's. The sum of the squares of these differences will be

$$\begin{aligned} & x_1^2 \Sigma \mu_1^2 + x_2^2 \Sigma \mu_2^2 + \dots + x_m^2 \Sigma \mu_m^2 \\ & + 2x_1 x_2 \Sigma \mu_1 \mu_2 + 2x_1 x_3 \Sigma \mu_1 \mu_3 + 2x_2 x_3 \Sigma \mu_2 \mu_3 + \text{etc.} \end{aligned}$$

On account of the condition (16) this expression may be transformed into

$$\begin{aligned} & \Sigma^2 A_{ij} (x_i - x_j)^2, \\ \text{where} \quad & A_{i,j} = -\Sigma \mu_i \mu_j. \end{aligned}$$

Here the sign Σ of summation extends to all the $\frac{n^m(n^m-1)}{2}$ products of μ having the same constant suffixes i and j .

Although the value of this expression admits of rigorous algebraic determination, it is doubtful whether there would be any advantage in computing it in any special case. I shall therefore seek only for a rough approximation to its probable value. Returning to the equation (15), we first note that the expression

$$\frac{\Sigma^2 w_i w_j (\eta_i - \eta_j)^2}{(\Sigma w_i)^2} \quad (i = 1, 2, \dots, n^m-1; j = i + 1, \dots, n^m)$$

is one-half a weighted mean value of $(\eta_i - \eta_j)^2$, the weight of each being the product $w_i w_j$, and the zero terms $(\eta_i - \eta_i)^2$ being allowed to enter with half weight in taking the mean. Instead of this weighted mean we may take the general mean formed by giving all the differences equal weight.

Now, when the number of observations treated is large, we may consider the amount by which any one value of η differs from the mean of all the values, or x , to be the result of an accumulation of accidental errors; namely, if we put

$$\frac{h_i^2}{k^2} \equiv q_i \quad (17)$$

(where h_i means the precision assigned to x_i), we shall have

$$\eta = q_1 x_1 + q_2 x_2 + \dots + q_m x_m,$$

while we have for x an expression of the form

$$x = q'_1 x_1 + q'_2 x_2 + \dots + q'_m x_m,$$

q'_i being the weighted mean of all the n^m values of q_i . If we now put, as before, ξ_i for the deviation of x_i from the general mean x , we have, on account of

$$\Sigma q_i = \Sigma q'_i = 1, \\ \eta - x = (q_1 - q'_1) \xi_1 + (q_2 - q'_2) \xi_2 + \dots + (q_m - q'_m) \xi_m.$$

Now, since k^2 is, in all cases, the sum of *some* m values of h^2 , the mean value of q given by (17) is $\frac{1}{m}$. The actual special values can never reach zero as their lower limit, and will seldom exceed $\frac{2}{m}$ as their upper limit. The range of value will, however, depend upon the range of values taken by the precisions h . Unless in extreme cases, the mean deviation of q from $\frac{1}{m}$ cannot exceed $\frac{1}{2m}$; that is, the mean value of $q - q'$ will, in general, be less than $\frac{1}{2m}$. Assigning this mean value, we may regard $\eta - x$ as made up of the probable accumulation of terms

$$\frac{1}{2m} (\pm \xi_1 \pm \xi_2 \pm \xi_3 \pm \dots \pm \xi_m).$$

If we put Δ^2 for the mean value of ξ^2 , then, by the theory of errors, we shall have, for the mean value of $(\eta - x)^2$,

$$\text{mean } (\eta - x)^2 = \frac{\Delta^2}{4m}.$$

Hence
$$\text{mean } (\eta_i - \eta_j)^2 = \frac{\Delta^2}{2m} = \frac{\Sigma \xi^2}{2m^2}.$$

To compare this with the last term of (15), let us suppose the most probable distribution of precisions with respect to their magnitude to be

$$\begin{array}{llll} m_1 & \text{precisions of value} & h_1; \\ m_2 & \text{"} & h_2; \\ \dots & \dots & \dots \\ m_n & \text{"} & h_n. \end{array}$$

We shall then have as a close approximation to the mean value of k^2 ,

$$k^2 = m_1 h_1^2 + m_2 h_2^2 + \dots + m_n h_n^2,$$

one-half the reciprocal of which may be taken for the last term of (15). Thus we may take for the amount of the evil, in all ordinary cases

$$E = \frac{1}{2} \frac{1}{m_1 h_1^2 + \dots + m_n h_n^2} + \frac{\Delta^2}{2m}.$$

As already shown, this evil will be the square of the "mean error" to be expected of the usual theory; so that we may take

$$\varepsilon = \pm \sqrt{E}$$

as the mean error to be expected.

I remark, in conclusion, that this theory and method may be extended to the case of several unknown quantities without any other difficulty than that of a resolution of the equations of condition with the *a posteriori* weights. We should first solve the equations if necessary, using equal weights for all, or such system of weights as might be deemed most probable. From the residuals thus obtained we should deduce the law of error, and in practice we should, in order to determine such law, combine with the residuals in question all others that astronomy could furnish pertaining to the same class of observations. Then we should re-solve the equations using the modified weights, which re-solution would give us the definitive result.