MISCELLANEA.

CONTENTS:

I.—*On the Probable Errors of Frequency-Constants.*

By PROFESSOR F. Y. EDGEWORTH.

Conspectus of Contents.

The "probable error" is a well-chosen index of the belief—or, rather, the credibility—that a value which has been obtained for a frequency-constant characterising a group of statistics has a certain degree of accuracy, will not differ by more than a certain extent from the result of continued observations *in pari materiâ*. The apparatus for testing this credibility—the received Law of Error, with a certain other law of great numbers—is exemplified by the following problems:—(1) Given a set of observations ranging under a normal error-curve, and given the coefficient of deviation for that curve, to determine the average to which the observations if indefinitely continued would tend; (2) Given a normal set of observations as before, and given the average to which they tend, to determine the coefficient of deviation; (3) Given a normal set of observations, but given neither the average nor the coefficient of deviation, to determine both those frequency constants; (4) Given two sets of observations ranging under a normal surface, and given the average and the coefficient of deviation pertaining to each set, to determine the coefficient of correlation; (5) Given, as before, two normal sets of observations, but not any of the frequency-constants, to determine them all; (6) Given one or more sets of observations ranging under any given (not in general normal) laws of frequency, to determine all the frequency-constants; (7) Given one or more sets of observations, but not given their laws of frequency, to determine the averages to which they tend (the Method of Least Squares); (8) To determine a coefficient of correlation between two given sets of observations for which the laws of frequency are not given (Yule's method). The mathematical treatment of credibility in Statistics is comparable to the mathematical treatment of utility in Economics.

THE "probable error" is not in favour with some high authorities. Mr. Galton denounces the term as a "cumbrous, slipshod, and misleading phrase."* He refers to Dr. Venn, who also regards the "probable error" as a "highly misleading term."†

* *Natural Inheritance*, p. 58.

† *Logic of Chance*, ed. 3, pp. 446—47.

"Such an error," he observes, "is not in any strict sense 'probable.' It is, indeed, highly improbable that in any particular instance we should happen to get this error. . . . Nor can it be said to be probable that we shall be *within* this limit of the truth, for by definition we are just as likely to exceed as to fall short." These and other eminent writers on Statistics and Probabilities who have protested against the use of the "probable error" are doubtless right with respect to the subject which they have in view, the normal law of error considered as a statistical fact. They are also well-advised in not altogether discarding the term against which they protest. As Dr. Venn says in the context of the passage above quoted, "it is best to stand by accepted nomenclature." The case, I think, is one in which there is a danger of incurring what Mill calls the "evil consequences of casting off any portion of the customary connotation of words."* In the weighty section of his Logic devoted to that topic, he points out that in the case of words which "in their original acceptation connoted a complication of outward facts and inward feelings," the latter portion of the meaning is apt to be obscured "by the incautious proceedings of mere logicians." The "inward feeling," the subjective phenomenon of belief or credibility, is not so well suggested by the term "quartile"—much less by the term "standard deviation," or some multiple thereof—as by a phrase of which the word "probable" is a part. It is not merely that the name recalls that species of psychical measurement which characterises the Calculus of Probabilities, but also the conception defined is peculiarly appropriate to that use of the Calculus. The "probable error" corresponds to the one definite notch in the scale of credibility, the point of complete uncertainty whether an event such as the occurrence of an error in excess of the assigned limit will or will not occur. In the scale of credibility this point has much the same significance as the point of "indifference"† in the scale of economic utility. The "standard deviation," of which the "probable error" is sometimes described as a mere appendage, has not this sort of advantage. The probability of an error exceeding (in one direction or the other) the standard deviation is about 0·317, a degree of no particular significance in the scale of credibility. This justification of the *adjective* in the phrase under consideration carries with it an apology for the substantive. The "error" which is defined as probable, or rather not improbable, is not the particular error which is just equal to the assigned deviation from the average, say q, nor yet an error in the immediate neighbourhood of q, say between q and $q + \Delta x$, where Δx is very small, the *minimum sensibile* on the abscissa along which q is measured; but an error of the class which is defined by excess above (or defect below) q. When Laplace says it is a million to one that the mass of Jupiter as deduced by him from certain observations "is not in error by

* *Logic*, Book iv, ch. iv, sec. 6 (Contents).

† The fundamental character of this conception in economic theory is particularly well shown by Professor Pareto.

1 per cent. of its value,"* he does not mean the error 0·02 to the exclusion of the next degree of error, say 0·021, but the whole class of errors which exceed 0·02. Dr. Venn, indeed, has anticipated this use of the term "error," and has directed against it an objection, above quoted, which is perhaps formally valid. That a future error shall be within the defined limit is not indeed probable, only not improbable. But the former statement differs by a negligible quantity from the latter, where, as usual, the extent of deviation varies, if not continuously, at least by very small degrees. The slight inaccuracy is fully excused by usage and the need of brevity.

Associated with the "probable error"—more useful in practice, if less appropriate for definition—is the conception of *improbable* error, that extent of deviation which is hardly credible. Such a measure of incredibility is afforded wherever the normal law of error prevails. Its perhaps most important use is to afford assurance that measurements effected by several observations, or more generally statistical determinations of frequency-constants, are trustworthy within assigned limits, may be relied on not to exceed those limits in future experience (*in pari materiâ*). The earliest and most familiar instance, the leading case, is the measurement of an objective quantity, such as the angular distance between two stars. Physicists require to have the sort of assurance which Laplace expresses in the passage above quoted that the measures which they have obtained will not differ from the thing measured by more than a certain extent. The methods which they successfully employ for this purpose are extended to the analogous case of types, such as the mean dimensions of men or crabs, which do not correspond to any one real objective thing. Though the thing measured has in this case, unlike the first case, no separate existence apart from the measurements, yet if these observations have the sort of stability, the sort of unity in the midst of plurality, which characterises fallible observations relating to one and the same real thing, we may still regard the mean value as a substantive thing about whose dimensions assurance is desiderated and obtained. This remark may be extended from the simpler kinds of averages to what may be called the secondary frequency-constants,† such as the coefficient of dispersion, or the coefficient of correlation, for the normal law of frequency, and other constants for other definite species of groupings. But a doubt may arise how far probability in the proper sense of the term as distinguished from objective statistical frequency, probability as understood by the older writers, is applicable to these newer results. In fact, with regard to the secondary frequency-constants at least, it is often not obvious where the normal curve occurs in virtue of which we are entitled to predicate probability, or improbability, of certain deviations. Even with regard to primary frequency-constants, and even with regard to the measurements of real objects, when the errors of

* *Théorie Analytique des Probabilités*, Supplement I.

† Compare Prof. Karl Pearson's distinction between "organs" and "constants" in a passage to be quoted below.

observation are not known—or are known, not—to obey the normal law, doubts have been expressed by high authorities as to the use of the Method of Least Squares and the normal law which is therewith implicit. It is, therefore, not otiose to enquire how far the classical conceptions of probability are applicable to the calculations of frequency-constants.

The somewhat metaphysical character of this enquiry need not alarm the mathematical reader. Our metaphysics are not of the kind which, according to Voltaire, are invented by philosophers to avoid the trouble of mathematical reasoning. "Plusieurs esprits ont mieux aimè rêver doucement que se fatiguer."* The philosophical principles here required are few and simple. The most recondite postulate is that the credibility of an event, such as the occurrence of an assigned deviation, is measured by the frequency with which in the long run such an event occurs.† The correspondence between the external measure and the "inner feeling" (in Mill's phrase) measured is indeed rough and loose, like the correspondence between the degrees of a thermometer and the sensation of heat: accepted not "semper," not "ab omnibus" not by every fevered patient, yet with sufficient generality to be of common use, for instance in the warming of a public library. There is not much difficulty about the psychophysical principle; the difficulty is mainly a mechanical one, how to regulate the thermometer, how to bring the normal law to bear upon the determination of frequency-constants.

In this investigation let us advance step by step from the simpler to the more arduous parts of the subject.

I.—The following is perhaps the simplest problem which the subject presents. Given a set of observations $x_1, x_2, \ldots x_n$, and given that they have been generated by divergence from an unknown point according to a given law of dispersion, a normal error-curve of given modulus, to find the most probable position of the unknown point. Let c be the known modulus, and x the sought abscissa of the unknown point. By familiar‡ reasoning the required value of x is that which makes the following expression a maximum:—

$$\left(\frac{1}{\sqrt{\pi}c}\right)^n e^{-\frac{(x-x_1)^2}{c^2}-\frac{(x-x_2)^2}{c^2}-\ldots-\frac{(x-x_n)^2}{c^2}}.$$

* Dictionnaire Philosophique *sub voce* "Metaphysique."

† In accordance with the explanations offered in the writer's paper on *The Philosophy of Chance* (Mind, 1884); which have the weighty confirmation of Prof. Karl Pearson's approval (*Grammar of Science*, p. 146).

‡ Among hosts of references justifying the description of this method the following may suffice:—Gauss, *Theoria Motus* ii, 3; interpreted by Czuber, *Beobachtungsfehler*, Teil ii, sec. 3. Laplace, *Theorie Analytique des Probabilités*, Liv. ii, ch. iv, Art. 23, *sub finem*, showing that the genuine inverse method is applicable if the distribution of errors is normal; reproduced by Todhunter, *History of Probabilities*, Art. 1,094. A simple statement of the method is given by Mansfield Merriman in his *Method of Least Squares*, ch. iii, sec. 41, *et seq*.

The solution is, of course, the Arithmetic Mean of the observations. The frequency with which different values of x correspond to the same set of observations $x_1, x_2, \ldots x_n$ (in the long run)* is assigned by a normal error-curve of which the centre is at the Arithmetic Mean of the observations, and the modulus is that which pertains to the observations, viz., c, divided by the square root of n; where n is the number of the observations. Thus if the continuous black curve in

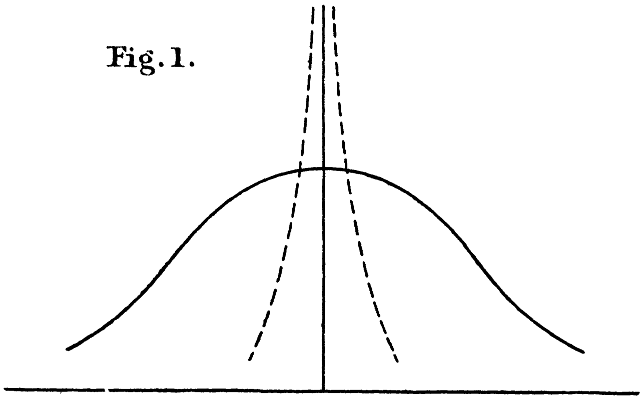

FIG. 1.

Fig. 1 represents the distribution of the observations, the distribution of the point from which they have diverged is represented by a normal curve like the dotted one, which becomes higher and slenderer as the number of observations is increased. (The apex of the attenuated shape is at too high an elevation to admit of its being shown in the figure.) This theorem, which, in its first intention, was applied to the case in which the point about which the observations cluster represents the size of a real thing, is equally applicable to the case in which the *quæsitum* is a type, a primary frequency-constant of the less objective kind above distinguished. The theorem is easily extended to the case in which the errors incident to different observations have not all the same modulus. It may be extended also to the case in which there are several dimensions, several unknown quantities x, y, z, . . . ; supposing the coefficients of correlation as well as the standard deviations given.

Simple and straightforward and generally accepted as this solution is, it is not free from difficulties, which may be stated under two heads.

1. The *quæsitum* enunciated is not that which the masters of the science, Laplace and Gauss, prescribe. They recommend as the proper combination of the given set of observations not that function thereof which is most frequently right, but that which,

* A reference to the illustration which I have given of "a long run" in the article on "Error (Law of)" in the Supplement to the Encyclopædia Britannica, ed. 9, may dispense with further explanations.

account being had of the extent and frequency of the errors incurred, in the long run, by its use—the distribution of deviations represented by our dotted curve—shall minimize the detriment incident to error. As the measure of that detriment, Laplace proposes the mean value in absolute quantity (taken positively) of the error incurred in the long run by the adopted formula. Gauss proposes the mean second power. But since the detriment incident to wrong measurement surely does not obey a law so simple, it seems proper to employ here a less determinate formula of the kind which mathematical economists employ to represent the disutility of labour or of loss. Accordingly, I have proposed (in the "Philosophical Magazine" for 1883)* employing the form $F(e)$ to denote the detriment of an error e in excess of the true quantity, and the form $f(e)$ to denote the detriment of an error of (absolute) magnitude e in defect; where nothing is known about F and f except that they continually increase in value with the increase of the variable. Combining this new datum with the other data we shall find the same solution as before. The most probable value for x is also (in the phrase of the classical writers) the "most advantageous." The true and the expedient coincide.

A theorem of nearly equal generality is given by Professor Czuber in his treatise on the Theory of Errors of Observation published in 1891.† The difficulty of saying anything new on such a subject is illustrated by the incident that Gauss in a letter to Bessel dated 1839—to which my attention was first called by a note in Professor Czuber's treatise‡—has proposed a general expression for the detriment of error very similar to that which is above mentioned. The reflections of the great mathematician on this branch of mathematical psychics deserve to be transcribed here:—"That the metaphysic employed in my *Theoria Motus Corp. Coel.*§ to justify the method of least squares has been subsequently allowed by me to drop (*Dass ich . . . habe fallen lassen*) has occurred chiefly for a reason that I have myself not mentioned publicly. The fact is, I cannot but think it in every way less important to ascertain that value of an unknown magnitude the probability of which is the greatest—which probability is nevertheless infinitely small—rather than that value by employing which we render the Expectation of detriment a minimum (*an welchen sich haltend man das am wenigsten nachteilige Spiel hat*). Thus if $f(a)$ represents the probability of the value a being assumed by (*für*) the unknown quantity x, it is not so important (*ist weniger daran gelegen*) that $f(a)$ should be a maximum as that $\int f(x)F(x-a)dx$, the integral extending over all possible values of x, should be a minimum; when for F is selected a function that is continually positive and continually increases in a due degree (*auf eine schichliche Art*) with the increase of the variable. That the Square is selected

* "Phil. Mag.," vol. xvi, p. 361.

† *Theorie des Beobachtungsfehler*, p. 266.

‡ *Op. cit.*, p. 289.

§ Gauss's first proof of the Method of Least Squares.

for this purpose is "purely arbitrary, and is in the nature of the subject that there should be this arbitrariness (Willkürlichkeit). Except for the well-known very great advantages* which the choice of the square secures, one might have chosen any other function satisfying the above conditions." Gauss does not, however, I think, show that the selection of any function satisfying the above conditions will lead to the same result as the determination of the most probable value of the unknown quantity; when the observations obey the normal law of frequency (or any symmetrical law). Indeed, his admission as to the character of detriment is, I think, destructive of what is most characteristic in his own contribution to the Method of Least Squares.†

If the method of dealing with the matter above suggested is accepted, this first difficulty, with regard to our first problem at least, is not very serious. Moreover, it is tenable that the scruple is less serious when the *quæsitum* is a type, a mere frequency-constant (our peculiar case) than when it is an objective quantity (the case principally contemplated by the classical writers).

2. A second scruple is excited by that premiss of the argument which is technically described as "*a priori*," or "antecedent,"‡ probability. It has been tacitly assumed that the *a priori* probability of any one value being the true one is the same as that of any other value. But this assumption cannot be universally admitted. It is, for instance, *a priori* improbable that the mean of a set of barometric observations should be in the neighbourhood of 31 inches. Nevertheless, I submit that very generally we are justified in assuming an equal distribution of *a priori* probabilities over that tract of the measurable with which we are concerned.§ And even when a correction is required by what is known about the *a priori* probability, this correction is in general, as I have elsewhere shown,‖ of an order which becomes negligible as the number of the observation is increased.

Upon the whole both the scruples which have been raised may be dismissed as regards the present problem. They will recur under other heads.

* The advantage here alluded to is presumably a convenience in the way of calculation other than that which I describe in the following note. Some light is thrown on Gauss's meaning by the interesting reminiscences of his lectures contributed by Dedekind to the Festscrift zur Feier des hundertfünfzigjährigen Bestehens der Kön. Gesell. der Wissen. in Göttingen.

† I refer to Gauss's "second proof" of the law of Error; of which it is characteristic that the method of minimising detriment is applicable, whatever the law of frequency obeyed by the observations, and however small their number (Cf. Czuber *op. cit.* p. 292). That the mean power of the sum of two or a few observations should equal the sum of the mean powers of each, for any law of frequency, is true in general only of the *second* power.

‡ *Cf.* J. S. Mill, *Logic*, Book III, ch. xviii, sec. 5.

§ On grounds stated in the article on the Philosophy of Chance to which reference has been made.

‖ *A priori Probabilities*, *Phil. Mag.*, 1884, vol. xvii., p. 204.

II.—The problem next in the order here adopted is to determine the coefficient of dispersion: given as before a set of observations subject to the normal law of frequency. The coefficient may be defined in any of the usual ways* as the *modulus*, or the *standard deviation*, that is the modulus divided by $\sqrt{2}$, or the squares of either of these coefficients, or the *precision* which is the inverse, or the *weight*,† which is the inverse square, of the *modulus*. I shall follow Gauss in employing the *precision*; and shall at first suppose that the normal law of frequency obeyed by each of the observations has the same precision. The problem may be split up by first supposing that the central point, which was the *quæsitum* of the last problem, is now given. Let the observations measured from this given point be $x_1, x_2 \ldots x_n$; and let $h\left(=\frac{1}{c}\right)$ be the sought *precision*. Then by a use of Inverse Probability analogous to that which was employed in the first problem the required value of h is that which renders the following expression a maximum:— $\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-x_1^2h^2 - x_2^2h^2 - \ldots - x_n^2h^2}$.

The required value of h is therefore $\sqrt{\frac{n}{2\Sigma x_t^2}}$; where Σ denotes summation extending from the first to the n^{th} observation. Call this value of h, h'; and put $h = h' + \mathrm{h}$. Then the law of distribution for the various values of h occurring in the long run is given by a curve which may be put in the form:— $z = \mathrm{G}e^{n \log (h' + \mathrm{h}) - (h' + \mathrm{h})^2 \Sigma x_t^2}$; where, in accordance with the received method of Inverse Probability, G is a coefficient such that the integral of z between extreme limits is unity. Now expand the index in ascending powers of h (divided by h'); and we have (substituting for Σx_t^2 its value in terms of h') $z = \mathrm{H}e^{-n\left(\frac{\mathrm{h}}{h'}\right)^2 + \frac{1}{3}n\left(\frac{\mathrm{h}}{h'}\right)^3 - \frac{1}{4}n\left(\frac{\mathrm{h}}{h'}\right)^4 + \cdots}$; where H is a coefficient which plays the same rôle as the G employed above. Now while h is within the order of magnitude $h'/\sqrt{n}$,‡ the ordinates of the above designated curve differ by a quantity which may be neglected from (some multiple of) the ordinates of a certain normal error-curve. And for values of h above that order of magnitude the ordinates and the area which is their integral are negligible. Therefore the portion of the curve between limits of the order $\pm h'/\sqrt{n}$ is approximately coincident with the normal error-curve whose modulus is $h'/\sqrt{n}$; the ordinates of the actual locus are (throughout that tract) not merely proportional to, but approximately equal to the ordinates of that normal error-curve.

Such is the substance of the reasoning which Gauss has applied

* See Encyclopædia Britannica, Article, "Error" (law of), p. 281, Note 4, as to the different forms of the coefficient.

† Laplace in one passage at least (Théorie Analytique des Probabilités, Supplement I.) gives the name of "Modulus" to the coefficient which is usually, after Gauss, I believe, called "Weight."

‡ That is h' *divided by* $\sqrt{n}$. I shall sometimes employ this modern notation, which saves room.

to the problem now before us,* and Laplace to other parts of the subject.† I have nothing material to add to their reasoning; but I submit that it might with advantage be put in a form which may be regarded as a *law of great numbers.* It might be described as an "inverse" or "subsidiary" law of that kind, in contradistinction to the *law of great numbers* which I have dealt with in former numbers of this *Journal.*‡ That law, it may be remembered, states that if numerous observations, each obeying (almost) any particular law of frequency, are taken at random, their sum (or more generally linear function, or approximation thereto) will approximately obey the normal law of error. A good way of contemplating this approximation is to put the frequency-curve pertaining to each observation in the *form* of the generalised law of error, as thus:—
$y = e^{-\frac{1}{3!}k_1 D + \frac{1}{4!}k_2 D^2 - \cdots} y_0$;§ where, the curve being referred to its centre of gravity as origin, y_0 is the normal error-curve which has for the square of its standard-deviation the mean-square of error pertaining to the curve under consideration, say k; D, D^2, stand for the symbols of differentiation, $\frac{d}{dx}$, $\frac{d^2}{dx^2}$, . . which are to be brought down from the index according to the usual rules of expansion, and brought to bear upon the function y_0. The constants k_1 k_2 . . are formed with the third mean power of deviation, the fourth mean power, and so on, for the particular law of frequency, according to the same rule as the constants in the generalised law of error. But an essential feature of the generalised law, that the constants should form a descending series, is here wanting. Thus the k's may now be supposed to be all of the order unity.‖ The sum of n such observations, each taken at random from the group to which it belongs, will obey a law of frequency, which has the same form as the above written expression for y, but differs in that for k_1, we have now K_1, the sum of all the k_1's each pertaining to a particular component, say $K_1 = \Sigma k_1$, for k_2 we have now Σk_2, and so on; with a corresponding change in the operand, now say Y_0,

* "Zeitschrift für Astron. u. verw. Wissenschaften," I, p. 185; referred to by Czuber *op. cit.* p. 125.

† As subsidiary to the proof of the Law of Error (in the manner which I have endeavoured to elucidate in my paper on the Law of Error in the "Cambridge Philosophical Transactions 1905," Part I, sect. 3); and in the attempt to apply the genuine inverse method where the law of error pertaining to the observations is not given, Théorie Analytique des Probabilités, Liv. II, ch. IV, Art. 23 (an attempt which will be considered in a subsequent section of the present paper).

‡ See in particular the paper on the Generalised Law of Error in the *Journal of the Royal Statistical Society*, 1906.

§ *Loc. cit.*, p. 498.

‖ The selection of a unit is a matter of mere convenience and convention. In former expositions I have taken that standard-deviation of the compound which is here of the order $\sqrt{n}$ to be of the order unity, and accordingly the standard deviation of a component which is here of the order unity to be of the order $1/\sqrt{n}$.

a normal curve which has now for modulus-squared Σk, $= K$. The compound curve thus formed will be much more distended than the original curve; a circumstance which may disguise the essential character of the compound, its tendency to approximate to the normal form. To exhibit that property it is convenient to furl up the compound curve by putting in its equation, for $x/\sqrt{K}$, the new variable X, whereby the mean-square-of-deviation for the compound becomes unity, of the same order as the mean-square-of-deviation for each of the components, each of those being supposed to be of the order unity. The approximation to the normal form becomes now visible. For in the transformation it is proper to put, for $\frac{d}{dx}\left(=\frac{d}{dX}\frac{dX}{dx}\right)$, $\frac{1}{\sqrt{K}}\frac{d}{dX}$. Accordingly D^t becomes transformed to $\left(\frac{1}{\sqrt{K}}\right)^t D^t$, where the first D stands for $\frac{d}{dx}$, the second D for $\frac{d}{dX}$. Thus every co-efficient K_t in the original form of the compound is now affected with a denominator $K^{(t+2)/2}$ where K_1, K_2 . . are each of the order n, agreeably to the convention that all the k's are of the order unity (K would be nk, K_t would be nk_t, if the component curves were identical). When the symbols descending from the index are brought to bear on the (transformed) normal function Y_o the long array of terms affected with K_1, K_2 . . , and powers and products of these co-efficients, grow down and shrink into insignificance, affected as they are with factors of the orders $\frac{1}{\sqrt{n}}$, $\frac{1}{n}$, $\frac{1}{n^{\frac{3}{2}}}$ . . The normal curve alone survives. That is, provided that the expressions multiplied by those factors, viz., the successive differentials of a normal curve with unit standard deviation, each divided by a corresponding factorial, may be treated as of the order unity; that is, for values of the abscissa not exceeding the order unity. Beyond those limits there is a jumble subject to no general law. The groupings of this kind composed by the nature of things may be compared to an unfinished frieze, the design nearly perfect at the centre, an unformed block at each extremity.

Similarly, the subsidiary law of great numbers states that the character of normality becomes attached to the curve which represents the probability that a number of random observations will satisfy a certain condition. The condition is not now that the sum of the observations is equal to an assigned magnitude, but that each of the observations is equal to an assigned magnitude. The proof of this law is simpler than the proof of the more familiar law. The (logarithm of the) ordinate, pertaining to the frequency-curve of any particular observation, is now to be expanded, not in the peculiar development proposed in the preceding paragraph, but according to the ordinary Taylorian law, as thus:—

$$y = Ge^{lx + hx^2 + h_1x^3 + h_2x^4 + \,..},$$

where $el\ h\ h_1 \ldots$ may conveniently be supposed to be of the order unity. The components are now put together, not by the method of cumulation proper to the genuine law of error, but by simple multiplication, that is, addition of logarithms, as thus:—

$$y = \mathrm{H}e^{x\Sigma l + x^2\Sigma h + x^3\Sigma h_1 + \ldots},$$

if, as before, Σ denotes summation extending over all the observations, and H is a co-efficient which secures that the area enclosed by the curve is unity. The effect of this composition is in general* to form a curve which towers above the components as a spire above ordinary buildings.† As this spire becomes higher and slenderer (the number of the components being increased) its shape becomes more nearly normal. This approximation is best exhibited by lowering the height of the spire without disturbing the relations between its parts. This will be effected by first taking the *mode* of the compound or derivative curve as the origin, and then *unfurling* the curve by a transformation the opposite of that which was just now employed in the statement of the ordinary law of error. Let the logarithm of y, the ordinate of the compound curve, be of the form $\mathrm{L} - \mathrm{N}x^2 + \mathrm{P}x^3 + \mathrm{Q}x^4 \ldots$, where x is a new variable referred to the *mode* (of the compound curve) as origin, and N, P, Q . . are each of the order n, say, respectively, $n\mathrm{h}$, $n\mathrm{h}_1$, $n\mathrm{h}_2$ . . Now unfurl the curve by substituting X for $\sqrt{\mathrm{N}}x$. The transformed curve will be of the form $\mathrm{Y} = \mathrm{H}e^{-\mathrm{X}^2 + \frac{\mathrm{h}_1}{\sqrt{n}}\frac{\mathrm{X}^3}{\mathrm{h}^3} + \frac{\mathrm{h}_2}{n}\frac{\mathrm{X}^4}{\mathrm{h}^4} \pm \ldots}$, where $\mathrm{h}, \mathrm{h}_1, \mathrm{h}_2 \ldots$ are of the same order as the h's, that is, the order unity. It is evident that as the number of the observations increases, the curve in question becomes more nearly normal.

Such then being the character of the law of distribution for the value of the *quæsitum* proper to this section, we see, not only that the value above assigned is the *most probable* value, but also that it is the *most advantageous* value (in the classical sense of the term above explained with reference to our *first* scruple), provided that the number of observations is so large that the law of distribution for the *quæsitum* does not differ significantly from a normal curve of error. Otherwise, the distribution being sensibly unsymmetrical, it may be that the *mode* of the curve in question will *not* be the point of least detriment. That point is more plausibly to be placed on one side of the mode, on the longer arm of the curve. We may conjecture, then, that the most advantageous value of the *precision* is somewhat greater than the most probable value above assigned.‡ The value of the modulus c, and of the other co-efficients relative to dispersion, may be deduced from the value of the precision h.

* Under conditions which will be considered with respect to the generalised problem, No. 6.

† Fig. 1 (above, p. 385) may serve to illustrate the relation between the compound locus and (any one of) its components, if it is conceived that the dotted curve is now only approximately, and the continuous curve, not even approximately, normal.

‡ It may suffice to point out that the centre of gravity for the curve representing the distribution of the values of h is *above* the mode.

We may also find the law of distribution for any of those co-efficients, *e.g.*, the modulus, directly by a calculation analogous to the above. But here arises a question suggested by our second scruple. In searching for the best value of c, as we have done for that of h, we should be taking for granted that one value of c is *a priori* as probable as another; just as we have above tacitly assumed that one value of h is *a priori* as probable as another. But these two assumptions are inconsistent. For if the values of h are evenly distributed, then as $c = 1/h$, $dh = -\frac{1}{c^2}dc$, the values of c are distributed in such wise that the *a priori* probability of any particular value of c is inversely proportional to its square; and conversely if the *a priori* probability of c is evenly distributed, that of h is not so. The apparent anomaly may recall the objection which Cournot brings against the prevalence of the normal law of error, viz., in effect, that if things generally obeyed that law, then the squares and other functions thereof which also often represent real things cannot obey that law. The answer is that the objection is only serious when the things are extraordinarily small (or large), and the functions not those occurring in ordinary practice.* Now in the case before us, where the (mean powers of) given errors are neither infinite or infinitesimal, it may be presumed that neither the modulus, nor its reciprocal, nor any multiples or powers which we may require, vanish. Accordingly the difficulty may be postponed to another section.

The excellence of the solution which has been obtained may be tested by observing its superiority over other solutions, such as those which are obtained from mean powers (of the given observation) other than the *second*, or by way of *percentiles*. It is shown by Professor Czuber, after Gauss, that each of these determinations is liable to a greater probable error than is the most probable value as (above) determined by genuine Inverse Probability.† It is a nice question whether this superiority could have been predicted prior to that verification.‡

So far in this section we have supposed all the observations, $x_1, x_2 \ldots$ (measured from a given centre), to range under a normal curve with the same modulus; and accordingly the value afforded by each for the modulus to be of equal weight. The solution is easily extended to the case in which it is given that the modulus pertaining to x_p is p times, to x_t t times, and so on, a certain common measure which is sought. An example occurs in my *Applications of the Calculus of Probabilities* in this journal.§ The *quæsitum* is the modulus pertaining to the ratio of male to female births for a group of 1,000 births. The data are the deviations from the average for the whole country (considered as a fixed and known centre) of forty counties with varying numbers of births.

* As I have elsewhere argued. "Phil. Mag.," 1892, vol. 34, p. 431, *et seq.*

† *Beobachtungsfehler*, Art. 52, *et seq.* referring to "true errors," the object of our Problem 2.

‡ See the remarks on the Method of Least Squares below.

§ Vol. lxi., 1898, p. 126.

III.—The next problem is a combination of the two preceding. Given a number of observations known to range under one and the same normal curve of error, to find the most probable (or advantageous) values both of the centre and of the modulus (or other co-efficient of dispersion, *e.g.*, the precision, which is a function of the modulus). The obvious course is to combine the two preceding solutions: to form from the data, by Inverse Probability, the surface representing the distribution of the frequency with which each particular value of the pair of *quæsita* would be in the long run associated with the given set of observations. Find the *mode* of that surface; say x' for the centre, and h' for the precision; and refer the surface to the mode as origin by substituting for x (measured from the origin in terms of which the observations have been given—say a point on the abscissa well to the left of the smallest observation) $x' + \mathrm{x}$, and for h (measured from zero) $h' + \mathrm{h}$. Then by the subsidiary law of error the surface representing the distribution of x and h approximates to the normal surface of two dimensions as the number of the observations is increased, with the origin (x' and h' in the original notation) as centre. That pair of values then constitutes the solution. For the required centre we have, as in Problem 1, the Arithmetic Mean of the observations; for the required precision the square root of the reciprocal of twice the mean square of errors; errors now measured not from a given centre, as in Problem 2, but from the centre which has been found—the sum of the so-called "apparent errors" or "residuals" divided by the number of the observations. I refer to my paper of 1883 on the Method of Least Squares for the argument in favour of this solution.*

It must give us pause, however, that the classical writers on Probabilities have not only followed a different procedure, but also reached a different result, namely, that the sum of the square of the residuals, in the formula for the precision, modulus, &c., should be divided not as above written by n, but by $(n-1)$. Professor Czuber also lends the weight of his authority to the classical formula.† In the paper to which reference has been made, I enquired whether the contradiction could be explained by taking as the probable error, not the most probable value of c (multiplied by 0·4769), but that error of x for which there is an even chance on an average of all the values that c may possibly assume—the "absolute" as distinguished from the "partial" probable error in phraseology subsequently introduced by Professor Pearson.‡ I found for this probable error an expression which appears to be of no great significance.§ But I would have hit upon the required

* "Phil. Mag.," vol. xvi (Series V), p. 367.

† *Beobachtungsfehler*, p. 152; with reference to "apparent errors."

‡ "Phil. Trans.," 1898, vol. 191A, p. 242.

§ It did not suggest the important theorem due to Prof. Karl Pearson that, in the case of a standard-deviation and the organ to which it refers, the "absolute" and "partial" probable errors are approximately equal. "Phil. Trans.," 1898, vol. 191A, p. 236.

explanation if I had gone one step further and enquired what is the probable error of c on an average of all the values that x may possibly assume. The values of c considered thus, irrespectively of the value assigned to x, are given by integrating, with respect to x, between extreme limits, the expression $Hh^n e^{-[n\text{x}^2 + \Sigma \text{x}_t^2]h^2}$, where H is a constant which secures that the (double) integral of the above expression with respect to both x and h is unity; x is measured from the Arithmetic Mean of the given observations as origin. When x is made by integration to disappear from the above-written expression there results, as the law of frequency for the values of h, $z = H'h^{n-1} e^{-h^2 \Sigma \text{x}_t^2}$, where H′ plays the same *rôle* as H in the preceding formula. Whence it is deducible, by reasoning on a par with that which has been applied to Problem 2, that the most probable value—in the sort of "absolute" sense which is now under consideration—of h is, in accordance with the classical formula, $\frac{(n-1)}{2\Sigma \text{x}_t^2}$. We may surmise that the "most advantageous" value of h is in this case, for the same reason as in the simpler case above considered,* greater than its most probable value. This consideration tends to move us back from $(n-1)$ towards n. Still, on the whole it may seem doubtful which of the formulæ is in general preferable. But the question is, I think, of merely theoretical interest. For the difference between n and $n-1$ can only be sensible when n is not large; and then an accurate determination is not to be expected, as both the scruples above indicated then come into force.

Before leaving this problem it should be noticed, that through the powerful methods introduced by Mr. Sheppard,† the solution admits of a verification similar to that which the solution of Problem 2 has received—logically similar, though mathematically more difficult.

IV. In the next problem the *quæsitum* is the co-efficient of correlation; given a set of (coupled) observations, corresponding to points in a plane, say (x_1, y_1), (x_2, y_2) . . . . . known to range under one and the same normal surface. We may begin by supposing known the other constants of the system, the two co-ordinates of the centre, the probable error or modulus of each variable considered by itself. Taking the centre as origin, and referring each of the variables measured therefrom to their respective moduli as units we find for the *quæsitum*, upon the principle above explained, that value of r which makes the following expression a maximum:—

$$\left(\frac{1}{\pi\sqrt{1-r^2}}\right)^n e^{-[\Sigma \text{x}_t^2 - 2r\Sigma \text{x}_t \text{y}_t + \Sigma \text{y}_t^2]/(1-r^2)}$$

where x_1, x_2, and likewise y_1, y_2 . . , are errors measured from the centre referred to unit modulus, and accordingly, t assuming every integer value from 1 to n, $\Sigma \text{x}_t^2 = \Sigma \text{y}_t^2 = \frac{1}{2}$. Equating the first differential (with respect to r) of this expression to zero, and observing that the second differential is (for the value of r

* Above, p. 391.

† "Phil. Trans.," 1899, vol. 192A, p. 131 *et seq.*

determined by that equation) negative, we obtain for the sought value of r, $2\Sigma x_t y_t$. Call this value r'. Put $r = r' + \text{r}$ and we obtain (by the subsidiary law of error) for the distribution of the various values of r (associated with the given set of observations in the long run) a normal error-curve with probable error, which has been given by Professor Karl Pearson.*

The application of the genuine inverse method to determine the co-efficient of correlation was introduced by Professor Pearson.† It is instructive to contemplate its superiority over other methods which may seem plausible. Such is the method which I once essayed† of averaging the ratios of the type y_t/x_t, a method which presents itself as natural when the co-efficient of correlation is defined as the most probable ratio between an assigned abscissa and the corresponding ordinate. It might have been and was foreseen that the simple averaging of such ratios did not afford the best value of r—if only from the interesting circumstance that the determination is improved by omitting some of the data. But it could not so easily have been foreseen, and is I think a curious circumstance, that in seeking the best variety of a method which is not the best I should have been guided to the absolutely best formula. I venture to reproduce the argument, premising that the x, y, and ρ_{12} of the following passage correspond respectively to x, y, and r as defined in the present section. "Regarding each assigned or [in Mr. Galton's phrase] 'subject' x, divided into the associated or 'relative' y as affording an observation-equation $\frac{y}{x} = \rho_{12}$, we see that the best combination of these data is obtained by affecting each observation $\frac{y}{x}$ with a weight inversely proportional to its modulus-squared. Now . . . every y, whatever the x with which it is associated has for the modulus of its fluctuation [the same quantity] $\sqrt{1-\rho^2}$. Accordingly . . . the weight of [each observation] $\frac{x}{y}$ is directly proportional to x^2. The best value of ρ_{12} is $\frac{\text{S}xy}{\text{S}x^2}$." ‡ Which agrees with the above result, each of the moduli being unity.

With reference to the first of the scruples raised in a former section, it may be worth while observing that the (complete) expression for the distribution of the values of r is not symmetrical, and that therefore a slight correction of n in the formula for r might be defensible.

A more serious difficulty is connected with the *second* scruple, relative to *a priori* probabilities. Since r, unlike h or c in the

* "Phil. Trans.," 1896, vol. 187A, p. 265.

† "Phil. Mag.," 1893, vol. 36, p. 100 *et seq.*

‡ The probable error for this value of ρ_{12} given in the context is not the same as that above alluded to for r, the latter being of the "partial" kind (the centre and moduli being in this section assigned), while the former agrees with the "absolute" probable error, viz., $0{\cdot}4769\ \sqrt{2}\ \sqrt{1-\rho^2}/n$ (*cf.* Pearson, "Phil. Trans.," 1898, vol. 191A, p. 242.

preceding problems, may very well be zero, it should seem that if the values of r are distributed with uniform probability, the values of r^2 cannot be, even approximately, so distributed. Let $p_1\Delta r$ be the probability of r having any particular value (say, between $a-\frac{1}{2}\Delta r$ and $a+\frac{1}{2}\Delta r$), where p_1 is a constant; then the probability of r^2 having any particular value is $p_2(\Delta r^2)$ where $p_2=p_1/2r$. Thus, when r is zero, the *a priori* probability of r^2 becomes infinite.

This deduction is not indeed absurd, as the "probability" p_2 is supposed to be multiplied by a differential $\Delta(r^2)$, but it is inconsistent with another equally plausible deduction. It is very natural to define the system, not by r, the co-efficient of correlation, but by the modulus of X, where X is one of the *principal* co-ordinates of the system, by transformation to which the normal surface appears as free from correlation. Let X be that one of the principal axes for which the modulus is greatest, being $\sqrt{1+r}$, say C (while the modulus for Y is $\sqrt{1-r}$). As C ranges from zero to $\sqrt{2}$, it may be presumed, on grounds above indicated, that in the neighbourhood of $C=1$ small powers (and other ordinary functions) of C are not very unequally distributed *a priori*. We may presume then that C^2—that is, $1+r$, and therefore r—is so distributed. We may also presume that C^4—that is, $1+2r+r^2$—is not very unequally distributed. But if the values of r, and also the values of r^2+2r, are distributed with approximate equality it seems to follow that the values of r^2 are thus distributed; which is contrary to the first deduction!

There seems to be here an irreducible element of arbitrariness; comparable to the indeterminateness which baffles us when we try to define a "random line" on a plane, or a "random chord" of a circle. It is a nice question how far such antinomies should give us pause when dealing with a value of r which is in the neighbourhood of zero.* For values in that neighbourhood there is, I submit, a more important *a priori* datum, namely, the presumption against correlation which exists in the case of attributes between which common sense cannot see any possible connection. A great many observations would be required to demonstrate the existence of the correlations which are the objects of psychical research.

V.—Next let us suppose not only the co-efficient of correlation, but also the other constants of the normal surface to be sought, the other data being the same. The procedure is a generalisation of that which has been already exemplified. There now comes into view the remarkable principle discovered by Professor Karl Pearson,

* It is to be remarked that similar antinomies may arise even with respect to our first problem, when the object under measurement is not an organ which cannot well be equal to zero, but such a quantity as *velocity*. If the velocities (in an assigned direction) of a set of molecules whose centre of gravity is at rest are distributed (as usually presumed) according to a law of frequency that is symmetrical about zero, is the probability that the *energy* (proportional to the *square* of the velocity) of a molecule taken at random from the set should be zero particularly great?

that in his words "if a random selection be made out of a population with regard to one organ, there will be tendencies for the variation of all other organs and the correlation between all organs to change also in certain directions, which can be definitely indicated so soon as the general population has been measured and the effect of the random selection on one organ has been obtained."* This theorem appears to be of great importance in its application to the science of evolution. Doubtless in this branch of Statistics, as often before in Physics, the development of mathematical theory has been stimulated by an ardent interest in the subject to which it is applicable. It would be superfluous to transcribe Professor Pearson's brilliant applications of this general principle to the particular present case. I go on to consider the more general case which is the object of his "General Theorem on the probable errors of frequency-constants."

(*To be continued.*)

II.—*Settlement and Agricultural Development of the North-West Provinces of Canada.* By ERNEST H. GODFREY.

OUT of that vast area, once known vaguely as the Canadian North-West, have recently been carved the two new provinces of Saskatchewan and Alberta, which, with Manitoba, now comprise the three North-West Provinces, each a federal constituent of the Dominion. The opening up of new fertile lands in these provinces by the construction of railways, coupled with a policy of vigorous advertisement, has, since the commencement of the twentieth century, attracted settlers in annually increasing numbers from other parts of Canada, from the United States, from the British Isles, and from the continent of Europe.

During the last seven years immigration into Canada has proceeded with tidal force and regularity. Every year in this period the number of immigrants has exceeded that of the previous one, as is shown by the following table, which, for each of the fiscal years 1901 to 1907, gives the number of immigrants into Canada, with the numbers destined for different parts of the Dominion, whether Eastern Canada, British Columbia, or the North-West Provinces. The destination of a small proportion of the total number is not shown.

* *Loc. cit.*, p. 235.