XI. *Mathematical Contributions to the Theory of Evolution.—X. Supplement to a Memoir on Skew Variation.**

By KARL PEARSON, *F.R.S., University College, London.*

Received May 22,—Read, June 20, 1901.

(1.) IN a memoir on Skew Variation published in the 'Phil. Trans.,' A, vol. 186, 1895, a series of frequency curves are discussed which are integrals of the differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-x}{c_1 + c_2 x + c_3 x^2} \quad \ldots\ldots\ldots\ldots \text{(i.)}.$$

(See p. 381 of the memoir.)

The discussion of four main types is given in detail, and a brief reference is made to various sub-types which may occur. The types considered in that memoir covered at the time all the frequency series, and they were fairly numerous, that I had had occasion to deal with. In the course of the last few years, however, I have been somewhat puzzled by frequency distributions for which the criterion $2\beta_2 - 3\beta_1 - 6$ (see p. 378) was positive, and therefore *à priori* a curve of the type

$$y = y_0 \frac{1}{\left\{1 + \left(\frac{x}{a}\right)^2\right\}^m} e^{-\nu \tan^{-1}(x/a)}$$

was to be expected, but which on calculation gave ν imaginary. The frequency distributions in question arose† occasionally in sociological statistics, but also in

* 'Phil. Trans.,' A, vol. 186, p. 343.

† Some other frequency distributions, which on first investigation fell under Types V. and VI. of the present paper, were found with improved values for the moments to fall under types already discussed. Mr. W. F. SHEPPARD'S values for the moments ('Lond. Math. Soc.,' vol. 29, p. 369, formula 30) should certainly be used in preference to those given by me ('Phil. Trans.,' A, vol. 186, p. 350) whenever we are calculating the moments of a curve from areas and not from true ordinates. I hope shortly to publish a paper on this point, which is one really of quadrature formulæ. Meanwhile for every true frequency curve *with high contact at both terminals* we ought to use

$$\mu_2 = c^2(\nu_2' - \nu_1'^2 - \tfrac{1}{12})$$
$$\mu_4 = c^4(\nu_4' - 4\nu_1'\nu_3' + 6\nu_1'^2\nu_2' - 3\nu_1'^4 - \tfrac{1}{2}(\nu_2' - \nu_1'^2) + \tfrac{7}{240}),$$

instead of the values given on p. 350, μ_3 remaining unchanged.

biological investigations. It seemed, therefore, desirable to enter a little more fully into the analysis of the cases in which the criterion was positive but ν imaginary, and discover what types of frequency curves had escaped my attention.*

The key to the solution lies in the fact noted on p. 369 of the memoir, namely, that even if the criterion be positive, there will still be a solution akin to Type I. and not to Type IV. if ϵ be negative. No frequency series satisfying these conditions had at that time come under my notice, and later, when collecting data of floral variability, my own remark as to ϵ had slipped from my memory. It is the object of this supplement to obtain an improved criterion of type, to discuss the nature of the curves which fill the gap observed, and to illustrate by one or two examples the fitting of such curves to actual statistics.

(2.) *The Two Criteria.*

Throughout this supplement the notation of the previous memoir will be assumed to be familiar to the reader.

Turning to p. 378 of that memoir, we note that since β_1 and $r - 1$ are necessarily positive, z if positive must be $> r^2$. Hence ν can only become imaginary if z be negative, or

$$\frac{\beta_1 (r-2)^2}{16\,(r-1)} > 1.$$

Substitute in this the value of r and it becomes

$$\frac{\beta_1 (\beta_2 + 3)^2}{4\,(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)} > 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(ii.)}.$$

Hence the complete condition that a curve of Type IV. shall give the distribution of frequency is not only

$$\kappa_1 = 2\beta_2 - 3\beta_1 - 6 > 0,$$

but also

$$\kappa_2 = \frac{\beta_1 (\beta_2 + 3)^2}{4\,(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)} < 1.$$

Turning back to p. 369, we see that ϵ being positive the complete conditions for a curve of Type I. giving the distribution of frequency are

$$\kappa_1 = 2\beta_2 - 3\beta_1 - 6 < 0,$$

* I was very loath to adopt Professor EDGEWORTH'S method of inventing new frequency curves by putting $x = f(x')$ in a normal frequency distribution, $y = y_0 e^{-cx^2}$. Besides strong theoretical objections to this process, I had found Equation (i.) so sufficient for a great variety of cases that I felt confident it must cover the newly discovered outstanding cases, and this confidence seems justified by the result.

and

$$\kappa_2 = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)} < 0.$$

The latter condition will be always satisfied since β_1 and $4\beta_2 - 3\beta_1$ are positive for any distribution whatever, and $2\beta_2 - 3\beta_1 - 6$ is negative by hypothesis.

Further, in the previous case κ_2 is seen to be essentially positive.

Hence the criteria written down cover all possible cases but those for which

$$\kappa_2 > 1.$$

Sub-cases which arise from transition curves just at the limits will, however, be likely to be of interest, What happens when $\kappa_2 = \infty$ and when $\kappa_2 = 1$? The only possibility for $\kappa_2 = \infty$ is $2\beta_2 - 3\beta_1 - 6$, or $\kappa_1 = 0$. But this curve has been fully treated under Type III. in the memoir.

We shall see later that $\kappa_2 = 1$ leads us up to a novel transition curve of considerable interest.

To ascertain something about the general case in which $\kappa_2 > 1$, let us return to the memoir again and examine the value of ϵ on p. 369. It can only be negative if

$$4 + \tfrac{1}{4}\beta_1(r+2)^2/(r+1) \text{ be } < 0,$$

where r is here

$$= \frac{6(\beta_2-\beta_1-1)}{3\beta_1-2\beta_2+6}.$$

Substituting, we find at once

$$\kappa_2 > 1,$$

which in itself involves $\kappa_1 > 0$.

Hence the missing gap corresponds to those cases in which ϵ is negative.

It will be clear that κ_2, although in form giving a more complex criterion than κ_1, is really more effective, as covering all the possible cases. We have then the following scheme :—

Criterion κ_2.	Corresponding frequency curve.
$\kappa_2 = \infty$	Transition curve, Type III. (Memoir, p. 373).
$\kappa_2 > 1$ & $< \infty$	Type VI. (see p. 448 below).
$\kappa_2 = 1$	Transition curve, Type V. (see p. 446 below).
$\kappa_2 > 0$ & < 1.	Type IV. (Memoir, p. 376).
$\kappa_2 = 0, \beta_1 = 0, \beta_2 = 3$. .	Normal curve.
$\kappa_2 = 0, \beta_1 = 0, \beta_2$ not $= 3$	Type II. (Memoir, p. 372).
$\kappa_2 < 0$	Type I. (Memoir, p 367).

The object of this supplement is to discuss the calculation of curves of Type V., and to consider those of Type VI. somewhat more at length, they being only briefly referred to on p. 369 of the memoir. It will be seen that Type I. of the memoir has now broken up into two divisions. One portion is the old Type I. passing into the normal curve on one side and Type III. on the other. This Type III. separates the second portion, Type VI., of the old Type I. from the first portion. Type VI. passes from Type III. to the new transition curve Type V., which, like Type III., will be found to have a range limited in one direction only. Finally this new Type V. is the transition to the old Type IV. bounded on the other side by the sub-curve, the old Type II., and beyond that the normal curve. Thus we see that Types I. and IV. do not pass directly into each other through Type III., as might be supposed by the criterion $\kappa_1 >$ or < 0, but that there are a series of intervening curves, two of which, Types V. and VI., require further consideration, if we are to complete the whole round of frequency distributions embraced under the differential equation (i.).

(3.) *On the Frequency Curve of Type V.*

Returning to the fundamental differential equation (i.), let us consider what transformation takes place when the denominator on the right has *equal* roots.* We may then write it in the form

$$\frac{1}{y}\frac{dy}{dx} = \frac{-x}{c_0(c_1+x)^2} = \frac{c_1}{c_0}\frac{1}{(c_1+x)^2} - \frac{1}{c_0(c_1+x)}.$$

Hence
$$\log y = -\frac{c_1}{c_0}\frac{1}{(c_1+x)} - \frac{1}{c_0}\log(c_1+x) + \text{const.}$$

Thus
$$y = y_0\, e^{-\frac{\gamma}{c_1+x}}(c_1+x)^{-p},$$

where, y_0 is a constant, $\gamma = c_1/c_0$ and $p = 1/c_0$. Thus changing the origin we may write the curve :

$$y = y_0\, x^{-p}\, e^{-\gamma/x} \quad \ldots\ldots\ldots\ldots \text{(iii.)},$$

where $x_{m0} = \gamma/p$ gives the distance of the mode from the new origin.

To find the moments about this origin, we notice that, p and γ being positive, $y = 0$ when $x = 0$ and when $x = \infty$. Thus as in the curve of Type III. we have a range limited at one end only.

To find the moments we have, if α be the area,

$$\alpha\,\mu'_n = \int_0^\infty y_0\, x^{-p+n}\, e^{-\gamma/x}\, dx \quad \ldots\ldots\ldots\ldots \text{(iv.)}.$$

* I owe to Miss AGNES KELLY, Ph.D., the suggestion that this type of frequency curve deserved fuller treatment.

Put $\gamma/x = z$, and we find

$$\alpha \mu'_n = y_0 \gamma^{n-p+1} \int_0^\infty e^{-z} z^{p-n-2} dz$$

$$= y_0 \gamma^{n-p+1} \Gamma(p-n-1) \quad \ldots\ldots\ldots \text{(v.)}.$$

Thus: $$\alpha = y_0 \gamma^{1-p} \Gamma(p-1) \quad \ldots\ldots\ldots \text{(vi)}.$$

$$\left.\begin{aligned} \mu_1' &= \gamma/(p-2) \\ \mu_2' &= \gamma^2/(p-2)(p-3) \\ \mu_3' &= \gamma^3/(p-2)(p-3)(p-4) \\ \mu_4' &= \gamma^4/(p-2)(p-3)(p-4)(p-5) \end{aligned}\right\} \quad \ldots\ldots \text{(vii.)}.$$

Transferring to the centroid we find

$$\left.\begin{aligned} \mu_2 &= \frac{\gamma^2}{(p-2)^2(p-3)} \\ \mu_3 &= \frac{4\gamma^3}{(p-2)^3(p-3)(p-4)} \\ \mu_4 &= \frac{3(p+4)\gamma^4}{(p-2)^4(p-3)(p-4)(p-5)} \end{aligned}\right\} \quad \ldots\ldots\ldots \text{(viii.)}.$$

$$\beta_1 = \mu_3^2/\mu_2^3 = \frac{16(p-3)}{(p-4)^2} \quad \ldots\ldots\ldots \text{(ix.)}.$$

$$\beta_2 = \mu_4/\mu_2^2 = \frac{3(p+4)(p-3)}{(p-5)(p-4)} \quad \ldots\ldots\ldots \text{(x.)}.$$

Eliminating p between β_1 and β_2 we find after some reductions:

$$\beta_1(\beta_2+3)^2 = 4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)$$

or, $$\kappa_2 = 1 \quad \ldots\ldots\ldots\ldots\ldots \text{(xi.)}.$$

Clearly, since this is the condition for Type V., that transition curve is none other than the curve obtained by making the denominator of the right-hand side of the differential equation have equal roots. The curve is clearly of considerable interest, and its existence had not been noticed in the previous series of frequency curves.

The manner of fitting it is now easily described.

Equation (ix.) gives us a quadratic to find $p-4$:

$$(p-4)^2 - \frac{16}{\beta_1}(p-4) - \frac{16}{\beta_1} = 0 \quad \ldots\ldots\ldots \text{(xii.)}.$$

The positive root of this is the required solution.

γ is then found from the first of equations (viii.), or if σ be the standard deviation $= \sqrt{\mu_2}$, then :*

$$\gamma = \sigma\,(p-2)\,.\,\sqrt{(p-3)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xiii.)}.$$

Then (vi.) gives :

$$y_0 = \frac{\alpha\gamma^{p-1}}{\Gamma(p-1)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xiv.)},$$

which determines y_0, the remaining constant for the shape of the curve.

For the position of the curve, we have for the distance from origin to mean, from the first of equations (vii.) :

$$\mu_1' = \gamma/(p-2) = \sigma\sqrt{(p-3)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xv.)}.$$

If d be the distance from mode to mean we have :

$$\begin{aligned} d &= \mu_1' - \gamma/p \\ &= \frac{\gamma}{p-2} - \frac{\gamma}{p} = \frac{2\gamma}{p\,(p-2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xvi.)}. \end{aligned}$$

Further, the skewness :

$$\text{Sk.} = d/\sigma = \frac{2\sqrt{(p-3)}}{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xvii.)}.$$

Thus the solution is completed.

(4.) *On the Frequency Curve of Type VI.*

Type VI., as we have seen, corresponds to the case in which Type I. of the memoir has its ϵ negative. Hence either m_1' or m_2' is negative and the curve transferring the origin takes the form

$$y = y_0\,(x-a)^{m_1}/x^{m_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xviii.)}.$$

Now it is possible that this curve falls under the limited range type of a frequency from $x = 0$ to $x = a$, but as we see that the criterion places Type VI. between two curves of range limited in one direction only, we expect Type VI. also to be of that character, and a complete solution is obtained by taking the range from $x = a$ to $x = \infty$; this indeed fills up the gap for $\kappa_2 > 1$ and $< \infty$, and (xviii.) with this range is seen to pass into one or other of the two transition curves

$$y = y_0\,x^p e^{-\gamma x},$$

or

$$y = y_0\,x^{-p} e^{-\gamma/x},$$

according as we allow the first or second factor to approach a limit.†

* The sign of μ_3 will determine the sign of γ, or, what may be taken as the same thing, the direction of the axis of x.

† Write : $y = \text{const.} \times x^{-m_2}\,(1 - x/a)^{m_1}$ and make $m_1 = \infty$, and $a = \infty$ but m_1/a finite.

Or, $y = \text{const.} \times (1 - a/x)^{m_1}/x^{m_2 - m_1}$, and make $a = 0$, $m_1 = \infty$, and $a \times m_1$ together with $m_2 - m_1$ finite.

Accordingly we shall write Type VI. in the form

$$y = y_0 (x - a)^{q_2}/x^{q_1} \quad \ldots\ldots\ldots \quad \text{(xix.)},$$

and take the range from a to ∞.

Differentiating to find the position of the mode we have

$$x_{mo} = \frac{aq_1}{q_1 - q_2} \quad \ldots\ldots\ldots \quad \text{(xx.)}.$$

For the moments about the origin:

$$a\mu'_n = \int_a^\infty y_0 \frac{x^n (x-a)^{q_2}}{x^{q_1}}\, dx.$$

Put $a/x = z$, hence

$$\begin{aligned} a\mu'_n &= \int_0^1 \frac{y_0}{a^{q_1-q_2-n-1}} z^{q_1-q_2-n-2} (1-z)^{q_2}\, dz \\ &= \frac{y_0}{a^{q_1-q_2-n-1}} \mathrm{B}(q_1 - q_2 - n - 1, q_2 + 1) \\ &= \frac{y_0}{a^{q_1-q_2-n-1}} \frac{\Gamma(q_1 - q_2 - n - 1)\,\Gamma(q_2 + 1)}{\Gamma(q_1 - n)}. \end{aligned}$$

Hence we deduce

$$a = \frac{y_0}{a^{q_1-q_2-1}} \frac{\Gamma(q_1 - q_2 - 1)\,\Gamma(q_2 + 1)}{\Gamma(q_1)} \quad \ldots\ldots \quad \text{(xxi.)}.$$

$$\left.\begin{aligned} \mu_1' &= \frac{a(q_1 - 1)}{q_1 - q_2 - 2} \\ \mu_2' &= \frac{a^2 (q_1 - 1)(q_1 - 2)}{(q_1 - q_2 - 2)(q_1 - q_2 - 3)} \\ \mu_3' &= \frac{a^3 (q_1 - 1)(q_1 - 2)(q_1 - 3)}{(q_1 - q_2 - 2)(q_1 - q_2 - 3)(q_1 - q_2 - 4)} \\ \mu_4' &= \frac{a^4 (q_1 - 1)(q_1 - 2)(q_1 - 3)(q_1 - 4)}{(q_1 - q_2 - 2)(q_1 - q_2 - 3)(q_1 - q_2 - 4)(q_1 - q_2 - 5)} \end{aligned}\right\} \quad \ldots \quad \text{(xxii.)}.$$

Now if we compare these results with those on p. 368 of the earlier memoir we see that the one set can be at once deduced from the other by writing $m_1 = -q_1$, $m_2 = q_2$. Thus with this interchange the whole of that solution holds, if we bear in mind that the range is now from $x = a$ to ∞.

We easily find:

$$r = -q_1 + q_2 + 2 \qquad \epsilon = 1 - q_1 + q_2 - q_1 q_2$$

and $1 - q_1$ and $q_2 + 1$ are the roots of

$$z^2 - rz + \epsilon = 0 \quad \ldots\ldots\ldots \text{(xxiii.)},$$

where r and ϵ are to be determined as in that memoir, pp. 368-369.

We have :

$$\mu_2 = \frac{a^2 (1 - q_1)(1 + q_2)}{r^2 (r + 1)} \quad \ldots\ldots\ldots \text{(xxiv.)},$$

where $1 - q_1$ and r are both negative. This gives a.* Thus q_1, q_2 and a are known, and from Equation (xxi.)

$$y_0 = \frac{\alpha\, a^{q_1 - q_2 - 1}\, \Gamma(q_1)}{\Gamma(q_1 - q_2 - 1)\, \Gamma(q_2 + 1)} \quad \ldots\ldots\ldots \text{(xxv.)}$$

we find the remaining unknown constant for the shape of the curve, y_0. As before, various approximations may be used to the values of the Γ functions when either q_1 or q_2 or both are large.†

We easily obtain for the distance between mode and mean

$$d = \frac{a(q_1 + q_2)}{(q_1 - q_2)(q_1 - q_2 - 2)} \quad \ldots\ldots\ldots \text{(xxvi.)},$$

and for the skewness :

$$\text{Sk.} = \frac{(q_1 + q_2)\sqrt{(q_1 - q_2 - 3)}}{(q_1 - q_2)\sqrt{\{(q_1 - 1)(q_2 + 1)\}}} \quad \ldots\ldots\ldots \text{(xxvii.)}.$$

(5.) A special case of some interest arises when the start of the curve is *à priori* known. Suppose its distance from the mean to be c and let (using moments about centroid)

$$\mu_2/c^2 = \gamma_2, \quad \mu_3/(2\mu_2 c) = \gamma_3 \quad \ldots\ldots\ldots \text{(xxviii.)}.$$

Then we easily find :

$$\gamma_2 = \frac{1 - q_1}{(1 + q_2)(-q_1 + q_2 + 3)}, \qquad \gamma_3 = \frac{q_1 + q_2}{(1 + q_2)(q_1 - q_2 - 4)}.$$

---

* $1 - q_1$ being negative, ϵ is negative, and accordingly by what goes before κ_2 lies between 1 and ∞.

† The value of y_0 for curves of Type I., if m_1 be small but m_2 large ('Phil. Trans.,' A, vol. 186, p. 369, foot-note), is

$$y_0 = \frac{\alpha}{b}(m_1 + m_2 + 1)\sqrt{\frac{m_1 + m_2}{m_2}}\, e^{\frac{1}{12}\left(\frac{1}{m_1 + m_2} - \frac{1}{m_2}\right)} \frac{m_1^{m_1} e^{-m_1}}{\Gamma(m_1 + 1)},$$

and this can be easily modified to suit (xxv.) above. A very convenient and exact formula for $\Gamma(n + 1)$, if n be large, is that given by FORSYTH ('B.A. Report,' 1883, p. 47) :

$$\Gamma(n + 1) = \sqrt{2\pi}\left(\frac{\sqrt{n^2 + n + \frac{1}{6}}}{e}\right)^{n + \frac{1}{2}},$$

the error being less than $\frac{1}{240 n^3}$ of the whole.

Whence we deduce to determine q_1 and q_2:

$$q_2 - q_1 = \frac{1 - 3\gamma_2 + 4\gamma_3}{\gamma_2 - \gamma_3} \qquad q_1 + q_2 = \frac{\gamma_3(1 + \gamma_2)(\gamma_2 - 1 - 2\gamma_3)}{(2\gamma_2 - \gamma_3 + \gamma_2\gamma_3)(\gamma_2 - \gamma_3)} \quad . \quad . \quad \text{(xxix.)},$$

and the solution proceeds as before.

(6.) *Illustrations.*—I propose to note a few distributions of frequency in which I have come across Types V. and VI.

(A.) *Statistics of Age of Bride at Marriage, the Bridegroom's Age being between 24 and 25 years.**

The observations given in the table, p. 454, are taken from PEROZZO's memoir: "Nuove Applicazioni del Calcolo delle Probabilità . . . ," 'Reale Accademia dei Lincei,' Anno CCLXXIX., 1881–2, Tavola I.

The total number of recorded marriages is 28,454. The moments were calculated by using SHEPPARD's corrections ('London Math. Soc. Proc.,' vol. 29, p. 369), and are as follows:—

Mean age of bride = 22·1877.

$$\mu_2 = 13{\cdot}3346$$
$$\mu_3 = 67{\cdot}8145$$
$$\mu_4 = 1224{\cdot}6342$$

Whence:

$$\beta_1 = 1{\cdot}9396$$
$$\beta_2 = 6{\cdot}8873$$
$$\kappa_1 = 1{\cdot}9558$$
$$\kappa_2 = 1{\cdot}1094$$

Thus by p. 445 we see that Type VI. is the frequency curve to be selected, but as κ_2 does not differ widely from unity, we shall probably get a good fit from Type V. as well.

Taking Type VI. first, we find:

$$r = -12{\cdot}11075, \qquad \epsilon = -317{\cdot}84987.$$

The quadratic (xxiii.) is accordingly:

$$z^2 + 12{\cdot}11075\, z - 317{\cdot}84987 = 0.$$

* I selected this example at random, as one out of several leading to the curve types it was my object to illustrate. There is so much tampering with statistics, however, whenever they refer to the ages of women, that it would probably have been better to have used the men.

Thus: $q_1 = 25{\cdot}88401, \qquad q_2 = 11{\cdot}77326.$

Hence by (xxiv.) $a = 8{\cdot}268{,}405,$

and by (xxv.) $\log y_0 = 24{\cdot}275{,}3032.$

We have accordingly for the equation to the curve:

$$y = 10^{24} \times 1{\cdot}884{,}965 \frac{(x - 8{\cdot}268{,}405)^{11{\cdot}77326}}{x^{25{\cdot}88401}}.$$

The distance from the origin to the mean is given by the first equation of (xxii.):

$$\mu_1' = 16{\cdot}98913,$$

or, the theoretical range starts with brides of $5{\cdot}198{,}570 + 8{\cdot}268{,}405 = 13{\cdot}466{,}975$ years. This is an excellent underlimit to the age of women marrying men of 24 to 25 in a country like Italy. Our first group is at 15·5, and the above start is just two base units before this initial group.

The skewness = ·498,953, and the distance from mode to mean = 1·822,004, or the mode is at 20·3657 years.

Turning now to Type V. we have the following results :—

$$16/\beta_1 = 8{\cdot}249{,}262.$$

Hence Equation (xii.) is:

$$(p - 4)^2 - 8{\cdot}249{,}262\,(p - 4) - 8{\cdot}249{,}262 = 0.$$

Thus the positive value of p is:

$$p = 13{\cdot}150{,}747.$$

Equation (xiii.) gives:

$$\gamma = 129{\cdot}73081.$$

Then (xiv.) gives:

$$\log y_0 = 22{\cdot}367{,}6952.$$

Thus the equation to the curve is:

$$y = 10^{22} \times 2{\cdot}331{,}821\, x^{-13{\cdot}150{,}747}\, e^{-129{\cdot}73081/x}.$$

To find the position of its start we have by (xv.):

$$\mu_1' = 11{\cdot}6343,$$

or, since the mean age of brides is 22·1877, the youngest possible theoretical bride is 10·5534 years. This is probably a worse determination of the underlimit than in the case of Type VI. At the same time I notice that out of about 180,000 women, 101

were married between 14 and 15 years of age, and all the curves begin with a sensibly finite ordinate at 14·5; it is accordingly possible that a somewhat lower age than 13·5 actually occurs in Italy.

Equation (xvi.) gives us for the distance from mode to mean:

$$d = 1{\cdot}7694,$$

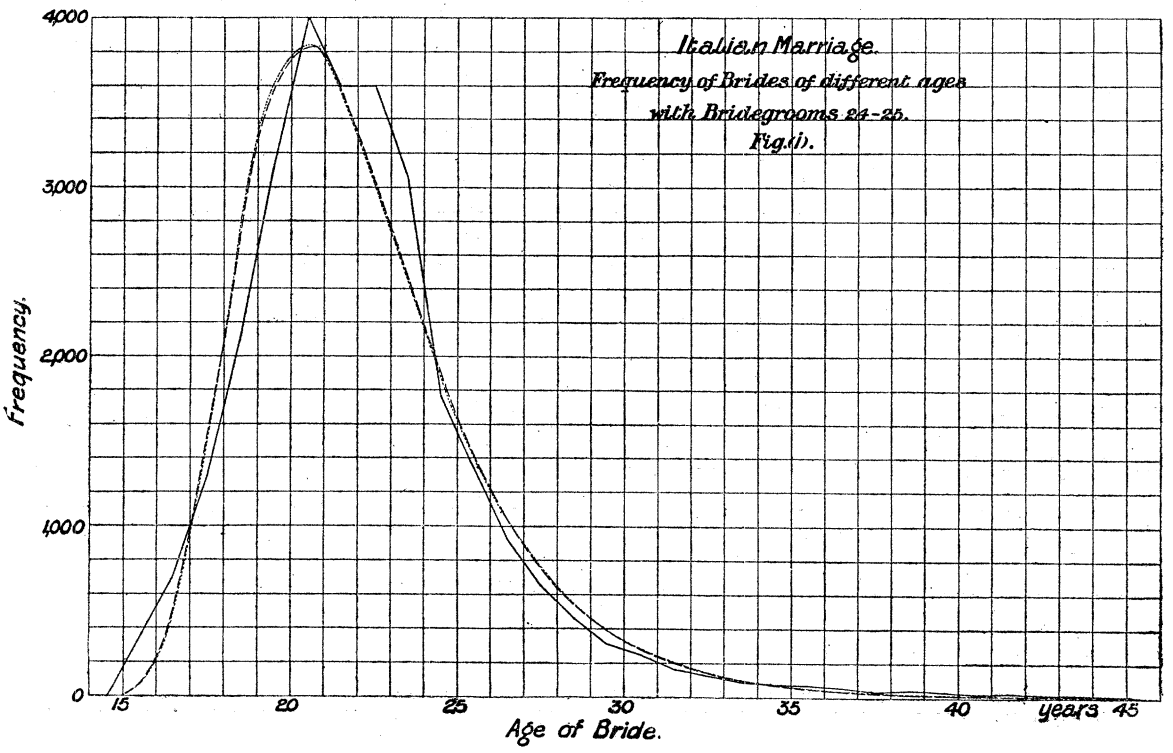
or the modal age at marriage is 20·4183 years. This is only about ·053 of a year or about 19 days different from the modal age as given by Type VI., a most satisfactory agreement.

For the skewness we have from Equation (xvii.):

$$\text{Sk.} = {\cdot}4845,$$

or, it differs by less than 3 per cent. from the skewness as given by Type (VI.).

The diagram (fig. i.) shows the two curves, and the table compares the results obtained from either with the observations.*



Type V. – – – – – , Type VI. . . . . .

It is clear that for all practical purposes the curve of Type V. is as good as that of Type VI. Indeed, there is practically no difference between them except for the

* The observation data are really areas, while to save lengthy calculations we have compared both in diagram and table the ordinates of the theoretical curves. This is in general legitimate, if, as in this case, the number of groups is very large.

ages 15 to 17. The fit is, however, not a very good one, and although it is indefinitely better than a normal curve, and we see why in the absence of these types the statistics could not be fitted with any of the first series of skew curves, yet we are compelled to consider that there are causes other than chance at work very definitely affecting the frequency of the *recorded* ages. Thus the bridegrooms being 24 to 25, the desire of the bride to be recorded as younger than her husband probably fully accounts for the bulk of the preponderance of observation over theory

TABLE of Observed and Calculated Frequencies.

Age.	Observed frequency.	Calculated frequency.		Age.	Observed frequency.	Calculated frequency.	
		Type V.	Type VI.			Type V.	Type VI.
15–16	367	70	49	30–31	256	281	282
16–17	717	514	489	31–32	164	201	198
17–18	1294	1538	1560	32–33	134	148	146
18–19	2121	2751	2800	33–34	94	104	105
19–20	3156	3591	3622	34–35	77	75	76
20–21	4009	3830	3831	35–36	68	55	55
21–22	3593	3577	3560	36–37	59	40	40
22–23	3604	3055	3034	37–38	33	29	29
23–24	3060	2456	2439	38–39	40	21	22
24–25	1774	1894	1884	39–40	27	16	16
25–26	1353	1419	1415	40–41	18	12	12
26–27	936	1044	1043	41–42	21	9	9
27–28	663	758	760	42–43	11	7	7
28–29	468	546	549	43–44	14	5	5
29–30	319	392	395	44–45	4	4	4

in the frequency of the brides of 22 to 24. The defect of brides between 17 and 20 may be again due to the tendency to state the age as over 21, and so free the woman from the need for parental sanction.* These causes, giving a false displacement of age frequency, are probably in themselves sufficient to account for the theoretical defect in brides of 15 to 17.

(7.) (B.) *On the Variation in the Number of Lips of the Medusa* P. Pentata.

My data are the following, taken from a paper by ALFRED GOLDSBOROUGH MAYER: "The Variations of a Newly Arisen Species of Medusa," 'Science Bulletin of the Museum of the Brooklyn Institute,' vol. 1, p. 1, 1901.

* I have found in England the statement of the bride's age in the marriage licence is for the same reason occasionally not in accordance with the year of birth as shown by the parish register.

	Frequency.	No. of lips
	2	1
	5	2
	18	3
	123	4
	798	5
	49	6
	1	7
Total	996	

Mr. MAYER (p. 12) notes the failure of my curve of Type IV. I find for the constants :

$$\text{Mean} = 4{\cdot}8685 \text{ lips.}$$
$$\mu_2 = {\cdot}309{,}006, \quad \sigma = {\cdot}55588$$
$$\mu_3 = -{\cdot}350{,}697$$
$$\mu_4 = 1{\cdot}181{,}718$$
$$\beta_1 = 4{\cdot}16834$$
$$\beta_2 = 12{\cdot}37598$$
$$\kappa_1 = 6{\cdot}24694$$
$$\kappa_2 = 1{\cdot}06594$$

Since κ_2 is so nearly unity we may use Type V.
Hence I find :

$$p = 8{\cdot}66184 \qquad \gamma = -8{\cdot}811634$$

(γ must be negative since μ_3 is negative)

$$\mu_1' = 1{\cdot}32270.$$

Thus the curve starts at 6·19118 lips, or the one medusa with seven lips is theoretically excluded. Here I have worked with the uncorrected moments because the lips are discontinuous variants. Working with SHEPPARD'S corrective terms the limit is about six lips, and with the corrective terms suggested in my memoir on skew variation the limit is 7·65. Further we have :

$$\log y_0 = 6{\cdot}829{,}3633,$$
$$\text{distance from mean to mode} = {\cdot}30541,$$
$$\text{Sk.} = {\cdot}54941.$$

The mode is thus at 5·17389, in good agreement with observation.

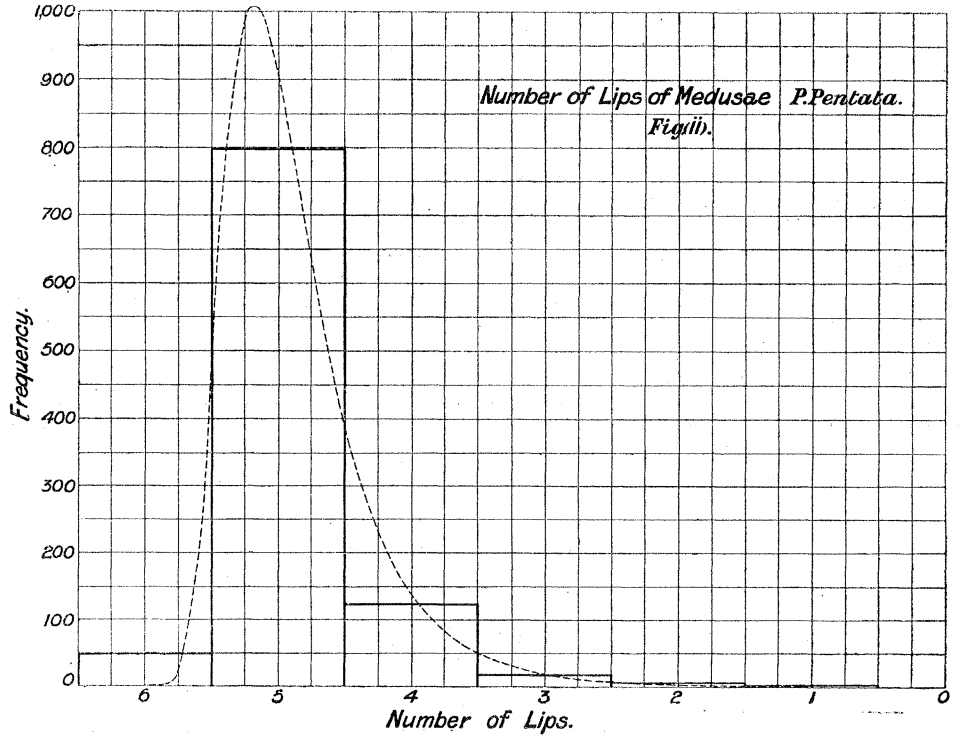
The equation to the curve is, taking x *positive* from 6·19118 lips towards lesser values :

$$\log y = 6{\cdot}829{,}3633 - 8{\cdot}66184 \log x - \frac{3{\cdot}826{,}8435}{x}.$$

This curve was drawn on a large scale and its areas read off with an integrator. The following theoretical frequencies were obtained:

No. of lips.	Observation.	Calculation.
6 and over	50	47
5	798	762
4	123	160·5
3	18	20
2	5	5
1	2	1·5

There would not be any serious divergence here, were it not for the group with four lips, which observation shows to be much under-represented. But it must be remembered that we have only seven groups, and that such a number is very insufficient for a good determination of the moments of a curve. Further, the variation is not really continuous, as indicated by the curve, but *discrete*. We have at present no clear statement as to how the moments of a discrete system of variation should be modified or corrected so as to give the best results for the moments of the continuous curve which is to theoretically represent the series. I am doubtful



whether SHEPPARD'S corrections—the best for continuous variation—are equally appropriate in this case. Above I have used merely the rough moments, but I

have found by considerable experience that in the case of discrete variables, to treat the system as a polygon and correct, as in my memoir on Skew Variation ('Phil. Trans.,' A, vol. 186, p. 350), appears to give the best results when the areas are compared with the discrete groups. The point wants further investigation; when we have a large number of groups it is of little importance, but it makes a considerable difference in these excessively skew distributions of discrete variables when the number of groups are small.*

Above all, the diagram (fig. ii.) shows how all important it is to compare *areas* and not merely the *ordinates* of the frequency curve with the blocks representing the discrete frequencies in such a case as this. The wide-spread custom among foreign investigators of comparing merely the ordinates of the theoretical frequency curve with the observed frequencies leads in such cases to most fallacious results.

(8.) (C.) *On the Distribution of Incidence of Scarlet Fever Cases with Age.*

It seems desirable to give an illustration of the method of dealing with a distribution which falls under the class dealt with in Section (5) of this paper. Dr. MACDONELL, in dealing with the intensity of incidence of different diseases at various ages, has come across in scarlet fever a good illustration of curves of the types now under consideration. The whole of the arithmetical work on the present example is due to him, and I have to thank him very heartily for allowing me to use it here.

The statistics are taken from the 'Report of the Metropolitan Asylums Board' (Statistical Part, 1899). They involve 39,253 *male* cases, distributed as follows:—

Year of life.	Frequency.	Year of life.	Frequency.
Under 1	443	20–25	926
1–2	1456	25–30	420
2–3	2631	30–35	215
3–4	3599	35–40	91
4–5	3862	40–45	45
5–10	15791	45–50	26
10–15	7359	50–55	17
15–20	2366	55–60	5
		60–65	1

The data being grouped partly in one and partly in five-year periods the moments had to be calculated with caution, separating the material into two pieces. Taking five years as the unit, Dr. MACDONELL found for the uncorrected moments:

* *E.g.*, petals of buttercups, teeth on the carapace of prawns, lips of medusæ, as compared with veins on chestnut leaves, florets on ox-eyed daisy, &c.

Mean age of incidence, 8·60975 years.

$$\mu_2 = 1{\cdot}369{,}345$$
$$\mu_3 = 3{\cdot}233{,}194$$
$$\mu_4 = 19{\cdot}143{,}575.$$

The moments were not modified by SHEPPARD'S corrections, for these suppose contact of a high order at both terminals of the curve, and it was quite apparent that the curve must rise at a finite angle on the birth side. The following additional constants were then determined:—

$$\beta_1 = 4{\cdot}071{,}222, \qquad \beta_2 = 10{\cdot}209{,}333,$$
$$\kappa_1 = 2{\cdot}205{,}000, \qquad \kappa_2 = 2{\cdot}813{,}783.$$

Thus κ_2 is >1 and $<\infty$ and the distribution is of Type VI. Now let us suppose the incidence of scarlet fever to start with birth, although there might, as in the case of enteric fever, be really some antenatal cases.*

Turning to Section (5) we have:

$$c = \text{distance from birth to mean} = 8{\cdot}60975 \text{ years} = 1{\cdot}72195 \text{ units}.$$

Hence we deduce

$$\gamma_2 = {\cdot}461{,}819, \qquad \gamma_3 = {\cdot}685{,}596.$$

And so from (xxix.)

$$q_2 - q_1 = -10{\cdot}532{,}485, \qquad q_2 + q_1 = 15{\cdot}417{,}281;$$

or,

$$q_1 = 12{\cdot}974{,}883, \qquad q_2 = 2{\cdot}442{,}398.$$

Then from

$$c = \mu_1' - a = a\,(q_2+1)/(q_1 - q_2 - 2)$$

we find

$$a = 4{\cdot}268{,}104,$$

and, finally, after determining y_0 from (xxi.),

$$\log y_0 = 13{\cdot}652{,}5078.$$

Thus the values of the frequency are given by

$$\log y = 13{\cdot}652{,}5078 + 2{\cdot}442{,}398 \log (x - 4{\cdot}268{,}104) - 12{\cdot}974{,}883 \log x.$$

The origin of the curve is thus 4·268,104 before birth. The mode is given by

$$x_{\text{mo}} = aq_1/(q_1 - q_2) = 5{\cdot}257{,}842.$$

Thus: $x_{\text{mo}} - a = {\cdot}989{,}738 = 4{\cdot}94869$ yrs.

* See 'Phil. Trans.,' A, vol. 186, p. 390. The remarkably sharp rise of the scarlet-fever distribution as compared with the enteric is, however, much against this.

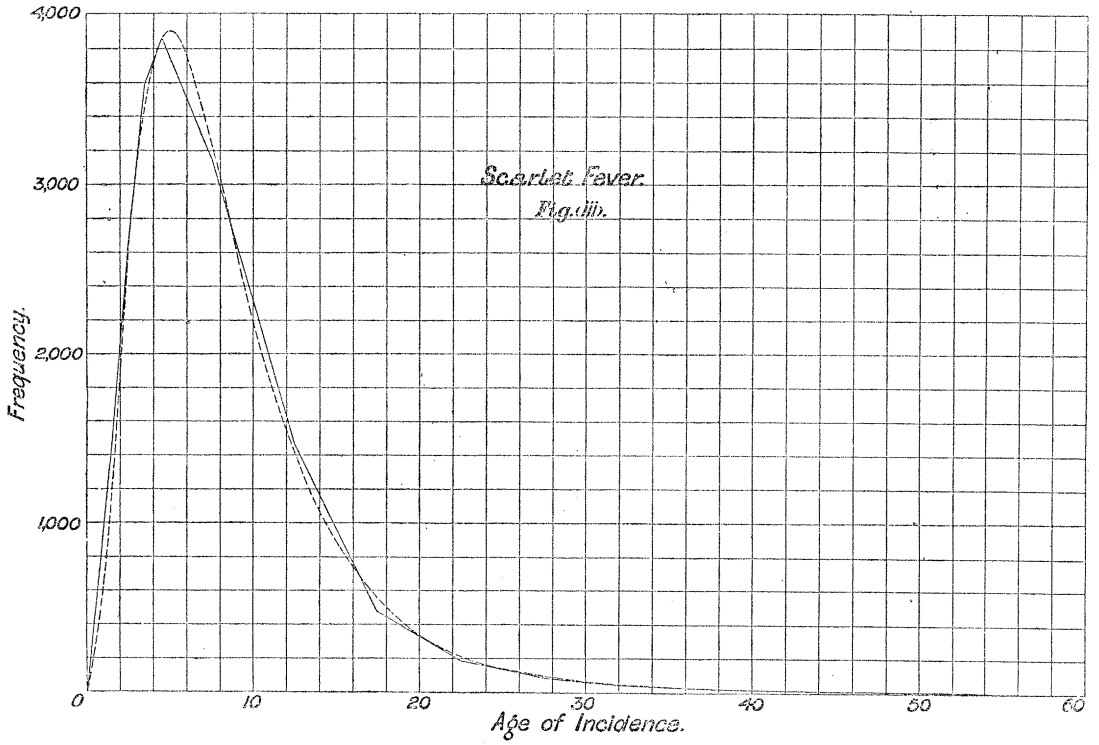
This gives for y_{mode} the value 3892.

Distance between mode and mean = 3·66106 yrs.

Whence we find for skewness the value

$$\text{Sk.} = \cdot 5347.$$

The diagram (fig. iii.) shows that the fit may be considered a good one.



(9.) The conclusions of this paper are, I think, of some interest from the general standpoint of scientific investigation. A certain number of frequency distributions had been found, not only by my co-workers and myself here, but by biologists in America, not to fit into the general system of skew distributions dealt with by me in my original memoir. The first conclusion was that however wide-reaching that system appeared to be, it was a failure for a few remarkably skew distributions. But on more careful investigation of the differential equation it appeared that two types of solution had been left out of consideration, and that these were precisely those needed in the recorded cases of failure.

I owe some apology to authors like Professor DAVENPORT and Dr. DUNCKER, who have recently issued text-books on the application of statistical methods to biological variation, because although we have known and used these curves for some years past, no account has hitherto been published of them, and, consequently, biological investigators* using their *résumés* of my methods have been, and I fear still may be, occasionally puzzled.

* *E.g.*, Mr. A. G. MAYER in the paper on Medusæ referred to above