"DAS FEHLERGESETZ UND SEINE VERALLGEMEINERUNGEN DURCH FECHNER UND PEARSON*." A REJOINDER.

By KARL PEARSON, F.R.S.

THERE is not much profit as a rule in complaining of the treatment one receives at the hands of critics, but still I think the needlessly hostile tone of Dr K. E. Ranke and Dr Greiner's review of my memoir on skew variation requires some protest on my part. As an illustration of the want of courtesy of which I complain I would cite for example the following statement (p. 323):

Er (Pearson) weist aber einmal darauf hin, dass zwar seine eingeschriebenen Kurven teilweise unbegrenzt werden können, das Gesetz selbst, die hypergeometrische Reihe aber nicht, und vertröstet auf eine spätere Arbeit, in der an Stelle der eingeschriebenen Kurven die Anpassung der Reihen selbst gegeben werden solle. Diese spätere Arbeit ist nie geschrieben worden.

This is not an isolated instance of the manner in which the authors criticise my work. It was quite open to them to have examined the customary sources of bibliographical information, or even to have written to me and asked if the memoir in question had been published. But here as elsewhere they assumed, without making proper investigation, that I could say nothing further and therefore had said nothing further. The memoir in question appeared so long ago as 1899 in a well known British scientific journal† from which the authors actually cite another paper of mine. Although my memoir is *nie geschrieben*

* The criticism of my work appeared in the *Archiv für Anthropologie*, Bd. II. pp. 295—331, 1904. The proper place to reply to an attack of the kind would be in the *Archiv* itself. Professor J. Ranke accepted a rejoinder and asked that it should be in German and not exceed 40 pp. I have heartily to thank Miss M. Lewenz for the labour of a translation, which I much regret I cannot make use of, because the Editors of the *Archiv* do not now see fit to publish this reply to K. E. Ranke's attack. As the reply was an endeavour to give an historico-critical account of the theory of skew variation it may interest readers of *Biometrika*, and will possibly reach in the course of time some readers of the *Archiv für Anthropologie*.

† "On certain Properties of the Hypergeometrical Series, and on the fitting of such series to Observation Polygons in the Theory of Chance," *Phil. Mag.* 1899.

worden according to my critics, it was actually written in 1895 and its substance given in academic lectures of the same or the following year; it was not published till some years afterwards owing to that want of leisure for preparing matter for press which every teacher who has to lecture four hours a day will appreciate. Dr Ranke says that had the paper been written, it would not have influenced his judgment. That is quite possible, and I only cite the matter here to indicate the tone adopted by my critics.

A very similar instance occurs in Dr K. E. Ranke's treatment of my fitting of Professor J. Ranke's data for 900 Altbaierisch crania. I used this example purposely because it had already been used by Stieda. I should not myself have mixed, even for the cephalic index, ♂ and ♀ data as Stieda did, but I wished to compare the results reached by the generalised curve with those reached by the Gaussian curve. I actually spoke of the resulting curves in my memoir, whether the generalised curve or the Gaussian, as being "quite good for this type of statistics*." My object of course was to show that the generalised method did not fail where the Gaussian succeeded, but surpassed it. Now how does Dr Ranke treat this instance? He cites an example (actually inserted† by me!) in a memoir by Palin Elderton giving the Ranke'sche Messungen as an illustration of my method of testing goodness of fit in the case of the normal curve. Undoubtedly as I said in 1894 the Gaussian curve is quite good for J. Ranke's data, but it does not follow that the Type IV. frequency curve does not give a *better fit*, and is not significant for constants which the Gaussian process cannot deal with. Now it was open to Dr Ranke to test the values given for the distance from mode to mean, the skewness and the other constants in the case of the Altbaierisch crania. I have given the probable errors of these constants in my memoir: *On the Mathematical Theory of Errors of Judgment*, etc., *Philosophical Transactions*, Vol. 198 A, see p. 278. Had Dr Ranke fairly tested my results he would have found that the asymmetry was not significant, and that the mode sensibly coincided with the mean, but that the constant β_2, which should equal 3 for the normal curve, has a value 3·65 with a probable error of only about ·11. Now this constant and its probable error have no relation at all to any particular theory of variation. They follow quite easily from the general Gaussian theory. It is accordingly extremely improbable that Ranke's measurements are truly given by a Gaussian distribution in all their features. The fact that β_2 is > 3 points to an emphasis

* p. 389.

† The same remark applies to the illustrations of goodness of fit given by Fawcett, cited in footnote Ranke u. Greiner, p. 326. The reference to Powys is inexact; his paper shows that in at least three cases the Gaussian curve is quite impossible. Dr Macdonell's work on the English skull shows that at least in 4 out of 13 cases the asymmetry is significant. "Die englische Schule," by which Dr Ranke refers to workers in my Biometric Laboratory has not discovered a truth which had escaped me; they have shown that the Gaussian curve is of wide applicability, but not of universal truth in anthropometric measurements. This result was reached with a view to testing whether the theory of inheritance, so far as it is based on the Gaussian theory, might be safely applied to human characters. In testing this validity of the Gaussian theory, it was of course needful to have a more general theory from which to determine the chief physical constants involved in non-Gaussian distributions.

of the modal frequency and to a reduction of the extreme frequencies which are inconsistent with the Gaussian curve. This is actually shown in my Plate 11 and referred to in my text, and corresponds to a sensible deviation from the Gaussian law. It was open to Dr Ranke to attribute this exaggeration of the modal at the expense of the extreme variates to heterogeneity in the material. But he had no right when the material for a judgment was before him in my memoirs to conclude that, because the *general* distribution of frequency was not on the average incompatible with the Gaussian law, the deviation of a particular constant of the distribution from its Gaussian value might not be most significant. This case of the deviation of the Ranke's measurements from the Gaussian form is of special interest, for it is not one of asymmetry, but of a non-Gaussian type of symmetry. Dr Ranke suggests that because my test of goodness of fit shows that the Gaussian curve is "quite a good fit," my generalised method of dealing with frequency is idle. Assisted by a mathematician he ought to have recognised that the expression $\beta_2 - 3$ (which measures whether the frequency towards the mean is emphasised more or less than that required by the Gaussian law) had a sensible value, and that my method not only led to the discovery of this deviation but provided a method of allowing for it in the description of the frequency. Had Dr Ranke read my memoir on errors of observation (*Phil. Trans.*, Vol. 198 A, pp. 274—286), he would have recognised that the two tests (*a*) whether special physical constants of the distribution satisfy the Gaussian law, and (*b*) whether the general distribution of frequency satisfies within reasonable limits the Gaussian law, are not necessarily identical. Finally had he concluded that (*b*) for the Altbaierisch crania was satisfied, but not (*a*), and that there was thus no necessary discrepancy between my memoir of 1894 and the statement in Palin Elderton's paper of 1900, he might indeed have fallen back on his customary assumption that when frequency is not Gaussian it is heterogeneous. But at any rate had he adopted this course he would have avoided the appearance of criticising his author without endeavouring to understand what the meaning of his investigations was, or striving to elucidate them by a study of his other memoirs on the same subject.

I do not wish to say anything further on this point. I want merely to indicate by these two out of several cases that the reader must not look for a really impartial statement of my position from Drs Ranke and Greiner. I am quite unable to account for the peculiar tone they have at times given to their criticism.

(2) With this preamble I should like to divide my reply under these headings:

(A) The need for generalised frequency curves, even in anthropological science.

(B) The nature of the assumptions made in the Gaussian theory and their insufficiency.

(C) The hypotheses made to generalise the Gaussian law.

(i) Poisson, Laplace and the early writers.

(ii) Francis Galton and D. McAlister.

(iii) Fechner.

(iv) Edgeworth and Kapteyn.

(v) General Results for Asymmetry*.

(D) The criticisms of my theory by Dr Ranke and Dr Greiner and the reply to be made to them.

A. *The need for Generalised Frequency Curves, even in Anthropological Science.*

I have already pointed out that even Prof. Ranke's measurements are not fully in accordance with the Gaussian theory—for the odds are great against a quantity exceeding its probable error more than five times. It is perfectly true that the English School have found that many characters, especially craniological characters, are for *practical* purposes sufficiently described by the Gaussian curve. But it is equally true that they have found other cases in which the deviation from the Gaussian curve is significant, and that they have only been able to measure this significance because they had a wider theory to base their researches upon. Dr Ranke entirely disregards the statements of Miss Fawcett and Dr Macdonell on this point. Both find a definite number of cases, the one in Egyptian skulls, the other in English skulls, in which the deviation from the Gaussian law is definitely significant†. Both conclude as I have done that in the case of many characters for a variety of practical purposes the Gaussian curve is sufficient; this is, however, not a *theoretical* justification of the Gaussian curve, but an argument in favour of its empirical use in a certain definite number of cases. Dr Ranke may of course say that the exceptions that we have found are due to heterogeneity of our material. If so he must face the difficulty that the *same* set of crania can be homogeneous and give the Gaussian curve for their length and be heterogeneous for their breadth, deviating therein largely from the Gaussian curve. If he asserts that this is quite possible then he must meet the further difficulty that they can be homogeneous for their cephalic indices, which are based upon the ratio of the supposed heterogeneous to the homogeneous material! The fact is that no unprejudiced observer can examine the constants by which we have defined the deviations from the Gaussian law without seeing that they present every variety of value, starting from the values to be expected on the Gaussian theory and rising to values which are absolutely incompatible with any Gaussian theory at all. In fact he must come to the conclusion that some theory is absolutely needful, which will provide a curve or series of curves capable of representing the

* I have left out of consideration the general method of Thiele, followed in Germany by Lipps, because I have dealt with these authors in a recent memoir.

† *Biometrika*, Vol. I. p. 443, and *Biometrika*, Vol. III. p. 227.

fundamental deviations of any distribution from the Gaussian curve and determining whether these deviations are significant or not. Looked at solely from this standpoint—which I am very far from accepting—my curves provide an empirical series which accurately measures the deviations from the Gaussian law and enables the enquirer to determine how far that law is applicable. Each one of them passes into the Gaussian curve if that curve is the better fit to the observations. This is not true of many of the other remedies which have been proposed to supplement what I venture to call the universally recognised inadequacy of the Gaussian law. They cannot as we shall see in the sequel effectively describe the chief deviations from the Gaussian distribution.

The chief physical differences between actual frequency distributions and the Gaussian theoretical distribution are:

(i) The significant separation between the mode or position of maximum frequency and the average or mean character.

(ii) The ratio of this separation between mean and mode to the variability of the character—a quantity I have termed the *skewness*.

(iii) A degree of flat-toppedness which is greater or less than that of the normal curve. Given two frequency distributions which have the same variability as measured by the standard deviation, they may be relatively more or less flat-topped than the normal curve. If more flat-topped I term them *platykurtic*, if less flat-topped *leptokurtic*, and if equally flat-topped *mesokurtic*. A frequency distribution may be symmetrical, satisfying both the first two conditions for normality, but it may fail to be *mesokurtic*, and thus the Gaussian curve cannot describe it.

The Gaussian curve is usually fitted from the mean square deviation, but it may also be fitted from the probable error, or the mean error, or again from the mean fourth power of the deviations—μ_4 in my notation. Whichever method is adopted we ought to get the same result within the errors of random sampling. When I first began to describe frequency data by the normal curve, I was startled to find the very large number of cases in which these different processes led to Gaussian curves, differing widely from one another, i.e. beyond all the limits of probable error. I was soon led to see that in actual statistics two distributions might have equal total frequency, be sensibly symmetrical, and have the same standard deviation and yet differ largely in their flat-toppedness. The mesokurtosis of the Gaussian curve is not a universal characteristic of frequency distributions.

When we test a theoretical distribution of frequency against observation, we may find an excellent fit for the *total* distribution and yet the distinction between mode and mean, the skewness, and the deviation from mesokurtosis may be most significant. The reason for this is that the test for goodness leaves a margin of variation which may be due to random sampling, or to the non-normal character of an important constant of the distribution. For example, 10 coins are tossed a hundred times, and the proportion of cases with five and more heads is somewhat in excess of the theoretical distribution $100(\frac{1}{2}+\frac{1}{2})^{10}$, but within the limits of a random sample. It is quite conceivable that if the returns for each individual coin were analysed it would be found that those of one exceeded in proportion of heads the limits of random sampling, and that the coin proved to be loaded when delicately tested. Thus as I have shown in my memoir on errors of observation, we have not only to test for general goodness of fit, but also to consider the probable errors of the fundamental constants of the distribution. Because the general distribution of frequency is given within the limits of random sampling by a normal curve it does not follow that the system will be mesokurtic.

Consider for example the two curves:

$$y = y_1\left\{1 - \frac{x^2}{2\sigma^2(m_1+\frac{1}{2})}\right\}^{m_1-1}, \quad \text{where:} \quad y_1 = \frac{N}{\sqrt{2\pi}\sigma}\frac{\Gamma(m_1+\frac{1}{2})}{\sqrt{(m_1+\frac{1}{2})}\,\Gamma(m_1)},$$

and

$$y = y_2\left\{1 + \frac{x^2}{2\sigma^2(m_2-\frac{1}{2})}\right\}^{-(m_2+1)}, \quad \text{where:} \quad y_2 = \frac{N}{\sqrt{2\pi}\sigma}\frac{\Gamma(m_2+1)}{\sqrt{(m_2-\frac{1}{2})}\,\Gamma(m_2+\frac{1}{2})}.$$

They are both symmetrical, they both for any value of m_1 or m_2 which is moderately large are indistinguishable in appearance from the Gaussian curve. If they represented actual observations, we should try to fit them (i) by finding the area, (ii) by finding the standard deviation. The former for both curves is N and the latter for both curves is σ. Hence we should fit them with

$$y = \frac{N}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}} \quad \text{..................................(i).}$$

But this in both cases would be incorrect. Both cases would only pass into the Gaussian curve when m_1 and m_2 are theoretically infinite, practically large. No Gaussian fitting could distinguish one of these curves from the other. Why?—Because it does not proceed further than the standard deviation. To measure the difference between either of the above distributions and the Gaussian curve we must proceed to higher moments. Let $N\mu_n$ be the nth moment about the mean, i.e. if $\bar{x}$ be the mean value of x,

$$N\mu_n = \int y(x-\bar{x})^n dx,$$

where the limits of the integral are those of the range. Then if $\beta_2 = \mu_4/\mu_2^2$, we easily find:

$$m_1 = \frac{3(\beta_2-1)}{2(\beta_2-3)} \quad \text{and} \quad m_2 = \frac{3(\beta_2-1)}{3-\beta_2},$$

or

$$\beta_2 - 3 = \frac{6}{2m_1-3}, \qquad 3-\beta_2 = \frac{6}{2m_2+3}.$$

Thus we reach one of the conditions for the Gaussian curve, i.e. $\beta_2 = 3$, in either case when m_1 and m_2 are considerable, but if β_2 be > 3, m_1 will be positive and if $\beta_2 < 3$, m_2 will be positive. Now since $\frac{m_1-1}{m_1+\frac{1}{2}}$ is always less than $\frac{m_2+1}{m_2-\frac{1}{2}}$, it is easy to show that in the neighbourhood of the origin y/y_1 is always greater than y/y_2 for the same value of x. In other words, the first curve is flatter topped than the second, and both lie on different sides of the corresponding Gaussian curve. The first curve type is platykurtic and the second leptokurtic.

Now there is nothing to prevent us fitting curves of the above types to any series of frequency observations. Supposing those observations are truly normal, then m_1 or m_2 will be so large that $\beta_2 = 3$ within the error of random sampling. Now the probable error of β_2 for a Gaussian distribution of total frequency N*:

$$= \cdot 67449 \sqrt{\frac{24}{N}},$$

and if β_2 differs from 3 by several times this probable error, it is absolutely impossible to treat the system as mesokurtic. In any such case one or other of the above curves *must* give a truer representation than the Gaussian curve. It is easy to show that for leptokurtic distributions the maximum frequency is greater than that given by the normal curve and for platykurtic distributions it is less. The Gaussian curve compels us to assert that the product of the maximum frequency into the standard deviation is a constant (i.e. $y_0\sigma = N/\sqrt{2\pi}$). This condition of mesokurtosis is unfulfilled—within the limits of random sampling—for a great variety of frequency distributions.

Further it is absolutely certain that divergencies from the Gaussian or normal curve are not exclusively in the direction of either platykurtic or leptokurtic distributions. Thus the *symmetrical* binomial is essentially leptokurtic, i.e. $\beta_2 < 3$, and therefore cannot be used for a great variety of distributions. In general all skew binomials with $p > \cdot 2113$ and $< \cdot 7887$ are leptokurtic; outside these limits they are platykurtic.

The test whether a curve satisfies the mesokurtic condition has nothing to do with my particular views on frequency, it is merely deduced from the general principles of probability and is a test of normal distribution. Of course there are many other conditions to be satisfied, e.g. μ_{2n} should equal $(2n-1)\mu_{2n-2}$. But as I have shown elsewhere the probable errors of the high moments increase so rapidly, that it becomes easier and easier to satisfy such conditions within the errors of random sampling, and without very large numbers they are of little practical value.

The following are significantly platykurtic distributions:

The Maximum Breadth in English ♂ skulls,

The Nasal Breadth in English ♂ skulls,

The Cephalic Index of Altbaierisch skulls,

The Auricular Height in ♀ Naqada skulls.

* Pearson: *Phil. Trans.* Vol. 198 A, p. 278.

As a rule the data available in craniological investigations are too sparse to give any real test of mesokurtosis, and this is the true reason why we must content ourselves with the Gaussian curve.

Again Mr Powys found out of twelve frequency distributions for the stature of men and women that 11 were leptokurtic and the twelfth essentially mesokurtic. This tendency to leptokurtic distributions—which can hardly be due to chance—is actually given by Ranke and Greiner as a case in favour of the Gaussian curve! (*Anmerkung* S. 326). They further cite Fawcett and Lee in the following manner:

In der an letzter Stelle zitierten Arbeit ist der Nachweis einer bestimmt gerichteten Asymmetrie für die Mehrzahl der Masse und zwar in der nach Fechner zu erwartenden Richtung besonders beachtenswerth.

They do not say that of the 24 curves given by Fawcett and Lee 14 are leptokurtic and that Fechner's curve can only represent platykurtic distributions. They do not draw attention to the fact that the Fechner curve would be impossible for the whole of Powys' stature data, and for 12 out of Macdonell's 26 curves for the English skull! In other words, if the *Abweichungen* of Fawcett and Macdonell and Powys' data are to be used as an argument at all, 38 out of these 62 distributions diverge from the normal curve in a manner which cannot possibly be represented by Fechner's theory!

If we turn from the condition for mesokurtosis to those for differentiation of mode and mean and for skewness we meet other considerations. So far we are not dependent for anything we have said on any theory of frequency other than the Gaussian. On that theory $\beta_2 = 3$, and if the difference $\beta_2 - 3$ be significant the distribution cannot be Gaussian. If we want to distinguish between the mode and the mean, we cannot start from the Gaussian theory, because that theory supposes the two values absolutely the same. On the other hand if we consider asymmetry, we ought to have, within the limits of random sampling, all the odd moments zero, i.e.

$$\mu_3 = \mu_5 = \mu_7 = \ldots = \mu_{2n-1} = 0.$$

Now it is of very little practical value testing the high moments because their probable errors are excessive. The probable error of μ_3 for the normal curve $= \cdot 67449 \sqrt{\frac{6}{N}}\sigma^3$ and of $\mu_5 = \cdot 67449 \sqrt{\frac{945}{N}}\sigma^5$, or in terms of σ as our unit is thirteen times as large. These are the *gross* errors; the percentage probable errors are of course infinite. As a rule it is hardly worth testing these conditions beyond μ_3. We determine whether the third moment is zero within the limits of random sampling. If we wish a relative magnitude we can take $\beta_1 = \mu_3^2/\mu_2^3$, a quantity which occurs over and over again in frequency discussions. The probable error of β_1 is obviously zero for the normal curve, because β_1 is of the square of the order of small quantities. The probable error of $\sqrt{\beta_1} = \cdot 67449 \sqrt{\frac{6}{N}}$, and

$\sqrt{\beta_1}$ for all truly normal distributions ought not to differ by more than two or three times the above expression from zero.

We can form other expressions involving β_1 and β_2 and ask what their value is for the Gaussian curve. We can calculate their probable errors, and determine whether the given distribution satisfies the Gaussian value within the limits of random sampling.

Thus I take the expression:

$$\chi = \frac{1}{2}\frac{\sqrt{\beta_1}(\beta_2+3)}{5\beta_2-6\beta_1-9} = \frac{1}{2}\frac{\sqrt{\beta_1}(6+\beta_2-3)}{6+5(\beta_2-3)-6\beta_1} \quad \text{.................(ii).}$$

Clearly this expression vanishes for the normal curve, and $=\frac{1}{2}\sqrt{\beta_1}$ nearly when $\sqrt{\beta_1}$ and β_2-3 are not very large, i.e. when we have not a very wide deviation from normality. The probable error of this expression, if the distribution be really normal, is $\cdot 67449\sqrt{\frac{3}{2N}}$.

Again, consider the expression:

$$d = \tfrac{1}{2}\sigma\frac{\sqrt{\beta_1}(\beta_2+3)}{5\beta_2-6\beta_1-9} \quad \text{..............................(iii).}$$

This is a length which vanishes, if the distribution be truly normal. Its probable error is $\cdot 67449\sqrt{\frac{3}{2N}}\,\sigma$ in the case of the Gaussian curve, and accordingly d should not differ from σ by more than two or three times the above probable error.

Now let us write $\eta = \beta_2 - 3$. Then it is absolutely impossible for any distribution to be looked upon as Gaussian unless χ, d and η are zero within the limits of random sampling. These limits being deduced from their known probable errors.

Now it will be asked why choose such an expression as χ instead of the simpler $\sqrt{\beta_1}$? The answer is quite simple. We want to determine whether the mode coincides with the mean or not, and we cannot do this on the basis of the Gaussian curve where no distinction is made between the two. We must take some curve which is not Gaussian to determine this important quantity from. Now the equation to the Gaussian curve is

$$y = y_0 e^{-\frac{(x-m)^2}{2\sigma_0^2}},$$

where m is the mean value of x and σ_0 the standard deviation, and we have for its differential equation:

$$\frac{1}{y}\frac{dy}{dx} = -\frac{x-m}{\sigma_0^2}.$$

Now if we assume that the actually observed character is not x but X, and that X is some function of x, we shall not in plotting the frequencies to X obtain a normal curve, but we ought if Y be the ordinate of this curve to have

$$YdX = ydx \quad \text{or} \quad Y = y\frac{dx}{dX}.$$

Taking logarithmic differentials

$$\frac{1}{Y}\frac{dY}{dX} = \frac{1}{y}\frac{dy}{dx}\frac{dx}{dX} + \frac{d^2x}{dX^2}\Big/\frac{dx}{dX}$$

$$= -\frac{1}{\sigma_0^2}(x-m)\frac{dx}{dX} + \frac{d^2x}{dX^2}\Big/\frac{dx}{dX}.$$

Assume $x - m = f(X)$ and we find:

$$\frac{1}{Y}\frac{dY}{dX} = -\frac{X}{\sigma_0^2 F(X)} \quad \text{..............................(iv)},$$

where $$F(X) = Xf'(X)/\{f(X)(f'(X))^2 - \sigma_0^2 f''(X)\}.$$

The form has been so chosen that the origin is the mode, i.e. dY/dX vanishes with X. The proposal to thus generally transform the Gaussian curve is due in a quite different form to Edgeworth*. Kapteyn following Edgeworth and without any acknowledgment takes:

$$f(X) = \beta(X+\kappa)^q$$

where β, κ and q are constants to be determined.

He therefore puts:

$$F(X) = \frac{X(X+\kappa)}{\sigma_0^2(q-1) - mq\beta(X+\kappa)^q - q\beta^2(X+\kappa)^{2q}}.$$

This is a somewhat complex expression. The resulting frequency curve is

$$Y = Y_0(X+\kappa)^{q-1}e^{-\frac{1}{2\sigma_0^2}\{(X+\kappa)^q - m\}^2} \quad \text{......................(v)},$$

and has been suggested by Kapteyn as a general form of the skew frequency curve. We shall consider it later.

Galton and McAlister as early as 1879 took

$$f(X) = b\log\frac{X}{a} - m,$$

where b and a are constants. Ranke and Greiner, without apparently knowing the history of research in this field, take the same value and attribute to Fechner the well-known Galton-McAlister curve of the geometric mean which results. We find

$$F(X) = \frac{X^2}{\sigma_0^2 + b^2\log\frac{X}{a}},$$

* He has developed it in a long series of papers published in the *R. Statistical Society's Journal.*

whence $$\frac{1}{Y}\frac{dY}{dX} = -\frac{1}{X}\left(1 + \frac{b^2}{\sigma^2}\log\frac{X}{a}\right),$$

and $$Y = Y_0 \frac{1}{X} e^{-\frac{b^2}{2\sigma_0^2}\left(\log\frac{X}{a}\right)^2} \text{ ....(vi).}$$

Edgeworth himself has made other suggestions as to suitable values for $f(X)$ and accordingly of $F(X)$. Now it is quite clear that assuming the character to be a definite function of another character which really obeys the normal law, there is no more reason for assuming one form of $f(X)$ than another, because we are in absolute ignorance of the nature of this function. Kapteyn's, or Edgeworth's, or Galton's are equally valid, and the only test of their relative suitability lies in the extent to which the resulting curves fit actual data. Clearly to assume $x = f(X)$ is to assume the actual frequency distribution to follow any law whatever. It is only screening the generality of the assumption

$$y = \phi(x),$$

where ϕ is unknown, by an appeal to the supposed universality of the Gaussian curve and by a perfectly arbitrary selection of the subsidiary function f.

But there is another manner of looking at this proposal. Returning to the equation

$$\frac{1}{Y}\frac{dY}{dX} = -\frac{X}{\sigma_0^2 F(X)},$$

and writing $\sigma^2 = \sigma_0^2 F(X)$, we see that it becomes identical in form with the normal equation, i.e.

$$\frac{1}{Y}\frac{dY}{dX} = -\frac{X}{\sigma^2}.$$

In other words the distribution of any frequency may be looked upon as given in the neighbourhood of any point by a normal curve of standard deviation $\sigma_0\sqrt{F(x)}$. Hence the conception arises that if the causes which produced variation in the immediate neighbourhood of any value x_0 of the character, were constants for the whole range of variation, we should have a normal curve of standard deviation $\sigma_0\sqrt{F(x_0)}$*. In reality there is a continuous and gradual change of the tendency to variability as we pass from one value of the character to a second†. Analytically

* This method of looking at the matter throws light on another point. If a curve be of limited range, it signifies that $\sigma = 0$ at certain points, or the curve stops because we have reached the limits of local variation. In a curve of unlimited range it is not the capacity for local variation but the absence of individuals to vary, which is the special feature.

† The matter is of such importance relative to some of Ranke's criticisms that I give another proof of equation (iv) here, based on the conception of an infinite number of infinitely small cause groups which Ranke considers can only lead to the normal curve. Let y_{r+1} be the $(r+1)$th term of a binomial, skew or symmetrical, say for simplicity the latter, i.e. $(\frac{1}{2}+\frac{1}{2})^n$. Then

$$\frac{y_{r+1} - y_r}{\frac{1}{2}(y_{r+1} + y_r)} = \frac{n+1-2r}{\frac{1}{2}(n+1)}.$$

Now let c_r be the distance between y_r and y_{r+1} used in plotting these ordinates to obtain a curve, and let it be related to some small value c_0 by the relation $c_r = c_0 \times$ function of $r = c_0 \times \phi(r)$. Let X_r be measured we may look at it in this way: If δx_0 be the variation in the neighbourhood of x_0, then δx_0 is not independent of x_0, *but correlated with it.* We may have a perfectly continuous population from dwarfs to giants, but it does not follow that the actual tendency to vary of dwarfs and of giants is identical. All proofs that I have seen of the normal curve fail in this respect. They assume that the character x is due to a number of increments, which are due to an indefinitely large number of *independent* cause groups. They assume that δx_0 is not correlated with the already accrued value x_0. All processes like those of Edgeworth, Galton, Kapteyn and Fechner are really devices for getting over this gradual change in the tendency to vary from point to point of the range. It appears to me best to directly acknowledge and face this difficulty by selecting a fitting function for $F(X)$.

If we drop now the distinction between X and x as unnecessary we reach as our frequency equation:

$$\frac{1}{y}\frac{dy}{dx} = -\frac{x}{\sigma_0^2 F(x)},$$

or if we use Maclaurin's theorem for $F(x)$:

$$\frac{1}{y}\frac{dy}{dx} = \frac{-x}{\sigma_0^2(1 + a_1 x + a_2 x^2 + a_3 x^3 + \dots)} \quad \text{.................(vii).}$$

Now I have shown* how to determine the successive constants σ_0^2, $\sigma_0^2 a_1$, $\sigma_0^2 a_2$, etc. Further all these constants but σ_0^2 are zero, when the distribution is normal, and the series will be found to converge rapidly, when the distribution is

from the largest term of the binomial, then $X_r = c_0\sqrt{(n+1)} \times$ function of $r = c_0\sqrt{(n+1)}\, f(r)$, say, and conversely $r =$ a function of $X_r/\{c_0\sqrt{(n+1)}\}$. Divide both sides of the above equation by c_r, which may be written on the left ΔX, and we obtain:

$$\frac{\Delta Y}{X\Delta X} = \frac{-X_r}{\frac{1}{4}(n+1)\, c_0^2 f(r)\, \phi(r)/\{r/(n+1) - \frac{1}{2}\}}.$$

Put $\sigma_0 = \frac{1}{2}\sqrt{n+1}\, c_0$ and $F(r)$ for the expression $f(r)\,\phi(r)/\{r/(n+1) - \frac{1}{2}\}$ which does not become infinite with $2r = n+1$, because X_r and therefore $f(r)$ vanishes for this value of r, and accordingly $f(r)$ contains $2r-(n+1)$ as a factor. We then have:

$$\frac{\Delta Y}{y\Delta X} = -\frac{X_r}{\sigma_0^2 F(X_r/\sigma_0)}.$$

Now make n infinite and c_0 vanishingly small, then we have if $\sigma_0 = \frac{1}{2}\sqrt{(n+1)}\, c_0$ be still finite

$$\frac{dY}{YdX} = -\frac{X}{\sigma_0^2 F(X/\sigma_0)},$$

a result in agreement with the above investigation. In other words this, and not the Gaussian curve, is the generalised frequency curve we reach if we directly abrogate the third Gaussian principle, that contributory increments of the variate are independent. Of course the first two Gaussian principles simultaneously disappear. This view of the matter occurred to me many years ago, when considering Hagen and Crofton's proofs of the Gaussian law. It was expressed in my memoir of 1894 by the statement that we require curves produced by conditions in which the contributory cause groups are not independent, i.e. in which an increment δx to the variate x depends upon the value of x, or is correlated with it. My method of reaching such curves, however, was a direct appeal to discrete series in which such a condition was fulfilled.

* "Mathematical Contributions to the Theory of Evolution, XIV." p. 6. Dulau and Co., London.

in the least approximately normal. Accordingly if we wish to get a good interpolation curve to determine the distance between the mode and the mean, we may assume

$$\frac{1}{y}\frac{dy}{dx}=\frac{-x}{\sigma_0^2(1+a_1x+a_2x^2)} \quad\text{.......................... (vii) } \textit{bis}.$$

In this case we discover with the previous notation, that

$$d=\frac{1}{2}\frac{\sqrt{\beta_1}(\beta_2+3)}{5\beta_2-6\beta_1-9}\sigma$$

is the distance between mean and mode, and that $\chi=\frac{\frac{1}{2}\sqrt{\beta_1}(\beta_2+3)}{5\beta_2-6\beta_1-9}$ is the ratio of this distance to the variability, or what I term the skewness, or the asymmetry relative to the variability.

If we leave out a_2 we find the skewness given by $\chi=\frac{1}{2}\sqrt{\beta_1}$ and the distance between mean and mode $d=\frac{1}{2}\sqrt{\beta_1}\sigma$. In practice these give fairly closely the same values as the fuller expressions above, and the fuller expressions are not numerically much modified if we include a_3. Shortly we have got a very fair mathematical process of determining the position of the mode and the degree of asymmetry.

Now the constants of such a curve as (vii) *bis* are absolutely determined by a knowledge of σ, β_1, and β_2; or looked at inversely they suffice to fix σ, β_1, and β_2. In other words the degree of kurtosis (β_2-3), the skewness χ and the distance between mean and mode—all most definite physical constants—are at once fixed by a knowledge of the constants of the curve, or on the other hand, being known, they fix those constants. It is of course allowable to replace any one of the three by the variability of the system. The actual position of the mode and the total magnitude of frequency suffice to fix the position and size of the curve. I have already called $\beta_2-3=\eta$ the degree of kurtosis; I call d the modal divergence. Then unless

η, the degree of kurtosis be zero, subject to probable error, $\cdot 67449\sqrt{24/N}$,

χ, the skewness be zero, subject to probable error, $\cdot 67449\sqrt{1\cdot 5/N}$,

d, the modal divergence be zero, subject to probable error, $\cdot 67449\sigma\sqrt{1\cdot 5/N}$,

no distribution can be legitimately described as normal or Gaussian.

It would be of interest to know how far Ranke and Greiner have applied such tests to any series containing a large number N of individuals. I think if they had done so, they must have come to the same conclusion as the majority of statisticians that the normal curve has only a limited range of application.

Of course if N be small, as in most craniological series, we find our probable errors so large, that it is not possible to say more than that for *short* series the Gaussian curve may roughly describe the result. But for long series in economics, sociology, zoology, botany and anthropometry the Gaussian curve over and over again fails. If in all these cases Ranke and Greiner assert that the material is heterogeneous they are merely arguing in a circle. The distributions are as continuous and smooth as those which occur in the case of the Gaussian curve, and *they occur for characters in the same group of individuals which present for other characters the normal distribution.*

Thus the length of meropodite of right claw in *Gelasimus pugilator** is quite sensibly normal, but the length of the carpodite of the right claw is almost as certainly platykurtic and skew. If two characters are normal, a third character which is their difference, whether they be correlated or not, should have a normal distribution, yet in the case of *Gelasimus pugilator* for the whole series of measurements the difference distributions are essentially platykurtic†.

The size of the disc in *Ophicoma nigra*‡ has a modal difference in the distribution of 1000 cases of ·271 mm. and the probable error on the basis of a normal curve is ·056. The deviation is thus five times its probable error and the asymmetry undoubtedly significant.

The outer diameter of *Arcella vulgaris*§ in 504 cases gives a modal difference of 3·226 mikrons, and the probable error of this difference is only ·211. The asymmetry is therefore undoubtedly significant.

The distance between the mean and mode in the case of the length of shell of *Nassa obsoleta*‖ from Lloyd Point, Long Island, U.S. was ·68 mm. for 368 individuals, the probable error of this modal difference was ·08. The asymmetry is therefore significant. Other characters of *N. obsoleta* were as definitely asymmetrical, while some from exactly the same individuals were sensibly normal.

The transverse arc in ♂ Naqada skulls has for 115 individuals a modal difference of 2·34 mm. and the probable error of this difference is ·78 mm., or it is probably significant. Yet the breadth of the male Naqada skulls is significantly symmetrical.

The height/length index of 117 English ♀ skulls has a modal difference of ·85 and a probable error of only ·22, the skewness is therefore significant.

The same is true of the distribution of many internal organs in man. For example, if we exclude recognised diseased hearts, we obtain a markedly skew distribution such as is given in the broken line of Fig. 1. This is for 1382 heart-weights. If this be supposed to be due to the great variety of ages, we have only to look at the continuous curve for hearts of 358 young adults, 25 to 35, to see the same asymmetry. This is drawn for four times the scale.

* G. Duncker: *Biometrika*, Vol. II. p. 313.

† There is another point to which I will only refer briefly here. If characters were always distributed according to the Gaussian law the regression curves must be *straight* lines. The generalised Mendelian theory of determinants I have developed makes them, however, hyperbolas, and I have given instances in a recent memoir of a variety of curved regression lines

‡ McIntosh: *Biometrika*, Vol. II. p. 470.

§ R. Pearl and F. J. Dunbar: *Biometrika*, Vol. II. p. 327.

‖ A. C. Dimon: *Biometrika*, Vol. II. p. 29.

Lastly if we consider that the question of health determines the skewness, we have in the dotted curve the weight-distribution of 699 hearts stated to be "healthy." We see that there is still the same essential skewness; the pathologist has merely cut off a small portion of the tail on the left and far too much of the tail on the right, i.e. unusually big hearts were discarded as necessarily "unhealthy." The form of the curve undoubtedly indicates that many of these large hearts are abnormal, but any continuous curve fitted to the remainder, the "healthy hearts," would not only be significantly skew, but would project a long way into the portion of the tail discarded as "unhealthy." The list of asymmetrical distributions might be indefinitely extended, but these must suffice to indicate that asymmetry cannot be lightly put on one side in the manner adopted by Ranke and Greiner.

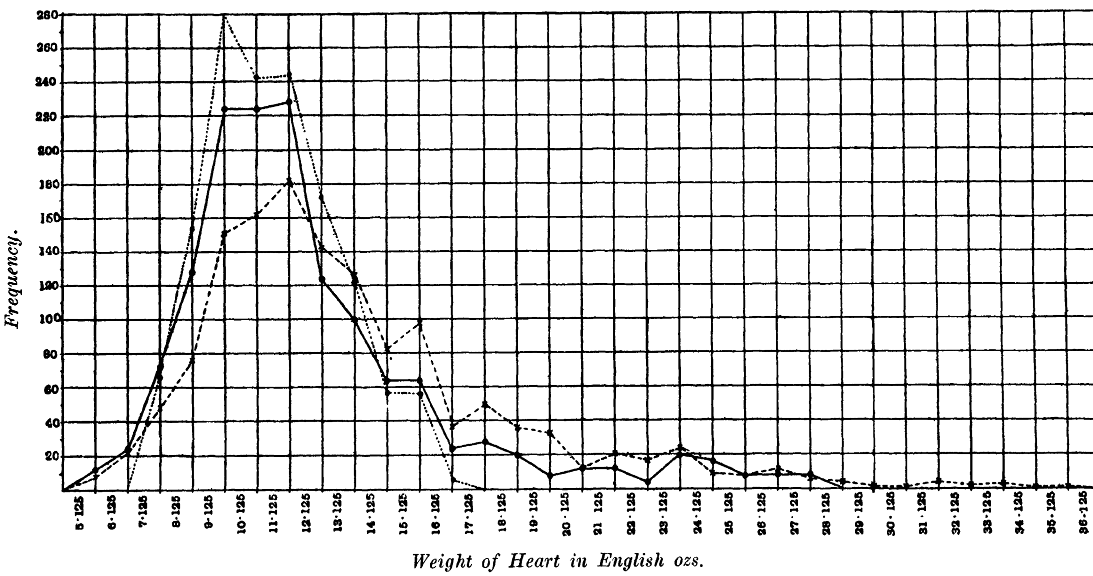

FIG. 1. Frequency Polygons of Weight of Heart in Males (Greenwood*).

I. *------* 25 to 55 years. Without specific disease of heart. Number of cases 1382.
II. O——O 25 to 35 years. Without specific disease of heart. Number of cases 358.
III. • ......• 25 to 55 years. Definitely sound hearts. Number of cases 669.

The scale is four times as great for II. and twice as great for III. as for I.

If we pass to discrete variates, we find as large a number, if not a larger number of distributions in which skewness is well marked, for example, fertility in the Aphis *Hyalopterus Trirhodus*†, fertility in man‡, fecundity in race-horses§, and fertility and fecundity in mammals generally. I illustrate this with an example of fertility in English mothers in Fig. 2. It will be seen at once that no normal curve could be used to describe this distribution. It is equally

* *Biometrika*, Vol. III. p. 45 *et seq.*

† Warren: *Biometrika*, Vol. I. p. 127.

‡ Pearson: *Phil. Trans.* Vol. 192 A, p. 257, and *The Chances of Death*, Vol. I. p. 63.

§ Pearson: *Biometrika*, Vol. I. p. 292.

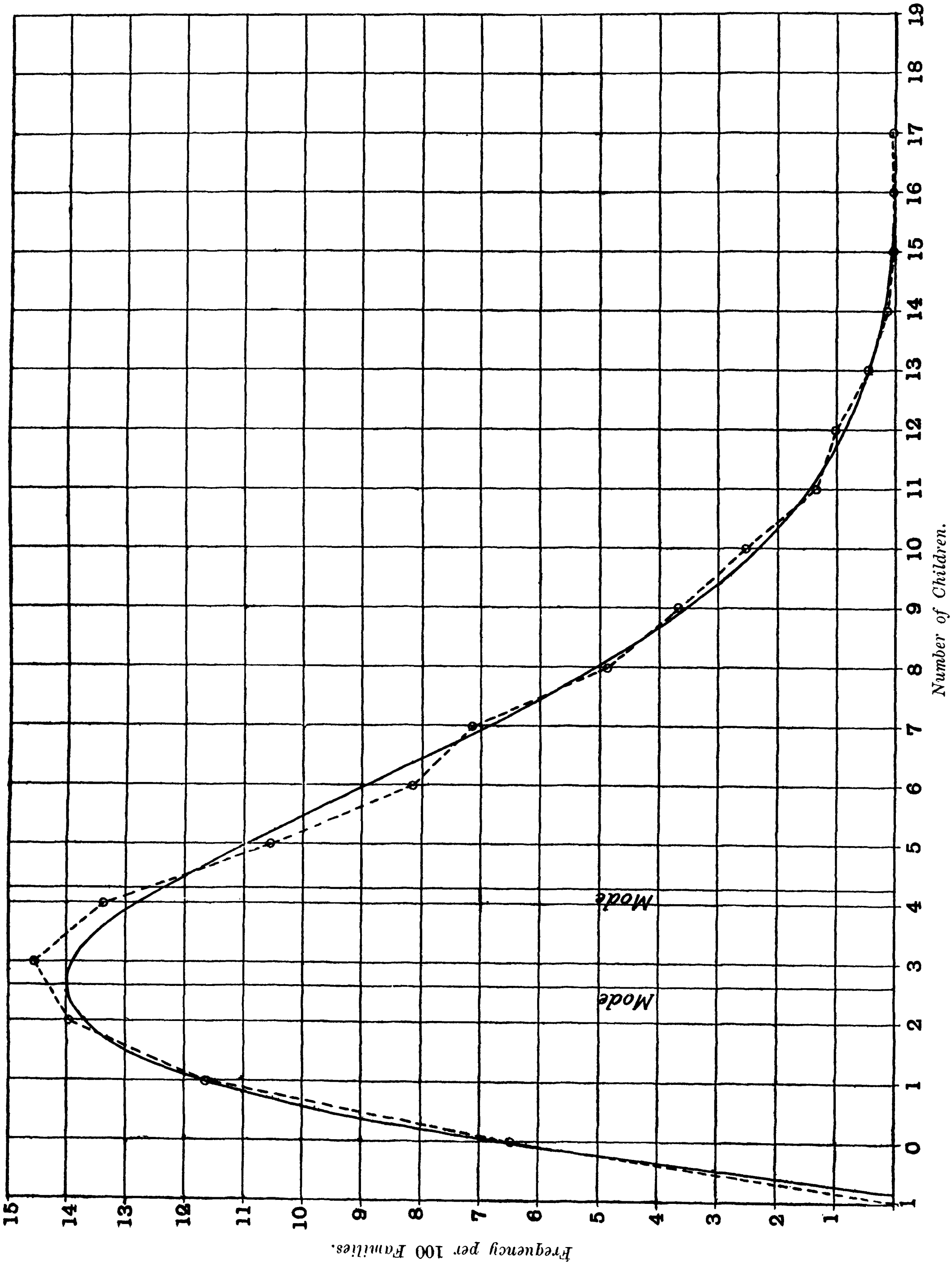

FIG. 2. Number of Children in English Families.

impossible also for a curve of the McAlister-Galton type, for the simple reason that that curve has high contact at both ends of the range. Now I contend that the anthropologist who either neglects such matters as human fertility, or confesses that he has no means of succinctly describing their distributions—and as long as he sticks to the Gaussian curve he certainly will not have—is simply putting on one side a fundamental factor in the science of man. Ranke lays great stress on homogeneity. He does not, however, clearly define what he means by the term. Apparently any series which follows a Gaussian curve is to him homogeneous, any other series is not. I should be glad if he would then consider any craniological series, say, of adult crania. This will involve crania of adults from perhaps 25 or 30 years of age to 50, and these are rather narrow limits considering our paucity of material. Now our data show that the correlation between head measurements and age may be of the order of about $-\cdot 1$ to $-\cdot 2$. After about 25 to 28 years of age in man, there is a continual shrinkage not only of stature but of skull capacity, brain-weight, circumferences and diameters of the skull. Under the circumstances, what right have we because the Gaussian curve is obtained to call this material "homogeneous"? I will go further; suppose we could, and we can, obtain the measurements on one or two thousand individuals of the same age; are these to be considered as a homogeneous distribution? My reply will be, in man the order of birth is an essential feature in determining the dimensions of the physical characters. My investigations show that physique and health are sensibly correlated with the position of a member in his own family. In mammals others have shown that the physique of an individual is sensibly correlated with the number of members born in that individual's litter. Now with these facts before us what stress can be laid on Ranke's conceptions of homogeneity? The practical anthropologist requires curves which will successfully *graduate* his data. Only on the basis of such graduations can he allow for the influence of disturbing factors like age, order of birth, season or special position of production in the organism. I take this very case: How is it possible to allow for the influence of order of birth, unless you know the size of families or the distribution of births within the community? It can only be achieved provided the distribution can be represented by a few simple constants which allow of definite mathematical handling.

Ranke and Greiner say with considerable asperity that my method of determining the range from given data can be of no service. Yet take this very case of size of families. In the English middle classes for 4390 instances, I find that the *observed* limit is 17, but fitting a skew curve the range is determined as 22 children running practically from just before 0 to over 21 births (Fig. 2). For Denmark by the same process in 34,000 cases the theoretical range is 26 and the observed range 22*. I then proceeded to take statistics for the Argentine Republic, and found for the town of Buenos Ayres, 27,510 births, that the range of the curve was from ·25 to 36·61 births, or 37 possible births. The maximum observed in these 27,510 births was 23. But among the South- and Mid-American

* *The Chances of Death*, Vol. I. *Reproduction Selection.*

populations cases of 24 (Trinidad), 26 (Cuba), 27 (Nicaragua) and 34 (Colombia), the women beginning to bear at 13 and continuing to 50, have been recorded! It will, I think, be realised by the impartial reader that Ranke's statements:

> Für den Anthropologen ist also nur das Gaussche Gesetz von Wichtigkeit. Für seine Probleme beansprucht es aber auch völlige Gültigkeit (S. 327)

and

> So haben anderseits auch die Werte von Variationsumfängen die bislang aufgefunden sind, keinerlei weitere Erkenntniss gebracht (S. 324)

fall wide of the mark. Ranke has either very much circumscribed the field of the anthropologist, or he has not in this, as in other cases noted in his paper, studied the literature of the subject, or finally he has disregarded the results reached.

It is quite true that the range cannot always be determined and being determined does not always give a very good result. The reasons for this are not far to seek. For example, If a discrete quantity has for its minimum 0 units the start of the curve must naturally fall on the negative side of the origin—since its area measures frequency between $-\cdot 5$ and $+\cdot 5$. Ranke would probably find something mysterious in this "reichen oft ins Negative." Actually it is to be expected, especially if due allowance be given for the probable error of the range. In most biometric statistics, we cannot as in the case of births deal with 20,000 to 30,000 cases and get small errors for our constants. We have only perhaps 500 to 2000 cases and even less than this in craniology. This may denote an error of 14 to 17 p.c. in the calculated range, and it is quite possible that the range may "reichen ins Negative." Take the case worked out by me* of the number of Müllerian glands in the forelegs of ♀ swine. The range theoretically calculated is 18 glands with a probable error of $\pm 2\cdot 54$; the start of this range ought not to have exceeded $-\cdot 5$. It is actually $-\cdot 82$ with a probable error of $\cdot 16$. The actual skewness of this distribution is $\cdot 31$ with a probable error of $\cdot 02$. The distribution is accordingly significantly asymmetrical.

I have cited these cases as sufficient for our present purpose, but there are many other cases in which the discovery of the range has been of biological or special anthropological interest, e.g. the earliest appearance of certain diseases in childhood, the range of cancer attacks, the first occurrence of signs of puberty, etc. It has been applied also effectively to a number of zoological and botanical data. A more striking case, perhaps, of usefulness is the limit to high barometric pressure obtained by dealing with the frequency statistics of barometric height at series of stations†. Throughout the whole of the stations of the British Isles dealt in, there is sensible skewness of distribution, and with one Irish exception, which is sensibly mesokurtic, the whole series of curves are platykurtic, and this deviation from normality cannot be chance, but is a significant character of the frequency distributions. In these cases the limit to high pressure has been found, and appears to be a constant of considerable physical importance for the local climate.

* Pearson and Filon: *Phil. Trans.* Vol. 191 A, p. 289.

† Pearson and Lee: *Phil. Trans.* Vol. 190 A, p. 423 *et seq.*

I again am forced to consider that Ranke has not been aware of what has been published, still less what has been done in this matter. He appears to base his conclusions chiefly on my first paper on skew-variation, and he has not noticed the fact that being the *first* paper much has to be corrected in the light of more recent work in the past ten years. Ranke speaks of the writer's:

Andwendung von allerlei grösseren und kleineren Änderungen in seiner Methode *ad hoc* in eine für den vorliegenden Zweck nicht zu unwahrscheinliche Form bringen (S. 324).

Now I contend that this gives a grossly unjust description of the paper in question. Had Ranke read recent literature, he would have been aware that the great difficulty with frequency distributions is to obtain the true values of the "moments" from records which merely give data for arbitrary "Spielräume," often far too large and usually selected by the observer without any regard to the needs of the computator. My method is one based on the method of moments, but to deduce the moments from given data is the real difficulty which Ranke never for an instant seems to grasp or at any rate refer to. The standard deviation (which he appears to consider sufficient for anthropologists) will vary, and often very sensibly, with the nature of the grouping of the data. This difficulty was very present in my mind in 1894, and is constantly referred to in my memoir, the "allerlei grössere und kleinere Änderungen in seiner Methode" are no changes in method at all but attempts to obtain some approximation to the true moments of the data. It was not till 1898 that Sheppard showed the correct manner of calculating the moments from the raw data in his important memoir on frequency constants*, for *one* type and one type only of frequency distribution. The curves calculated by Sheppard's method, now in general use, would give better results undoubtedly than are to be found in my memoir of 1894. Further, however, Sheppard's method applies only to curves with high contact with the horizontal axis at both ends. It leaves us still in doubt as to how to find the moments of curves, which cut the axis at the end of the range or are asymptotic at one or both ends to the vertical axis. At such ends of the range, the real solution lies in recording the frequency for very small elements, but this was not provided in any of the statistics which were then before me. It is just these cases of limited range at one or both ends which present difficulty in the determination of the moments. The difficulty will be familiar to all statisticians, if it has escaped Ranke. To some extent it is met in my memoir on the systematic fitting of curves issued in April, 1902†. Yet granting all these difficulties what do we find in my memoir of 1894? An analysis of the cases in which range is dealt with seems justified by the charges made:

Example I. Range determined of Cambridge Barometric Heights. There is nothing physically improbable in the result.

Example VI. Range found for enteric fever runs from $-1{\cdot}35$ years to about 385 years. The probable error of the range is not given, but the whole difficulty

* *Proc. London Math. Society*, Vol. XXIX. p. 353 *et seq.*

† *Biometrika*, Vol. I. p. 265.

turns upon the *great* changes introduced into the range by different methods of calculating the moments. More recent investigations, in which the sexes are separated, the moments more accurately determined, and larger numbers dealt with, give far better results for zymotic diseases. I presume that one character being age, however, Ranke and Greiner would dismiss these data from consideration under any circumstances.

Example VII. Guesses at 9 tints. Possible range 1 to 9, i.e. curve to run from ·5 to 9·5. Observed guesses run from 1 to 8. Theoretical range of 11 instead of 9. The paucity of the observations gives a probable error of at least 20 to 30 per cent. in the determination of the range, and the result is rather better than might have been anticipated.

Example VIII. Ratio of forehead to body length in *Carcinus moenas,* observed range 30, calculated range 51. This range is probably not very close but it is not in any way that I can see impossible. The material is probably dimorphic.

Example XI. H. de Vries's data for *Ranunculus bulbosus.* Actually observed range 5 to 10 petals. Calculated range 5 to 11 petals.

Example XII. H. de Vries's data for a race of *Trifolium repens.* Actually observed range 0 to 10 high blossoms. Theoretical range in complete agreement.

Example XIV. Pauperism percentages for 632 cases. Observed range 18 for the year 1891 dealt with. Calculated range 31. This range gives 2 units of negative pauperism. Its probable error is, perhaps, 14 per cent.

It will be seen that out of the seven examples in which range is calculated only three *reichen ins Negative,* and that this *reichen* is well within the limits of the errors arising on the one hand from random sampling and on the other from the defective methods of determining the moments, which were alone available in 1894. While quite appreciating the honour done me when other workers use my methods, I must decline to be responsible in any way for their application of my formulae. I have so often found that their failure to fit my curves is due to a misapprehension of my methods or to actual errors in arithmetic, that I have long given up any attempt to set such matters right. The frequent assumption made that statistical methods can be applied without adequate mathematical training is the source of most of the slips in this matter*.

So far then I think we may conclude that Ranke is completely unjustified both in his statement that the Gaussian curve fully describes all the frequency that is of importance to the biologist, and in his attempt to discredit any result of scientific value which flows from endeavouring to measure such differences from the Gaussian law as we find in the distance between mode and mean, the skewness, the kurtosis and range of many actual frequency distributions.

* A good illustration, by no means unique, of this is F. Reinöhl: *Die Variation im Andröceum der Stellaria Media,* 1903. He finds it impossible to fit certain distributions with my curves, owing to ignorance of the full literature and to faulty determination of the moments. He then argues from this want of fit to biological conclusions.

All the leading statisticians, from Poisson to Quetelet, Galton, Edgeworth, and Fechner, with botanists like de Vries, zoologists like Weldon have realised that asymmetry must be in some way described before we can advance in our theory of variation. In innumerable cases the important quantities measured by η, χ and d actually exist; these have each their physical significance and they must be found. It is perfectly open to Ranke and Greiner to criticise my method of determining these quantities, but that they should shut their eyes to their existence appears to me only compatible with a very small acquaintance with the data of variation.

Let us see now how various authorities have met this difficulty of skewness.

B. *The Gaussian Curve**.

Gauss proceeds from the axiom that: *The arithmetical mean of a series of observations gives their most probable value,* i.e. the mean is the value of maximum frequency. This result is not axiomatic. It can only be a result of experience, and if it were true it would make the normal curve as much a result of experiment, i.e. an empirical result, as any other proposed curve of frequency. Gauss's proof demands, however, something more than this first statement. It involves (i) the equal probability of errors in excess of the mean and of errors in defect, (ii) the continuity of magnitude in the errors, and (iii) the independence of all the small contributions to the total error.

Experience shows that Gauss's fundamental axiom as to the mode and mean coinciding is not universally true. It is not true of errors of observations, it is not true of variations in living forms. Gauss reaches a differential equation which leads to the normal curve. His proof seems to me, as it has done to many others, quite invalid, because the equal probability of errors in defect and excess of the mean is not demonstrated, the possible dependence of contributory elements is not discussed, and the question of continuity of errors is not considered.

C. (i) *Laplace and Poisson.*

Laplace and after him Poisson took, I venture to think, much firmer ground. They did not assume (i) and (ii), but they did not realise the importance of (iii). They proceeded by evaluating the terms of the binomial:

$$(p+q)^N.$$

* In writing for Germans I naturally spoke of the Gaussian curve. But I am not clear that precedence is to be given to Gauss. Gauss first gave a proof of the well-known equation $y = y_0 e^{-\frac{1}{2}x^2/\sigma^2}$ in his *Theoria Motus Corporum Coelestium* of 1809. This was three years before the publication of Laplace's *Théorie Analytique des Probabilités* of 1812. But to give absolute priority to Gauss is to disregard Laplace's earlier memoirs, particularly those of 1782, "Sur les approximations des Formules qui sont fonctions des très-grands nombres," and its *Suite du Mémoire* of 1783. On p. 433 of the latter memoir Laplace actually suggests the importance of forming a table of the probability integral $\int e^{-t^2}dt$. The *Théorie des Probabilités* reproduces the substance of this memoir, and on this account some writers have post-dated Laplace's work. Gauss stated that he had used the method of least squares in 1795, but this does not necessarily involve a knowledge of the probability integral, and if it did, it is ten years after Laplace. On the whole my custom of terming the curve the Gauss-Laplacian or *normal* curve saves us from proportioning the merit of discovery between the two great astronomer mathematicians.

Using Stirling's theorem they showed that if the mth be the largest term in the binomial, then the sum p_r of all the terms from $m-r$ to $m+r$ is very nearly given by:

$$p_r = \frac{2}{\sqrt{2\pi}\sigma}\int_0^r e^{-\frac{x^2}{2\sigma^2}}dx + \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{r^2}{2\sigma^2}} \quad \text{..................(viii)},$$

where $\sigma = \sqrt{Npq}$.

Here we have the first appearance of the probability integral as representing a series of *discontinuous* binomial terms. In fact when N is fairly large Laplace and Poisson show that sums of terms of the binomial are closely given by the areas of the probability curve. It is an approximate result based upon Stirling's theorem, and it does not for a moment involve making N infinitely large, or the spacing apart of the binomial terms very small. This representation of a number of finite terms by the probability integral seems to be unfamiliar to Ranke and Greiner, but no practical statistician would calculate the sum of r terms in the binomial $(p+q)^N$ for even moderate values of N. He would simply calculate the standard deviation $\sigma = \sqrt{Npq}$ of the binomial and turn up tables of the probability integral. This fundamental property of the normal curve, i.e. that it closely represents a *discontinuous* series, is passed over in silence by my critics. It is the very purpose for which the probability integral was originally introduced by Laplace. In other words it arises without any consideration of (i) continuity of variation, or (ii) equal probability of negative and positive deviations.

It will be observed that the above approximation to the binomial, i.e. to $(p+q)^N$ is symmetrical, but we can easily allow for some degree of asymmetry. Still writing $\sigma = \sqrt{Npq}$, and for the binomial

$$\beta_1 = (1-4pq)/(Npq), \quad \eta = \beta_2 - 3 = (1-6pq)/(Npq).$$

I have shown*, y_0 being the maximum term in the binomial, that the rth term from the maximum is given by:

$$y_r = y_0 e^{-\frac{r^2}{2\sigma^2}(1-\beta_1+\frac{1}{2}\eta) - \frac{1}{2}\sqrt{\beta_1}\frac{r}{\sigma} + \frac{1}{6}\sqrt{\beta_1}\frac{r^3}{\sigma^3}(1-\frac{5}{2}\beta_1+\frac{3}{2}\eta) - \text{etc.}} \quad \text{............(ix)}.$$

The term in r/σ was, I believe, first added by Poisson, and expresses his attempt to allow for asymmetrical variation. Edgeworth expanding the exponential has adopted for his asymmetrical curve, a form easily deduced from (ix),

$$y_r = y_0 e^{-\frac{r^2}{2\sigma^2}}\left\{1 - \frac{1}{2}\sqrt{\beta_1}\left(\frac{r}{\sigma} - \frac{1}{3}\frac{r^3}{\sigma^3}\right)\right\} \quad \text{........................(x)}.$$

It will thus be seen that the normal function and the probability integral arise naturally from the expression for a single term or a series of terms of the binomial polygon. This is their historical origin and the historical origin of the conception of asymmetrical variation. Instead of the complex form given above resulting from Stirling's theorem, I approached the subject by looking at the relation of the

* *Phil. Trans.* Vol. 186 A, p. 348, footnote.

normal curve to a symmetrical binomial in a totally different manner. I succeeded in showing that the ordinates and areas of the normal curve gave exceedingly closely the terms and sums of terms of the symmetrical binomial even for relatively small values of n. This had already been done by Laplace. The reader will realise this if he looks at the closeness of the normal curve with

$$\sigma = \sqrt{Npq} = \tfrac{1}{2}\sqrt{N},$$

and the binomial $(\frac{1}{2}+\frac{1}{2})^{10}$ with $N=10$ in the accompanying Figure 3. But my method enabled me to give a simpler expression to the asymmetrical binomial

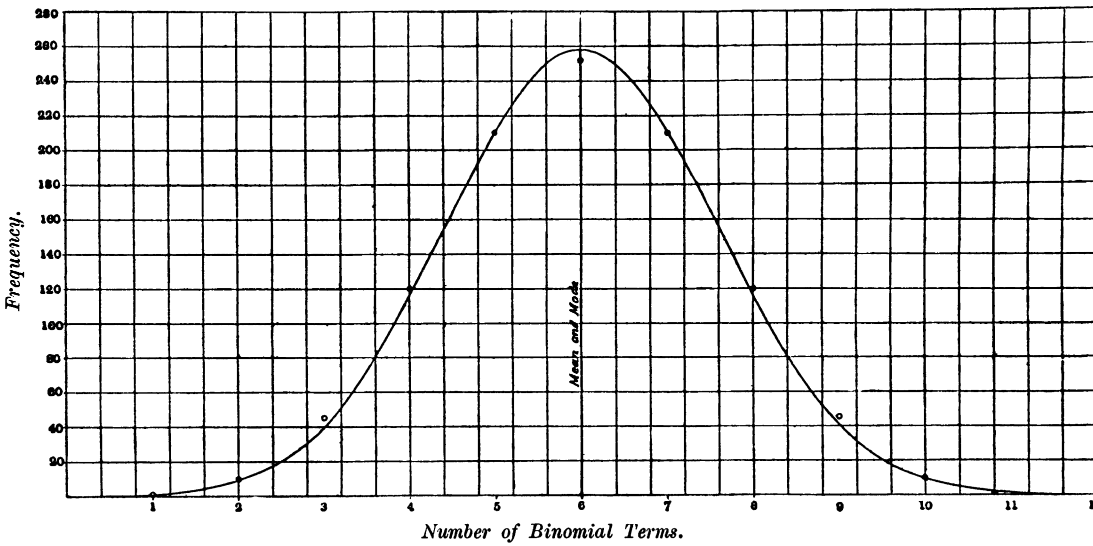

FIG. 3. Comparison of Point Binomial $1024\,(\frac{1}{2}+\frac{1}{2})^{10}$ with the Gaussian Curve.
S.D. $=\sqrt{Npq}=1\cdot5811$. Maximum Ordinate $258\cdot35$.

than had been obtained by Poisson or Edgeworth using Stirling's theorem. Figure 4 shows how closely the terms of the asymmetrical binomial $5000\,(\frac{1}{6}+\frac{5}{6})^{12}$ and the sums of terms are reproduced by my curve of Type III., i.e.

$$y = 1536\cdot54\left(1+\frac{x}{5}\right)^{14} e^{-3x}$$

I had no higher ambition—nor could I have had one higher—than Laplace had when he discovered the normal curve. I wanted to find a close mathematical expression for the terms of the asymmetrical binomial for relatively small values of N.

Now Laplace and Poisson had both retained the last of Gauss's limiting conditions, i.e. they had by adopting the binomial supposed each increment of the deviation to be independent of previous increments. It seemed needful to me to get rid of this condition, and I therefore introduced instead of the binomial the hypergeometrical series. Here the successive increments are correlated. In order to place this new representation on the same footing as the symmetrical binomial to which Laplace approximated with the normal curve, I deduced as I had done for the symmetrical and the asymmetrical binomials, curves which gave the hypergeometrical series and the sum of its terms as closely as Laplace's normal curve gave the symmetrical binomial. This is the complete history of the development of my skew curves. Before I proceed to discuss Ranke and Greiner's criticisms, I must remark that their attack on this point does not concern me only. Every practical statistician uses Laplace's representation of the point binomial by the

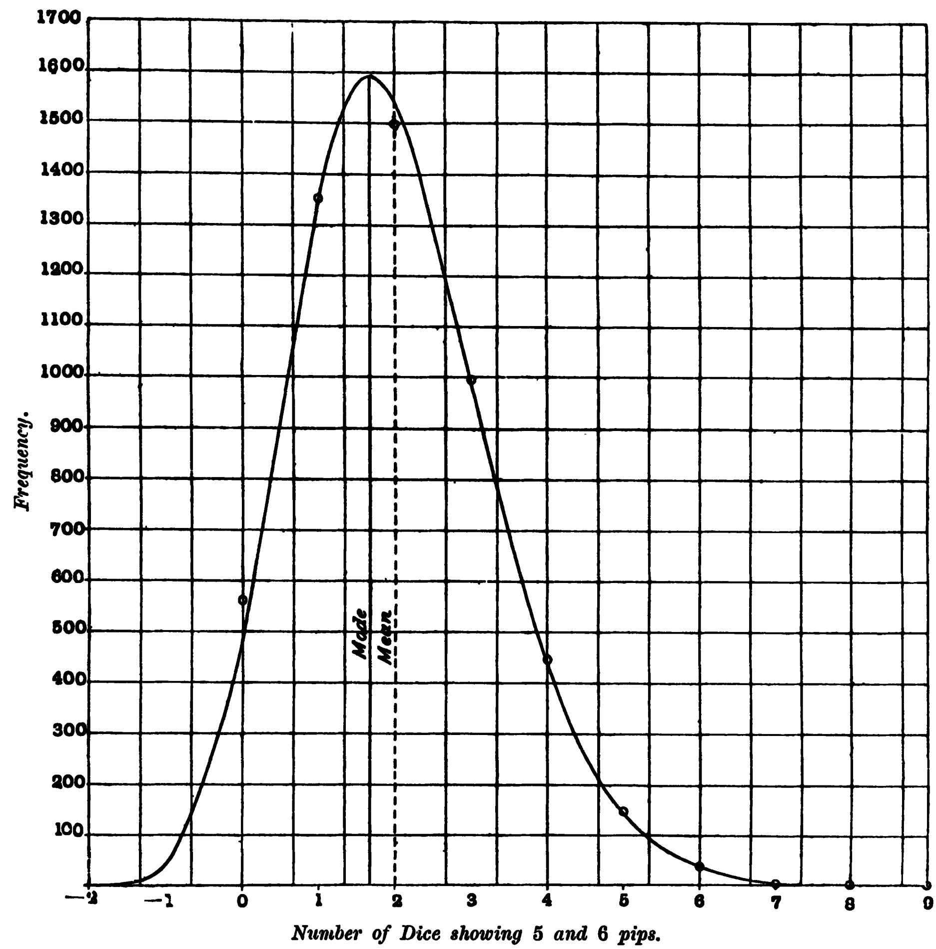

FIG. 4. 5000 Throws with 12 Dice.

○ ⊙ Points of the Binomial $(\frac{1}{6}+\frac{5}{6})^{12}\times 5000$.

—— Curve $y=1536{\cdot}54\left(1+\frac{x}{5}\right)^{14}e^{-3x}$.

Origin at the mean 2. Mode of curve at $x=-\frac{1}{3}$.

Frequency of each number of 5 and 6 pips	−1	0	1	2	3	4	5	6	7	8	9
Ordinates of Binomial	0	561	1346	1480	987	444	142	33	6	1	0
Areas of Curve	557		1340	1493	982	443	139	37	8	1	0

probability integral, when he is discussing the probable errors of death rates, class indices and a multitude of other problems, and this when the binomial is skew and N relatively small! It is the whole theory of current statistics which Ranke and Greiner are tilting at when they object to the use of what is equivalent to the Euler-Maclaurin theorem, i.e. the mathematical representation of a finite sum of terms by a definite integral*.

There is another point also which may be noted here before we leave the binomial. The quantity β_1 is a measure of the asymmetry. Now consider the ratio β_1/η, or the ratio of this measure of asymmetry to the kurtosis. For the asymmetrical binomial we have $\beta_1/\eta = (1-4pq)/(1-6pq)$. Since p and q if positive must give a product lying between 0 and $\frac{1}{4}$, this ratio cannot take any value between 0 and + 1. Hence any curve which gives for β_1/η a value less than unity cannot possibly diverge from the normal curve in the direction of a binomial series. We shall see the application of this later.

C. (ii) *The Galton-McAlister Curve.*

I have already referred to the attempt of Poisson to give the skew binomial by an extra term applied to the Gauss-Laplacian probability integral. Quetelet endeavoured to meet the asymmetry of frequency distributions by placing graphically skew binomials on top of the frequency polygon—a very rough and somewhat deceptive process. The next step in the advance was taken by Francis Galton, who in 1879 suggested that the geometrical mean and not the arithmetical mean is likely to give the most probable result in many vital phenomena†. Galton refers to Gauss's assumption that errors in excess or in defect are equally probable, and says "this assumption cannot be justified in vital phenomena." He cites especially the cases of errors in human judgment, guessing at temperatures, tints, pitch, etc. He appeals to Fechner's law in its simplest form as evidence to the contrary, and placing the matter in the hands of D. McAlister‡, the law of frequency

$$y = y_0 \frac{h}{\sqrt{\pi} x} e^{-h^2\left(\log \frac{x}{a}\right)^2}$$

was deduced, and methods for fitting this curve were discussed. The curve is well-known in England and also on the continent §. It is therefore curious to find Ranke and Greiner attributing this curve to Fechner's work which was not published till 18 years later. It was not till I had made a fairly complete set of experimental determinations of the kind supposed to give this curve, that I finally discarded it. Thus I asked audiences of 100 to 300 persons to match tints in several ways, I asked them to guess heights, to determine mid-lengths, to state which figures in randomly distributed series were most closely circles,

* See Lacroix: *Traité du Calcul différentiel et intégral*, Tom. III. p. 136.

† *R. S. Proc.* Vol. 29, p. 365 *et seq.* "The Geometric Mean in Vital and Social Statistics."

‡ *R. S. Proc.* Vol. 29, p. 367 *et seq.* "The Law of the Geometric Mean."

§ It is cited by Kapteyn, for example.

squares, equilateral triangles, etc. etc. In all these results I found the distribution asymmetrical, but the most probable value was *not* the geometrical mean, nor the distribution the Galton-McAlister curve. One of the striking defects of the curve was its high contact at both ends. The distributions clearly often corresponded to curves in which the contour cut the axis at a finite angle. Another point was that the skewness was *in the opposite direction* to that presupposed by the reasoning from which the curve is deduced. It was precisely this experience which showed me that putting $x = \beta \log (\xi/a)$ in the Gaussian curve is not a sufficient generalisation.

Ranke and Greiner, to say nothing of Fechner himself, are remarkably vague as to the accurate determination of the position and constants of the Galton-McAlister curve. McAlister gives no clear description of how the curve is to be placed if neither the mode nor start of the range is known. I think it desirable therefore, having regard to the inferences I wish to draw, to give the fitting by my method of moments. I write the curve:

$$y = \frac{y_0}{x} e^{-\frac{1}{2c^2}\left(\log \frac{x}{a}\right)^2} \quad \text{.................................(xi).}$$

Differentiation shows us at once that the distance x_{mo} of the mode from the origin is given by:

$$x_{mo} = ae^{-c^2} \quad \text{...................................(xii).}$$

Integrating the expression $N\mu_n' = \int_0^\infty yx^n dx$ we find if $N =$ total frequency:

$$y_0 = N/(\sqrt{2\pi}c) \quad \text{.............................(xiii),}$$

and generally:

$$\mu_n' = a^n e^{\frac{n^2}{2}c^2} \quad \text{...................................(xiv).}$$

Thus the distance from the origin to the mean, x_{me}, is given by

$$x_{me} = ae^{\frac{1}{2}c^2} \quad \text{...................................(xv).}$$

Now write $e^{c^2} = \lambda$ and we have if μ_n be a moment coefficient about mean:

$$\left.\begin{aligned} \mu_2 &= \mu_2' - \mu_1'^2 = a^2\lambda(\lambda - 1) \\ \mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = a^3\sqrt{\lambda}(\lambda^4 - 3\lambda^2 + 2\lambda) \\ \mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = a^4\lambda^2(\lambda^6 - 4\lambda^3 + 6\lambda - 3) \end{aligned}\right\} \quad \text{...(xvi).}$$

Forming the usual constants of frequency we have:

$$\beta_1 = \mu_3^2/\mu_2^3 = (\lambda - 1)(\lambda + 2)^2 \quad \text{.......................(xvii),}$$

$$\eta = \beta_2 - 3 = (\lambda - 1)(\lambda^3 + 3\lambda^2 + 6\lambda + 6) \quad \text{..............(xviii),}$$

where $\beta_2 = \mu_4/\mu_2^2$;

$$d = x_{me} - x_{mo} = a(\sqrt{\lambda} - \lambda^{-1}) \quad \text{.......................(xix),}$$

$$\chi = \frac{1 - \lambda^{-\frac{3}{2}}}{\sqrt{\lambda - 1}} \quad \text{...(xx).}$$

We see then that the mean and second and third moments must be found to determine this curve. From the second and third moments we have β_1, whence by equation (xvii) λ is determined. The first equation of (xvi) then gives

$$a = \sigma/\sqrt{\lambda(\lambda-1)}.$$

Then equation (xv) gives the distance of the start of the curve from the known mean. Further since $\lambda = e^{c^2}$, we determine c and finally y_0 is determined by equation (xiii). There is no obscurity or difficulty with the fitting, if we use the method of moments. The cubic equation (xvii) is solvable at once either by Lill's, Mehmke's or Reuschle's mechanisms.

But what are the objections?

(i) The curve touches the axis at the end of the range. Skew curves extremely often cut it at a finite angle.

(ii) The skewness has a definite direction, which to be logically consistent we ought not to neglect, i.e. since λ is always > 1, d remains always of one sign.

(iii) Since $\lambda > 1$, η, the kurtosis, is always positive and the curve can only represent platykurtic distributions. It can never give a curve which deviates from the Gaussian curve in the direction of the Laplace-Poisson skew binomial for $p > {\cdot}2113 < {\cdot}7887$, because this is essentially leptokurtic.

(iv) The range of skewness given by χ is very limited. Differentiating χ we find it is a maximum for $\lambda = 1{\cdot}7200$ and this gives $\chi = {\cdot}2075$. The Galton-McAlister curve cannot therefore describe any curve whose skewness does not lie between 0 and ·2. A cursory examination of the observational results reached, shows that the skewness in all kinds of data over and over again exceeds ·2.

(v) β_1 and η are both functions of λ only*. Hence there is a relation between them or between η and χ. That is to say the kurtosis is determined by the skewness. The kurtosis must vanish with the skewness. But experience shows that many distributions are sensibly symmetrical and yet have far from zero kurtosis, e.g. nasal breadth in English women, etc. etc.

Finally consider the ratio β_1/η. If we approach the normal curve as the limit to a point binomial $(p+q)^N$ we have seen that

$$\beta_1/\eta = (1-4pq)/(1-6pq) \quad \text{......................(xxi)},$$

and this equals nothing if we take the symmetrical binomial. Otherwise it has a finite value depending upon the particular binomial along which we reach the Gaussian curve. The Galton-McAlister curve, if we make $a\sqrt{\lambda}$ infinite, but $a\sqrt{\lambda}\sqrt{\lambda-1}$ finite, approaches the Gaussian curve.

* Actually it is

$$\beta_1^4 - 12\beta_1^3 + 156\beta_1^2 + 64\beta_1 - \eta^3 + 12\eta^2 - 36\eta + 18\beta_1^2\eta - 6\beta_1\eta^2 - 117\beta_1\eta = 0.$$

I have to thank my assistant, Mr J. Blakeman, for much aid in the analysis of this section and the following section of this memoir.

Now for the Galton-McAlister curve

$$\beta_1/\eta = \frac{(\lambda+2)^2}{\lambda^3+3\lambda^2+6\lambda+6},$$

and this approaches the limit 9/16, its maximum value, when λ approaches unity. If we take $(1-4pq)/(1-6pq)=9/16$, we get imaginary values for p and q. Thus while the normal curve itself gives an indeterminate value for $\beta_1/\eta=0/0$, and as Laplace has shown describes with fair accuracy any slightly skew binomial with large power, the Galton-McAlister curve cannot describe even approximately any skew binomial, however near to a normal distribution.

On all these grounds we see that the "law of the geometric mean" fails to supply the fundamental need of describing the modal difference, the kurtosis and the skewness of actual frequency distributions. It cannot describe these physical characteristics of the frequency.

C. (iii) *Fechner's Double Gaussian Curve**.

We have noted that the Gaussian curve was first deduced by Laplace to represent a finite number of the terms of a binomial expression, and that Gauss deduced it on hypotheses which amount to the following:

(i) The arithmetic mean is the most probable value.

(ii) Deviations in excess and defect of the mean are equally probable if of the same magnitude.

(iii) The facility of an increment is the same for all values of the character.

Now every one of these assumptions is negatived when the double Gaussian curve is used, and yet the Gaussian curve *which is only deduced by aid of them* is adopted to describe what conflicts with its fundamental axioms. This proceeding is the reverse of logical. However, if the double Gaussian curve be adopted, there is absolutely no reason why we should adopt the rough process by which Fechner determines the mode and obtains the constants of the distribution. The fitting by my method of moments is perfectly straightforward, and as it leads to the points we have to consider it will be indicated here. Let the two half curves be:

$$\left.\begin{aligned} y_1 &= \frac{n_1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\frac{x^2}{\sigma_1^2}}, \quad x>0 \\ y_2 &= \frac{n_2}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\frac{x_2^2}{\sigma_2^2}}, \quad x<0 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xxii)}.$$

Then, since the modal value is common, $\sigma_1/n_1=\sigma_2/n_2$. Further, the total frequency $N=\frac{1}{2}(n_1+n_2)$. Now write $\kappa=\sqrt{2/\pi}$ and $u=\sigma_1-\sigma_2$, $v=\sigma_1\sigma_2$. Then taking moments round the mode we easily find:

* Here again it is historically incorrect to attribute these curves to Fechner. They had been proposed by De Vries in 1894, and termed "half-Galton curves," and Galton was certainly using them in 1897. See the discussion in Yule's memoir, *R. Statist. Soc. Jour.* Vol. LX. p. 45 *et seq.*

$$\mu_1' = \kappa u, \quad \mu_2' = u^2 + v, \quad \mu_3' = 2\kappa u (u^2 + 2v), \quad \mu_4' = 3\{u^2 (u^2 + 3v) + v^2\}.$$

Transferring to the mean we deduce:

$$\mu_2 = u^2 + v - \kappa^2 u^2 \quad \text{.........(xxiii)},$$

$$\mu_3 = \kappa u (v - u^2 + 2\kappa^2 u^2) \quad \text{.........(xxiv)},$$

$$\mu_4 = 3\{u^2 (u^2 + 3v) + v^2\} - \kappa^2 u^2 \{2u^2 + 10v + 3\kappa^2 u^2\} \quad \text{.........(xxv)}.$$

μ_2, μ_3, μ_4 will be known quantities as soon as the frequency distribution is known.

Now determine $\beta_1 = \mu_3^2/\mu_2^2$ and write $\lambda' = u/\sqrt{\mu_2}$, we easily find:

$$\left(\frac{6}{\pi} - 2\right) \lambda'^3 + \lambda' - \sqrt{\frac{\pi}{2}}\, \beta_1 = 0 \quad \text{.........(xxvi)}.$$

This cubic* gives by its real root the value of $u = \sigma_1 - \sigma_2$. We then easily deduce

$$\left.\begin{aligned} \sigma_1 &= \tfrac{1}{2}\sqrt{\mu_2}\,(\{4 + (4\kappa^2 - 3)\lambda'^2\}^{\frac{1}{2}} + \lambda') \\ \sigma_2 &= \tfrac{1}{2}\sqrt{\mu_2}\,(\{4 + (4\kappa^2 - 3)\lambda'^2\}^{\frac{1}{2}} - \lambda') \end{aligned}\right\} \quad \text{.........(xxvii)}.$$

These determine the different variabilities of the two halves. Then

$$n_1 = \frac{2N\sigma_1}{\sigma_1 + \sigma_2}, \qquad n_2 = \frac{2N\sigma_2}{\sigma_1 + \sigma_2} \quad \text{.........(xxviii)}$$

give the frequencies in each Gaussian curve, while

$$\mu_1' = \kappa \sqrt{\mu_2}\,\lambda' \quad \text{.........(xxix)}$$

fixes the position of the origin relative to the known mean value of the system. Thus the complete solution depends on a knowledge of the mean, and the second and third moment coefficients. As before the cubic is readily solved by Lill, Reuschle or Mehmke's mechanisms.

The analysis is now a little more complex than in the case of the Galton-McAlister curve. Write $\epsilon = v/u^2 = \sigma_1\sigma_2 (\sigma_1 - \sigma_2)^2$. Then we have:

$$\left.\begin{aligned} \beta_1 &= \kappa^2 (\epsilon - 1 + 2\kappa^2)^2/(1 - \kappa^2 + \epsilon)^3 \\ \beta_2 &= \{3(1 + 3\epsilon + \epsilon^2) - \kappa^2 (2 + 3\kappa^2 + 10\epsilon)\}/(1 - \kappa^2 + \epsilon)^3 \end{aligned}\right\} \quad \text{.........(xxx)}.$$

Thus again we see that β_1 and β_2 are both functions of ϵ only, or the skewness is not independent of the kurtosis†. Whenever the skewness is zero, the kurtosis must also be zero or the curve be normal.

Now consider the expression $1 - \kappa^2 + \epsilon$ which we will write γ, or,

$$\gamma = \cdot 36338 + \sigma_1\sigma_2/(\sigma_1 - \sigma_2)^2.$$

The last term is positive or γ must be $> \cdot 36338$.

* This cubic was, I believe, first given by Edgeworth.

† The actual relation is:

$$29521\eta^2 + 62500\beta_1^2 - 110506\beta_1\eta + 13468\eta^3 - 11345\beta_1 + 15925\eta = 0,$$

which, as in the Galton-McAlister case, has no obvious physical significance.

Now:

$$\eta = \beta_2 - 3 = \frac{\cdot 45352}{\gamma} - \frac{\cdot 05003}{\gamma^2},$$

and this with the above limitation to the value of γ can never become negative. Hence the double Gaussian curve is, like the Galton-McAlister curve, invariably platykurtic. Now consider the value of

$$\beta_1 = \frac{\cdot 63662}{\gamma} - \frac{\cdot 11477}{\gamma^2} + \frac{\cdot 00518}{\gamma^3}.$$

Hence the ratio $\frac{\beta_1}{\eta}$ tends as γ increases to take the value 1·40374. Equating this to $(1 - 4pq)/(1 - 6pq)$ we see that the double Gaussian curve approaches the normal curve along the particular platykurtic binomial $p = \cdot 8985$, $q = \cdot 1015$, or it cannot in the neighbourhood of the normal curve represent any skew binomial but this.

Lastly it may be shown that β_1 has its maximum value when $\gamma = \cdot 36338$ or its minimum value. Thus we find that the maximum possible value of β_1 is about ·99. In the same way the maximum skewness is 1·3236. These values are sufficiently high to cover the great bulk of cases, but I have found $\beta_1 = 4\cdot 071$ for scarlet fever incidence, $= 1\cdot 9396$ for age of brides who marry men in their 24th year and $= 4\cdot 1683$ for the distribution of lips in the medusa *P. pentata*. These exceptions suffice to show that the curve is not general enough.

Summing up we conclude that the double Gaussian curve is not satisfactory because theoretically

(i) It starts by denying the very axioms from which alone we can reach the Gaussian curve;

and empirically because

(ii) It can describe no frequency distribution which cuts the axis at a finite angle, and such distributions constantly occur.

(iii) It is essentially platykurtic. Therefore it is not available for leptokurtic curves, nor even for any but very special skew binomials, i.e. those in which p does not lie between ·2113 and ·7887. As we approach close to the normal curve we get nearer and nearer to one definite point binomial, i.e. that in which $p = \cdot 8985$.

(iv) There is always a relation between the skewness and the kurtosis, or these important physical constants are not independent. In particular we cannot have any form of symmetry but the mesokurtic.

(v) The range of β_1 and of the skewness is fairly large, but frequency distributions actually occur markedly outside this range.

(vi) Lastly, and of much import, the kurtosis can never exceed ·8692, or the maximum value of $\beta_2 = 3\cdot 8692$. This degree of kurtosis is exceeded in a great number of distributions. Thus in the lips of *P. pentata*, in tint guessing, in the breadth of male English skulls, in the nasal breadth of female English skulls, in no less than eight of Duncker's series in the case of *Gelasimus pugilator* and in various other distributions. In all these cases the platykurtosis is significant, and the double Gaussian curve fails us hopelessly.

C. (iv) *The Edgeworth-Kapteyn Curves.*

Kapteyn*, without recognizing Edgeworth's priority, has proceeded in the manner indicated on p. 178 above. He assumes that some quantity x obeys the normal distribution

$$y = \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}x^2/\sigma^2}.$$

He then takes $x = F(X) - M$ and reaches the frequency distribution:

$$Y = \frac{N}{\sqrt{2\pi}\sigma} F'(X)\, e^{-\frac{1}{2}\{F(X)-M\}^2/\sigma^2} \quad\text{......................(xxxi).}$$

Thus far (as we have already shown) nothing has been achieved, because this equation may by a proper choice of $F(X)$ represent any curve whatever. As Kapteyn himself says, following Edgeworth, "as $F(x)$ may represent any function, we see that the equation may be made to represent any curve whatever. Therefore it must be the most general form of frequency curve possible" (p. 17). There is, however, one point to be raised here. What is x of which the observed character X is a function? Is it, as in the explanatory illustrations cited by Kapteyn, another characteristic of the organism? If so we ought in some cases to be able to determine it. What is the character which obeys the normal law? For example, sagittal arc in English women is almost exactly normal in its distribution, and nasal breadth is very asymmetrical. Shall we take x = sagittal arc and X = nasal breadth and make

$$x = F(X) - M?$$

Now every biologist knows that such a relation is not in the least true. No two characters in an organism are in any way connected by a mathematical function, such that when one is given the other is determined. The relation is always of the loose kind that we term association or correlation. X does not fix x, and a multitude of x's with varying degree of probability are associated with a given x. This correlation is often of a very low order. Between any two characters of a given organism, no such relation of perfect correlation as that involved in Edgeworth or Kapteyn's relation has ever been discovered. Very imperfect correlation or at any rate all degrees of correlation have been invariably demonstrated to exist. The function x has no real existence as a biological entity. It is only a mechanism for introducing the normal curve, and is not a true character of the organism at all. Supposing, as in English female crania, nasal breadth is asymmetrical, what is the quantity which is symmetrically distributed of which nasal breadth is a function? It has no reality in the organism at all, and Kapteyn proceeds to make it still more impossible in the following manner. If x has

* *Skew Frequency Curves in Biology and Statistics*, Groningen, 1903.

existence at all, its limits lie between $-\infty$ and $+\infty$, i.e. the whole range of the normal curve. But in order to get a range limited at one end, not the whole series of values of X corresponding to x are taken. A value is selected for X, such that X becomes impossible after a certain value of x. In other words, x is a character which although following the normal curve is abruptly terminated as far as X is concerned at a value with a finite frequency! Kapteyn takes*

$$x = (X + \kappa)^q - M,$$

where κ, q and M are to be determined from the data.

For example, in Professor Weldon's data for the measurement of foreheads of *Carcinus moenas*, Kapteyn (p. 39) finds that with our notation

$$\begin{aligned} \sigma &= \cdot 002{,}204, \\ M &= \cdot 002{,}561, \\ \kappa &= -0 \cdot 5781, \\ q &= 2 \cdot 21. \end{aligned}$$

Thus when $X = -\kappa$, we have $x = -M = -\sigma$ roughly, or the Gaussian frequency curve for x is to be abruptly cut off, and about 15 per cent. of its tail discarded. If it be said that this could be achieved by natural selection of foreheads, the reply is the simple question: Please show what physical character in a crab is given by an abruptly truncated normal curve! The fact is no such character has ever been met with, and it must be recognised that x represents a wholly fictitious variable having no physiological relation to the character X at all, but introduced solely to reduce the frequency by hook or by crook to that fetish distribution the Gaussian curve.

We can now sum up the objections to Kapteyn's method.

Theoretical:

(i) There is no justification whatever for assuming that some character x actually exists which obeys the normal law of distribution, and that the observed character is a function of this. Some characters are found as a rule in any organism which obey the normal law, but no two characters in an organism have ever been found to be the one a mathematical function of a second, they are always imperfectly correlated.

(ii) Kapteyn's hypothesis involves if his normal character were a physiological entity, that distributions of organic characters should occur which would be represented by fragments of Gaussian curves, or such curves abruptly curtailed. We have no experience of such distributions in actual vital statistics. If they did exist they would contradict the first two axioms on which the Gaussian law itself is based, and would thus deprive that law of the sole justification for its application. As it cannot be supposed that all skewly distributed characters X in an organism are functions of one and the same x, for in this case they would be

* This becomes the Galton-McAlister curve for the limit $q=0$.

perfectly correlated with each other, which is contrary to experience, it must follow that if Kapteyn's hypothesis were correct large quantities of characters distributed in truncated Gaussian curves ought to appear when we deal with variation. The total absence of such characters is evidence that the x-characters are shadow variables and of no biological import.

(iii) The previous statements reduce Kapteyn's special choice of

$$F(X) = (X + \kappa)^q$$

to a mere artifice adopted to get an empirical curve of variation by Edgeworth's hypothesis. Many other functions are *à priori* equally valuable, and might be adopted to get curves of limited range, e.g.

$$F(X) = (X + \kappa)^q (X - \kappa')^{q'}.$$

The hypothesis gains nothing in logical consistency by its appeal to the Gaussian curve; that appeal is one adopted for convenience of fitting, and the sole test of Kapteyn's curve is empirical goodness of fit.

Practical:

(iv) *Every frequency curve should be a graduation formula.* Kapteyn's method of fitting is by equating certain *total* frequencies in order to determine his four constants. They thus fail to successfully smooth any special causes tending to exaggerate any particular frequency group. Such screening of special causes of frequency deviation is far less likely to occur when we use the method of moments, which is a true method of graduation*.

(v) *We ought in every law of frequency distribution to be able to judge of the effect of the unit of grouping on the values of the constants.* This has been satisfactorily achieved for, perhaps, the bulk of cases, when the method of moments is used by Sheppard's corrections†. In Kapteyn's process we have no means of ascertaining the extent to which the size of the unit of grouping influences the constants of his distribution.

In his Example II., for instance, he takes his curve to accurately reproduce the total area of the group of houses under £10 annual value. What difference would

* Thus Kapteyn deals with some statistics of the values of house property in England fitted by me (*Phil. Trans.* Vol. 186 A, p. 396). I specially state that £20 was the limit to taxable value, and that accordingly the frequency of houses immediately below this value will be exaggerated. Kapteyn's method fails to indicate such a source of *à priori* recognised irregularity. For example, one of his conditions is that the houses of value less than £10, i.e. more than half the total frequency, shall be *identical* in his result with the observed frequency. He thus cuts away at once any possibility of smoothing this group or allowing for the large probable error in it due to random sampling even. His method leads to a limiting house value of £2. 2*s*., while mine leads to £4. 4*s*. Mine corresponds to a weekly rent of about 2*s*.; his to a weekly rent of 1*s*. The latter rent hardly occurs in England unless the house is given in part payment of wages, or in charity. Kapteyn says that his distribution starts with a zero frequency, and mine with an infinite ordinate. "It seems hardly admissible that the latter solution can be in accordance with nature (*sic*) in this particular." Why not? An infinite ordinate may and does in my case give a finite frequency.

† See reference, p. 187.

be made if the first group had included only houses under £5? We are unable to answer this question.

(vi) *Every frequency curve should be determined by constants of which the probable errors are easily deducible.* The method of moments admits of the probable errors of the moments being easily determined (see *Biometrika*, Vol. II. pp. 273 *et seq.*). My system of skew curves gives all the constants in terms of moments whose probable errors are known.

The moments in Kapteyn's theory depend on the integration of:

$$\int_{-M}^{\infty} (x+M)^{\frac{n}{q}} e^{-\frac{1}{2}x^2/\sigma^2} dx,$$

and there is no means of readily evaluating this integral. *In fact the arithmetical mean*, the standard deviation, the skewness and the kurtosis, and the modal divergence are unobtainable from the constants of Kapteyn's theory.* This seems to me sufficient to deprive the method of any practical significance even as an empirical representation.

It has further been shown by Sheppard that the probable errors of constants determined by class frequencies (partial areas) are higher than when these constants are determined by the method of moments. We may give the above statement a separate paragraph as:

(vii) The fundamental physical constants of the frequency distribution are not determinable from Kapteyn's empirical curve.

To illustrate the results of this want of a knowledge of the probable errors, I turn to the three illustrations given by Kapteyn.

Example (i). Observations on the *Threshold of Sensation.* Kapteyn himself shows that his solution is hardly less satisfactory if he uses the Galton-McAlister curve (our equation (xi) p. 194). He does not therefore know whether $q = \cdot 00$ and $q = -\cdot 04$ differ within the probable error of q.

Example (ii). *Valuation of House Property.* Kapteyn fits this with a Galton-McAlister curve for his q comes out ·00. Owing to the difficulty in calculating moments, we cannot do more than approximate to the value of χ the skewness in these data. I make it 1·8. It is certainly well over unity. We have already seen that it is impossible for a Galton-McAlister curve to give a skewness above ·21. The apparent agreement Kapteyn finds for the frequencies is not therefore sufficient evidence that the fundamental constants of the distribution will be really given by reasonable values.

Example (iii). Foreheads of *Carcinus moenas.* Kapteyn fits these first with $q = 2\cdot 21$. "The agreement seems satisfactory." Then with $q = 0$, or a Galton-McAlister curve, "The representation is hardly less satisfactory." Then with

* "The arithmetic mean of all the X's cannot be *generally* found in a simple and rigorous way," Kapteyn, p. 44.

$q = \infty$, "The representation, though sensibly less satisfactory than that by the [previous] solutions, is still pretty close."

A method by which a fundamental constant of the distribution may take any value between 0 and ∞ and still give a "pretty close" representation, must I think condemn itself. Such a statement demonstrates effectively that the author has not yet determined numerically or even approximately in his own mind the probable errors of the constants he uses.

It will be seen that as far as the three illustrations Kapteyn himself gives go he has not advanced the matter beyond the Galton-McAlister curve. That curve fits reasonably (according to Kapteyn) all his three series. But the skewnesses of the three series are respectively ·72, 1·8 and ·32. I have calculated these roughly, but I think there can be no doubt of their approximate correctness. In every one of these cases the skewness sensibly exceeds the maximum limit of skewness, i.e. ·21, possible for the Galton-McAlister curve which Kapteyn applies to them*.

C. (v) *The General Results which flow without the Third Gaussian Axiom.*

It seems to me accordingly that very grave objections can be raised not only from the theoretical but from the practical standpoint to the methods I have discussed which attempt to allow for asymmetry, i.e.

(i) The Galton-McAlister Geometrical Mean Law

(ii) The Galton-Fechner use of Half Gaussian Curves,

(iii) The Edgeworth-Kapteyn use of transformed Gaussian Curves.

All these experienced statisticians differ *in toto* from the opinion of Ranke and Greiner—that we need not trouble about descriptive curves for asymmetrical distributions—but their methods seem to me unsatisfactory theoretically and insufficient practically, because they still make a fetish of the Gaussian axioms. They do not return to the Laplace-Poisson method of replacing those fundamental axioms by more general conceptions. If a Gaussian curve does not fit, they will consent to deduce their own curves from a truncated Gaussian curve, which some shadow variable of the mathematician is supposed to follow, and of which we have no experience in any organic characters hitherto measured. Indeed if we had such experience, it would at once negative the very axioms on which the Gaussian curve is based.

Now it seems to me that all these attempts, whether embodied in the general method of Edgeworth or in the special hypotheses of Galton-McAlister or Kapteyn, amount to abolishing the third of the Gaussian assumptions, namely that small increments of the variable or the character are independent of the total already reached. That is to say that they amount to saying that increments of the

* I am unable to say how far the general form of Kapteyn allows for the requisite range of skewness and kurtosis, because neither the modal difference, nor the standard deviation, to say nothing of the higher moments, can in general be evaluated.

variate are *correlated* with the value of the variate already reached*. Galton and Fechner made the increment proportional to the variate. But in our ignorance of the actual nature of variation in organisms, we have no reason at all for making such a narrow assumption. We can to please our critics put the matter as I have already indicated in the Gaussian form. We simply assume that if the causes of variation in the immediate neighbourhood of the character x_0 remained the same for the whole range they would give a normal curve, hence we should have a relation of the form:

$$\frac{1}{y}\frac{dy}{dx}=-\frac{x}{\sigma^2}.$$

They do not, however, remain the same; the tendency to vary at x is a function of x, in other words $\sigma=\sigma_0\sqrt{F(x)}$, where $F(x)$ is an arbitrary function. We have then

$$\frac{1}{y}\frac{dy}{dx}=-\frac{x}{\sigma_0{}^2F(x)}.$$

This is a result as general as Edgeworth's and more so than Kapteyn's or Fechner's. We now take the simplest possible functional series for $F(x)$, i.e.

$$=a_0+a_1x+a_2x^2+\ldots.$$

The coefficients a_0, a_1, $a_2 \ldots a_n$ can be found at once in terms of the moments†, and my special curves result if we stop at a_2. Against going to higher powers are the objections I have raised in my memoir on skew correlation‡, namely (i) that the higher powers involve moments of the 5th and higher orders and their probable errors are very large, (ii) that it has not yet been shown that going to a_2 does not suffice to describe all the types of frequency which occur in common practice.

The above is the simplest and most general form into which I would put my theory of asymmetrical frequency for those who feel compelled to approach all frequency from the Gaussian standpoint.

D. *Specific Criticisms of Ranke and Greiner on my Theory.*

I think these may be summed up as follows:

(*a*) That all distributions of variates are continuous, and that accordingly no curves, however closely they may approximate to finite discontinuous series like the binomial and the hypergeometrical series, can be applicable to variation in nature.

* Suppose we draw r cards from a pack, and wish to consider the chance of s being of one suit, we may do so by drawing one card at a time, observing it and returning it, and then drawing again. Here there is not correlation between the successive contributions to r. Or we may draw the r cards, without replacing the individual; here the successive contributions are correlated with the previous contributions, and the third Gaussian principle is upset.

† "Mathematical Contributions to the Theory of Evolution, XIV. On the General Theory of Skew Correlation and Non-Linear Regression," *Drapers' Company Research Memoirs.* Biometric Series, II. (Dulau and Co. 1904) p. 6.

‡ *Ibid.* p. 7.

(*b*) That what I term the "number of cause-groups" must be infinite in number, for without such infinity it is impossible to reach continuity.

(*c*) That looked at from the standpoint of binomial or hypergeometrical series the constants of some one or more of my curves may become unintelligible.

Now not one of these objections has any application to the method which has been used in this paper to deduce the differential equation to my curves. But I still think there are very grave objections to every one of the above statements.

To begin with (*a*). We meet in an immense variety of living forms with discrete variates. For example, the number of teeth on the rostrum of a prawn, the number of lips of a medusa, the number of veins in a leaf, the number of glands in a swine's foot, the number of tentaculocysts in *Ephyra*, the number of individuals in a litter, the number of bands on snails' shells, the number of somites in the body of an earthworm, the number of petals or sepals in a flower, etc. etc. Are we to put all these distributions of variation on one side because Ranke and Greiner hold that all distributions of variates are continuous? We have in these cases probably continuous causes producing discontinuous distributions. Are we not to use the areas of a continuous curve to give the frequency of such discrete variates?

Consider for example the function given by:

$$y = N \times \overset{n}{\underset{0}{S}} \left\{ p^{n-r} q^r \frac{\lfloor n}{\lfloor n-r \lfloor r} \frac{1}{\sqrt{2\pi}\,\sigma_r} e^{-\frac{1}{2}(x-rc)^2/\sigma_r^2} \right\}.$$

This is compounded of $\overline{n+1}$ normal curves, the area of the $(r+1)$th normal curve being $Np^{n-r}q^r \dfrac{\lfloor n}{\lfloor n-r \lfloor r}$, i.e. the $(r+1)$th term of the binomial $N(p+q)^n$, and this $(r+1)$th normal curve has σ_r for its standard deviation. The origin of the system is at the mode of the normal curve corresponding to $r=0$, and the means of these normal curves are spaced equal distances c apart.

When every $\sigma_r = 0$ we have discrete variation. When σ_r is small, less say than $\frac{1}{8}c$, it would probably be difficult to distinguish the result from discrete variation. Enlarging σ_r we pass on till we get a system which it would be practically impossible to distinguish from continuous variation, even if n were only moderate in magnitude. I lay no stress whatever on the above expression because I am in no sense pledged to any Gaussian curve, but it illustrates well what I want to express: namely, in actual nature the frequency might fundamentally fall on certain values of the character, but that the effect of nurture, environment, and growth may well scatter the values of the variable round the fundamental value, so that continuity of variation is all that can be actually observed. The number of somites in an annulose animal is discrete and probably inherited, but the length of the body may appear as a continuous variate. I do not think for a moment that the distinction made by Ranke between discrete and continuous variation, and the further statement that variation in man is essentially continuous, is at all valid. Our units of grouping for the numbers available are not very fine, we can hardly successfully classify a few hundred observations into more than 20 groups, and with this unit of grouping it would be practically impossible to distinguish between the apparently continuous distribution and a discrete distribution of a similar number of classes with the variates modified by growth, nurture or any other scattering tendencies. Ranke appears entirely to have overlooked the current biological theory of inheritance summed up in the words *inheritance by determinants.* Such theories, whether they be those of Weismann or Mendel, lead us directly to discrete variation*. The discreteness of the variation will be more or less, in many cases probably entirely, obscured by the environmental influence. In such cases the number of fundamental cause-groups is not infinitely great, and Ranke is overlooking current biological views when he asserts that we *must* take "die Anzahl der Elementarursachen selbst als unendlich gross und die Grösse der Wirkung der einzelnen Ursache als unendlich klein." If the number of determinants which fix a character is finite, that character would correspond to a discrete variation of limited range. If the number of determinants be very large, the distribution would by Laplace's theorem be represented more and more closely by the normal curve.

I toss ten coins into the air and for every head in the result I pay a gramme of gold-dust, the frequency distribution of gold-dust would closely be given by the terms of the binomial $(\frac{1}{2}+\frac{1}{2})^{10}$ as in the points of our Fig. 1. But suppose instead of weighing my gramme of gold-dust accurately, I give a "handful" of sugar. If 6 heads turn up I give six handfuls of sugar, but each of these will not be exactly my standard mean handful. I am unlikely to give either five or seven standard handfuls as my six approximate handfuls, but in some cases even these might be possible; we pass in fact from discrete to continuous variation, and the multimodal character of the discrete variates will disappear with the roughness of the handfuls, or have the peaky appearance of random sampling. The total area up to any midpoint between two discrete groups s and $s+1$ will be given by the continuous integral which represents the first s terms of the binomial. If we have two such total areas, one up to the midpoint between groups s and $s+1$ and the other up to the midpoint between groups $s+1$ and $s+2$, then an interpolated area between these values as given by the continuous integral will be sensibly the same as if, c being the unit of discrete difference, we determined a curve corresponding to the mean binomial frequency in the *Spielraum* c, i.e.

$$\frac{N}{c}p^{n-r}q^{r}\frac{\lfloor n}{\lfloor n-r\,\lfloor r},$$

by simply fractionising r, i.e. we replaced the factorials by Stirling's theorem or used Γ functions, and supposed r to change continuously from s to $s+1$. This is

* Thus I have shown that a generalised Mendelian theory leads directly to skew binomial distributions of characters in the general population. *Phil. Trans.* Vol. 203 A, pp. 53—86.

merely another way of looking at the change from discrete to continuous variation, due to the influence of a multitude of causes on the discreteness of the variates which fall into a given *Spielraum*. I still find nothing absurd in the statement that the actual effect of the scatter is sensibly equivalent to a fractionising of the indices. It is simply equivalent to the statements, (i) that the ordinates of the Gaussian curve closely give, even for small values of n, the terms of the binomial, (ii) that the ordinate of the Gaussian curve between two terms of the binomial closely gives a fractionised binomial term (owing to Stirling's theorem being true for fractionised factorials or Γ functions), (iii) that we have no knowledge of how the "scatter" within the *Spielraum* may be distributed so as to give a continuous effect*. Now these points are not in the least needful for the deduction of my skew curves, they are merely given here because in our complete ignorance of the nature of the causes, hereditary and environmental, which produce continuous variation, I think we have no warranty for saying that a limited number of cause-groups is impossible, or that no such limited number of fundamental cause-groups could give a continuous variation. In the present state of our knowledge we cannot agree with Ranke in sweeping away as impossible all the discreteness which follows from determinantal theories of inheritance. We cannot afford to be dogmatic as to the continuous or discontinuous character of the ultimate sources of variation and any effective theory must like the Laplace probability integral be equally applicable to the sum of a discontinuous series as well as to the areas of a continuous curve.

(*b*) Any finite series of cause-groups, Ranke tells us, must lead to discontinuity.

I have endeavoured to show above that the discontinuity may be as real and yet as undetectable as the distribution of lengths, say, of the vertebral columns of sharks which yet depends on the number of discrete vertebrae, with a scatter of their individual sizes. But Ranke's argument in itself is a false one, many discontinuous systems lead at once to continuous distributions. In our ignorance of the exact sources of variation, all we can do is to show that a limited number of cause-groups can quite well lead to continuous variation. To take a perfectly arbitrary illustration, suppose that a character can only take values lying between a_1 and a_2 and that this character is to be settled by the determinants derived from $s+1$ ancestors, i.e. suppose all but $s+1$ to be cast out in the successive divisions of the germ-cell. Then it by no means follows that the character will be a *blend* of these $s+1$ determinants, one or other of them may be *dominant*. It does not follow that the dominant one represents either the one with the least or the greatest value of the character. It might be the one with $s-r$ determinants below and r above it.

* For example I have a variety of Binomial machines or "Quicunxes" like that figured in my memoir, *Phil. Trans.* Vol. 186 A, Plate I. Fig. 2, p. 414. It is quite possible to arrange a quicunx in which there are only a limited number of compartments, but in which the top of the seed in these compartments is not horizontal, but gives a continuous curve, e.g. the greater air draft of the greater frequency might be used to pile up the material in any receptacle on the side of the greater frequency.

In this case the frequency of individuals with character x would be given by the curve

$$y = y_0 (x - a_1)^{s-r} (a_2 - x)^r.$$

This is a perfectly continuous curve, and one of my series of skew curves. Thus it is quite conceivable that a finite number of fundamental cause-groups should lead to an absolutely continuous distribution*.

Now how does Ranke treat this illustration? He first states that all continuity must involve an infinite number of cause-groups, or variation in man being continuous, must be associated with an infinite number of cause-groups. He had this very case before him, and yet he writes:

Die Analyse der Elementarursachen ergibt uns also unweigerlich die bisher immer angenommene unendliche Anzahl derselben, die unendliche Kleinheit der Wirkung jeder einzelnen Ursache und die Kontinuität der moglichen Wirkungsgrade.

Sie ergibt also wirklich die Verhältnisse, die wir zum Verständniss der kontinuerlichen Variationskurve ganz unumgänglich nötig haben. Denn wie soll eine kontinuerliche Kurve sich aus der Kombination endlicher Bausteine ergeben (S. 321)?

Ranke only gets out of the difficulty by asserting that since the number of causes is finite, but must be infinite for variation, my continuous curve based on a finite number of cause-groups cannot represent variation! A more remarkable specimen of circular reasoning can hardly be conceived. The fact is that Ranke suffers from the old third Gaussian axiom, i.e. the supposition that the increments that go to build up the variate are independent of each other. The fundamental cause-groups are by no means *Bausteine* in the sense that the total variate is the sum of these *Bausteine* placed on top of each other! The causes determine the magnitude of the variate, but not at all necessarily by their sum.

(*c*) Ranke asserts that some of my curves have constants which if we endeavour to interpret them from the standpoint of the binomial give impossible or improbable values for the constants.

The answer to this is that the series were only the scaffolding to deduce the curves. The differential equation to the curves contains the limit to a good many other frequency systems which directly diverge from the fundamental axioms of Gauss. I used the original series as a means of dispensing with the Gaussian axioms in familiar cases, but the result reached involves a good deal more than can be interpreted by the original series. Ranke can only see absurdity in a binomial with a negative p or q. But the nature of the sources of variation is so little known to us that we cannot possibly assert the absurdity of such values. We may not indeed be able to directly interpret them in the case of man, say, but they occur and recur in chance investigations. I will illustrate this in one case only, but such will demonstrate the required possibility and dispose at once of Ranke's argument as to absurdity.

* Making r and a_2 infinite, but $s-r$ finite, we get the curve I have deduced as the limit to a binomial of finite power. In other words, that curve is also shown to correspond to a possible continuity.

Suppose an organ to require the conjunction of exactly n determinants of one kind to fix it, but that the size of the organ depends on how soon this conjunction takes place. Let D be the necessary kind of determinant and π the chance that it is left in the right position after each operation, say a cell-division or cell-fusion. Let D' be any other kind of determinant and κ its chance of appearing, so that $\kappa + \pi = 1$. Then if D appears n times in the first n operations, we have a certain size for the organ, if in the first $\overline{n+1}$ operations another size, and if only in the first $n+r$ operations a third size. But the chances of these respective appearances are the terms of the series

$$\pi^n + n\pi^n\kappa + \frac{n(n+1)}{2}\pi^n\kappa^2 + \ldots + \frac{n(n+1)(n+2)\ldots(n+r-1)}{\underline{|r}}\pi^n\kappa^r + \ldots$$
$$= \pi^n(1-\kappa)^{-n} = \left(\frac{1}{\pi} - \frac{\kappa}{\pi}\right)^{-n} = (p-q)^{-n}, \text{ say,}$$

where $p - q = 1$. Here p and q have lost the condition that they are both to be less than unity. I do not for a moment suggest that this is the real interpretation of a binomial with a negative q. I only assert that because we are as yet unable to certainly interpret such expressions, the absence of interpretation does not involve the absurdity which Ranke postulates.

There are many other matters to which I might justifiably take exception in Ranke and Greiner criticism*, but I think I shall have said sufficient to convince the impartial reader of the following points:

(i) The great bulk of modern statisticians are agreed that the Gaussian law is absolutely insufficient to describe observed facts. They may disagree as to the method of supplementing it. I do not think that the opinion of Ranke and Greiner can possibly weigh against those of Poisson, Quetelet, Galton, Edgeworth, Fechner and Kapteyn—all authorities who have had to deal for years with statistical data.

(ii) The original use of the probability integral (the areas of the Laplace-Gaussian curve) as introduced by Laplace was to represent the sum of terms of a discontinuous series. To the mathematical mind there is no absurdity in this replacement of discontinuity by continuity; it is the basis of the Euler-Maclaurin theorem.

(iii) The dogmatic assertion of Ranke that variation in man is due to an infinite number of infinitely small fundamental cause-groups, simply neglects the

* For example, all the discrete variates mentioned on p. 205 have been dealt with by biometric writers, and many others besides, yet Ranke speaks as if such writers had not dealt with discrete variation. He speaks of Ludwig's multimodal curves for flowers as if there had been no controversy as to the actuality of the "Fibonaccizahlen" modes, when due regard is paid to homogeneity of season and environment. He speaks as if Johannsen had demonstrated normal variation in his "Erbsenspopulation," when he has really applied no valid criterion whatever to test for asymmetry, etc. In short he seems to me to have neglected a great deal of the modern literature of the subject, and, if I may venture to say so, to write over-dogmatically on what he has read.

whole determinantal theory of inheritance. A complete theory of asymmetrical frequency must describe in the manner of Laplace's probability integral either continuous or discontinuous variation.

(iv) The apparent or practical continuity of many variation data may be due either to real continuity or to discontinuity effectually masked by: (*a*) the relative paucity of material and roughness of our measurements, which compels us to divide it into groups of the same order of number as the number of determinants, (*b*) the influence of age, nurture and environment superposed upon the pure hereditary results, or (*c*) the fact that many of the characters measured by us are built up of a larger or smaller, but not necessarily an infinite, number of simple organs or characters, which may possibly individually have discontinuous variation*.

(v) The assertion of Ranke that a finite number of fundamental cause-groups *must* lead to discontinuity is disposed of by illustration. It is quite possible to invent a great variety of determinantal systems—the number of the determinants being finite—which lead to continuous variation. In our present ignorance of the sources of variation, especially of the mechanism of inheritance, it would be idle to lay weight on any special interpretation of this kind. It is important, however, to observe that continuity or discontinuity of variation are not, as Ranke asserts them to be, associated with the finite or infinite number of the cause-groups.

(vi) The absurdity which Ranke finds in the values taken by some of the constants of my curves, exists only when a very narrow view is taken of the sources of organic variation. A binomial series with negative power or with negative p or q is capable, as is shown in this paper, of perfectly rational interpretation. But in the present state of our knowledge it would be idle to specify any particular interpretation as the correct one†.

(vii) The problem of variation can be looked at in the following manner without the least loss of generality. Modify Gauss by replacing his third axiom, the independence of contributory increments to the variate, by the postulate that the increments are *correlated* with previous increments‡. Start with any binomial and we reach the generalised probability curve for an infinite number of cause-groups:

$$\frac{1}{y}\frac{dy}{dx} = \frac{-x}{\sigma_0^2 f(x/\sigma_0)},$$

where f is an arbitrary function. This theory covers Galton, Edgeworth, Kapteyn and Fechner. Expanding $f(x/\sigma_0)$ in a series of ascending powers of x/σ_0 we have

* Ranke has quite overlooked the work by Galton and myself on the *discontinuity* of the series of individuals even when the population obeys the Gaussian law. See *Biometrika*, Vol. I. pp. 289—299.

† Ranke apparently considers that $(p+q)^n$ with p, q and n positive *is* interpretable. A little philosophical consideration will show that it is merely "familiar," not really intelligible. There is no physiological meaning in p, q, n, and we cannot as yet associate them with any true organic mechanism.

‡ This postulate of course abrogates the first two axioms of the Gaussian theory as well.

my generalised probability curves*. A very few terms of the expansion, however, suffice for describing practical frequency distributions. If we keep only three terms, we see that the same system of curves suffices to describe continuous and discrete variates—an important point. If I lay stress upon this method here, it is because Ranke insists on an infinity of cause-groups and supposes no continuity can arise without them—

...“a truth
Looks freshest in the fashion of the day.”

(viii) The important physical constants of a frequency distribution are those which can be determined with the least probable errors. The probable errors of the moment coefficients increase rapidly with the moments. Hence the important physical constants are those which depend on the low moment coefficients, i.e. on the early terms of the expansion of $f(x/\sigma)$. Now these physical constants are (*a*) the mean, (*b*) the modal difference or distance of mean from mode, (*c*) the skewness, and (*d*) the kurtosis. We may replace either (*b*) or (*c*) by the standard deviation. Experience shows that these four physical constants are certainly independent. The constants of my skew curves directly give them and we are able to determine by their probable errors whether they are significant or not.

(ix) With regard to the other theories discussed I have shown:

(*a*) That the Galton-McAlister curve, ascribed by Ranke to Fechner, is not applicable to a great number of cases, for its kurtosis is a function of its skewness and its skewness cannot exceed ·21.

(*b*) That the double Gaussian curves due to Galton and Fechner are illogical, because they reach a Gaussian result by rendering invalid every one of the Gaussian principles. Further, the skewness is always a function of the kurtosis and the kurtosis cannot exceed ·87, a degree which is exceeded in a great variety of data.

(*c*) That the Edgeworth curves as developed by Kapteyn fail from the logical standpoint, for they appeal to a truncated Gaussian distribution which has never been observed in experience. They are not true graduation formulae, and are obtained in such manner that it is not possible to determine any one of the chief physical constants or evaluate their probable errors. Further in the examples given by Kapteyn they all sensibly reduce to the Galton-McAlister curve. But this curve has in every one of the cases dealt with by Kapteyn a skewness significantly less at a maximum than is required by every one of the statistical series involved.

Finally it seems to me that all discussion of asymmetrical frequency must turn in one form or another on the proper form to be given to $F(x)$ in the equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-x}{\sigma_0^2 F(x)}.$$

* See “Mathematical Contributions to the Theory of Evolution, XIV.” Dulau and Co., London.

If we assume it to be $F(x) = \Sigma(a_r x^r)$ as I have done, we fall back on the normal curve when a_r for $r > 0$ is zero within the limits due to its probable error. Ranke, if he wishes to demonstrate the Gaussian law as general, must show this to be the case. It has been over and over again demonstrated that a_1, a_2, etc. differ significantly from zero for a great variety of series. Another advantage of the form $F(x) = \Sigma(a_r x^r)$ is that it covers as I have shown discrete as well as continuous variation. Considering $\sigma = \sigma_0 \sqrt{F(x)}$ as the standard deviation of the "instantaneous Gaussian curve," we see that the "instantaneous Gaussian curve" varies from one position to a second, like the "instantaneous ellipse" of the astronomers. A reasonable first hypothesis to make is that the local mean square deviation σ^2 is independent of x, we obtain the Gaussian curve. A next assumption is that it is a linear function of x—perhaps it would be better to say that its *mean* local value is a linear function of x, i.e. the mean square of the local variability σ^2 is *correlated* lineally with x. This gives my curve of Type III. The next easiest assumption is to suppose the regression line of σ^2 on x to be parabolic. In this case we obtain the remainder of the curves treated in my II. and XI. memoirs. If we stop at a_q we have what I have termed the skew frequency curves of the qth order*, and we see that this involves a regression curve between the square of the mean local variability and the character of the qth order*. I see, however, at present no *practical* necessity for proceeding beyond skew curves of the 2nd order, although I propose shortly to publish a discussion of skew curves of the 3rd order illustrating some *theoretical* points which arise in their discussion.

To sum up I think Ranke's criticism fails (α) because he has disregarded the universally recognised need of modern statistical science for asymmetrical frequency curves, (β) because he has not appreciated the mathematical transformation by which a number of finite terms are replaced by an integral expression, (γ) because he has not realised that modern theories of heredity lead directly to discontinuous skew distributions, (δ) because continuity does not depend upon infinity of fundamental cause-groups, and lastly (ϵ) because, and this may be due to my fault in the first deduction of my curves, he has quite failed to see either their scope or their real generality.

* "Mathematical Contributions to the Theory of Evolution, XIV. On the Theory of Skew Regression." Dulau and Co.