PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Law of Error and the Combination of Observations

Harold Jeffreys

Phil. Trans. R. Soc. Lond. A 1938 **237**, 231-271
doi: 10.1098/rsta.1938.0008

THE LAW OF ERROR AND THE COMBINATION OF OBSERVATIONS

By Harold Jeffreys, F.R.S.

(*Received* 18 *December* 1937)

1. The normal or Gaussian law of error rests partly on a particular hypothesis about the nature of error, that the error of any individual observation is the resultant of a large number of comparable and independent components; and partly on comparison with frequencies in actual series of observations. Both arguments are defective. The theory does not prove that the law is true for errors of any magnitude, even if its fundamental hypothesis is accepted. It involves a number of approximations, and when the effects of these are examined it is found that the law should hold only up to a moderate multiple of the standard error. This is obvious in the simplest case of the binomial distribution, since the normal law predicts a finite though small chance of an error of any amount, but the binomial law is rigidly limited at each end. The assertion of the normal law for errors of any magnitude as the limit of the binomial really assumes that the number of components is infinite, and that all have the same infinitesimal range. This is a very remarkable hypothesis. There may be some inductive reason, in cases where many sources of variation have already been determined and allowed for, to suppose that there are several others just below the magnitude that can be detected separately; but there is every reason to suppose that these are not infinite in number, but merely the largest members of a convergent series, and in such a case there is no reason to suppose that their resultant will tend to the normal law. Whittaker and Robinson give a striking example to the contrary (1924, p. 178), where the components all follow the median law and their magnitudes diminish. Indeed, if there is one dominant component the law for the resultant will approximate to the law for that component separately. Even if there are several equal components the proof may fail. Fisher has shown (1922, p. 321) that if several errors follow the law

$$P(dx \mid h) \propto (1+x^2)^{-1}\,dx,$$

the mean of any number will follow the same law with the same constants. The sum of m will therefore also follow this law except that the scale is altered in the ratio 1 to m. If we use the range between the inflexions as a standard of scale, it increases in proportion to m, not to $m^{\frac{1}{2}}$ as in the composition of normal errors, and the form of the law does not approach the normal at all as the number of components increases. The theory can therefore be regarded only as an indication that in suitable circumstances the normal law may be a good approximation up to a few times the standard error; but in the very conditions where this approximation is likely to be at its best there is no reason to assert the law for all values of the error. Even in the case where the components are equal and independent there may be no approach to the normal, as the above example shows. This is because the transformations involved in the proof are possible only subject to a hypothesis about the rate of convergence of the series expansion of the moment-generating function, which is violated by the above form.

K. Pearson showed long ago (1900) that some series of residuals published in support of the law showed as great departures from it as would warrant the rejection of any scientific hypothesis; in a later paper (1902) he made a study of actual observations by himself and three collaborators, and again found substantial departures. The law continues, however, to be generally applied. Even some of those who explicitly reject it appear to regard it as an indication of special virtue to retain all observations at equal weight and to take their mean as the best value; they often, however, take the average error without regard to sign or some other substitute in place of the standard deviation as their estimate of accuracy. But the three hypotheses (1) that the normal law is correct, (2) that the mean is the best value, (3) that the mean square residual provides the best standard of accuracy, are all equivalent; any one of them implies the other two. If the normal law is wrong, it is also wrong to use the arithmetic mean; if the arithmetic mean is the best summary value to adopt, its uncertainty is correctly estimated from the mean square residual and from no other statistic. The use of the average error without regard to sign is sometimes recommended because it is less influenced by large outstanding residuals than the standard deviation is; but this involves a confusion. Subject to a certain condition of convergence, if we adopt the mean as our estimate, then whether the normal law is true or not the uncertainty of the true value, given the mean, is found from the standard deviation and nothing else. If the average residual is used instead, there will usually be little difference beyond a slight increase of the uncertainty if the normal law is true; but if the normal law is true and there happen, in a particular case, to be a few exceptionally large residuals, or if it is untrue and there are residuals beyond the range predicted by it, the average error will lead to an under-estimate of the uncertainty of the true value as estimated from the mean. In fact, the case where the use of the average error is recommended is precisely the one where its use is most undesirable. The large residuals, in this treatment, have already received full weight in estimating the mean, and if they have led to error in the mean they must also be given full weight in estimating the standard deviation.

If the normal law is wrong, on the other hand, neither the mean nor the standard deviation is the best estimate to use in finding the true value and its uncertainty, and we should try to find out in what respects it is wrong and correct the postulate of the arithmetic mean accordingly. The rejection of the postulate of the mean is not the same as the rejection of observations. The latter means that every observation is either retained at full weight and used to form the mean, or else absolutely rejected. The decision with respect to a single doubtful outlying observation may easily affect the position of the mean by its standard error, and such an effect of the position of the limit of retention may fairly be described as intolerable. If we adapt our discussion to any continuous law of error, on the other hand, the weight will be a continuous function of the residual, and no great change will be made by a slight alteration in the parameters in that law. In discussing seismological observations I have found that the data lead to a satisfactory determination of the law of error (1936), which departs widely from the normal, but when this is used to estimate weights and the data are summarized accordingly the agreement between different sets of estimates is very good (1937*d*). In most physical observations the departure from the normal law is not so great as it was in these seismological ones, but it is appreciable. It is often remarked that large residuals occur oftener than the normal law would indicate, and the actual excess may be larger than is recorded, since a "bad observation" may be rejected at sight without being recorded at all. There is, however, no hard and fast line between a good observation and a bad one. With a correct system of weights these outlying observations would receive low weight and it would make little difference whether they were retained or not; but the uncertainty found will be genuine, which is not the case when the arithmetic mean of all observed values within certain arbitrary limits is the only estimate considered. Even if, as may well be true, the normal law is correct up to a certain multiple of the standard error, and this range includes most of the probability distribution, this does not justify the use of equal weights except within that range. If the law, with respect to the unknown a, is that the chance of an observation in a particular range dx is $f(x-a)\,dx$, the correct equation to estimate a is

$$\Sigma \frac{f'(x-a)}{f(x-a)} = 0. \tag{1}$$

This is equivalent to the rule of the arithmetic mean

$$\Sigma(x-a) = 0 \tag{2}$$

only if the law of error is normal at all values of x whatever. If the chance falls off with x less rapidly at the tails than the normal law suggests, the terms in (1) for large $x-a$ are smaller in comparison with those in (2) than for moderate $x-a$, and this is equivalent to giving reduced weight to the large residuals. If it falls off more rapidly, the weights must be increased. Cases of both types are known. For the former, we have the case given by Fisher, which would be realized if lines, equally likely to lie in any direction, were drawn through a point in a plane, and we had to estimate the position of the point, using as data the positions of the intersections of the lines with a known line. In this case the mean is no more accurate than one observation, but a suitable method of fitting will give high accuracy. Fisher (1922, pp. 348–51) gives the rectangular distribution as an instance of the second type; that is, the case where $f(x)$ is uniform between certain values of x, and zero outside those values. Here the extreme observations contain the whole of the relevant information in the sample. The mean square error of the mean of the two extreme observations decreases with the number of observations like n^{-1}, that of the mean of all n only like $n^{-\frac{1}{2}}$. It has been argued that with any symmetrical law of error the mean is valid, because it will be expected to tend to the true value when n is large; but this is not necessarily the case, and even if it is the case there may be other functions of the observations that will tend to the true value more rapidly, and should be used if the best use is to be made of a finite number of observations.

2. The testing of the normal law and the determination of corrections to it requires the use of long and homogeneous series of observations. These are difficult to find. The published data that are capable of providing comparisons fall into four main classes, which must be carefully distinguished:

(1) Genuine binomial distributions. In these the departure is the resultant of several strictly equal variations, all independent and each capable of only two values, which are the same for all components.

(2) Distributions resembling binomial ones except that the components are not strictly equal and may not all follow the same law; but it is known that they are numerous and comparable in amount.

(3) Real variations of a magnitude, to which error of observation makes a minor or negligible contribution.

(4) Residuals whose variation is not accounted for by any known cause except human inaccuracy.

Evidence derived from one of these types of data cannot be applied to another type without special reason. If binomial distributions are found to agree with the normal law, that affords no ground for applying the law to human inaccuracies unless these also are made up in the binomial way, which there is no reason to believe.

(1) With regard to the binomial distribution, which approaches the normal asymptotically when the number of components is large, it appears to be supposed in some presentations that the derivation of the normal law in this way requires the components to be infinite in number and each infinitesimal in amount. This is not the case. The approximation is very good for quite small numbers of components. Thus for three component departures ± 1 from the mean, the standard error σ is $\sqrt{3}$, and the following are the calculated expectations according to the binomial law and according to the normal law with the same value of σ, for eight observations in all. To make the data comparable the observations with the normal law are supposed rounded to the nearest odd number:

	< -4	-3	-1	$+1$	$+3$	$> +4$
Binomial	0	1	3	3	1	0
Normal	0·084	0·908	3·008	3·008	0·908	0·084

For four components and sixteen observations the expectations in ranges about the even numbers are as follows:

	< -5	-4	-2	0	$+2$	$+4$	$> +5$
Binomial	0	1	4	6	4	1	0
Normal	0·10	0·97	3·86	6·13	3·86	0·97	0·10

In neither case do the probabilities of one observation falling in a particular range differ by more than 0·012. To see what this means with regard to testing the normal law, let us suppose that n observations are made and that we are given only the totals by ranges to compare by the χ^2 test with the postulate that they are derived from the normal law; and that in fact they are derived from a binomial law with three components. According to the usual rule for applying the test when some of the expectations are small, the terminal groups cannot be taken separately until the expectations in them, according to the hypothesis to be tested, reach 5. Until this happens they will be combined with the adjacent groups before χ^2 is evaluated. Thus they will not appear separately in the test until n reaches $8\times 5/0{\cdot}084$ = about 480. Until then the expectations, for given n, will be

	<-2	-1	$+1$	$>+2$
Binomial	$0{\cdot}125n$	$0{\cdot}375n$	$0{\cdot}375n$	$0{\cdot}125n$
Normal	$0{\cdot}124n$	$0{\cdot}376n$	$0{\cdot}376n$	$0{\cdot}124n$

Now if the actual frequencies agreed exactly with the binomial expectations we should compute χ^2 from the usual formula for frequencies

$$\chi^2 = \Sigma \frac{(m_r - n_r)^2}{n_r},$$

where m_r is the observed number and n_r the expectation in each group. Thus we should have

$$\chi^2 = 0{\cdot}001^2 n^2\left(\frac{2}{0{\cdot}124n}+\frac{2}{0{\cdot}376n}\right) = 2\times 10^{-5}n$$

roughly; and the contribution to it from the departure, for $n\leqslant 480$, cannot exceed 0·01. There is no possibility, therefore, of detecting the difference with less than 480 observations if the test is applied in the usual way.

For more than 480 observations the expectations in the terminal groups, according to the binomial law, will exceed 5 and they can be taken separately. The expectations will be just $n/8$ times the values for $n = 8$. The contribution to χ^2 will therefore be

$$n\left\{2\times\frac{0{\cdot}0105^2}{0{\cdot}0105}+2\times\frac{0{\cdot}0115^2}{0{\cdot}1135}+2\times\frac{0{\cdot}001^2}{0{\cdot}376}\right\} = 0{\cdot}0233n,$$

of which $0{\cdot}0210n$ comes from the terminal groups.

If the mean and standard error are given there are 5 degrees of freedom, and the 5 % limit is at $\chi^2 = 11{\cdot}07$. The expectation of the contributions of random variations to χ^2 is 5; hence if we are to have any reasonable chance of detecting a systematic departure it must contribute about 6 to χ^2. This means that n must be about 260. Thus the test would detect the departure as soon as it became applicable to the terminal groups, but not before. It would take 480 observations to show by this method that the normal law did not hold, even though the number of component errors is as small as 3, and the evidence would not come from discrepancies within the range where observations exist but from their absence where we should expect them.

In practice, if we were given the individual observations for such a series, we should discard the normal law on account of the fact that the observations are all separated by integral multiples of a constant. This would suggest a variation of the binomial type. But if instead of three sources of error, each capable of giving only two values, we had three each implying a uniform distribution of the chance of error between two fixed limits, the observations would not reveal their discreteness in this way, and also the difference between the true law and the normal would be less, and would take more observations to reveal it.

This can be seen in a still more extreme case by considering the triangular distribution of error

$$P(dx) = \tfrac{1}{2}(1-\tfrac{1}{2}x)\,dx \quad 0<x<2,$$

$$P(dx) = \tfrac{1}{2}(1+\tfrac{1}{2}x)\,dx \quad -2<x<0,$$

which has standard error $\sqrt{\tfrac{2}{3}}$. With this form the probabilities in ranges grouped about the integers, and the corresponding ones for the normal law with the same second moment are

	−3	−2	−1	0	1	2	3
Triangular	0	0·03125	0·2500	0·4375	0·2500	0·03125	0
Normal	0·0011	0·0320	0·2370	0·4597	0·2370	0·0320	0·0011

If we are testing normality, and the actual distribution is triangular, the end groups will not come into consideration till there are about 4500 observations. Combining them with the ranges centred on ± 2, we have the contributions to χ^2

$$n\left[2\times\frac{0{\cdot}00185^2}{0{\cdot}0331}+2\times\frac{0{\cdot}0130^2}{0{\cdot}2370}+\frac{0{\cdot}0222^2}{0{\cdot}4597}\right] = 0{\cdot}0027n.$$

As before, the systematic departure would have to contribute about 6 to χ^2 before it would have any appreciable chance of being detected. Thus about 2200 observations would be needed. But the triangular distribution is the result of combining only two uniform distributions between fixed limits. Lest it should be thought that the laws are so similar that we should be entitled to combine observations according to them in the same way, we notice that the equation for determining the mode, if the law of chance is $f(x-a)\,dx$, is

$$\Sigma\,\frac{f'(x-a)}{f(x-a)} = 0$$

taken over the observations. But $f'(x-a)$ is constant on each side, while $f(x-a)$ diminishes steadily to zero at ± 2. The most important terms therefore come from the extreme observations. Indeed, if we simply took the mean of the extreme observations by themselves we should get almost as good an estimate as by taking the mean of all, though neither makes full use of the data. To adapt a method given by Fisher (1922,

p. 348), the chance of an observation within a small distance y of 2 is $\frac{1}{8}y^2$. The probability that no observation out of n lies in this range is $(1-\frac{1}{8}y^2)^n = \exp(-\frac{1}{8}ny^2)$ nearly. The probability that the extreme one of n lies between y and $y+dy$ is therefore the differential of this, or $\frac{1}{4}nydy\exp(-\frac{1}{8}ny^2)\,dy$. Then we find easily that the expectation of the distance of the end observation from 2 is $\surd(2\pi/n)^{\frac{1}{2}}$, and that of the square of the distance is $8/n$. If then we simply go $\pm(2\pi/n)^{\frac{1}{2}}$ beyond the extreme observation at each end, we fix the positions of the ends with standard error $1{\cdot}31/\surd n$. The mean of the results will have standard error $0{\cdot}926/\surd n$. But the mean of all the n observations will have standard error $\sigma/\surd n = 0{\cdot}816/\surd n$. It is impossible to give a general formula for the standard error when the whole of the data are used; in any particular case it will be given by

$$\sigma_a^{-2} = \Sigma\frac{1}{2-(x-a)^2},$$

but the sum becomes so large and varies so rapidly with the distances of the terminal observations from ± 2 that no convenient approximation is possible. The difference is even more marked when we consider the scale of the distribution. If the terminal observations are corrected in the above way the standard error of the difference is $1{\cdot}85/\surd n$, out of 4, giving a proportional error of $0{\cdot}46/\surd n$. For the normal law the best estimate of the standard error has a standard error of $0{\cdot}71\sigma/\surd n$. The probability distribution of the range between the two extremes of n observations derived from the normal law has been studied by E. S. Pearson (1926). His results, for $n=200$, 500, and 1000, give for the ratios of the standard error of the range to its expectation the values 0·103, 0·086, 0·077; these can be written as $1{\cdot}46/\surd n$, $1{\cdot}93/\surd n$, $2{\cdot}43/\surd n$, so that the uncertainty of an estimate from the extremes would be 3 to 5 times that for the triangular distribution. I am indebted to Mr H. O. Hartley for these values.

Thus with laws practically indistinguishable observationally from the normal law the appropriate treatments of the data to give estimates of the parameters will differ drastically. This fact makes it all the more necessary that we should try to find out what the law of error really is.

The use of the normal law as an approximation, when the actual error is the resultant of several comparable components, does not rest, therefore, on the artificial introduction of one of those limiting processes that delight the hearts of mathematicians. It is not necessary that the number of components should be infinite. With quite a small number of components the law will approach the normal so closely that hundreds or thousands of observations will be needed to detect any difference. This is the true justification of the law in these cases, so far as it goes. But it does not justify the supposition that the law is right at the tails, and therefore does not justify the use of the arithmetic mean.

(2) A specimen of the second type is given by the number of letters on the lines of a page of print. Letters and spaces are not all of the same length, and their distribution among the lines gives rise to a variation in the number that a line will hold. Such a variation is therefore the resultant of a number of small and nearly independent

variations of comparable magnitude, and the usual proof of the normal law is applicable (Whittaker and Robinson 1924, p. 168; Jeffreys 1937*a*, pp. 56–60). Many actual errors are of this form, notably all where the data are not separate observations but linear combinations of several observations of the same type. Such a case is provided by Bullard's determinations (1936) of gravity in Africa by comparison of the periods of pendulums swung locally with others swung at the same time in Cambridge. Two swings in each place are combined to give a mean difference; thus the error of the comparison is the resultant of four comparable components, each probably capable of continuous variation, and in these conditions we should expect the normal law to hold with considerable accuracy.

The proof of the normal law in such cases rests on several approximations, and the law cannot be asserted from it without some evidence, such as that just mentioned, that these are valid in the particular case. The proof fails if one of the components contributes most of the variation, the argument leading in that case back to a law approaching the error law for that component by itself. If the number of components is k the law will fail at deviations more than about $\sigma\sqrt{k}$, sometimes less (Jeffreys 1935, p. 209). This is true even for the binomial law; we see that it forbids departures more than this altogether, whereas the normal law gives a finite though small chance. The range within these limits includes most of the chance; but, unfortunately, the principle of the arithmetic mean does not require the truth of the normal law at deviations only up to $\sigma\sqrt{k}$, but at all deviations whatever, and the difference will be larger if the components are unequal. The treatment of outlying observations must depend on a fuller discussion of the law of error at the extremes.

(3) and (4) In these two types of cases there is little or no reason to expect the normal law. There may be an overwhelming component variation that does not satisfy the law. The distinction between them is as follows. In some types of observation we can repeat the conditions of observation and can place limits to the variation that occurs when the conditions, so far as we know, are the same. The variation is then called observational error. But when we come to apply the same methods, the limits of observational error being already known, to data containing another possible variation, we find a much greater range. For instance, we may be able with a simple technique to measure the stature of a man and find the results consistent to a millimetre or so; but when we apply it to different men we may find variations of 30 cm. We then infer systematic differences between different men. Their further treatment is a matter for the special subject. Observational error is merely the unexplained residuum when we have allowed for all known systematic variations. If other systematic variations are afterwards discovered they can be allowed for, and the outstanding variation is correspondingly reduced, but it never wholly disappears. Our question is, then, when we have done all we can to avoid or eliminate systematic variations, does the remaining variation satisfy the normal law? The most conspicuous case obviously does not satisfy it, being the rounding off of a reading to the nearest scale interval, which gives a rect-

angular distribution for the chance of error.* To reduce the importance of this we need cases where the observations cover several scale intervals, and the variation is limited as far as possible to what human inaccuracy can produce. It is of no use to consider types where most of the variation is genuine in the sense that it would be repeated if the observations were repeated. The existence of genuine normal variations is not the subject of this paper, but for the special subjects where they arise; the question is whether errors of observation satisfy the normal law, and to answer it we need cases where these errors account for most of the variation. The conditions needed are approached in some astronomical observations. Unfortunately they do not appear to have been published in any suitable form. Thus Brunt (1931, p. 33) quotes a series of residuals from Bessel, also given by Chauvenet, which appear to satisfy the law closely. Positive and negative residuals are combined, so that asymmetry cannot be tested; but comparison of the observed numbers and the expectations as they stand gives $\chi^2 = 3\cdot6$, based on eleven groups. The whole number of observations and the standard error have been determined to fit the data, thus removing 2 degrees of freedom. As the data are combined without regard to sign it seems unnecessary to allow for variations of the mean, which would only produce a second order effect on the combined numbers. Thus we can take the effective number of degrees of freedom as 9. From Fisher's table we find that a smaller χ^2, with 9 degrees of freedom, would be expected in 7 % of the cases if the law was true. The agreement is therefore surprisingly good, indeed a little too good, because it suggests that this series has been selected because it gives an unusually good agreement with the law and that others that may have disagreed violently with it may have been suppressed. This danger of selection makes it undesirable to make use of published series to test the normal law, unless there is some definite reason to believe that no suppression has taken place.

3. After some search it appeared to me that the conditions could be satisfied only if some special device was available to magnify the inaccuracy and the error could be found more accurately by some other method. Otherwise the rounding-off error would be the dominant one and we should be no farther forward. Three series of suitable data were found, which satisfied the conditions and did not appear to have been selected. The first two were by K. Pearson and his collaborators (1902); the evidence against selection is that the observations were made to test, not the normal law, but the correlations between different observers and other somewhat subtle points. The normal law was discussed, apparently, only as an afterthought when the observations had already been made. In the first type of observation three sheets of paper were marked each with two pinholes at the same distance apart, and three observers were asked to bisect by eye with a pencil the line joining them. The distances were afterwards measured carefully and the errors were found. In the second type a bright line reflected

* The same might be expected to apply to estimation to the nearest tenth, but Yule (1927) has shown that this may be far from being the case.

by a pendulum moved over a screen between two fixed marks. At a certain instant in its passage a bell rang, and at that instant the observers divided a line with a pencil in the same ratio as the bright line divided the interval between the fixed marks. The records were afterwards measured and compared with one another and with an automatic record. The two types of reading resemble the conditions in the determination of the declination and right ascension of a star with the transit circle. Bond's series were given in his book (1935). An illuminated slit was covered with a ground-glass screen and viewed with a travelling microscope slightly out of focus. The slit was kept fixed, but the microscope was moved well outside the range of vision after each reading, so that the settings would be as far as possible independent. The variation of the readings measures the variation of the judgment about the position of the centre of a fuzzy object. The conditions resemble those in the measurement of a spectrum line or (apart from the shape of the object) that of a star image on a photographic plate. Bond's published data were grouped without regard to sign, but on request he was able to supply the distribution with regard to sign, so that asymmetry also can be tested. We have therefore six series of about 500 observations each and one of about 1000.

Bond's data were given as an example of the normal law, but on examination they showed departures from it that were clearly systematic. Pearson in every case found a departure from the normal law, both in the sense of either excess or deficiency of observations at the tails and of asymmetry. His methods of fitting the data and his tests of significance for the departures are, however, somewhat defective, and a rediscussion is needed.

In each case Pearson fitted laws of the types associated with his name by the method of moments. According to these laws the chance of a deviation between x and $x+dx$ is ydx, where y satisfies the equation

$$\frac{1}{y}\frac{dy}{dx} = -\frac{x-a}{b_0+b_1x+b_2x^2}, \tag{1}$$

and the arbitrary factor is adjusted to make the integral of y through the permitted range equal to 1. Their merits are, first, that if b_1 and b_2 are 0 the law reduces to the normal law about $x=a$; second, that they involve two more parameters than the normal law does, and are therefore capable of being adjusted to a wider range of data; third, that y has at most one maximum or minimum, and this condition appears to be satisfied by errors of observation. The solutions separate into three main types and a number of degenerate cases. If the zeros of the denominator are imaginary, the solution is of Type IV, and may be written

$$y \propto \left(1+\frac{x^2}{2m\sigma^2}\right)^{-m} \exp\left(-2p\tan^{-1}\frac{x}{\sigma\surd(2m)}\right), \tag{2}$$

if the origin is chosen suitably. If p is 0 the law reduces to the symmetrical one of Type VII. x may have any real value.

If the zeros are real and the admissible values of x lie between them, the law is of Type I, which may be written

$$y \propto \left(1-\frac{x^2}{2m\sigma^2}\right)^m \left(\frac{1-x/\sigma\surd(2m)}{1+x/\sigma\surd(2m)}\right)^p \tag{3}$$

for x between $\pm\sigma\surd(2m)$. If p is 0 the law becomes the symmetrical one of Type II. If one of $m\pm p$ is negative, y is a monotonic function; if both are negative, y has a minimum; if both are positive, it has a maximum. Neither can be -1 or less.

If the zeros are real and the admissible values of x do not lie between them, we have two cases according as the admissible range includes a or not. If it does not, y is a monotonic function of x within the range permitted, which is from the nearer root to infinity; if it does, there is a maximum at a, and the admissible values of x stretch from the nearer root to infinity. These are Pearson's Type VI. Pearson makes his classification according to the signs of the roots of the denominator in (1) as they stand, but if they are of opposite signs they could be made to have the same sign by simply adding a constant to x, thus displacing the curve without altering the form or scale. His classification therefore does not rest on invariant properties, and one based either on the position of a with respect to the roots, or on whether the admissible values of x lie between the roots or not, appears to be preferable.* These two cases do not arise in the present work, in which Types I and IV (with positive indices), with their special cases II and VII, cover the ground.

Pearson's method of fitting these laws was to calculate the expectations of the first four moments and to adjust the parameters a, b_0, b_1, b_2 so that the calculated and observed moments agreed. This method has been severely criticized by Fisher on the ground that it does not make adequate use of the information contained in the observations, except in the particular case when the law reduces to the normal one. It is useful only for bell-shaped curves (those with a single maximum) and does not give the maximum accuracy for these. For J-shaped curves (the monotonic ones) the extreme observation gives more accurate information than any of the moments; for U-shaped ones (those of Type I with both indices negative) the two extreme observations do; the same holds for the rectangular law, mentioned above, which is a degenerate case of Type I, constituting a transition from a bell-shaped to a U-shaped curve. For laws of Type VII, with $m\leqslant\frac{5}{2}$, the fourth moment as calculated is infinite, whereas the calculated one is necessarily finite; consequently the method of moments with this law will always give $m>\frac{5}{2}$, whatever the true value may be. In studying these series of observations Pearson several times found values of m near 4, and it seemed possible that the bias introduced by the method of moments in such cases might be even more important than the increase of the random error pointed out by Fisher.

Fisher has pointed out (1922, p. 356) that for small departures from the normal law the equation (1) is equivalent to that satisfied by the exponential of a quartic, and that

* Better than either, I think, would be to introduce the powers occurring in (3) and its analogues directly, since their signs separate bell-shaped, U-shaped, and J-shaped curves at once.

in these conditions the method of moments should be fully efficient. This is correct for Type I; but for Type IV the coefficient of the fourth power of x is negative and the convergence of the integrals is only saved by higher powers. In any case, however, it seems that departures from the normal law, with practicable numbers of observations, can be established only if they are large enough for the moment series to diverge. This point will be considered further later.

The practical difficulty of the method of maximum likelihood, in application to the bell-shaped Pearson curves, has been the labour of calculation. A device given later in this paper enables this to be reduced considerably, at any rate when the law is nearly symmetrical, as in these cases. The method of maximum likelihood is practically equivalent to the principle of inverse probability in problems of estimation (Jeffreys 1938*a*); Fisher does not advocate it for that reason, but as I accept the principle of inverse probability I do so.

It may be noticed that in the laws of Types I and IV we can associate x with an additive constant, which will be the parameter of location; σ is the parameter of scaling; m and p are numbers, m expressing excess or deficiency of chance at the tails and p the asymmetry. It is necessary to associate with σ^2 a multiplier of order m in order to keep the scale of the same order of magnitude when σ remains constant and m varies; otherwise, when m tends to infinity the entire area of the curves would be concentrated at one value of x. This leads to an interesting consequence. Let m tend to infinity, σ remaining finite. Then both forms tend to

$$y \propto \exp\left(-\frac{x^2}{2\sigma^2}-\frac{2px}{\sigma\sqrt{(2m)}}\right) \propto \exp -\frac{1}{2\sigma^2}\left(x+\frac{2p\sigma}{\sqrt{(2m)}}\right)^2, \tag{4}$$

so that in the limit both are normal distributions about $x = -2p\sigma/\sqrt{(2m)}$. It follows that if m is large enough, the scale of the curve (as measured, for instance, by the distance between the inflexions) remaining finite, it will be impossible to separate the parameter of location from p by means of any observed data. This will be true even if p becomes large at the same time as m; indeed, if it does not become large of order $m^{\frac{1}{2}}$ the displacement of the mode will tend to zero; and it must be less than m for bell-shaped laws of Type I, otherwise y has no maximum in the permitted range. Thus we cannot hope to detect asymmetry unless we can show that m is finite; for if m could be infinite any possible effect of p could equally be interpreted as the effect of a permissible change of location. Asymmetry is not worth considering unless we can first show that there is a symmetrical departure from the normal law.

In the present cases Pearson finds for every series of observations both a symmetrical and an antisymmetrical departure from the normal law, and his wording towards the end of the paper, where he always speaks of the non-normal distributions as "skew distributions", places the emphasis on the asymmetry.* The nearest approach to

* Any function $f(x)$, where x may have either sign, may be expressed as the sum of an even function $\frac{1}{2}f(x)+\frac{1}{2}f(-x)$ and an odd one, $\frac{1}{2}f(x)-\frac{1}{2}f(-x)$. I speak of the former as the "symmetrical" part and the latter as the "antisymmetrical" part. Where the latter is not zero I call the function "asymmetrical".

a separate test for asymmetry, however, is provided by the ratio $\surd\beta_1 = \mu_3/\mu_2^{\frac{3}{2}}$, where μ_2 and μ_3 are the second and third moments about the mean. The test for the finiteness of m is based on the ratio $\beta_2 = \mu_4/\mu_2^2$, μ_4 being the fourth moment. The expectations of $\surd\beta_1$ and β_2 on the normal law are 0 and 3. Pearson finds their mean square random variations, still assuming the normal law; if the observed values exceed these substantially both are asserted as genuine, so that m is taken as finite and p as different from zero. It would presumably be legitimate to assert one of these propositions on this ground, but not both, unless it is also shown that the first departure asserted will not explain the second. This danger of introducing too many parameters at a time is very serious. In this case β_2 is found to exceed 3 in 4 cases out of 6, indicating a law with an excess of observations at the tails in comparison with the normal law. The actual number is small, but it is just these observations that contribute most of the third moment. The tails could not contribute much if the normal law was true, but given that there are more outlying observations than the normal law indicates, a new random error comes in. With a symmetrical departure that increases their frequency, there is a sampling error in the difference of the numbers of outlying observations on the two sides. This hardly arises with the normal law, since the whole expectation in these ranges is small; but with an increase in the total expectation the expectation of the sampling error rises too, and the standard error of $\surd\beta_1$ as found from the normal law ceases to be applicable. It therefore requires to be shown, before we can assert skewness, that $\surd\beta_1$ or some other standard of skewness is not explicable as due to a sampling error in the outlying observations, rendered possible by the existence of a symmetrical departure. The values of $\surd\beta_1$ and β_2, indeed, answer the same question: are the data reasonably consistent with the normal law? They do not, however, entitle us to introduce two new parameters, until it has been shown that one, expressing only a symmetrical departure from the normal, will not suffice to explain both departures. m must in any case be considered before p, because, as shown above, p is totally indeterminate if m is infinite.

A special complication was shown by Pearson by taking the means of groups of about 25 consecutive observations, which were found to vary more than they should if all the observations were independent (unless indeed the law of error was of Type IV or VII with m near 1). They are no steadier than the means of 2 to 15 independent observations should be. Such a correlation between consecutive observations suggests that all estimates of uncertainty based on the hypothesis of independence will be too low; it turns out, further, that it may explain a large part of the variation of m between different series and possibly also the asymmetry. Consequently I have not thought it worth while to try to estimate the asymmetry accurately. It may be suggested that the discussion should be limited to cases where the condition of independence is more accurately satisfied, but I know of no evidence that there are any such cases. The existence of systematic personal errors is well known to observers, and astronomers, in particular, achieve their high standards of accuracy by arranging their work so that these will cancel as far as possible. The results are more nearly independent, but they are differences of observed values, and their law of error will be that of the resultant of several errors and should approach the normal more closely than that of an individual observation. This will make the departure from normal more difficult to estimate, and we should still have the problem of inferring the law for one observation from that of a difference.

In testing the normal law Pearson began by finding a mean and a standard deviation, the mean being interpreted as personal equation; the expectations by ranges according to the normal law with these parameters were compared with the observed ones by the χ^2 test. The respective probabilities of larger values of χ^2, given the normal law, were given as 0·65, 0·28, 0·21 for the absolute judgments in the bisection series, as 0·0006, 10^{-10}, and 0·29 for the bright-line series. Four of these six values would not be taken as evidence against the normal law by most statisticians; but it is not clear how the observations have been grouped to find χ^2. The numbers of groups quoted are such that in several cases it would be impossible for the expectations in all of them to have reached 5, and Pearson cannot have applied the test in the way that he afterwards recommended, in which the smaller expectations are combined so that the total in each group is at least 5. The later rule has the advantage that the probability of the contribution to χ^2 from each group is more nearly normally distributed, and therefore that the probability of a given total χ^2, according to the hypothesis to be tested, will follow the usual rule more closely. It has the disadvantage, illustrated by the comparison of the binomial and normal laws that has been given above, that if the expectation in a range, on the hypothesis tested, is small enough, that range may never appear in the test except in combination with an adjacent one, where the systematic departure has the opposite sign. Thus if only observations in a given range are sufficiently unlikely on the hypothesis their presence may fail to be taken as evidence against the hypothesis. This is a serious defect of the test, and was possibly avoided, at some cost, by the earlier form, which permitted groups with smaller expectations. I consider, however, that the χ^2 test by itself, applied to a large number of groups, is unsatisfactory because the random error of χ^2 is so large that a genuine systematic departure may contribute less than the random error of χ^2 when it would be obvious if tested by itself. The use of χ^2, in my opinion, is simply that the distribution of the larger contributions to it may suggest what particular systematic departures, if any, are worth special attention; these departures, however, can be asserted only when a test specifically adapted to them has been applied. If groups can be combined in such a way that only one new parameter could be determined from them, the value of χ^2 for this grouping is an estimate of the square of the ratio of that parameter to its standard error, and when the expectations in the groups are all large this is a very efficient method. It is, however, inferior to a direct estimate of the parameter and its standard error by the method of maximum likelihood, though the difference is small when the trial distribution is nearly uniform.

It may be worth while to call attention to one type of spurious increase of χ^2. If the observations are measures grouped by unit ranges, a decision must be made about where to put a reading ending in 0·5. Usually a convention is made to round off always to the nearest even integer to avoid systematic error in the mean. Thus values, really from 1·45 to 2·55, are recorded as 2, and the expectations in the even and odd ranges are respectively raised and lowered by 10 %. Thus this convention makes a contribution to χ^2 equal to

$$\Sigma \frac{(0\cdot 1 m_r)^2}{m_r} = 0\cdot 01 \Sigma m_r,$$

which is not negligible if there are some hundreds of observations. If several ranges are taken together, these systematic changes cancel and there is no trouble.

4. *Tests by grouping*. As a standard of comparison, I first took the normal law with the same mean and second moment. In comparison with this law, a Type VII one is higher at the mode and at the tails, lower on the flanks. The opposite is true for Type II. It appeared, therefore, that a test of a symmetrical departure could be obtained by comparing the whole numbers of observations in the ranges where the departure suggested is positive or negative with their expectations according to the normal law. The law of Type VII

$$y = \frac{1}{\sqrt{(2\pi)}\,\sigma} \frac{(m-1)!}{(m-\frac{3}{2})!\,(m-\frac{3}{2})^{\frac{1}{2}}} \left(1+\frac{x^2}{2(m-\frac{3}{2})\,\sigma^2}\right)^{-m} \tag{1}$$

has the same second moment as the normal law

$$y = \frac{1}{\sqrt{(2\pi)}\,\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{2}$$

and to order m^{-1} it is equivalent to

$$y = \frac{1}{\sqrt{(2\pi)}\,\sigma} \left\{1+\frac{1}{8m}\left(3-6\frac{x^2}{\sigma^2}+\frac{x^4}{\sigma^4}\right)\right\} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{3}$$

The addition needed to express the departure from the normal law along the Type VII series is therefore given by a Hermite function, with a coefficient proportional to m^{-1}. It vanishes at

$$x/\sigma = \pm 0\cdot 742 \quad \text{and} \quad \pm 2\cdot 335. \tag{4}$$

For comparison we may take the large departure given by $m = \frac{5}{2}$. With this value the expectation of the fourth moment is infinite, and it would be impossible to determine m by the method of moments at all. Yet even then the values of y according to (1) and (2) agree at $x/\sigma = 0\cdot 68$ and $2\cdot 64$, which are not very different from the points of agreement for small departures. It is only at large deviations that we need consider the departure of the change from proportionality to m^{-1}. The same is easily seen to be true for departures of Type II.

A test of significance for a symmetrical departure could therefore be made by comparing observation and expectation in ranges separated by the points of agreement. But if we do this we lose accuracy, and it is better to omit the groups near the zeros of the departure and use only ranges about the extreme departures. In other cases it has been found (Jeffreys 1938*b*) that efficiencies of about 90% can be got by using ranges that include about $\frac{2}{3}$ of the expectation. Hence for the intermediate ranges we should sacrifice $\frac{1}{6}$ of the total expectation at each end, and for the terminal ones $\frac{1}{3}$ of it at the beginning. Now

$$\operatorname{erf} 0{\cdot}742/\sqrt{2} = 0{\cdot}5419; \quad \operatorname{erf} 2{\cdot}335/\sqrt{2} = 0{\cdot}9804;$$

and subdividing the ranges as indicated we find that we should use ranges given by

$$x/\sigma = \pm 0{\cdot}470, \quad \pm 0{\cdot}869 \text{ to } \pm 1{\cdot}681, \quad \pm 2{\cdot}480 \text{ to } \pm\infty.$$

We denote the central range by the figure 0, the two side ones by ± 1, and the two tail ones by ± 2. Since a symmetrical departure would not affect $O_1 - O_{-1}$ or $O_2 - O_{-2}$, our data will be O_0, $O_1 + O_{-1}$, and $O_2 + O_{-2}$. The total number of observations and the standard error having been found as for the normal law, we have 1 degree of freedom left to estimate m, and the evidence about m will be tested by the contribution to χ^2 from these three sets.

For a small antisymmetrical departure, on which a symmetrical one will always be superposed, the additional terms needed, if the expectation of the mean is to be unaltered, will be proportional to the Hermite function

$$\phi_3(x) = \left(\frac{3x}{\sigma} - \frac{x^3}{\sigma^3}\right) \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right), \tag{5}$$

which vanishes at $x/\sigma = 0$ and $\pm\sqrt{3}$. The ranges to retain are as follows:

Range number	x/σ
-2	$-\infty$ to $-1{\cdot}915$
-1	$-1{\cdot}185$ to $-0{\cdot}193$
1	$+0{\cdot}193$ to $+1{\cdot}185$
2	$+1{\cdot}915$ to $+\infty$

Given the normal law these would suffice to estimate the number of observations, the mean and the standard error, leaving apparently 1 degree of freedom to test asymmetry. But we have already seen that if there is no evidence of a symmetrical departure there cannot be any for asymmetry; and χ^2 found from these groups, in comparison with the normal law, would contain a contribution from the symmetrical departure. It would therefore always be illusory. On the other hand, if m is finite the expectations in these ranges will always be equal in pairs for the symmetrical laws; an asymmetry will be shown if, when we have adjusted the mean, the signs of $O_2 - O_{-2}$ and $O_1 - O_{-1}$ are opposite and their magnitudes significant. The standard of comparison, however, will not be the expectation according to the normal law, since these types imply different values for $O_{-2} + O_2$, and the differences may be considerable. With a suitably chosen asymmetry it will be possible to fit all four groups without altering the mean; hence the number of degrees of freedom is 1 in comparison with a symmetrical law, the latter being adjusted to the observations. Then according to such a law the expectation of O_2 would be $\frac{1}{2}(O_{-2}+O_2)$ and its residual $\frac{1}{2}(O_2-O_{-2})$, with analogous relations; and the contribution to χ^2 from asymmetry is

$$\frac{2\{\frac{1}{2}(O_2-O_{-2})\}^2}{\frac{1}{2}(O_2+O_{-2})}+\frac{2\{\frac{1}{2}(O_1-O_{-1})\}^2}{\frac{1}{2}(O_1+O_{-1})}=\frac{(O_2-O_{-2})^2}{O_2+O_{-2}}+\frac{(O_1-O_{-1})^2}{O_1+O_{-1}}. \tag{6}$$

If we had the individual observations, this evaluation would be straightforward. Actually, however, the observations are grouped in ranges of about $\frac{1}{4}$ of the standard error. The result is that if we retain groups up to and including the suggested points of separation adjacent ranges will often be retained and some observations put into the wrong side of the scale. It becomes necessary, therefore, to make it an absolute rule to reject any range where the suggested correction term changes sign. The mean, again, does not come at the end or the mid-point of a range of the table. The extended ranges used for the tests are therefore asymmetrically situated and expectations for them according to any symmetrical law are not symmetrical. This gives no trouble in testing a symmetrical departure, since we have the expectations according to the normal law for the actual ranges used. But for the antisymmetrical one there is a serious complication, since the expectations of O_2 and O_{-2} on the symmetrical law are rendered unequal by this adjustment, and we cannot estimate their difference until m has been evaluated. But since the departure from normality is small we can find C_2 and C_{-2}, the expectations according to the normal law, and suppose that these are still approximately correct. Then we can allow for the difficulty of asymmetrical grouping by replacing O_2-O_{-2} by $O_2-C_2-O_{-2}+C_{-2}$ before evaluating the contribution to χ^2 from possible asymmetry.

In this way the contributions to χ^2 from symmetrical and antisymmetrical departures will be determined and tested for significance separately, with possibly only a trifling loss of accuracy. If desired a more accurate solution can be constructed later with a finer grouping, but what we want to know first is whether there is a *prima facie* case for asserting the existence of the departures at all.

4·1. In the first place the means and standard deviations for Pearson's six series were found; the observed numbers were then compared with expectations according to the normal law by the χ^2 test, terminal groups being combined so as to keep an expectation in every range of at least 5. The results are as follows:

Bisection:	(1)	Mean $-0·01230$,	$\sigma = 0·02455$.	Observer, Dr Alice Lee.
	(2)	Mean $+0·00495$,	$\sigma = 0·03065$.	Observer, Professor Pearson.
	(3)	Mean $+0·00377$,	$\sigma = 0·02625$.	Observer, Mr Yule.
Bright line:	(1)	Mean $+0·2909$,	$\sigma = 4·7566$.	Observer, Professor Pearson.
	(2)	Mean $-4·5993$,	$\sigma = 4·6915$.	Observer, Dr MacDonell.
	(3)	Mean $-1·8054$,	$\sigma = 7·2859$.	Observer, Dr Lee.

For convenience I subtracted 0·005 from all the bright line readings and multiplied by 4 before making my determinations. The results are in agreement with Pearson's grouped ones (1902, p. 252). His results (grouped) for the bisection experiments all agree numerically, but all the means are given with the opposite sign. (His sign convention for these changes during the paper.) The detailed comparisons are as follows:

Bisection series (all readings multiplied by 100)

	(1) Obs.	Calc.	$O-C$	χ^2	(2) Obs.	Calc.	$O-C$	χ^2
−9	1	0·58	+0·42		0	(0·84)		
−8	4	1·91	+2·09	4·19	1	1·44	−0·44	0·47
−7	8·5	5·3	+3·2		3	3·35	−0·35	
−6	12	12·5	−0·5	0·02	11	7·0	+4·0	2·28
−5	13·5	25·2	−11·7	5·44	14·5	13·2	+1·3	0·13
−4	45	43·0	+2·0	0·09	21·5	22·2	−0·7	0·02
−3	61	62·4	−1·4	0·03	30	34·1	−4·1	0·49
−2	76	76·8	−0·8	0·01	47	46·7	+0·3	0·00
−1	90·5	80·4	+10·1	1·27	51·5	57·4	−5·9	0·61
0	74·5	70·1	+4·4	0·28	72	64·0	+8·0	1·00
+1	50	53·8	−3·8	0·27	65·5	63·8	+1·7	0·04
2	30·5	34·4	−3·9	0·45	53	57·6	−4·6	0·37
3	21·5	18·7	+2·8	0·42	50·5	46·6	+3·9	0·33
4	7	8·65	−1·65	0·31	28·5	34·0	−5·5	0·89
5	3	3·35	−0·35	0·63	27	22·2	+4·8	1·04
6	2	1·12	+0·88		13·5	13·0	+0·5	0·02
7	0				7·5	7·0	−0·5	0·04
8	0			13·4	0	3·32	−3·32	
9	0	0·41			1	1·42	−0·42	
10	0				0	0·55	−0·55	1·18
+11	0				2	0·19	+1·81	
						(0·08)		
								8·9

	(3) Obs.	Calc.	$O-C$	χ^2
		(0·67)		
−7	1	1·53	−0·53	0·77
−6	7·5	4·10	+3·40	
−5	9·5	9·5	0·0	0·00
−4	22	19·2	+2·8	0·41
−3	40·5	33·3	+7·2	1·56
−2	43·5	50·3	−6·8	0·92
−1	51	66·8	−15·8	3·55
0	68·5	74·7	−6·2	0·52
+1	75	73·4	+1·6	0·03
2	70·5	62·7	+7·8	0·97
3	61	46·1	+14·9	4·82
4	25·5	29·5	−4·0	0·54
5	13·5	16·3	−2·8	0·48
6	10	7·8	+2·2	0·62
+7	1	3·22	−2·22	3·12
		(1·70)		
				18·3

O = observed; C = calculated.

Bright line series

	(1) Obs.	Calc.	$O-C$	χ^2	(2) Obs.	Calc.	$O-C$	χ^2
+25	1	(0·01)	+ 0·99		0			
23	0				0			
21	0				1	0·00	+1·0	
19	1	0·04	+ 0·96	0·50	0			
17	0	0·20	− 0·20		0			
15	1	0·78	+ 0·22		0	0·01	−0·01	
13	6	2·56	+ 3·44		0	0·08	−0·08	0·77
11	4	7·1	− 3·1		0	0·38	−0·38	
9	12	16·6	− 4·6	1·27	0	1·39	−1·39	
7	22	32·4	−10·4	3·35	3	4·33	−1·33	
5	57	53·5	+ 3·5	0·28	8	11·1	−3·1	0·87
3	71	73·6	− 2·6	0·09	31	24·0	+7·0	2·03
+ 1	97	85·4	+11·6	1·58	35	43·5	−8·5	1·67
− 1	85	83·5	+ 1·5	0·03	73	65·4	+7·6	0·88
− 3	69	68·1	+ 0·9	0·01	76	82·8	−6·8	0·56
− 5	56	46·9	+ 9·1	1·77	96	87·2	+8·8	0·89
− 7	23	27·2	− 4·2	0·65	79	77·1	+1·9	0·05
− 9	7	13·2	− 6·2	2·91	60	56·7	+3·3	0·19
−11	4	5·4	− 1·4		30	35·0	−5·0	0·72
−13	1	1·90	− 0·90		17	18·1	−1·1	0·07
−15	1	0·53	+ 0·47	0·12	5	7·78	−2·78	
−17	0	0·13	− 0·13		3	2·81	+0·19	
−19	1	0·03	+ 0·97		1	0·84	+0·16	
−21	0	(0·01)	− 0·01		0	0·21	−0·21	0·24
−23	0				0	0·04	−0·04	
−25	0			12·6	0			
−27	0				1	0·00	+1·00	
								8·9

	(3) Obs.	Calc.	$O-C$	χ^2
		(0·04)		
+25	1	0·07	+ 0·93	
23	0	0·18	− 0·18	
21	0	0·44	− 0·44	4·28
19	0	0·99	− 0·99	
17	0	2·05	− 2·05	
15	1	4·04	− 3·04	
13	12	7·27	+ 4·73	3·07
11	19	12·2	+ 6·8	3·78
9	18	19·0	− 1·0	0·05
7	26	27·6	− 1·6	0·09
5	42	36·7	+ 5·3	0·77
3	48	45·6	+ 2·4	0·12
+ 1	44	52·6	− 8·6	1·40
− 1	48	56·3	− 8·3	1·22
− 3	46	55·8	− 9·8	1·72
− 5	60	51·5	+ 8·5	1·40
− 7	42	44·0	− 2·0	0·09
− 9	36	34·9	+ 1·1	0·04
−11	36	25·7	+10·3	4·14
−13	20	17·6	+ 2·4	0·33
−15	12	11·1	+ 0·9	0·08
−17	5	6·51	− 1·51	0·35
−19	2	3·55	− 1·55	
−21	1	1·79	− 0·79	2·11
		(1·45)		
				25·0

The respective values of χ^2, the numbers of groups (reduced by 3 to convert to degrees of freedom) and the corresponding values of $P(\chi^2)$ from Fisher's table are as follows. Against them I give Pearson's values:

		H. J.			K. P.		
		χ^2	n'	P	χ^2	n'	P
Bisection	1	13·4	10	0·22	13·3	14	0·65
	2	8·9	13	0·78	22·0	17	0·28
	3	18·3	11	0·08	20·3	14	0·21
Bright line	1	12·6	9	0·18	42·8	15	0·0006
	2	8·9	9	0·44	83·5	13	0·0000
	3	25·0	15	0·05	21·8	17	0·2933

Comparing the determinations of P we see that there are great differences, presumably owing to the fact that I have grouped more of the regions with small expectations together. None of my values of P would usually be taken as decisive evidence against the law tested. Pearson, for the second "Bright-line" series, remarks that the P would differ from zero only in the tenth place; I get a perfectly normal value. He also gets a value of χ^2 for the first bright-line series that would imply rejection of the hypothesis; mine is within the expectation and its standard error. The reason is presumably that these series contain outlying observations at distances where deviations would be practically impossible according to the normal law. Pearson's earlier practice makes allowance for their improbability; his later one, which I have adopted here, lumps them with the larger expectations and their importance is disguised. It appears that the earlier method may be the better. But what the comparison really shows is that when the expectation becomes very small in some ranges the χ^2 test, in any form, is so sensitive to the arbitrary method of grouping that it is practically useless. The only suitable method is one that tests the suggested departures separately. (My contingency test (1937*c*) is applicable however small the expectations are.)

The bisection series consist of 500 observations each, the bright-line ones of 519. Bond's series contains 1026. The unit is 0·01 mm. displacement from an arbitrary zero. Grouped without regard to sign, as in his original summary, they are as follows. The standard error is 4·601. Residuals from 1·0 to 2·0 and $-2·0$ to $-1·0$ are grouped together and entered as 1·5. Comparison with the calculated expectations gives $\chi^2 = 16·8$, from 13 groups. A slight error in the mean would hardly affect the expectations grouped without regard to sign, and we may take the number of degrees of freedom as 11. From Fisher's table (1936), $P(\chi^2) = 0·12$, which is normal. Yet inspection of the signs shows a variation of the most systematic kind possible. If the true law was of Type VII and we did our best to fit a normal law to the data, then we should expect the signs of the $O-C$ values to show just the distribution that they do, and there are no exceptions whatever. The signs as they stand would indeed suggest a Type VII law, differing from the normal by so much that the difference in every range is more than the sampling error; yet the random error of χ^2 introduced by using numerous ranges reduces the sensitivity so much that the whole matter is left in doubt.

Error	Obs.	Calc.	$O-C$	χ^2	Calc. (2)	$O-C$ (2)	χ^2
0·5	184	176·5	+ 7·5	0·32	186·1	− 2·1	0·02
1·5	178	168·4	+ 9·6	0·55	176·2	+ 1·8	0·02
2·5	54	153·5	+ 0·5	0·00	157·9	− 3·9	0·10
3·5	138	133·1	+ 4·9	0·18	133·8	+ 4·2	0·13
4·5	122	110·3	+11·7	1·24	107·4	+14·6	1·98
5·5	69	87·2	−18·2	3·77	81·5	−12·5	1·92
6·5	60	65·7	− 5·7	0·50	58·7	+ 1·3	0·28
7·5	33	47·3	−14·3	4·32	39·9	− 6·9	1·19
8·5	29	32·3	− 3·3	0·34	25·7	+ 3·3	0·42
9·5	16	21·3	− 5·3	1·32	15·6	+ 0·4	0·00
10·5	19	13·2	+ 5·8	2·55	9·0	+10·0	
11·5	9	7·9	+ 1·1	0·15	5·0	+ 4·0	
12·5	5	4·5	+ 0·5		2·6	+ 2·4	
13·5	5	2·5	+ 2·5		1·2	+ 3·8	
14·5	2	1·25	+ 0·75	1·51	0·6	+ 1·4	
15·5	1	0·64	+ 0·36		0·25	+ 0·75	
16·5	1	0·29	+ 0·71		0·11	+ 0·89	
17·5	1	0·13	+ 0·87		0·04	+ 0·96	
				16·75			6·06

It has already been seen that the theoretical arguments that lead to the normal law suggest that it should break down at a moderate multiple of the standard error. If the true law was a smoothed binomial one, there would be a deficiency at the tails. Most series of observational errors, however, show an excess at the tails, as in this series. It is interesting, therefore, to see whether a normal law would fit better if it was determined only from the residuals up to about $\pm 2\sigma$. This can be done approximately by considering frequencies. There are 654 observations up to ± 4, 983 to ± 10. These give the equation

$$\frac{\operatorname{erf} 4h}{\operatorname{erf} 10h} = \frac{654}{983} = 0{\cdot}6654$$

to determine the precision constant. Then

$$h = 0{\cdot}1662; \quad \sigma = 4{\cdot}254(1 \pm 0{\cdot}027).$$

The values given as Calc. (2) are $1001{\cdot}8\, \Delta \operatorname{erf} hx$, with this value of h. They give $\chi^2 = 6{\cdot}06$ for 8 degrees of freedom, from the values up to ± 10. This is perfectly satisfactory and would leave no room for a significant systematic departure within this range. The normal law therefore appears to hold up to about $2{\cdot}5\sigma$. Beyond that range, however, it breaks down utterly. Beyond $\pm 13{\cdot}0$ there are ten observations; the normal law would predict 2·2.

4·2. *Tests by extended groups*. The ranges used in testing the first bright-line series for a symmetrical departure from the normal law are as follows:

	Range covered	Obs.	Calc.	$O-C$
−2	$+\infty$ to +13	9	3·59	+ 5·41
−1	+ 7 to + 5	79	85·9	− 6·9
0	− 1 to + 1	182	168·9	+13·1
+1	− 5 to − 7	79	74·1	+ 4·9
+2	−13 to $-\infty$	3	2·60	+ 0·4

Thus the combinations to be used are:

	Calc.	$O-C$	χ^2
-2 and $+2$	6·19	$+$ 5·81	5·45
-1 and $+1$	160·0	$-$ 2·0	0·02
0	168·9	$+$13·1	1·01
			6·48

For an antisymmetrical departure we have:

	Range covered	Obs.	Calc.	$O-C$
-2	$+\infty$ to $+11$	13	10·69	$+$ 2·31
-1	$+5$ to $+3$	128	127·1	$+$ 0·9
$+1$	-3 to -5	125	115·0	$+$10·0
$+2$	-9 to $-\infty$	14	21·2	$-$ 7·2

$$(O-C)_2-(O-C)_{-2}=-9\cdot5: \quad O_2+O_{-2}=27$$
$$(O-C)_1-(O-C)_{-1}=+9\cdot1: \quad O_1+O_{-1}=253$$
$$\chi^2=3\cdot63.$$

The second bright-line series gives for the symmetrical departure:

	Range covered	Obs.	Calc.	$O-C$		Calc.	$O-C$	χ^2
-2	$+\infty$ to $+7$	4	6·20	-2·20	-2 and $+2$	10·1	-1·1	0·12
-1	$+3$ to $+1$	66	67·5	-1·5	-1 and $+1$	120·6	-7·6	0·48
0	-3 to -7	251	247·1	$+3$·9	0	247·1	$+3$·9	0·06
$+1$	-11 to -13	47	53·1	-6·1				0·66
$+2$	-17 to $-\infty$	5	3·90	$+1$·10				

and for the antisymmetrical one:

	Range covered	Obs.	Calc.	$O-C$
-2	$+\infty$ to $+5$	12	17·3	-5·3
-1	$+1$ to -3	184	191·7	-7·7
$+1$	-7 to -11	169	168·8	$+0$·2
$+2$	-15 to $-\infty$	10	11·7	-1·7

$$\chi^2=0\cdot77.$$

Bright line (3). *Symmetrical departure*

	Range covered	Obs.	Calc.	$O-C$		Calc.	$O-C$	χ^2
-2	$+\infty$ to $+17$	1	3·76	$-$ 2·76	-2 and $+2$	5·98	$-$ 3·98	2·65
-1	$+11$ to $+5$	105	95·5	$+$ 9·5	-1 and $+1$	184·8	$+$24·2	3·17
0	$+1$ to -5	198	216·2	$-$18·2	0	198	$-$18·2	1·53
$+1$	-9 to -15	104	89·3	$+$14·7				7·35
$+2$	-21 to $-\infty$	1	2·22	$-$ 1·22				

Antisymmetrical departure

	Range covered	Obs.	Calc.	$O-C$
-2	$+\infty$ to $+13$	14	15·07	-1·07
-1	$+7$ to $+1$	160	162·5	-2·5
$+1$	-3 to -11	220	211·9	$+8$·1
$+2$	-17 to $-\infty$	8	13·30	-5·30

$$\chi^2=1\cdot12.$$

Bisection (1). *Symmetrical departure*

	Range covered	Obs.	Calc.	$O-C$		Calc.	$O-C$	χ^2
−2	−∞ to −8	5	2·68	+ 2·32				
−1	−5 to −4	58·5	68·2	− 9·7	−2 and +2	4·21	+ 2·79	1·85
0	−2 to 0	241·0	227·3	+13·7	−1 and +1	121·3	−10·8	0·96
+1	+2 to +3	52	53·1	− 1·1	0	227·3	+13·7	0·83
+2	+6 to +∞	2	1·53	+ 0·47				3·64

Antisymmetrical departure

	Range covered	Obs.	Calc.	$O-C$
−2	−∞ to −6	25·5	20·48	+5·0
−1	−4 to −2	182	182·2	−0·2
+1	0 to +2	155	158·3	−3·3
+2	+4 to +∞	12	13·12	−1·12

$\chi^2 = 1{\cdot}03.$

Bisection (2). *Symmetrical departure*

	Range covered	Obs.	Calc.	$O-C$		Calc.	$O-C$	χ^2
−2	−∞ to −8	1	2·28	−1·28				
−1	−5 to −3	66	69·5	−3·5	−2 and +2	4·52	−0·52	0·06
0	−1 to +2	242	242·8	−0·8	−1 and +1	138·7	−3·7	0·10
+1	+4 to +6	69	69·2	−0·2	0	242·8	−0·8	0·00
+2	+9 to +∞	3	2·24	+0·76				0·16

Antisymmetrical departure

	Range covered	Obs.	Calc.	$O-C$
−2	−∞ to −6	15	12·63	+2·37
−1	−3 to −1	128·5	138·2	−9·7
+1	+2 to +4	132	138·2	−6·2
+2	+7 to +∞	10·5	12·56	−2·06

$\chi^2 = 0{\cdot}82.$

Bisection (3). *Symmetrical departure*

	Range covered	Obs.	Calc.	$O-C$		Calc.	$O-C$	χ^2
−2	−∞ to −7	1	2·20	− 1·20	−2 and +2	7·12	− 5·12	3·68
−1	−4 to −3	62·5	52·5	+10·0	−1 and +1	144·4	+18·1	2·27
0	0 to +1	143·5	148·1	− 4·6	0	148·1	− 4·6	0·14
+1	+3 to +5	100	91·9	+ 8·1				6·09
+2	+7 to +∞	1	4·92	− 3·92				

Antisymmetrical departure

	Range covered	Obs.	Calc.	$O-C$
−2	−∞ to −5	18	15·8	+ 2·2
−1	−3 to −1	135	150·4	−15·4
+1	+2 to +4	157	138·3	+16·7
+2	+6 to +∞	11	12·72	− 1·7

$\chi^2 = 4{\cdot}06.$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

On inspection we see that out of six series of observations the contributions to χ^2 from the symmetrical departure are, in turn, 6·5, 0·7, 7·4, 3·6, 0·2, and 6·1. Four exceed their expectation, unity, three of them considerably. There can be little doubt that these values cannot be attributed to random variation. But we may notice that the range of random variation of χ^2, as found directly from a series with 11 degrees of freedom, would be approximately represented by a standard error of 4·7. This would be quite sufficient to obscure completely systematic contributions of the amounts found. Out of the four that give departures that are probably genuine, two express excesses and two deficiencies of observations at the tails in comparison with the normal law, corresponding to laws of Types VII and II. This was unexpected, since the usual tendency of errors of observation is usually believed to be towards excesses at the tails.

To put the matter in the way recommended by Fisher (1936, pp. 104–17), we could combine the six determinations and regard them as a contribution to χ^2 from 6 degrees of freedom, its amount being 24·5. This is far beyond his 1 % limit, which comes at $\chi^2 = 16{\cdot}8$.

The total contribution from the antisymmetrical departures is 11·43, which would give $P = 0{\cdot}08$, and is only a shade beyond the standard error of the expected χ^2. It would not be accepted as significant by any test.

On the whole, therefore, we have evidence for departures from the normal law, but no evidence that this departure is asymmetrical; or rather, perhaps, we may say that the whole of the asymmetry is summarized adequately by the personal equation, which displaces each distribution bodily in comparison with the true value. The abnormalities are not in the same sense for the same observer in both types of observation. Dr Lee's observations deviate in the sense of Type VII for the bisection series, II for the bright-line series; Pearson's are normal for the bisections and Type VII for the bright-line observations. Also each set includes one apparently normal series, one of Type II, and one of Type VII. There is no basis, therefore, for any generalization about the type of departure.

Bond's data, as published, do not suffice to test for an antisymmetrical departure. To test for a symmetrical one we take range 0 from 0 to 3, 1 from 4 to 8, and 2 from 11 to $+\infty$. The totals in these give $\chi^2 = 5{\cdot}58$, $P = 0{\cdot}018$, the departure at the tails being an excess. But since this series was the longest homogeneous series of observations that I could find, subject to the condition that human inaccuracy would be expected to be the dominant source of error, I asked Dr Bond for his original readings so that asymmetry could be tested. He kindly supplied these and gave much valuable help in their discussion. The mean was not at zero, but at $+0{\cdot}18$. In the table P and N are the numbers of positive and negative residuals of equal magnitude; the column headed "mean" gives the expected effect on $P-N$, on the hypothesis of the normal law, due to the fact that the mean is not at zero. This is subtracted from $P-N$ to give differences analogous to the $(O_n-C_n)-(O_{-n}-C_{-n})$ used in discussing Pearson's data.

	P	N	$P-N$	Mean	$P-N$ (corrected)
0·5	91	93	− 2	+0·7	− 2·7
1·5	76	102	−26	2·0	−28·0
2·5	67	87	−20	3·1	−23·1
3·5	62	76	−14	4·0	−18·0
4·5	65	57	+ 8	4·2	+ 3·8
5·5	39	30	+ 9	4·1	+ 4·9
6·5	32	28	+ 4	3·4	+ 0·6
7·5	22	11	+11	3·0	+ 8·0
8·5	17	12	+ 5	2·5	+ 2·5
9·5	10	6	+ 4	1·7	+ 2·3
10·5	9	10	− 1	1·2	− 2·2
11·5	4	5	− 1	0·8	− 1·8
12·5	3	2	+ 1	0·5	+ 0·5
13·5	4	1	+ 3	0·3	+ 2·7
14·5	0	2	− 2	+0·1	− 2·1
15·5	1	0	+ 1	0·0	+ 1·0
16·5	1	0	+ 1	0·0	+ 1·0
17·5	1	0	+ 1	0·0	+ 1·0

The ranges used in testing asymmetry are: −1 and +1, 1·5 to 5·5 inclusive; −2 and +2, 9·5 to 17·5 inclusive. For these we find

	$P-N$ (corrected)	$P+N$	χ^2
±1	−60·4	845	4·31
±2	+ 2·4	59	0·10
			4·41

Thus χ^2 is rather large. It is clear on inspection, however, that there would be no justification for inferring that the distribution is of Pearson's Type IV. This would give practically a normal law at deviations up to over the standard error, and the asymmetry would be shown at the tails. Here much the greatest part of χ^2 comes from the first four groups on each side of the origin, covering residuals up to less than the standard error. There is no evidence for asymmetry in the tails at all. A good fit could therefore be obtained only by going outside the range of the Pearson functions, and it does not seem worth while to do so on the strength of a value of χ^2 that would be doubtfully significant in any case. Mr Hey's solution for Type VII gave $\chi^2 = 27·37$ on 22 degrees of freedom, $P = 0·24$; for Type IV, 25·62 on 22 degrees of freedom, $P = 0·27$.

Yet these departures at small residuals are larger than we should normally expect, and I enquired of Bond whether he could suggest any explanation. If the observations were not strictly independent, as, for instance, if the microscope was not moved far enough away after each reading, there might be a tendency to repeat readings on account of backlash in the screw. It appeared that such dangers had been foreseen and that all precautions had been taken. He said, however, that before he had completed the series another man had interfered with the apparatus and altered the focusing, which Bond had restored to its original state. This might have been done imperfectly, so that the data might really refer to two superposed normal laws with different modes and standard errors. Such superposition, even if there was no change of the mode, would give the excess of observations at the tails characteristic of Type VII, and we might be left with no evidence against the normal law.* Mr G. B. Hey, who helped me with the arithmetic in this part of the work, tested this by fitting the sum of two normal laws to the data; but he found that if they were to account for the hump, their parameters must differ so widely that it seemed incredible that such a change could have occurred without Bond's having noticed it. Accordingly I think that even if these observations are not strictly a homogeneous series they depart from one so slightly that we may accept the inference that there is a symmetrical departure from the normal law but probably not an asymmetrical one. In any case there would be no justification for preferring Type IV to Type VII, since the former would make little difference in the range where there seems to be a possible departure from the latter.

5. *The fitting of Pearson curves of Types I, II, IV, and VII.* Most of the difficulty of fitting Pearson curves by the method of maximum likelihood appears to arise from the fact that in their usual form they are stated in terms of parameters whose errors are not independent. We have seen that if the scale of a law of Type II or VII is to remain of the same order of magnitude for all values of m, a factor of order m must be associated with σ^2. But if this factor is simply m, the effect of increasing σ is to lower the curve at small deviations and raise it at large ones (the outside factor being adjusted to make the area unity). But the effect of reducing m for Type II, or increasing it for Type VII, is qualitatively just the same; the departure only changes sign at a different deviation. The determination of m requires that these two effects, which are strongly correlated, must be separated, and consequently a large number of figures must be kept in the computations. Solution would be expected to become much more expeditious if we could state the laws in such a way that the effects of small variations in the unknowns are nearly orthogonal. It appears that this can be done by modifying the parameters of location and scaling. In Type I, for instance, we may write

$$y \propto \left(1-\frac{(x-a)^2}{2M\sigma^2}\right)^m \left(\frac{1-(x-a)/\sigma\surd(2M)}{1+(x-a)/\sigma\surd(2M)}\right)^p, \tag{1}$$

where M will be a function of m and p, tending to m when m is large; a is a parameter of location. Suppose that we have a set of trial values of a, σ, m, and p and denote departures from these by accents. Then the effects of changes of m and σ on y are symmetrical, those of changes of a and p antisymmetrical. Hence, apart from random errors, the logarithm of the likelihood will contain no terms in $a'\sigma'$, $a'm'$, $p'\sigma'$, or $p'm'$. But if $M = m$, there will be terms in $m'\sigma'$, and there will always be terms in $a'p'$. But if we choose M suitably it may be possible to make the terms in $\partial M/\partial m$ cancel those in $\partial(\log L)/\partial m$ that involve σ', and $\log L$ will contain no terms in $m'\sigma'$. Similarly if we put

$$a = b + k\sigma\surd(2M), \tag{2}$$

* Except, of course, the type of evidence relating to the survival value of the Bread-and-Butter-Fly. "But that must happen very often", Alice remarked thoughtfully. "It always happens", said the Gnat.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

where k is a suitable odd function of p, and write down the equations of maximum likelihood for b and p instead of for a and p, it may be possible to choose k so that there will be no terms in $b'p'$ in $\log L$, and therefore no term in p' in the equation for b' and no term in b' in the equation for p'. Then the equations of maximum likelihood will determine all of b', σ', m' and p' separately, and the errors of these estimates will be independent. Thus even though no strictly sufficient statistics exist a small change in any of the unknowns will introduce only a second order change in the estimate of any other, and it may be possible to approximate rapidly by iteration. It is not possible to assign M and k in advance so that the product terms will cancel exactly for all distributions of observations, but it is possible to assign them so that the expectations of these terms, on the hypothesis that the law used is correct, will vanish. The expectation of $\frac{\partial^2}{\partial\sigma\,\partial m}(\log L)$, for instance, will be the expectation of $\Sigma\partial^2\log y/\partial\sigma\,\partial m$ from the various observed values of x, and therefore is

$$n\int y\frac{\partial^2\log y}{\partial\sigma\,\partial m}\,dx,$$

where n is the number of observations. If we can choose M to make this integral vanish for all values of m and σ we can satisfy our conditions. The product terms in $\log L$, as found from any actual set of observations, will represent only random error and may be omitted; they will be of order $n^{\frac{1}{2}}$ and will give only errors of order n^{-1} in the estimates.

Consider first the symmetrical law of Type II, which we take in the form

$$y=\frac{(m+\frac{1}{2})!}{(2\pi M)^{\frac{1}{2}}\,m!\,\sigma}\left(1-\frac{(x-a)^2}{2M\sigma^2}\right)^m \qquad -\surd(2M)\,\sigma<x-a<\surd(2M)\,\sigma. \tag{3}$$

The outside factors contribute nothing to $\partial^2\log y/\partial\sigma\,\partial m$; we find

$$\frac{\partial^2}{\partial m\,\partial\sigma^2}\log y=\frac{(x-a)^2}{2M\sigma^4-\sigma^2(x-a)^2}-\frac{2mM'(x-a)^2}{\{2M\sigma^2-(x-a)^2\}^2}, \tag{4}$$

and
$$\int y\frac{\partial^2\log y}{\partial m\,\partial\sigma^2}\,dx=\frac{1}{2\sigma^2}\left\{\frac{1}{m}-\frac{M'}{M}\frac{m+\frac{1}{2}}{m-1}\right\}. \tag{5}$$

Hence
$$\frac{M'}{M}=\frac{m-1}{m(m+\frac{1}{2})}=\frac{3}{m+\frac{1}{2}}-\frac{2}{m}, \tag{6}$$

$$M=(m+\tfrac{1}{2})^3/m^2, \tag{7}$$

if our conditions are to be satisfied. The arbitrary factor can be absorbed into σ.

The integral diverges at the limits if $m\leqslant 1$, and the method fails. But for $m=1$ the curve cuts the axis of x at a finite angle (being then a parabola); for $0<m<1$ it cuts it normally; for $m=0$ it is rectangular; and for $-1<m<0$ it is U-shaped. Below $m=-1$ the parameters cannot be adjusted to give the curve a finite area. For all values from

$m = -1$ to 1 the extreme observations give estimates whose errors decrease with n more rapidly than $n^{-\frac{1}{2}}$, and it is best then to introduce the termini directly as unknowns. The present method therefore covers all cases where accuracy of order $n^{-\frac{1}{2}}$ is as much as is attainable.

The appropriate form for Type II is therefore

$$y = \frac{(m-\frac{1}{2})!}{\{2\pi(m+\frac{1}{2})\}^{\frac{1}{2}}(m-1)!\,\sigma}\left\{1-\frac{m^2(x-a)^2}{2(m+\frac{1}{2})^3\sigma^2}\right\}^m. \tag{8}$$

It remains significant down to $m = 0$, when the outside factor is 0 and the permitted range is infinite. The permitted range, for given σ, is a minimum for $m = 1$, the parabolic law, which is the extreme of the range where this adjustment is useful.

For Type VII we may take similarly

$$y = \frac{(m-1)!}{(2\pi M)^{\frac{1}{2}}(m-\frac{3}{2})!\,\sigma}\left\{1+\frac{(x-a)^2}{2M\sigma^2}\right\}^{-m}. \tag{9}$$

Then
$$\frac{\partial^2}{\partial m\,\partial\sigma^2}\log y = \frac{(x-a)^2}{2M\sigma^4+\sigma^2(x-a)^2} - \frac{2mM'(x-a)^2}{\{2M\sigma^2+(x-a)^2\}^2}. \tag{10}$$

To perform the integration we make the substitution

$$\frac{(x-a)^2}{2M\sigma^2+(x-a)^2} = z, \tag{11}$$

and find
$$\int y\frac{\partial^2 \log y}{\partial m\,\partial\sigma^2}\,dx = \frac{1}{2\sigma^2}\left\{\frac{1}{m}-\frac{M'}{M}\frac{m-\frac{1}{2}}{m+1}\right\}, \tag{12}$$

whence
$$\frac{M'}{M} = \frac{m+1}{m(m-\frac{1}{2})} = \frac{3}{m-\frac{1}{2}}-\frac{2}{m}, \tag{13}$$

$$M = (m-\tfrac{1}{2})^3/m^2. \tag{14}$$

We therefore take the law in the form

$$y = \frac{m!}{\{2\pi(m-\frac{1}{2})\}^{\frac{1}{2}}(m-\frac{1}{2})!\,\sigma}\left\{1+\frac{m^2(x-a)^2}{2(m-\frac{1}{2})^3\sigma^2}\right\}^{-m}. \tag{15}$$

The integral in (12) always converges for $m > \frac{1}{2}$, and therefore over the whole range of the type.

A similar treatment was applied to the general law of Type I, which was taken in the form

$$y = \frac{(2m+1)!}{2^{2m+\frac{3}{2}}\sigma\surd M(m+p)!\,(m-p)!}\left(1-\frac{x-a}{\sigma\surd(2M)}\right)^{m+p}\left(1+\frac{x-a}{\sigma\surd(2M)}\right)^{m-p} \tag{16}$$

$$-\sigma\surd(2M) < x-a < \sigma\surd(2M).$$

All the integrals are reduced to complete Beta-functions by the substitution

$$x-a = \sigma\sqrt{(2M)}\,(2\theta-1). \tag{17}$$

It is found that $$\int y\frac{\partial^2 \log y}{\partial\sigma\,\partial m}\,dx = \frac{2m+4p^2}{m^2-p^2} - \frac{M'}{M}\frac{(2m+1)\,(m-1+p^2)}{(m-1)^2-p^2}, \tag{18}$$

whence the condition that the errors of σ and m shall be independent is

$$\frac{1}{M}\frac{\partial M}{\partial m} = \frac{2m+4p^2}{m^2-p^2}\frac{(m-1)^2-p^2}{(2m+1)\,(m-1+p^2)}. \tag{19}$$

This does not integrate in any convenient form. If p^2 is neglected it reduces to (13). If we retain p^2/m but neglect p^2/m^2, we find, approximately,

$$M = (m+\tfrac{1}{2})^3\, m^{-2} \exp(-3p^2/m). \tag{20}$$

The maximum of y and the expectation of x are respectively at

$$x-a = -\frac{p\sigma\sqrt{(2M)}}{m}; \quad x-a = -\frac{p\sigma\sqrt{(2M)}}{m+1} \tag{21}$$

exactly. But let us make the substitution (2) and try to adjust k so that the errors of b and p will be independent. Our condition will be

$$\int y\frac{\partial^2 \log y}{\partial b\,\partial p}\,dx = 0. \tag{22}$$

We shall have $$\frac{\partial \log y}{\partial b} = -\frac{\partial \log y}{\partial x}, \tag{23}$$

but it must be remembered that k and M are both functions of p. The result is that k must satisfy the equation

$$(m-1)\left(\frac{\partial k}{\partial p} + \frac{k}{2M}\frac{\partial M}{\partial p}\right) = \frac{(m-1)^2-p^2}{m^2-p^2} + \frac{p}{2M}\frac{\partial M}{\partial p}. \tag{24}$$

But $\partial M/\partial p$ is of order p; hence if we neglect p^2 we have

$$k = \frac{m-1}{m^2}p, \tag{25}$$

and $x-a$ in (1) is best replaced by $x-b-(m-1)\,p\sigma\sqrt{(2M)}/m^2$. The best parameter of location is therefore b, which is near the mean, differing from it by a quantity of order m^{-3}; it differs from the mode by a quantity of order m^{-2}. Considerations of convergence indicate, as for Type II, that one or both termini should be introduced as unknowns directly if $m+p$ or $m-p$ is $\leqslant 1$.

A closer approximation may be attempted by using (20), which gives

$$\frac{1}{2M}\frac{\partial M}{\partial p} = -\frac{3p}{m}, \tag{26}$$

and (24) becomes
$$(m-1)\left(\frac{\partial k}{\partial p}-\frac{3pk}{m}\right) = \frac{(m-1)^2-p^2}{m^2-p^2}-\frac{3p^2}{m}. \tag{27}$$

Put
$$k = \frac{m-1}{m^2}p+u. \tag{28}$$

Then
$$(m-1)\left(\frac{\partial u}{\partial p}-\frac{3pu}{m}\right) = \frac{(m-1)^2-p^2}{m^2-p^2}-\frac{3p^2}{m}-\frac{(m-1)^2}{m^2}+\frac{3p^2(m-1)^2}{m^3}$$
$$= O(p^2/m^2), \tag{29}$$

all terms of higher order in m cancelling. Thus u will be of order p^3/m^3, and we have already neglected p^2/m^2 in obtaining (20). We cannot therefore improve on (25) without a much more elaborate treatment, which does not appear worth while.

To determine a law of Type I, which does not depart too far from symmetry, we may therefore take the origin of location at the theoretical mean and define M by (20). The errors of b, p, σ and m will then be nearly independent. Similar considerations will apply to Type IV. Detailed attention to the unsymmetrical laws does not, however, seem worth while for the present purpose, since (1) the estimates of p will have large uncertainties in any case, (2) the true uncertainties will be larger on account of failure of the hypothesis of independence, (3) there is a systematic personal error of the mean, so that asymmetry about the true value is already established, and asymmetry about the mean will hardly affect the mean if this is taken as the standard.

The practical fitting of Types II and VII can now be done as follows. We begin by estimating the mean and σ as for the normal law; then, for Type II, the equation to estimate m is

$$\frac{\partial}{\partial m}\Sigma \log y = 0, \tag{30}$$

or
$$n\left[\frac{d}{dm}\log(m-\tfrac{1}{2})!-\frac{d}{dm}\log(m-1)!-\frac{1}{2(m+\frac{1}{2})}\right]$$
$$-\Sigma\log\left(1-\frac{m^2(x-a)^2}{2(m+\frac{1}{2})^3\sigma^2}\right)+\Sigma\frac{m^2(m-1)(x-a)^2}{2(m+\frac{1}{2})^4\sigma^2\{1-m^2(x-a)^2/2(m+\frac{1}{2})^3\sigma^2\}} = 0. \tag{31}$$

The digamma functions were taken from the British Association Tables. The functions of m can be computed irrespective of σ. The function on the left can then be computed for a set of trial values of m, using the approximate values of a and σ. The value of m that makes it zero is then found by interpolation. There is one slight complication, since the asymptotic expression for the factor multiplying n is $\frac{3}{8m^2}-\frac{1}{8m^3}+O(m^{-4})$; and this gives a fair approximation down to quite small values of m. The result is that all terms vary rapidly with m, and linear interpolation would be dangerous. This can be avoided by multiplying the equation by m^2. There are other reasons for this device. The change in y for moderate values of $(x-a)/\sigma$ is proportional, nearly, to $1/m$. Let us then put $\mu = 1/m$. The problem of fitting a distribution of chance to a series of observed frequencies is fundamentally a generalized one of sampling. The ultimate problem is to estimate the chances in different ranges of the argument. We could get an estimate of m by simply comparing the number of observations in the range where its effect is positive with the whole number, and this would be a problem of simple sampling. But in sampling we ordinarily take the prior probability of a chance as uniformly distributed, and as the larger chances here are linear in μ we should take that of μ as uniformly distributed if we are to be consistent. Hence the posterior probability density for μ would be proportional to $\Pi(y)$, and would be greatest where

$$\frac{\partial}{\partial\mu}\Sigma \log y = 0, \tag{32}$$

which is equivalent to multiplying (31) by $-m^2$. The probability distribution for μ, given the observations, will be nearly normal. This does not apply to m, especially when m is large. The standard error of μ will then be given by

$$\frac{1}{\sigma^2(\mu)} = -\Sigma \frac{\partial^2}{\partial\mu^2} \log y, \tag{33}$$

which is easily estimated by differencing.

For Type VII the equation is

$$n\left(\frac{d}{dm}\log m! - \frac{d}{dm}\log(m-\tfrac{1}{2})! - \frac{1}{2(m-\frac{1}{2})}\right) - \Sigma \log\left(1+\frac{m^2(x-a)^2}{2(m-\frac{1}{2})^3\,\sigma^2}\right) + \frac{m^2(m+1)\,(x-a)^2}{2\sigma^2(m-\frac{1}{2})^4\{1+m^2(x-a)^2/2(m-\frac{1}{2})^3\,\sigma^2\}}. \tag{34}$$

The asymptotic expansion of the coefficient of n is $-\dfrac{3}{8m^2}-\dfrac{1}{8m^3}$, and it is again desirable to multiply by m^2 before interpolation. In the following table the sign of μ is reversed for laws of Type II, since these laws give departures from normal in the opposite sense from those of Type VII.

		μ	m	$\mu^2/\sigma^2(\mu)$	χ^2	m (by moments)
Bisection	1	$0{\cdot}111 \pm 0{\cdot}037$	9·0	9·0	3·6	16·6
	2	$0{\cdot}04? \pm 0{\cdot}04$	25??	1?	0·2	132
	3	$-0{\cdot}225 \pm 0{\cdot}057$	4·5 (II)	15·6	6·1	4·4
Bright line	1	$0{\cdot}230 \pm 0{\cdot}057$	4·3	16	6·5	5·2
	2	$0{\cdot}163 \pm 0{\cdot}050$	6·1	10·6	0·7	5·0
	3	$-0{\cdot}080 \pm 0{\cdot}049$	7·3 (II)	2·7	7·4	4·6
Bond		$0{\cdot}123 \pm 0{\cdot}051$	8·2	5·8	5·6	8·67

For comparison I give the values of χ^2 for the symmetrical departure found earlier in this paper by extended grouping, and the values of m found by the method of moments. The first six were taken from Pearson's paper, the last from a discussion by Mr Hey. In this a Type VII law was fitted directly.

The column $\mu^2/\sigma^2(\mu)$ gives the square of the ratio of μ to its standard error as found by the most accurate method; χ^2 gives the same ratio as it would be found by using the totals for the ranges where the departure from normal keeps the same sign, the ranges being chosen to give the maximum efficiency. It is seen that the former values are larger in all cases but one, in four cases considerably so. It appears, therefore, that the method of extended grouping involves considerable loss of accuracy when the trial distribution of chance is far from uniform. It has, however, been thought desirable to give the results of the method as an illustration. The estimates for the central and flank groups should be good ones. It was doubtful whether the treating of all observations beyond $\pm 2{\cdot}48\sigma$ together would sacrifice much of the information contained in them. It appeared at first that most of this information would have been used in computing the standard error, and that the rest might be no more important than when the trial distribution is uniform, when the effect of grouping in this way would only reduce χ^2 by about 10%. Apparently this is not the case. Two deviations of 3σ and one of 4σ would produce about the same effect on the calculated standard error, and therefore on the expectations in the central and flank groups, but apparently still produce very different effects on the estimated departure from normal. The departure shown by this method for the second bright-line series is entirely due to the separate allowance for the two extreme observations.* It appears that χ^2 is necessarily unsatisfactory when the expectation is small, however it is used; its status is that of a rough test, which is easy to apply and often supplies as much information as is needed, but at best it is an approximation and at worst an extremely bad one.

The values of m found by Pearson all allow for asymmetry. If we had a law of Type I or IV we could reduce it to a symmetrical one with the same m by taking the geometric means of the chances for equal and opposite values of $x-a$. Hence if there is asymmetry there will be a second order bias in the estimate of m found in this way; we should use the geometric mean of the chances at the tails, but we do use the arithmetic one, which is larger, and the result is that we overestimate the symmetrical departure from normal if the law is of Type IV and underestimate it if it is of Type I. Pearson's values of m may therefore be expected to be larger than mine when I find Type VII, smaller when I find Type II. There are some signs of this, but they are not decisive. Hey, using the method of moments on Bond's data, got $m = 9{\cdot}22$, $2p = 2{\cdot}40$ for Type IV, $m = 8{\cdot}67$ for Type VII. The difference here is in the expected direction. Incidentally Pearson gives his values of m to six or seven figures, of which the first is doubtful and the second practically meaningless. In the above work, except for the preliminary calcula-

* These are so far from the main group that one might consider the presence of an abnormal source of error. But this would reintroduce the question of rejection of observations, and would say nothing about what we are to do in the practical case, where such observations are liable to occur.

tion of the functions of m, which was done to five figures, four-figure accuracy was found ample. More figures would be needed for laws that approach the normal more closely, but the number of observations needed to establish a departure from the normal law would have to be much larger. The digamma functions were originally found to seven decimals, of which the first two, for $m = 10$, disappeared on computing the function actually needed.

I expected originally to find that Pearson's values of m, for the larger departures from normal of Type VII, would be strongly biased, since the method of moments must fail altogether at $m = \frac{5}{2}$. But the relevant cases are the first and second bright-line series, and for these his solutions differ from mine in opposite directions. The differences are indeed rather small on the whole. If we take μ from Pearson's m, the greatest difference from my solution is for the first bisection series. The two values of μ are $0{\cdot}111 \pm 0{\cdot}037$ and $0{\cdot}060$, so that the difference is $1{\cdot}4$ times my standard error. This happens, however, to be the second smallest departure from normal. Pearson's own uncertainties are much too small. Thus he shows that the standard error of β_2 involves the eighth moment, and this becomes infinite for $m \leqslant 4{\cdot}5$ for Type VII. But he does not give standard errors as found from the actual moments or even from his inferred values of the parameters; he gives them as for random variations from the normal law. Hence his uncertainties for β_2, while they afford evidence against the normal law, are not those of the estimates. These estimates of uncertainty are still extensively used, but are quite misleading.

The method of estimating m makes its error independent of that of σ to the first order, but not to the second. Consequently it is desirable to test the series that give the extreme values of μ to see whether any appreciable error has accumulated in the estimate of σ. For the first bright-line series (Type VII) I took $m = 4{\cdot}5$ and solved for σ. The result was $\sigma = 4{\cdot}658 \pm 0{\cdot}175$. The first approximation, assuming the normal law, was $4{\cdot}7566$, so that the difference is under $0{\cdot}6$ of the standard error. For the third bisection series (Type II) I took $m = 4{\cdot}5$ and got $\sigma = 2{\cdot}636 \pm 0{\cdot}068$. The first approximation was $2{\cdot}625$, and the difference is $\frac{1}{6}$ of the standard error. It is therefore unnecessary to revise the estimates of σ for the smaller departures from normality, and as any further correction to m will be proportional to the square of the error of σ it is also unnecessary to revise m.

With regard to Fisher's statement that the method of moments is efficient for small departures from the normal law, it needs to be noticed that the approximation depends on the expansion of $\log y$ in a series of powers of μ and p; but on inspection we see that this is also expansion in powers of x. To estimate the uncertainties correctly we must go to the second order in μ and p, and this introduces the fifth and sixth moments. But what is even more serious is that the logarithmic series diverges if $|x-a| \geqslant \surd(2M)\,\sigma$, and converges slowly before this limit is reached. To simplify the arithmetic a little, let us take the laws in the forms

$$y \propto \left(1+\frac{x^2}{m}\right)^{-m}; \quad y \propto \left(1-\frac{x^2}{m}\right)^{m}.$$

Let us suppose that the degree of slowness of convergence that we shall tolerate is given by $x = k$, where $k/\sqrt{m}$ is moderate. The chance of an observation beyond $\pm k$, for Type VII, is greater than that for the limiting normal law, which is approximately $\exp(-k^2)/k\sqrt{\pi}$. This will remain a good approximation for Type II if the index is large. The chance that at least one observation out of n will lie beyond $\pm k$ will exceed $\frac{1}{2}$ is

$$n > \tfrac{1}{2}\sqrt{(\pi m)}\,(\tfrac{5}{4})^{m-1}\log 2$$

if we take $k = \frac{1}{2}\sqrt{m}$ as our limit. On the other hand, if m is to be shown to be finite $\beta_2 - 3$ must exceed its standard error substantially. In this case we may use Pearson's uncertainty; we have

$$\beta_2 = 3\frac{m-\frac{3}{2}}{m-\frac{5}{2}}; \quad \sigma(\beta_2) = \sqrt{\left(\frac{24}{n}\right)},$$

whence we may reasonably require, to establish departure from the normal, that n shall exceed $3m^2$ considerably. The following table gives some representative values showing n_1, the maximum number of observations that may be expected to avoid slow convergence, and $n_2 (= 3m^2)$ the number that must be exceeded if any determination of such a value of m is to be possible.

m	n_1	n_2
9	11	343
25	650	1875
49	193,000	7203
100	$2 \cdot 4 \times 10^{10}$	30,000

If in fact $m = 25$, we shall need well over 2000 observations to detect it; but in 650 observations there will ordinarily be some beyond the range where the series gives a satisfactory approximation. It will be only at such large values of m that the number of observations needed to show that m is not infinite will be 10,000 or more that the approximation by series will be of any use. The trouble is that it is biased in dealing with the outlying observations, which are precisely the ones of most use in finding m.

We can illustrate this from the above series. The danger line is at $x - a = \frac{1}{2}\sigma\sqrt{(2M)}$, and the series will diverge beyond $\sigma\sqrt{(2M)}$. We tabulate the latter and the departure of the extreme observation from the mean.

		$\sigma\sqrt{(2M)}$	Extreme
Bisection	1	9·7	8
	3	9·2	7
Bright line	1	11·5	25
	2	14·4	25
	3	30·6	27
Bond		17·0	17

The second bisection series is omitted as being practically normal. We see that in the first and second bright-line series the extreme deviation is about twice what would permit convergence of the logarithmic series at all; for Bond's series it is a shade beyond the limit; for the first bisection series it is nearly up to it and much larger than could be considered to give satisfactory convergence. These are all cases of Type VII. For Type II observations beyond the range of convergence are impossible, but in the third bisection and third bright-line series, which are of this type, the extremes approach the limit. It appears therefore that none of these series is near the normal in the sense required for Fisher's highly qualified approval of the method of moments; yet one of them has been published as an instance of the normal law.

6. *Departure from independence*. Pearson gives for each series of observations the means by groups of 25 to 37 consecutive ones, and finds that the means fluctuate much more than would be expected on the hypothesis of independence. In an illustration of a test for independence (Jeffreys 1938*c*) I have confirmed this statement. It seemed possible, however, that this might be due to failure of the usual rule for the uncertainty of a mean, analogous to Fisher's result for errors derived from a Type VII law with index 1, since some of the series agree with this type with rather small indices. On comparison, however, I found just the opposite result to what this would imply. In the following table I give the results for μ, with regard to sign, arranged in descending order, and the corresponding values of γ^2, the ratio of the square of the standard error of a group mean to that of one observation. For independence γ^2 would be 0·04 or less.

	μ	γ^2	r
Bright line 1	+0·230	0·066	0·16
2	+0·163	0·100	0·24
Bisection 1	+0·115	0·093	0·23
2	+0·04?	0·362	0·57
Bright line 3	−0·080	0·140	0·32
Bisection 3	−0·225	0·550	0·72

The order of decreasing μ is practically that of increasing γ^2. In other words, the more nearly the condition of independence is satisfied, the further the law of error recedes from the normal along the Type VII series. To test the matter further, I subtracted 0·04 from each value of γ^2, this being approximately its expectation on the hypothesis of independence, and took the square root, thus obtaining estimates of the range of the non-random variation. These roots are given as r above. The correlation between μ and r has the astonishing value of $-0·92$; it is particularly surprising since the values of both have quite appreciable uncertainties, and the true correlation may be practically perfect. It looks as if drifting of the personal equation may account for practically the whole of the variation of μ.

It can be seen qualitatively that the effect would be of this type, if we consider two superposed normal distributions of equal totals about different means. If we have a normal law about $x = d$ with standard error σ, the second moment about $x = 0$ is

σ^2+d^2 and the fourth $3\sigma^4+6d^2\sigma^2+d^4$. If we have two about $\pm d$, each with total area $\frac{1}{2}$, their totals will be the same. Thus

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3(\sigma^2+2d^2)\,\sigma^2+d^4}{(\sigma^2+d^2)^2} = 3-\frac{2d^4}{(\sigma^2+d^2)^2}.$$

Now for a Type II distribution $\beta_2 = 3\dfrac{m+\frac{3}{2}}{m+\frac{5}{2}}$,

and therefore if we try to fit a Pearson law to a compound distribution of this type we shall infer that it is of Type II. Further, our estimate of μ_2 will really be σ^2+d^2, while the γ^2 of the present work will be $d^2/(\sigma^2+d^2)$. The extreme value of γ^2 is about 0·5, and therefore we may take $d^2=\sigma^2$, and therefore $\beta_2=\frac{5}{2}$. This would be interpreted as a Type II law with $m = 3{\cdot}5$, which is beyond the range of m actually found. If we take the extreme value $\mu = -0{\cdot}225$ and add $\frac{2}{7}$ to allow for the possible effect of non-independence we get $\mu = +0{\cdot}061$. Further, if the total probabilities in the two component laws are unequal we shall find a third moment different from zero and infer that the law is unsymmetrical; so that lack of independence may also explain asymmetry, if any.

A further precaution is needed before we accept this explanation, since γ^2 is inferred from several groups, and if the probabilities of the group means were distributed normally the resultant law would again be normal. We may therefore have overestimated the effect. But the distribution of the group means is not random. If we subtract from each the mean for its series and examine the signs, there are altogether 67 persistences and 40 changes, indicating that the fluctuation affects more than one consecutive group. Six changes have been introduced by the allowance for the mean; on correcting for this the numbers become 67 and 34. This suggests a possibly irregular but nevertheless continuous variation, which cannot possibly give the kind of distribution of magnitudes that the normal law does; it would give concentrations at definite distances from the mean. The previous discussion gave τ, the standard deviation of the group means. If these varied harmonically the range between the extremes would be $2\sqrt{2}\tau$; if they came from a rectangular distribution of chance it would be $2\sqrt{3}\tau$. The following table shows the comparison, in Pearson's original units.

		Range	τ	$2\sqrt{2}\tau$	$2\sqrt{3}\tau$
Bisection	1	0·027	0·0075	0·022	0·026
	2	0·055	0·0185	0·053	0·064
	3	0·064	0·0195	0·055	0·068
Bright line	1	1·066	0·304	0·86	1·06
	2	1·333	0·371	1·04	1·29
	3	2·34	0·682	1·94	2·37

The ranges between the extreme values are therefore in very good agreement with what we should expect from a rectangular distribution; they are both smaller and in a much steadier ratio to τ than we should expect from a normal one. Let us therefore repeat the previous work on the supposition that the mode fluctuated between $\pm k$ according to the rectangular law. For a given element of the range of the mode, da, the contributions to the zero, second, and fourth moments will be

$$\frac{1}{2k}da, \quad (\sigma^2+a^2)\frac{da}{2k}, \quad (3\sigma^4+6\sigma^2a^2+a^4)\frac{da}{2k}$$

and on integration we shall have

$$\mu_2 = \sigma^2+\tfrac{1}{3}k^2; \quad \mu_4 = 3\sigma^4+2\sigma^2k^2+\tfrac{1}{5}k^4.$$

But $$\tau^2 = \tfrac{1}{3}k^2;$$

and therefore $$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3-\frac{6}{5}\frac{\tau^4}{(\sigma^2+\tau^2)^2}.$$

With $\gamma^2 = \frac{1}{2}$, therefore, we should find $\beta_2 = 2{\cdot}700$, and this would be interpreted as coming from a Type II law with $m = 7{\cdot}5$. The correction to μ will therefore be $+0{\cdot}133$ and the corrected result for the third bisection series $-0{\cdot}092$. Uniform distribution of a fluctuating personal error over a range will therefore not account directly for the whole of the variation of μ between these series. Nevertheless the value of μ is so highly correlated with the degree of independence that it appears to be legitimate to try to extrapolate to complete independence ($r = 0$) on the hypothesis of linearity. The line of regression of μ on r, taken to $r = 0$, gives $\mu = +0{\cdot}306$. But this has uncertainties from the means of μ and r, and from the uncertainty of the correlation coefficient, which is not symmetrically distributed since it is so near -1; Fisher's transformation (1936, p. 200) suggests that limits corresponding to the standard error in a normal distribution would be at $0{\cdot}77$ to $0{\cdot}97$, giving the extrapolated μ as $0{\cdot}262$ to $0{\cdot}321$. These may be taken as the correct limits, since their range is three times that given by the uncertainties of the means. If the relation inferred above between independence and the value of μ is general, therefore, we may say that the law of error, if complete independence held, would be a Type VII one with index probably between 3 and 4.

7. This would have some remarkable consequences. If a set of observations appears to satisfy the normal law, it would be evidence that they are not independent and that all uncertainties found from them by the usual formulae are too low. It is, of course, well known that such a set *may* give too low an uncertainty on account of dependence; but if this suggestion is right it *must*. Now it often happens that two estimates by different methods or even by the same method differ by larger amounts than would be expected from their apparent standard errors, and the difference is usually attributed to some systematic error. But if the law of error is as above, and the estimates are those adapted to the normal law, many of these differences may quite well be random error. It is said that the probability of the error of the mean will be normally distributed, even if those of the separate observations are not, but the argument that leads to this involves the same approximations as are made in the proof of the normal law for one observation.

There is no reason to suppose that it is valid for large errors, merely that it will usually be right for moderate ones. An extreme case to the contrary is provided by the law $(1+x^2)^{-1}$; for which the mean of any number of observations has the same probability law as one observation. But for any law of the form $(1+x^2)^{-m}$ the even moments of order $2m-1$ and more are infinite, and those of the same orders for the mean of any finite number of observations will also be infinite. Hence large random errors of the mean must occur more often than the normal law indicates. For the scale of the law the difference is even greater. Fisher shows (1922, p. 342) that the efficiency of the second moment for determining the scale is, in his notation, $1-\frac{12}{r(r+1)}$, where his r is my $2m-2$; so that the efficiency is $1-\frac{3}{(m-1)(m-\frac{1}{2})}$. We have the following values: $m\leqslant 2\cdot 5$, 0; $m=3$, 0·40; $m=3\cdot 5$, 0·60; $m=4$, 0·72. If the number of observations is not very large, therefore, the scale may be quite wrongly determined. This is just the case where the posterior probability distribution of the true value differs most greatly from the normal, even when the fundamental law is normal; the reason being that this distribution contains contributions from all values of σ, and the integration with regard to σ seriously alters the form of the law. Large differences of the true value from the mean are more probable than they would be if the estimate of σ was an exact determination, because it may be too low, and if it is, large errors of the mean, in relation to the estimated scale, will be more probable. This point was first noticed by "Student" (1908), and I have rediscussed it elsewhere (1937*c*). But if in addition the law is such that the error of the scale may be magnified by $2\cdot 5^{\frac{1}{2}}$ or $(5/3)^{\frac{1}{2}}$ by inefficient methods of fitting, the effect will be greatly increased. As a rough illustration, we may consider the values of t, the ratio of a deviation to its estimated standard error, when the law of error is normal, given in Fisher's table (1936). For an infinite number of observations the 5 % limit is at $t=1\cdot 96$; for 6 observations ($n=5$ in the notation of the table) it is at $t=2\cdot 57$, the difference being due to allowance for the uncertainty of the standard error. If this has to be increased 30 % the limit will be about $t=2\cdot 75$. The corresponding values for the 1 % limit will be 2·58, 4·03, 4·46. This extrapolation is quite rough, but it is enough to show that with the usual methods of fitting, not much confidence need be placed in a discrepancy of about four times the estimated standard error, if the number of observations is small; in other words, in a discrepancy 1·8 times the estimated standard error of one observation.

Returning to 5 (9) we see that the equations for determining a and σ by the method of maximum likelihood are

$$-\frac{\partial}{\partial a}\log L = -\Sigma\frac{\partial}{\partial a}\log y = \frac{m}{M\sigma^2}\Sigma\frac{x-a}{1+(x-a)^2/2M\sigma^2}=0, \tag{1}$$

$$-\frac{\partial}{\partial \sigma}\log L = -\Sigma\frac{\partial}{\partial \sigma}\log y = \frac{n}{\sigma}-\frac{m}{M\sigma^3}\Sigma\frac{(x-a)^2}{1+(x-a)^2/2M\sigma^2}=0. \tag{2}$$

It is obvious that if these are used the errors of the estimates will follow much more closely the laws found from the normal law. They are equivalent to giving weights inversely proportional to $1+(x-a)^2/2M\sigma^2$; thus the importance of the outlying observations is much reduced. But it was the large contributions made by these to the estimates of the mean and standard error by the usual methods that produced the infinite $(2m-1)$th and higher moments for the mean, and the large uncertainty of the second moment. If these estimates are used, the probability of the errors of the estimated a should follow "Student's" rule much more closely. The equations can be solved, for an adopted m, either by using trial values and interpolating, or by starting with the usual solution and proceeding by successive approximation. The results of this paper suggest $m = 3\cdot5$ as a reasonable value. Then $M = 3^3/3\cdot5^2 = 2\cdot204$, and the weight will drop to $\frac{1}{2}$ at about $x-a = 2\cdot1\sigma$. When some hundreds of observations are available it will, of course, be possible to make an independent determination of m and also of the degree of independence of the observations. The latter point is the more important, since it is known that a first order change in the distribution of the weights produces only a second order one in the uncertainties of the estimates.

To estimate the uncertainties we have the rules, since those of a and σ are independent to the first order,

$$\frac{1}{\sigma^2(a)} = -\frac{\partial^2}{\partial a^2}\log L; \quad \frac{1}{\sigma^2(\sigma)} = -\frac{\partial^2}{\partial \sigma^2}\log L. \tag{3}$$

These can be evaluated directly for any actual series; or we can use their expectations as approximations. We find

$$-\int y\frac{\partial^2 \log y}{\partial a^2}dx = \frac{m^3}{(m+1)(m-\frac{1}{2})^2\sigma^2}, \tag{4}$$

$$-\int y\frac{\partial^2 \log y}{\partial \sigma^2}dx = \frac{2m-1}{(m+1)\sigma^2}, \tag{5}$$

whence
$$\sigma^2(a) = \frac{(m+1)(m-\frac{1}{2})^2\sigma^2}{nm^3}; \quad \sigma^2(\sigma) = \frac{(m+1)\sigma^2}{2n(m-\frac{1}{2})}. \tag{6}$$

The infinite uncertainty of the latter for $m = \frac{1}{2}$ is due to the fact that the Type VII law becomes impossible then. These expressions are valid over the whole range of the law, but direct calculation from the observations is desirable in individual cases until more is known about the extent of the random variations.

The second moment μ_2 is given by

$$\mu_2 = \int y(x-a)^2\,dx = \frac{(m-\frac{1}{2})^3\sigma^2}{m^2(m-\frac{3}{2})}, \tag{7}$$

and the standard error of $\bar{x}$ as estimated from it would be $(\mu_2/n)^{\frac{1}{2}}$. Comparing this with (6) we have

$$\frac{\sigma^2(a)}{\sigma^2(\bar{x})} = \frac{(m+1)(m-\frac{3}{2})}{m(m-\frac{1}{2})} = 1-\frac{3}{2m(m-\frac{1}{2})} = 1-\frac{6}{(r+1)(r+2)}, \tag{8}$$

which agrees with Fisher's expression (1922, p. 341). The standard error as estimated from μ_2 is therefore unnecessarily large by a factor of $(\frac{10}{7})^{\frac{1}{2}}$ for $m = 2{\cdot}5$; $(\frac{5}{4})^{\frac{1}{2}}$ for $m = 3$; $(\frac{7}{6})^{\frac{1}{2}}$ for $m = 3{\cdot}5$; $(\frac{28}{25})^{\frac{1}{2}}$ for $m = 4$. This, however, is only an average rule. The real danger of using μ_2 is that the probability distribution of μ_2 for *one* set of observations, given all the parameters in the law, falls off much less rapidly than for the normal law and consequently large differences from expectation will occur oftener.

It will be noticed that the expectation of μ_2 is a little greater than σ^2, so that successive approximations will ordinarily decrease the estimate of the latter slightly. The standard error of a is a little less than $\sigma/\sqrt{n}$. But these are only average rules, and in individual cases the differences may be much larger and in either direction.

It may be remarked that if observations of different precisions satisfying the normal law are combined, the effect is to give β_2 greater than 3, and therefore would imitate a law of Type VII without the normal law being in fact wrong. I think, however, that the onus of proof, since Pearson's paper, is on those who say that it is ever right, and in any case such a consideration would affect nothing in the treatment of actual observations. We could make no use of the fact that a series of observations was compound unless we knew which observations belonged to each component law. If we did, we should treat the laws separately; but when, as is certainly the usual case, we do not, it is necessary to use the rules for the type of distribution as we actually find it. In either case it is wrong to treat all observations as of equal weight.

In conclusion I must express my thanks to the late Dr W. N. Bond, whose observations first attracted me to this problem, and whose assistance in the discussion of them was most valuable; to Mr G. B. Hey, who carried out the solution by moments for these data; to Dr Wishart, who repeatedly gave me the benefit of his experience with Pearson curves; and to Mr Yule, who first drew my attention to Pearson's data, which form the main subject-matter of this paper.

Summary

The limitations of the theoretical grounds for accepting the normal law of errors of observation are discussed, and seven series of observations capable of providing tests of its truth are examined. It is found that the χ^2 test, as usually employed, is not sufficiently sensitive to establish departures from the normal law. A wider grouping, however, reduces the random error of χ^2 sufficiently to show the departures clearly, though it is still less sensitive than the ratio of the maximum likelihood solution for the departure to its standard error. It appears that no form of the test is of much use when the law to be tested implies very small expectations in some of the groups.

An approximation to the method of maximum likelihood for Pearson laws of Types II and VII is developed, and extensions to Types I and IV are suggested. The approximation does not involve an excessive amount of labour or the retention of a large

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

number of figures. It is found that the various series of data give laws ranging from Type II with index 4·5 to Type VII with index 4·3, and that the index is closely correlated with the degree of correlation of the errors within groups of successive observations; an extrapolation using this correlation suggests that genuinely independent observations would follow a law of Type VII with index between 3 and 4. Methods of combining observations following such a law and determining their uncertainties are provided. It appears that a number of discrepancies in astronomy and physics that have been accepted as systematic may turn out to be random, since with such a law large random discrepancies may occur more often than with the normal law if the mean and the mean square deviation are still used as estimates.

References

Bond, W. N. 1935 "Probability and Random Errors." London: Arnold.
Brunt, D. 1931 "The Combination of Observations." Camb. Univ. Press.
Bullard, E. C. 1936 *Philos. Trans.* A, **235**, 445–531.
Fisher, R. A. 1922 *Philos. Trans.* A, **222**, 309–68.
— 1936 "Statistical Methods for Research Workers." Edinburgh: Oliver and Boyd.
Jeffreys, H. 1935 *Proc. Camb. Phil. Soc.* **31**, 203–22.
— 1936 *Bur. Centr. Int. Séism. Assn., Trav. Sci.* **14**. Strasbourg.
— 1937*a* "Scientific Inference." Camb. Univ. Press.
— 1937*b* *Proc. Roy. Soc.* A, **160**, 325–48.
— 1937*c* *Proc. Roy. Soc.* A, **162**, 479–95.
— 1937*d* *Mon. Not. R. Astr. Soc., Geophys. Suppl.*, **4**, 165–84, 225–50.
— 1938*a* *Ann. Eugen., Camb.*, **8**, 146–51.
— 1938*b* *Proc. Roy. Soc.* A, **164**, 307–15.
— 1938*c* *Proc. Roy. Soc.* A, **165**. *In the Press.*
Pearson, E. S. 1926 *Biometrika*, **18**, 174–94.
Pearson, K. 1900 *Phil. Mag.* **50**, 157–75.
— 1902 *Philos. Trans.* A, **198**, 235–99.
"Student" 1908 *Biometrika*, **6**, 1–25.
Whittaker, E. T. and Robinson, G. 1924 "The Calculus of Observations." London and Glasgow: Blackie.
Yule, G. U. 1927 *J. Roy. Statist. Soc.* **90**, 570–87.