Calculated as above the composition of the undiluted black-damp was thus:—

Nitrogen...............	85·30
Carbonic acid............	14·70

This result differs little from that given by the other samples.

II. "Mathematical Contributions to the Theory of Evolution. II. Skew Variation in Homogeneous Material." By KARL PEARSON, University College, London. Communicated by Professor HENRICI, F.R.S. Received December 19, 1894.

(Abstract.)

PART I.—*Theoretical.*

In the deduction of the normal curve of frequency from the symmetrical point binomial, three conditions are usually assumed to be true:—

(*a*.) The chances of any "contributory cause" giving its unit of deviation in excess or in defect are presumed to be equal.

(*b*.) The number of "contributory causes" are supposed to be indefinitely great.

(*c*.) The "contributory causes" are all supposed to be independent.

(*c*) amounts to the assumption of a binomial form $(p+q)^n$, (*a*) to the equality of p and q, (*c*) to the indefinitely great value of n.

It is shown in the paper that there is an important geometrical relation between the normal frequency curve

$$z = z_0 e^{-x^2/2\sigma^2},$$

and the symmetrical point binomial, $(\frac{1}{2}+\frac{1}{2})^n$, which is true independently of the magnitude of n. Thus the condition (*b*) is not necessary to the very close fitting of symmetrical point binomials to normal curves for even very small values of n, such, for example, as 8 or 10. This has been long recognised in statistical practice if its source has not been noted.

We can remove the condition (*a*) from our *à priori* limitations by finding a curve which is related to the skew binomial $(p+q)^n$ in precisely the same manner as the normal curve is related to the symmetrical binomial $(\frac{1}{2}+\frac{1}{2})^n$. The equation to this curve is

$$z = z_0 \left(1+\frac{x}{a}\right)^p e^{-\gamma x}.$$

If α be the total frequency, and $\mu_r\alpha$ the rth moment of the frequency curve about its centroid vertical, then for this curve

$$2\mu_2(3\mu_2^2-\mu_4)+3\mu_3^2=0.$$

This relation must be satisfied or nearly satisfied if a series of observations or measurements is to be fitted with the skew curve, which is related to the skew point-binomial as the normal curve to the symmetrical point-binomial. For fitting a skew point-binomial we must have

$$\mu_4 < 3\mu_2^2+3\mu_3^2/(2\mu_2).$$

For the normal curve $\mu_4=3\mu_2^2$. But a great number of statistical returns—especially in anthropometry and zoometry—give

$$\mu_4 > 3\mu_2^2+3\mu_3^2/(2\mu_2).$$

Hence they differ from the normal curve in the opposite direction to the skew point-binomial and its corresponding frequency curve.

After the complete theory of the fitting of skew binomials and this special skew curve has been discussed with examples, the memoir proceeds to the generalisation of the frequency curve by withdrawing the limitation (c) above. Just as the symmetrical binomial and normal curves are illustrated by the tossing of a group of n coins, and the skew binomial and its skew curve by the spinning of a group of n m-sided teetotums, so we can arrive at a series of curves in which the contributory causes are interdependent, by considering the withdrawal of r cards from a pack of ns cards containing s suits; or, again, by drawing a definite amount of sand from a vessel containing two kinds of sand.

For discontinuous series the solution is a hypergeometrical series. If now a curve be formed which is related by the same fundamental geometrical relation to this hypergeometrical series as the normal curve to the symmetrical point-binomial, or the first skew curve to the skew point-binomial, we obtain a generalised frequency curve which contains both those hitherto considered as special or limiting cases.

It is not suggested that the hypergeometrical series or its corresponding curve is the only case in which the *à priori* condition (c) of dependence of "contributory causes" is replaced by an interdependence. But it is suggested that it is one of the most important cases, and one which naturally occurs at the commencement of our investigations. That it is probably quite sufficient is evidenced by the fact that the author has hitherto failed to find any group of homogeneous and skew statistics which cannot be closely expressed by the curves which correspond to the hypergeometrical series.

The differential equation to the generalised frequency curve is shown to be of the form

$$\frac{1}{z}\frac{dz}{dx} = -\frac{x}{\beta_1 + \beta_2 x + \beta_3 x^2}.$$

If we put $\beta_3 = 0$ we have the curve corresponding to the skew-binomial; if we put $\beta_2 = \beta_3 = 0$ we have the normal curve. In the most general case we are led to two principal types of curves

(i.) $$z = z_0 \frac{1}{\left(1 + \left(\frac{x}{a}\right)^2\right)^m} e^{-\gamma \tan^{-1}\frac{x}{a}}.$$

(ii.) $$z = z_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}.$$

The second of these curves is marked by a limited range and skewness. Its theory—method of fitting to actual statistics and its geometrical properties—are discussed, and the curve is shown to involve in fitting only the use of a table of Γ-functions—a table which already exists.

The first of these curves has skewness but no limit to range. This unlimitedness of range is not, however, necessarily significant. There is a limit to the height of adult males, or at any rate to the ratio of their sitting to standing height, but we do not hesitate to express the results in terms of the normal curve. The fact is that both normal curve and generalised curve are only close approximations to series—point-binomial and point-hypergeometrical series—which can themselves give a limited range, and we ought to fit these series rather than the curves to our observations.*

The criterion to distinguish between the application to any special case of curves (i) or (ii) is the negative or positive value of

$$2\mu_2(3\mu_2^2 - \mu_4) + 3\mu_3^2,$$

which we have seen vanishes for the curve corresponding to the skew point-binomial.

The complete treatment of curves of the first kind is shown to depend on a certain integral called a G-function. This G-function has been discussed in a recent paper by Dr. Forsyth, to whom the author had referred for information with regard to it. It is built up of Γ-functions with imaginary arguments. The function has not yet been tabulated, but various formulæ are given for its evaluation, and it is hoped that its values may shortly be calculated for the range

* The fitting of the first series is discussed in this memoir; the fitting of hypergeometrical series is reserved as the memoir is already of considerable length.

of arguments having more special practical interest.* The theory of the whole system of skew curves and their limiting cases is then discussed.

The author regrets that while he has obtained a criterion for each species of skew curve, he has hitherto failed to find one which will distinguish a compound curve, *e.g.*, heterogeneous material, from a skew curve resulting from skew variation in homogeneous material. He does not despair, however, of ultimately finding such a criterion. The test of actual fitting is generally sufficient, but is, of course, laborious.†

PART II.—*Illustration.*

The second part of the memoir provides the minimum of illustration, which the author considers absolutely necessary, in order to demonstrate that the generalised curves reached are capable of the widest application to every variety of practical statistics. The illustrations show that from the slight amount of skewness usually neglected by statisticians—although of vital import when we come to consider variation with growth, as in statistics of child-variation with growth—even to extreme cases in which the curve is asymptotic to the ordinate of maximum frequency, a good fitting generalised frequency curve can be found. Although the number of illustrations is considerable, and is only a part of the author's material, yet he hesitates at present to make any dogmatic statements with regard to the relations of skewness in variation to secular evolution; but he believes that the persistent recurrence of certain types of curves in zoometry and of certain directions of skewness in anthropometric statistics will be found, as sufficient material accumulates, to justify broad generalisations, although at present they can only be treated as suggestions for further investigation.

The special illustrations given are: barometer variation (Venn), variation in crabs and prawns (Weldon), in height of American recruits (Baxter), American school girls (Porter), in length-breadth index of Bavarian skulls (Ranke), frequency of enteric fever (Metropolitan Asylums Board), guesses at tints (Gresham College), divorce statistics (Willcox), variation in house-value (Goschen), variation in buttercups and clover (De Vries), variation in pauper percentages (Booth), and resolution of the English male mortality curve (Ogle) into skew components.

The memoir concludes with some general remarks on the modifications required in the theory of correlation by the use of generalised curves, but reserves for the present its complete discussion.

* The British Association have kindly given a grant for this purpose.

† It is noteworthy that all cases of compoundedness dealt with hitherto by the author give $2\mu_2(3\mu_2^2-\mu_4)+3\mu_3^2$ positive.