II.—*On the Probable Errors of Frequency-Constants* (*Contd.*).*

By Professor F. Y. EDGEWORTH, D.C.L.

VI.—Most of the preceding investigations can be subsumed under the general problem which may be enunciated as follows: There are given numerous observations relating to the attributes or "organs"† of particular cases or individual specimens; each observation being of the type $(x_t, y_t, z_t \ldots)$, where $x_t, y_t, z_t \ldots$ are concurrent values of the variables $x, y, z \ldots$, which represent the attributes or organs that are under measurement. It is given also that if the set, or "series"‡ of observations were prolonged indefinitely, under unaltered conditions, the group of attributes thus constituted would (tend to) conform to a frequency-function, or "surface" of any number of dimensions, of which the form is given: say, $w = f(x, y, z \ldots; c_1, c_2 \ldots)$; $c_1, c_2 \ldots$ being constants, in general not given in magnitude. Such being the *data*, the *quæsita* are as follows. It is required to determine the most probable, or best available, values of the *primary* constants, the averages (mode, arithmetic mean, &c.) of the organs. It is required also to determine the constants $c_1, c_2 \ldots$ which constitute the *secondary* frequency-constants. The probable errors incident to all these determinations form a third class of *quæsita*, with which we are particularly concerned.

For the discussion of this problem in all its generality, and its illustration by splendid examples, the reader is referred to the *fourth* of Professor Karl Pearson's "Mathematical Contributions to the Theory of Evolution."|| The following reflections are largely suggested by his "general theorem," on the probable errors of a system of frequency-constants. I have also to make grateful reference to that section of Laplace's Theory of Probabilities,¶ in which he employs, or at least shows how to employ, the genuine inverse method in order to determine the most probable value of an object under measurement.

Let us begin with the simplest case in which there is only one variable, say x, and the secondary frequency-constants are all given. In this case the frequency-function which in general corresponds to a surface of many dimensions reduces to a curve of which the equation may be designated $y = f(x)$. In the case contemplated by the earlier writers on Probabilities, the leading case, as it may be called, in which the observations relate to a real external object,

* See the June number of the *Journal of the Royal Statistical Society*.

† The term employed by Professor Karl Pearson in his parallel enunciation.

‡ The technical term employed by Dr. Venn in his *Logic of Chance*.

|| *Phil. Trans.*, A, vol. 191 (1898).

¶ *Théorie Analytique*, liv. ii, ch. iv, sec. 23. The feature which is germane to my present purpose, the character of inverse probability, is obscured by the exposition, in the same section, of the doctrine of *greatest advantage*, a doctrine to which I have adverted in another connection (*ante*, p. 386).

the most probable value of that object is determined by a familiar method of reasoning which it may be well to give in the words of a recognised authority. Todhunter, paraphrasing that section of Laplace to which I just now referred, reasons as follows:—"Suppose that observations assign values a_1, a_2, a_3, to an unknown element; let $\phi(z)$ be the function of facility of an error z, the function being supposed the same at every observation. Let us now determine the probability that the true value of the element is x, so that the errors are $a_1 - x, a_2 - x, a_3 - x$ . . . at the various observations.

"Let $P = \phi(a_1 - x) . \phi(a_2 - x) . \phi(a_3 - x)$ . . . Then by the ordinary principles of inverse probability, the probability that the true value lies between x and $x + dx$ is $\frac{Pdx}{\int Pdx}$, the integral in the denominator being supposed to extend over all the values of which x is susceptible. Let H be such that, with the proper limits of integration $H \int Pdx = 1$, and let $y = H\, \phi(a_1 - x) . \phi(a_2 - x) . \phi(a_3 - x) \ldots$"

So far Todhunter interpreting Laplace. It is a short step further, one suggested by Laplace,* to put for Todhunter's ϕ, or for our own symbol f, e^{ψ}; and accordingly, substituting for Todhunter's a_1, a_2 . . our own x_1, x_2 . . , we have $y = He^{\psi(x_1 - x) + \psi(x_2 - x) + \cdots}$ This expression (multiplied by Δx) gives the probability that any particular point should be the sought true point. Accordingly that *quæsitum* which we will designate x', must satisfy the condition $\Sigma \frac{d}{dx'}\psi(x_t - x') = o$; where t receives n values corresponding to the n given observations.

This reasoning is equally applicable to the case in which the *quæsitum* does not correspond to an external thing, but is a mere frequency-constant, such as the height of *l'homme moyen.* In fact we have already applied the reasoning in Problem I to the particular case in which ψ reduces to $- x^2$.

Where, as in the case contemplated by Laplace, the function is symmetrical, no question arises as to the position of x' relatively to the frequency curve. But when that limitation is removed—the principal averages, the Arithmetic Mean, the Mode, the Median, being no longer coincident—what does x' stand for? The physicist measuring an objective magnitude will now have to answer the question, *which* average of the observations tends (as the number thereof is increased) to correspond to the true point sought.† Likewise the statistician will have to choose the primary frequency-constant most suited to his purpose. For example, let the given statistics be the ages at death of persons who came under observation in the year 1850, when they were all forty years old. For the

* See the section cited. But note that Laplace's ψ does not, like ours, involve (odd powers of) x, but only x^2, the distribution of the observations being, he supposes, symmetrical.

† I have pointed out the necessity of making this choice in my Paper on "The Method of Least Squares" in the *Philosophical Magazine* for 1883, vol. xvi, p. 373.

purpose of ascertaining whether there is a material difference between the average mortality of a group of, say a thousand, temperate persons and an equal group of intemperates, is it better to employ the Arithmetic Mean or—assuming that the observations conform to a known function, such as the Gompertz-Maheham law—the Mode; better theoretically, and abstracting the difficulty of calculating the Mode?

Whatever answer is given to this question, whichever primary constant is selected as our goal, there arises the further question: May x' stand for that frequency-constant? Or is the process above indicated specially adapted to some particular sort of average, say the Mode? And should we first determine that average by inverse probability; and then proceed by way of the given (secondary) constants to the average which may be our ultimate *quæsitum*, say the Arithmetic Mean? For example, let $f(x) \equiv H(x-l)^p e^{-\gamma(x-l)}$, which represents a curve belonging to Professor Pearson's Type III, referred to an arbitrary origin, O, outside the curve. The distance of that origin from the point where the curve strikes the abscissa is l; p and γ are given constants; H is a constant determined by the condition that the area enclosed by the curve equals unity. The curve is shown in Fig. 2, of which the general shape is copied from one of Professor Pearson's diagrams.* Ought

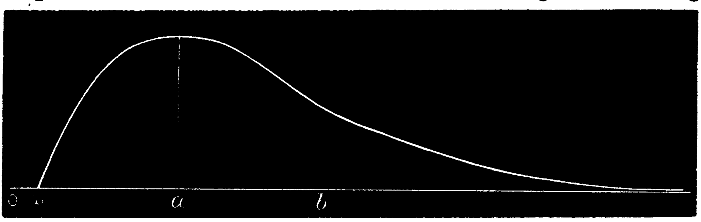

FIG. 2.

we first to find by the inverse process the position of the mode (a) in the figure, and then to proceed to the arithmetic mean (b), supposing that to be the ultimate *quæsitum*, by means of the equation† $Ob = Oa + 1/\gamma$? Or may we equally well begin with b? The answer is: We may begin with *any* mean value, or more generally with any point fixed relatively to the curve. For instance, x', the *quæsitum*, may stand for the (finite) extremity of the curve, ω, which cannot with propriety be termed a mean. We have then to substitute x' for l in the equation of the curve and to equate to zero, $\Sigma \frac{d}{dx'} \log H\,(x_t - x')^p\, e^{-\gamma(x_t - x')}$, where t receives n values, each corresponding to a given observation. Now this aggregate is approximately (the observations being numerous) equal to $n \int_{\omega}^{\infty} y \frac{d}{dx} \log y \, dx$, where y is the ordinate of the theoretic

* Subtype V, in Plate 9, Fig. V, of "Contributions to the Mathematical Theory of Evolution," No. II. *Phil. Trans.*, vol. 186 (1895) A.

† Pearson, *loc. cit.*, p. 373.

curve to which the observations tend to conform. And, as pointed out by Professor Pearson,* $\int y \frac{d}{dx} \log y = \int \frac{dy}{dx} dx$; the integration to extend from one extremity of the locus to another, in the case before us from $+\infty$ to ω. Thus, if x' has received its true value the above-written condition of maximum probability will be satisfied (since y_∞ and y_ω are each zero); and if instead of taking for x' its true value, say l, we take a neighbouring value, say $l \pm \Delta x$, then it will be found that the expression which is to be a maximum becomes less by approximately $K\Delta x^2$, where K is positive, being (n times) the integral of positive elements, viz., $dx\left(\frac{dy}{dx}\right)^2/y$.†

The most probable position of x' having been thus determined we may proceed by equations which Professor Pearson has given to the position of any required average.

Just so in our first problem we might have taken as the *quæsitum* the abscissa of either point of inflexion, or any other point fixed relatively to the theoretic curve of frequency, and then determined by inverse probability the most probable position of the point which is at the distance $1/\sqrt{2}$ times the given modulus from the centre of the curve. But in the case of both problems there is a certain propriety, I think, in taking the *Mode* as the *quæsitum*. It fulfils particularly well a condition which the above-written equation involves, that as we proceed step by step from one extremity of the curve to the other the sum of the (vertical‡ heights of the) steps upwards is just equal to that of the steps downwards.

The condition, of course, presumes that we start and finish at the same level. The condition would not be fulfilled in a case like the following. Given the statures of a regiment which has been formed by taking men at random from a certain province and rejecting all below a certain limit; to find the probable height of *l'homme moyen* in the province. We must not now equate $\Sigma \frac{d}{dx'} \psi(x_t - x')$ to zero, but to zero corrected by the sum of the steps which have been omitted. Say the mean height is about 5 feet $8\frac{1}{2}$, or 5 feet $8\frac{3}{4}$, inches, and that the minimum admissible is 5 feet 6 inches. Then the modulus (which is supposed given) being, say 3·7 inches, there has been omitted (in the left side of the equation) a quantity about equal to

* *Phil. Trans.* A, vol. 191, p. 232.

† This integral ($\times n$) may in general be substituted for the quantity here termed K, if, as usual, there is fulfilled a condition laid down by Professor Pearson ("Mathematical Contributions," XIV, Drapers' Co. Research Memoirs, Biometric Series II, p. 5) as appropriate to curves of frequency, viz., that $\frac{dy}{dx}$ should vanish at each extremity. *Cf.* the Appendix to the present paper.

‡ To be sure the condition is fulfilled by any point on the curve if "up" and "down" are taken in what may be called their algebraical sense—the sense in which they must be taken, even with regard to a mode, when there is more than one mode.

the ordinate at the point of inflexion of the normal curve representing statures.

But we have not much to do in this study with discontinuous loci. A considerable degree of continuity must be postulated in order to secure the further condition of a maximum that, when x' is changed to $x' + \Delta x'$ in the expression $\Sigma \psi'(x_t - x')$, while the first term of expansion in powers of $\Delta x'$ vanishes, the second should be finite and negative. To secure a maximum, indeed, it is not absolutely necessary that the second term should be negative; it might be zero, provided that the *fourth* term* is negative. But for the purpose of our theory about "probable error" the narrower condition is appropriate.

Frequency-curves can often be put in a form which exhibits the postulate as very simple and reasonable; namely, a variant of the form recommended by Demorgan as appropriate to represent errors in general.† Demorgan's general type may be presented in the form $e^{-P}Q$ where P and Q are rational integral functions. This form he proposes to abridge by omitting all the terms in P after the first power of the variable (or after the second in the case of an even function). As a variant suited to the present purpose I propose to omit all terms involving the variable in Q, to reduce Q to a constant, while P retains the form—

$$-[A + Bx + Cx^2 + Dx^3 + Ex^4 + Fx^5 + \ . \ .].$$

Thus for $\frac{1}{2}\psi''$ we have—

$$-[C + 3\,Dx + 12\,Ex^2 + 60\,Fx^3 + \ . \ .],$$

and for $\frac{1}{2}\Sigma\psi''$ approximately—

$$-n[C + 3\,Dx^{(1)} + 12\,Ex^{(2)} + 60\,Fx^{(3)} \ . \ .];$$

where $x^{(p)}$ denotes the mean pth power of the variable measured from an assumed origin, the position of the average (or other primary constant) which is taken as the *quæsitum*, and n the number of observations is large. Now it may be postulated with respect to concrete frequency-curves ‡ that the mean powers of deviation from the Arithmetic Mean (of an indefinitely large group), and thus the mean powers measured from a point which is at a finite distance from the Arithmetic Mean, are finite.§ Therefore $\frac{1}{2}\Sigma\psi''$ is finite supposing that none of the coefficients A, B, C . . become infinite, and that they are not, or at least may be treated as not, infinitely numerous.

Taking for granted that $\Sigma\psi''$ is finite and negative, and making certain other assumptions which are commonly and probably fulfilled by frequency-curves, we may transform the equation of the curve

* Or the fourth term also vanishing, that the sixth should be negative; and so on.

† Article on "Theory of Probabilities" in the *Encyclopædia Metropolitanal*, vol. ii, sec. 88.

‡ I take this as the fundamental postulate for the genesis of the Law of Error (Camb. *Phil. Trans.*, 1905).

§ Demorgan makes the postulate with respect to "errors" measured from the true point, the real value of the magnitude under measurement.

(to which the observations tend to conform) to x' as origin—x' being the value found for the sought primary constant—by putting $x = x' + \text{x}$; and expanding $\Sigma\psi(\text{x}_t - \text{x})$—($\equiv \Sigma\psi(x_t - x)$)—in ascending powers of x we may neglect terms involving powers above the second, according to the subsidiary law of error which has been set forth.* Thus we obtain for the distribution of the frequency with which the *quæsitum* occurs at any assigned distance x from x', the normal curve whose *weight*, or inverse modulus squared is $-\frac{1}{2}\Sigma\psi''(\text{x})$, which comes to the same as $-\frac{1}{2}\Sigma\psi''(x)$.

The probable error thus determined for one point fixed relatively to the curve is equally applicable to any other fixed point of which the abscissa differs from that which has been found by a known function of the (as yet supposed) given constants, as in the example above given.

The example suggests the question: What are we to do when the equation for x' obtained by the inverse method is impracticable, or at least troublesome, *e.g.*, for the Mode in the example chosen $-\Sigma p \frac{1}{x_t - x' + a} + n\gamma = 0$; where $a = p/\gamma$, n is the number of the observations, ($x_t - x' + a$ is always positive). The answer is that we must employ what I have called in the enunciation the best available method; presumably in the case proposed to equate the Arithmetic Mean of the given observations to the sought true "centroid" of the curve. It will be noticed that this best available unit is not the "most advantageous" in the sense of that term which Laplace and Gauss, as above explained, opposed to the "most probable." The most probable value of the *quæsitum* ascertained by the inverse method proper is also the most advantageous in that sense; the law of frequency for the *quæsitum* being a normal law of error of the spire-shaped kind above described.† The best available in practice is not theoretically so good. For it cannot be questioned, I think, that the genuine inverse method, taking account of the given distribution of the observations in connection with what is given as to their origin, forms a better rule theoretically and abstracted from practical difficulties, and would give more accurate results in the long run of its application, than a summary method which does not utilise the *à priori* data.

* *Ante*, p. 389; purporting to be a re-statement of reasoning employed by Gauss and Laplace. See, with reference to the present problem, Laplace, *Théorie Analytique des Probabilités*, liv. ii, sec. 23, p. 368, ed. 1844-47 (p. 336, ed. 1814):—"Suppose the number of observations s to be very great, and let us determine a by the equation $\text{N} = 0$ [corresponding to the equation on p. 501, above, four lines from foot], which gives the condition for y being a maximum; then we have $y = \text{H}e^{-\text{M} - \text{P}z^2 - \text{Q}z^3 - \&c.}$ [putting e for Laplace's symbol]. M, P, Q, &c., are of the order s; thus, if z is very small, of the order $\frac{1}{\sqrt{s}}$, $\text{Q}z^3$ becomes of the order $\frac{1}{\sqrt{s}}$, and the exponential expression $e^{-\text{Q}z^3 - \&c.}$ is reducible to unity . . ." Compare Pearson, *Phil. Trans.* A, vol. 191, p. 246, paragraph 1, and note.

† *Ante*, p. 385 and p. 391.

There is more room for doubt about the answer to the question whether the probable error of a frequency-constant determined by the proper inverse method is necessarily smaller than the probable error incident to some other methods. This is a question which we postponed at an earlier stage,* when it was noticed that in a particular case the probable error incident to inversion proper, was in fact smaller than that of other determinations which were compared therewith.

The compared method must, of course, be "other." *Cadit quæstio* when the compared method of combining the observations is the very formula prescribed by inversion proper. This happens more frequently than may be supposed, with respect to the Arithmetic Mean. It happens in the case of the normal curve as we have observed in the solution of Problem I. In the symbols introduced in this section we have now $\psi(x) = -x^2/c^2$ (c being the modulus of the normal curve); $\frac{1}{2}\psi''(x) = -1/c^2$; $-\Sigma\frac{1}{2}\psi'' = n/c^2$, which is identical with the inverse square of that modulus which measures the accuracy of the determination obtained by taking the Arithmetic Mean of the observations.

There is a similar identity between the prescription of inversion proper (based on all the data) and the summary method of taking the Arithmetic Mean in other simpler cases of observations formed by the fortuitous concurrence of independent causes.† Here is an urn containing an immense number of black and white balls mixed up in some unknown ratio. *A priori* the values π_1, π_2 . . for the proportion of the number of white balls to the total number are equally probable. There are given n "observations" as to the constitution of the urn, each consisting of the proportion of white balls in a batch numbering m drawn at random from the urn (with replacement after extraction). Say the observations are $p_1, p_2 \ldots p_n$. And let $p = \Sigma p_t/n$. Now let us compare in respect of accuracy the summary method which takes account only of a simple function, the Arithmetic Mean of the observations p_1, p_2, &c., and the complete method which takes account of the individual observations in connection with what is known as to their genesis. To find the relative probability of the causes, we have by the summary method the proportions

$$\frac{n!}{np!\,n(1-p)!}\,\pi_1^{np}(1-\pi)^{n(1-p)} : \frac{n!}{np!\,n(1-p)!}\,\pi_2^{np}(1-\pi_2)^{n(1-p)} : \cdot \cdot \cdot ;$$

and by the complete method—

$$\left(\frac{n_1!}{n_1p_1!\,n_1(1-p_1)!}\right)\left(\frac{n_2!}{n_2p_2!\,n_2(1-p_2)!}\right) \cdot \cdot \cdot \pi_1^{np}(1-\pi_1)^{n(1-p)}$$

$$: (\; . \; . \; . \; . \;)(\; . \; . \; . \; . \; . \;) \; . \; . \; \pi_2^{np}(1-\pi_2)^{n(1-p)}$$

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

---

* *Ante*, p. 392.

† As observations filling the normal law are presumably formed.

The same proportions are given by the two methods. The complete method has no advantage over the summary method.*

Thus the property holds true not only of the limiting case where the number of independent elements which go to an observation is indefinitely great, the case in which the normal law of error is set up, but also of cases in which the number of elements, the m of the preceding example, may be small. Seeing that an intermediate position between the case of indefinitely numerous and very few independent elements to each observation is occupied by the Generalised Law of Error which has been set forth in former numbers of this *Journal*, it may be expected that for this law of frequency also the summary process of taking the Arithmetic Mean is identical with the process prescribed by the method of inversion proper; for Binomial elements at least, if not generally.†

When we leave the precincts of the Law of Error, the hypothesis of observations formed by independent elements, we can no longer expect the Arithmetic Mean (or other summary method such as the Median) to concur with the method of inversion proper.

Certainly it is very natural to associate increased accuracy in the ordinary sense of the term with increased *precision* in a technical sense. But the following objection occurs. Grant that the increased knowledge obtained by taking account of all the data affords a more accurate determination of the probabilities that the observed event, the given set of observations, should have resulted from each of the possible causes, the different values of the *quæsitum*. But what if, in this corrected distribution of probabilities, the outlying causes as distinguished from the central become relatively more probable; and accordingly the "spread" of the curve of frequency for the values of the *quæsitum* is increased!

The objection is specious only while there is ignored what is known about the applicability of the normal curve. The following answer may suffice. Consider any particular set of observations, $x_1, x_2, \ldots x_n$, forming one of a series of sets, such as are encountered in practice. The probability that any particular point should be the true one is given by inversion proper as above; the most probable value being a root of the equation, $\Sigma \frac{d}{dx} \psi(x_1 - x) = 0$, say $\phi(x_1, x_2 \ldots x_n)$; and the probabilities of other points being disposed about that maximum in conformity with a normal curve. Now consider some other formula, some other function of the observations known to coincide with the true value of the *quæsitum*, in the long run formed by a series of sets; *e.g.*, $(x_1 + x_2 + \ldots + x_n)/n$, the Arithmetic Mean. The point designated by this formula being generally different (for any particular set) from

* The coincidence between the two methods of determining the sought primary constant may be compared to the coincidence which has been noticed (*ante*, p. 395), with respect to a secondary constant (the coefficient of correlation), between the formula which is prescribed by inversion proper and that which on a first view of the subject suggested itself to the present writer as natural and convenient.

† See the Appendix to this Paper.

$\phi(x_1\ x_2 \ldots x_n)$, the centre of the normal curve assigning the frequency with which each point is the true value of the *quæsitum*; it follows that, in the long run formed by the different originations of the particular set, the mean square of deviation from the true point for $(x_1+x_2+x+x_n)/n$ is greater than the mean square of deviation for $\phi(x_1, x_2, \ldots x_n)$. The like is true of any other particular set. It is therefore true for the whole series that the Mean Square of deviation from the true point, and accordingly the probable error, is less for the formula given by inversion proper than it is for the Arithmetic Mean, and, by parity of reasoning, for any other rival method, say, $\chi(x_1, x_2, \ldots x_n)$. If then* we take numerous sets of observations, each set numbering n, and form for each set the value ϕ and also χ, while both series—that of the ϕ's, and that of the χ's—will fluctuate according to a normal law of frequency, the probable error for the ϕ's will be less than what it is for the χ's.

The proof might have been put more simply; indeed the proposition may appear to some self-evident. But I think it well to examine the foundations of a theorem, on which an enormous weight of inference is to be rested. For the theorem may be used to support not only conclusions of interest in Probabilities, but also mathematical propositions which are not so easily proved otherwise. For as many as are the formulæ which may be substituted for inversion proper, so many complicated mathematical propositions are there, affirming that the mean square of deviation pertaining to the former is greater than the mean square of deviation pertaining to the latter. Take for instance a particularly simple case where the curve to which the observations tend to conform is symmetrical about its (single) Mode, and the rival method is the Arithmetic Mean. Here the mean square of deviation for the rival curve is $\int_{-\infty}^{\infty} yx^2dx$; while for inversion proper, upon the usual assumption as to the extremities of the curve,† it is $1/\int_{-\infty}^{\infty} \frac{\left(\frac{dy}{dx}\right)^2}{y} dx$. The former expression exceeds (when it does not equal)

* If any hesitation is felt as to the connection between this and the preceding statement, it may be removed by the following illustration. Imagine a long line of soldiers shooting bullets at a wall-shaped target parallel to the long line; each man aiming at a point on the target straight in front of him. The deviation, measured horizontally, of the bullets fired by each man, from the point he aimed at, obeys the same law of frequency, namely a normal error-curve with one and the same probable error. Considering any particular shot-mark on the target, let us determine by inverse probability the most probable position of the man that fired that shot. The probable error affecting this determination is the same as the probable error shown by the dispersion of the bullets fired by any particular man (supposing that the distance between two adjacent men is small, compared with the probable error in question). In this parable a single shot stands for a combination of n observations, such as $\phi(x_1, x_2 \ldots x_n)$, or $\chi(x_1, x_2, \ldots x_n)$; each of which is known, by the Law of Error, to fluctuate according to a normal law of frequency.

† Above, p. 502, note.

the latter; a proposition which ought to be true, not only for continuous values of y, but also for combinations of different curves, if joined together without discontinuity of tangent. The reader may amuse himself by trying to find a combination which will defeat the proposition. A promising construction is offered by a curve which makes no contribution to ψ'', viz., a curve of the form $y = be^{-cx}$. Let this curve be taken as the locus on the positive side of a central point, from infinity up to the distance τ from the centre; and on the negative side the same form, with the sign of x changed, from $-\infty$ up to the point whose abscissa is $-\tau$. Let these branches be connected by a concave curve, say an ordinary parabola or a normal error-curve, having a maximum at the central point and a common tangent with each of the (infinite) branches at the points $+\tau$ and $-\tau$ respectively. As we are at liberty to take τ as small as we please, and thus secure that for almost the whole of the locus ψ'' should be null, while the compared coefficient, the inverse mean square of deviation in the case of the Arithmetic Mean, is substantial—it might seem for a moment that the trick was done. But not so!

Here is another instance of a proposition thus proved *a priori*, or by the logic of probabilities. Under the conditions above defined, if P is the maximum ordinate, $2\,P^2 < \frac{1}{2}\int_{-\infty}^{\infty}\frac{\left(\frac{dy}{dx}\right)^2}{y}\,dx$. For $2P^2$ is the *weight* or reciprocal of twice the mean-square-of-error incident to the use of the Median; while the expression on the right of the inequation is the weight of the determination belonging to inversion proper.

A purely mathematical proof of these propositions which Professor Love has kindly supplied will be printed in the Appendix to this Paper.

The reader may be advised to verify these propositions by simple examples; for instance, the curve $y = He^{-x^4}(H = 2/\Gamma(\frac{1}{4}))$, a simple specimen of the class recommended by Demorgan as proper to represent error-curves. Here is another simple example, belonging to one of the types formulated by Professor Karl Pearson:* $y = H(1-x^2)^2\,(H = \frac{15}{16})$. A form common to the schemes of both mathematicians has been already used to illustrate the inverse method. The example shows that no difficulty arises, with reference to the present issue, from the circumstance that the inverse method may be primarily directed to the determination of a point different from that which is determined by some compared method, *e.g.*, the use of the Arithmetic Mean.

Like propositions are true of secondary constants. We are entitled to presume *a priori* that the combination of the observations which gives the most probable value of a constant—the modulus for example †—is subject to a less probable error than other formulæ. As between any two of such other formulæ we are entitled to prefer

* *Phil. Trans.*, vol. 186 (1895) A.

† *Ante*, p. 392.

the one which is subject to less probable error, as approaching more nearly to the character of that determination which if we took account of all the data would be the best.

The determination of secondary frequency-constants appears to be in one respect simpler than that of primary-constants. The distinction between the different kinds of average—the Mode, the Arithmetic Mean, and so forth—is not so serious as in the case of primary-constants. The difference between the several averages is in general for the secondaries of the same order as the probable error incident to the determination, say, the order $1/\sqrt{n}$ (n being the number of observation); whereas the difference between the respective averages may be of a higher order, which may be called unity, in the case of primary-constants.

With reference to the case of many dimensions—each corresponding to a different primary-constant—it may be noticed that, as we have seen in the determination of the modulus,* there is a distinction between the "absolute" and "relative" values of constants, as well as between the "absolute" and "relative" values of the probable errors to which they are liable. But the latter distinction is I think by far the more important.

Once more committing the investigator of this subject to Professor Pearson's guidance, I go on to a problem which is distinguished from the preceding by the comparative paucity of its data.

VII.—So far we have supposed the law of frequency to which the observations tend to conform to be given. This datum is now withdrawn. With no knowledge of the shape to which the observations (if indefinitely multiplied would tend to) conform, we still seek the Arithmetic Mean or Mode, or other primary frequency-constant pertaining to that form. Let us begin with the simple case of a single dimension and observations believed to be of equal worth.

Considering what definite results have been obtained from almost indefinite data in the theory of Error, it cannot be regarded as a hopeless enterprise to attack this problem on the lines of the method proper to the preceding problem. And in fact we can advance a certain distance on those lines. We reach the position that the *quæsitum* is the centre of a certain tapering normal error curve. But as Laplace says, "our complete ignorance of the law [e^{ψ}] of error for each observation prevents us from forming the equation" [$\Sigma\psi'(x_t - x) = o$].†

Repulsed in this frontal attack we have recourse to a second best method which is described by Laplace as relating to "observations not yet made." The contemplation of observations in this stage—like seeing General Wade's roads "before they were made"—is not free from difficulty; and the commentators are not agreed as to the explanation of the Method of Least Squares. I trust that the interpretation which I have elsewhere offered,‡ though

* *Ante*, p. 392, paragraph 1.

† See the passage referred to above, p. 499, note.

‡ Article on Error (Law of) in the *Encyclopædia Britannica* Supplement to ed. 9, sec. 26; abridged from the writer's pamphlet *Metretiké* (1887).

not commanding universal assent, may yet be in the position of minimum dissent. According to this view deficiency of data in this section generally reduces us to the same position as inability to deal with data sometimes reduced us in the preceding section. We now are ignorant of what we then ignored. We must be content with a method which, as contrasted with genuine inversion was above described as "summary."

This method is characterised by the construction of an auxiliary normal system whereby to determine the *quæsitum.* The simplest conception of such a system is to regard the Arithmetic Mean of a given set of n observations as a single observation conforming to—forming a sample of—a normal group of extremely small probable error. It comes to the same for the purpose to which the Method of Least Squares is ordinarily applied, to break up a given set of n observations into several, say ν, parcels each numbering m observations where m is a number of such magnitude that the Arithmetic Mean of m observations (supposed of like character) fluctuates in approximately normal fashion. The determination obtained by taking the Arithmetic Mean of these ν compound observations is the same, and is subject to the same probable error, as the single more highly compound observation first described.

The singular and the plural arrangement are not quite identical when we go on to what I have called the Method of Least Squares *plus* cubes.* The auxiliary system is now constructed, not with the normal law of error, but with the Generalised Law of Error of the second order, account being taken of mean cubes of deviation. We are to suppose each of the ν compounds above described to obey a law of that kind. A correction of the Arithmetic Mean is thereby obtained which might be of some avail where the *quæsitum* is the Mode. But there is something arbitrary in the construction as the correction will vary with the size of m, the number of original observations which go to a compound one.

VIII.—In the preceding section we have incidentally performed an operation which forms the special subject of the present section: the determination of *secondary* frequency-constants for an auxiliary normal (or more generally Generalised) error-function. The secondary constants are of a more substantive character in the following problem. The data are observations in two or more dimensions, not known—or even known not—to conform to the normal law; and the *quæsita* are coefficients of correlation which may enable us to test the degree of causal connexion between the variations in the different organs or attributes.

Let the observations be of the type (x_1,y_1), (x_2,y_2), . . (x_n,y_n). And suppose that they can be broken up into ν compound observations (ξ_1,η_1), $\xi_2,\eta_2)$ . . . (ξ_ν,η_ν) approximately conforming to a normal surface; each compound aggregating m of the original data as follows:—

* Article on Error in *Encyclopædia Britannica*, sec. 27.

$$\begin{cases} \xi_1 = (x_1 + x_2 + \ . \ . \ . \ + x_m)/\mu \\ \eta_1 = (y_1 + y_2 + \ . \ . \ . \ + y_m)/\mu \end{cases}$$
$$\begin{cases} \xi_2 = (x_{m+1} + x_{m+2} + \ . \ . \ . \ + x_{2m})/\mu \\ \eta_2 = (y_{m+2} + y_{m+2} + \ . \ . \ . \ + y_{2m})/\mu \end{cases}$$
$$\begin{array}{ccc} . & . & . \\ . & . & . \end{array}$$
$$\begin{cases} \xi_\gamma = (x_{(v-1)m+1} + \ . \ . \ . \ + x_{\nu m})/\mu \\ \eta_\nu = (y_{(\nu-1)m+1} + \ . \ . \ . \ + y_{\nu m})/\mu ; \end{cases}$$

where μ is a coefficient to be assigned. For this normal system of observations form the coefficient of correlation according to the usual formula $r = \Sigma\frac{\xi_t\eta_t}{\sigma_1\sigma_2}$, where σ_1, σ_2 are respectively the standard-deviations for the group of the ξ's, and the group of the η's. Observing the formation of any of the products which constitute r, *e.g.*, $\xi_2\ \eta_2$ we shall find that it consists of m items which have an average value different from zero, when there is a real connection between the organs, or attributes, *e.g.*,

$$x_{m+1}y_{m+1} + x_{m+2}y_{m+2} + \ . \ . \ . \ + x_{2m}y_{2m} ;$$

and $m(m-1)$ products which do each tend to hover about zero and vanish, upon the usual presumption that any one of the original observations, *e.g.* (x_a, y_a) is *independent* (in the sense of the term proper to Probabilities) of any other observation, *e.g.* (x_b, y_b). Thus the coefficient of correlation for the auxiliary system is found from the same products as the coefficient would be found from the original data on the supposition that those data were normal. It only remains to ascertain the σ_1 and σ_2 pertaining to our compound observation, and to assign the coefficient μ. The square root of the mean square of deviation for the sum of m x-observations divided by μ will be $\frac{\sqrt{m}}{\mu}$ × the corresponding coefficient for the original uncompounded observations. The like is true of the compound y-observations. As for the coefficient μ I suggest as an appropriate value—not unity as usual in the proof of the law of error, which is here required to establish the normality of our compound system of observations, not m as would be natural if our only object was to determine a primary frequency-constant—but a value intermediate between those two, viz., $\sqrt{m}$. We shall thus have converted the original set of observations into a representative normal system of which the *fluctuation** is the same as that of the original not normal system.

But indeed any value of the coefficient is sufficient for one of the principal aims of this section: to show that Mr. Yule's method of treating skew material as if it were normal for the purpose of obtaining a secondary frequency-constant admits of the same justification

* I have proposed this term to denote (twice) the mean square of deviation from the Arithmetic Mean. It has the convenience of being applicable not only to observations obeying the normal law, but also, as here, to the makings of such groups, abnormal observations which by aggregation form normal compounds.

as the similar procedure in the received Method of Least Squares with respect to primary constants.

To pursue the analogy between the present and the preceding problem, let us construct an auxiliary generalised Error Surface of the second order. By parity of reasoning we are justified in proceeding *as if* the material were adapted to the construction, as if the mean third powers of deviation (referred to in the corresponding powers of the respective moduli) were small. How to proceed in that case I have shown in my papers on the Generalised law.* I take the opportunity of here introducing a simplification. Having found in the passage referred to that the corrected locus of correlation was a conic section, I asked: is there any reason for presuming that this curve of the second degree is in general one kind of conic rather than another? I now answer, Yes. The curve, it will be remembered, was determined as the locus of that point at which the frequency, z, or what comes to the same, log z, is a maximum for each assigned value of x. Now the (Napierian) logarithm of z is of the form $-\frac{x^2 - 2rxy + y^2}{1 - r^2} - (ax + by + cx^3 + dx^2y + cxy^2 + fy^3)$ where a, b, . . are linear functions of the mean third powers of deviation presented by the given set of observations when each of the moduli is taken as unity. Accordingly the required locus is found by equating to zero the first differential (with respect to y) of the above expression, that is $\frac{+ 2rx - 2y}{1 - r^2} - (b + dx^2 + 2exy + 3fy^2)$. Since y is approximately $= rx$, the coefficients b, d, e, f being by hypothesis small, it is legitimate to substitute rx for y in the terms $2exy$ and $3fy^2$. Thus the equation is reducible to the form $y - \tau = rx + \gamma x^2$ where τ and γ are small; the equation of a *parabola* which approaches the normal line of correlation $y = rx$. It will be noticed that this parabola is not identical with that which Professor Karl Pearson has employed for the same purpose.† This parabola is constructed entirely with mean third (and second) powers, whereas he employs a fourth power.

These examples of probable error might have been much extended if I had included a case to which the classical writers on Probabilities, intent on physical observations, have paid much attention; the case in which the given observations are known to differ from each other in worth (otherwise than merely as being sums, or Arithmetic Means, of different numbers of equally good observations). But, if indeed this case is important in physics, it is not so I think for the purposes of statistics in general. It is therefore omitted here, along with the case of numerous dimensions and other complications.

(*To be continued.*)

---

* *Journal of the Royal Statistical Society*, 1906, p. 515.

† "Mathematical Contributions," No. XIV.