PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Mathematical Contributions to the Theory of Evolution. XIX. Second Supplement to a Memoir on Skew Variation

Karl Pearson

Phil. Trans. R. Soc. Lond. A 1916 **216**, 429-457
doi: 10.1098/rsta.1916.0009

IX. *Mathematical Contributions to the Theory of Evolution.—XIX. Second Supplement to a Memoir on Skew Variation.*

By KARL PEARSON, *F.R.S.*

Received February 2,—Read February 24, 1916.

[PLATE 1.]

(1) IN a memoir presented to the Royal Society in 1894, I dealt with skew variation in homogeneous material. The object of that memoir was to obtain a series of curves such that one or other of them would agree with any observational or theoretical frequency curve of positive ordinates to the following extent:—(i) The areas should be equal; (ii) the mean abscissa or centroid vertical should be the same for the two curves; (iii) the standard deviation (or, what amounts to the same thing, the second moment coefficient) about this centroid vertical should be the same, and (iv) to (v) the third and fourth moment coefficients should also be the same. If μ_s be the s^{th} moment coefficient about the mean vertical, N the area, $\bar{x}$ be the mean abscissa, $\sigma = \sqrt{\mu_2}$ the standard deviation, $\beta_1 = \mu_3^2/\mu_2^3$, $\beta_4 = \mu_4/\mu_2^2$, then the equality for the two curves of N, $\bar{x}$, σ, β_1 and β_2 leads almost invariably in the case of frequency to excellency of fit. Indeed, badness of fit generally arises from either heterogeniety, or the difficulty in certain cases of accurately determining from the data provided the true values of the moment coefficients, *e.g.*, especially in J- and U-shaped frequency distributions, or distributions without high contact at the terminals; here the usual method of correcting the raw moments for sub-ranges of record fails.

Having found a curve which corresponded to the skew binomial in the same manner as the normal curve of errors to the symmetrical binomial with finite index, it occurred to me that a development of the process applied to the hypergeometrical series would achieve the result I was in search of, *i.e.*, a curve whose constants would be determined by the observational values of N, $\bar{x}$, σ, β_1 and β_2.

The hypergeometrical series was one not only arising naturally in chance problems, but covering in itself a most extensive range of functions. The direct advantage of the hypergeometrical series is that it abrogates the fundamental axioms on which the Gaussian frequency is based. The equality in frequency of plus and minus errors of the same magnitude is replaced by an arbitrary ratio, the number of contributory

 [Published June 6, 1916.

causes is no longer indefinitely large, and the contributions of these causes are no longer independent but correlated.*

Since β_1 and β_2 are by nature positive we can represent all possible values of β_1 on a chart in which β_1 and β_2 are the co-ordinates of a point in the positive quadrant. But a little consideration shows that β_2 must be greater than β_1, thus one-half the area of the quadrant, that above the line $\beta_2 = \beta_1$ is removed from the field of possible occurrences. Further, there is a limit to the application of the series of curves discussed when β_2 gets large, for the high moments of two of the types of curves, *i.e.*, Types IV. and VI., or

$$y = y_0 \frac{e^{-\nu \tan^{-1} x/a}}{\left(1+\dfrac{x^2}{a^2}\right)^{\frac{1}{2}(r+2)}} \quad \text{and} \quad y = y_0 \frac{(x-a)^{q_2}}{x^{q_1}},$$

become infinite when the order of the moment is greater than r, or the probable error of the fourth moment would become indefinitely large for $r = 7$, *i.e.*, we are practically limited by the line $8\beta_2 - 15\beta_1 - 36 = 0$. The first four moments of the curve remain finite, but from the fifth onwards they can become infinite, the lines corresponding to these, however, lying outside the above line.† For curves corresponding to points below this line it is fitting to take as differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{b+x}{c_0+c_1x+c_2x^2+c_3x^3}, \quad \ldots \ldots \ldots \quad \text{(i)}$$

or a slightly more general form which is related to the higher hypergeometrical $F(\alpha, \beta, \gamma, \theta, \epsilon, 1)$ as the present series of curves to the simple hypergeometrical $F(\alpha, \beta, \gamma, 1)$. The whole theory of curves of the above type has been worked out for some time past, but has remained unpublished, for we failed to find any definitely homogeneous data by which it could be effectively illustrated, and for this reason *heterotypic* curves have for the time being been left in abeyance. We may, however, notice the following point. If we take our generalised hypergeometrical to be

$$1 + \frac{\alpha \cdot \beta \cdot \gamma}{\theta \cdot \epsilon \cdot \zeta} + \frac{(\alpha+1)(\beta+1)(\gamma+1)\,\alpha \cdot \beta \cdot \gamma}{(\theta+1)(\epsilon+1)(\zeta+1)\,\theta \cdot \eta \cdot \zeta} + \ldots$$

$$= y_0 + y_1 + y_2 + \ldots .$$

Then

$$\frac{y_{x+1}}{y_x} = \frac{(\alpha+x)(\beta+x)(\gamma+x)}{(\theta+x)(\epsilon+x)(\zeta+x)},$$

and this will correspond to the ordinary form if $\zeta = 0$, *i.e.*, $F(\alpha, \beta, \gamma, \theta, \epsilon, 1)$.

* Just as values of the binomial $(p+q)^n$ with negative n and $p>1$ very often give good fits to frequency distributions, so we have recently found that hypergeometricals $F(\alpha, \beta, \gamma, 1)$ with imaginary α and β are of fairly common occurrence in frequency distributions, and when applied to individual samples from real hypergeometrical populations may give better fits than the theoretical series, *i.e.*, in card drawings.

† See RHIND, 'Biometrika,' vol. VII., p. 133.

We have

$$\frac{y_{x+1}-y_x}{\frac{1}{2}(y_{x+1}+y_x)} = \frac{2\{\alpha\beta\gamma-\theta\epsilon\zeta+x(\alpha\beta+\beta\gamma+\gamma\alpha-\theta\epsilon-\epsilon\zeta-\zeta\theta)+x^2(\alpha+\beta+\gamma-\theta-\epsilon-\zeta)\}}{\{\alpha\beta\gamma+\theta\epsilon\zeta+x(\alpha\beta+\beta\gamma+\gamma\alpha+\theta\epsilon+\epsilon\zeta+\zeta\theta)+x^2(\alpha+\beta+\gamma+\theta+\epsilon+\zeta)+2x^3\}},$$

and accordingly we get the curve approximating to the hypergeometrical of the higher order by putting

$$\frac{1}{y}\frac{dy}{dx} = \frac{\text{quadratic function of } x}{\text{cubic function of } x},$$

$$= \frac{a_0+a_1x+a_2x^2}{c_0+c_1x+c_2x^2+c_3x^3}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(ii)}$$

where the six independent constants can be expressed in terms of the original six, α, β, γ, θ, ϵ, ζ. It will be seen that a hypergeometrical of the second order will, in general, have *two* modes, the exception being when

$$\alpha+\beta+\gamma = \theta+\epsilon+\zeta; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(iii)}$$

in which case (ii) coincides with (i) the general equation to the fourth approximation of curves when β_1 and β_2 fall into the heterotypic area. It will thus be noted that such curves approximate to hypergeometric series of the second order when the special condition (iii) holds; always assuming the unimodal character of homogeneous material. It seems probable that for the most part bimodal frequencies would be those that lead to values of β_1 and β_2 lying in the heterotypic region, and such are excluded from practical statistics.

In the original paper* four types of curves were dealt with beside the Gaussian curve corresponding to an isolated point. A supplementary memoir issued in 1901† dealt with two further types, which had been overlooked until actual experience demonstrated their existence. I have now to confess the omission of five further types, not to speak of a horizontal straight line, as sub-groups of the J-section of curves, which are themselves in practice so rare, that the region of the β_1, β_2 plane in which they occur had not been very fully investigated. My attention was drawn to these curves while considering the frequency curves for the correlation of small samples. If we take a sample of four from uncorrelated material, the sample is equally likely to have every correlation from -1 to $+1$.‡ In this case, $\beta_1 = 0$, $\beta_2 = 1{\cdot}8$, and the frequency curve is a horizontal straight line. What would my series of curves give in this case? I discovered that they also gave a rectangle of frequency or a horizontal straight line, and this discovery led me to a closer investigation of the sub-groups of curves in the neighbourhood of the J-curve area. The point in the

* 'Phil. Trans.,' A, vol. 186 (1895), pp. 343–414.

† 'Phil. Trans.,' A, vol. 196 (1901), pp. 443–459.

‡ 'Biometrika,' vol. VI., p. 306, and vol. X., p. 312.

$\beta_1\,\beta_2$ plane for which $\beta_1 = 0$, $\beta_2 = 1{\cdot}8$, I term the *rectangle-point* and denote by R. (See folding diagram, Plate 1, at end of paper.)

The rectangle-point is the point of contact with the axis of β_2 of the biquadratic

$$\beta_1(8\beta_2-9\beta_1-12)(\beta_2+3)^2 = (4\beta_2-3\beta_1)(10\beta_2-12\beta_1-18)^2,$$

which bounds the area of J-curves. The novel curves are in part limiting curves which occur when the point β_1, β_2 lies on this biquadratic, *i.e.*, transition curves from J-curves to U-curves and from J-curves to limited range curves, and in part a limiting curve which exists along the line $5\beta_2-6\beta_1-9 = 0$ which passes through the rectangular point and never again meets the biquadratic in the loop in the positive quadrant. It would be convenient to speak of this line as the axis of the biquadratic loop, but unfortunately the loop is not symmetrical about it, and to avoid misunderstanding I term it the R-line.

Up to the present the minimum limit to the area of U-curves had not been given. Since β_2 is $> \beta_1$, half the positive quadrant was impossible, but a recent observation shows that frequency curves above the line $\beta_2-\beta_1-1 = 0$ are impossible. This limit was suggested in the following manner. When samples of three are taken from an indefinite population, the frequency curves for the correlation of any two variates of the three individuals sampled are U-shaped frequency curves, but when samples of two are taken the correlation must be either positive or negative, and accordingly the frequency is collected into two lumps or blocks as a limiting case of a U-shaped distribution. But for two such lumps $\beta_2-\beta_1-1 = 0$. In other words, along the line $\beta_2-\beta_1-1 = 0$, the U-shaped frequency either brings all frequency to an end, or passes through a transitional case. The former is the true state of affairs, for β_2 cannot be less than β_1+1. To demonstrate this,* let $s_p = \mathrm{S}(x_n^{\,p})$, and let there be n quantities x_u. Clearly, $s_0 = n$, and $s_1 = 0$. Now by BURNSIDE and PANTON, 'Theory of Equations,' vol. II., p. 35,

$$\sum_{\substack{r>s>t\\ r,s,t=1}}^{n}\{(x_s-x_t)^2(x_t-x_r)^2(x_r-x_s)^2\} = \begin{vmatrix} s_0, & s_1, & s_2 \\ s_1, & s_2, & s_3 \\ s_2, & s_3, & s_4 \end{vmatrix}$$

$$= \begin{vmatrix} n, & 0, & s_2 \\ 0, & s_2, & s_3 \\ s_2, & s_3, & s_4 \end{vmatrix} = n(s_2s_4-s_3^2)-s_2^3$$

$$= s_2^3\left(\frac{s_4/n}{s_2^2/n^2}-\frac{s_3^2/n^2}{s_2^3/n^3}-1\right) = s_2^3\left(\frac{\mu_4}{\mu_2^2}-\frac{\mu_3^2}{\mu_2^3}-1\right)$$

$$= s_2^3(\beta_2-\beta_1-1),$$

* I owe this neat proof to the kindness of Mr. G. N. WATSON.

which must therefore be either zero or a positive quantity. Thus we see that the whole area covered by my frequency curves is limited above by the line $\beta_2 - \beta_1 - 1 = 0$, and below by the line $8\beta_2 - 15\beta_1 - 36 = 0$. The first line limits all frequency; the second line limits my types.*

(2) Before proceeding further, let us examine the limit to all frequency. Consider the line $\beta_2 - \beta_1 - 1 = 0$.

The form of the curve is†

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}.$$

Now,

$$m'^2 - rn' + \epsilon = 0,$$

where

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

therefore

$$r = 0 \quad \text{and} \quad \epsilon = \tfrac{1}{4} r^2 / (1 - \kappa_2) \text{ also} = 0.$$

Hence

$$m'_1 + m'_2 = 0 \text{ and } m'_1 m'_2 = 0, \quad \text{or} \quad m_1 = -1,\ m_2 = -1.$$

The form of the curve is accordingly

$$y = \frac{y_0}{\left(1 + \dfrac{x}{a_1}\right)\left(1 - \dfrac{x}{a_2}\right)},$$

or, apparently, U-shaped. Now

$$b = \tfrac{1}{2}\sigma \{\beta_1 (r+2)^2 + 16(r+1)\}^{1/2}$$
$$= \sigma \{\beta_1 + 4\}^{1/2},$$

and is finite. But

$$y_0 = \frac{\mathrm{N}}{b} \frac{m_1^{m_1} m_2^{m_2}}{(m_1+m_2)^{m_1+m_2}} \frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\,\Gamma(m_2+1)},$$

$$= \frac{\mathrm{N}}{b} \frac{(m_1+1)(m_2+1)}{m_1+m_2+2} \frac{m_1^{m_1} m_2^{m_2}}{(m_1+m_2)^{m_1+m_2}} \frac{(m_1+2)(m_2+2)}{(m_1+m_2+3)} \times \frac{\Gamma(m_1+m_2+4)}{\Gamma(m_1+3) \times \Gamma(m_2+3)},$$

$$= \frac{\mathrm{N}}{b} \text{ limit of } \frac{(m_1+1)(m_2+1)}{m_1+m_2+2} \frac{4 \times \Gamma(2)}{\Gamma(2) \times \Gamma(2)}.$$

But

$$\text{limit of } \frac{(m_1+1)(m_2+1)}{m_1+m_2+2} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2},$$

* It is not accurately correct to say it limits my types of skew curves. What it actually does is to cut off an area in which the probable errors of the constants of Types IV. and VI. curves can be very great. The curves may give a good fit, but the constants cannot be cited as characteristics of the frequency distribution as they are unstable.

† The notation throughout is that of my original 'Phil. Trans.' memoirs of 1895 and 1901.

when both $\lambda_1\lambda_2$ are to be made vanishingly small, being m_1+1 and m_2+1 respectively. Thus the limit

$$= \frac{1}{1/\lambda_1+1/\lambda_2} = \frac{1}{\infty} = 0.$$

Hence y_0 vanishes or y is zero at all points, but $x = -a_1$ and $x = a_2$ where it is undetermined.

Since $m_1/a_1 = m_2/a_2$, we have $a_1 = a_2$, and the frequency really consists of two concentrated groups at $-a_1$ and a_2, or at $\pm\frac{1}{2}b$.

If μ'_1 and μ''_1 be the distances of the centroid from the two ends of the range,

$$\frac{\mu'_1}{\mu''_1} = \frac{n''}{n'},$$

where n' and n'' are the frequencies concentrated at the range terminals. But $\mu'_1 = b\,(m_1+1)/(m_1+m_2+2)$, or we have $\mu'_1/\mu''_1 = (m_1+1)/(m_2+1) = \lambda_1/\lambda_2$, or is the finite quantity which marks the ratio of the vanishing of m_1+1 and m_2+1; this, therefore, is equal to n''/n'.

Clearly

$$\mu'_1 = (n''-n')/(n''+n')\,\frac{b}{2},$$

$$\mu'_2 = (n''+n')/(n''+n')\,\frac{b^2}{4} = \tfrac{1}{4}b^2,$$

$$\mu'_3 = (n''-n')/(n''+n')\,\frac{b^3}{8},$$

$$\mu'_4 = (n''+n')/(n''+n')\,\frac{b^4}{16} = \tfrac{1}{16}b^4,$$

and

$$\mu_2 = b^2\,(n'n'')/(n'+n'')^2,$$

$$\mu_3 = b^3 n'n''(n''-n')/(n'+n'')^3,$$

$$\mu_4 = b^4 n'n''\,(n'^2+n''^2-n'n'')/(n'+n'')^4.$$

Thus

$$\beta_1 = \frac{n''}{n'}+\frac{n'}{n''}-2, \quad \beta_2 = \frac{n''}{n'}+\frac{n'}{n''}-1,$$

giving as verification $\beta_2-\beta_1-1 = 0$.

Thus the whole problem is solved if we know the magnitude of the two frequencies n' and n'' concentrated at $-\frac{1}{2}b$ and $+\frac{1}{2}b$.

As special cases the point on the β_2-axis gives $\beta_1 = 0$, $\beta_2 = 1$, and represents two equal concentrated frequency lumps $n' = n'' = \frac{1}{2}\mathrm{N}$. The point at ∞ on the line $\beta_2-\beta_1-1 = 0$, or $\beta_1 = \beta_2 = \infty$ represents a single frequency lump, for which $n' = 0$, $n'' = \mathrm{N}$. I speak of these concentrated frequency lumps lying on the line $\beta_1-\beta_2-1 = 0$, as block-frequency, and represent them by the letter B; they correspond to points on the B-line. (See Diagram, Plate 1.)

The most remarkable limiting case of this kind has been already referred to. It will be shown in practical examples in a memoir on "small samples," now nearly ready for press, that the correlation between two variates may be determined by sampling these populations in pairs, and merely observing, which can be usually done without measurement, whether the pair is positively or negatively correlated. The ratio of the two frequency "lumps" easily provides the correlation.*

(3) Let us now consider the nature of the frequency on the loop of the biquadratic. Taking the form of the curve to be

$$y = y_0 (1+x/a_1)^{m_1} (1-x/a_2)^{m_2}$$

we know that m_1 and m_2 are the roots of the quadratic

$$m^2 - m(r-2) + \epsilon - r + 1 = 0,$$

where

$$r = 6(\beta_2 - \beta_1 - 1)/(3\beta_1 - 2\beta_2 + 6),$$

and

$$\epsilon = \frac{r^2}{4 + \frac{1}{4}\beta_1 (r+2)^2/(r+1)}.$$

Now $\epsilon - r + 1 = 0$ provides the biquadratic

$$\beta_1 (8\beta_2 - 9\beta_1 - 12)(\beta_2 + 3)^2 - (10\beta_2 - 12\beta_1 - 18)^2 (4\beta_2 - 3\beta_1) = 0;$$

actually

$$\epsilon - r + 1 = \frac{(4\beta_2 - 3\beta_1)(10\beta_2 - 12\beta_1 - 18)^2 - \beta_1 (\beta_2 + 3)^2 (8\beta_2 - 9\beta_1 - 12)}{(3\beta_1 - 2\beta_2 + 6)\{\beta_1 (\beta_2 + 3)^2 + 4\beta_1 (4\beta_2 - 3\beta_1)(3\beta_1 - 2\beta_2 + 6)\}}.$$

Now β_1, $4\beta_2 - 3\beta_1$ and $\beta_2 + 3$ are by their nature essentially positive. Hence, provided $3\beta_1 - 2\beta_2 + 6$ is positive, *i.e.*, as long as we deal with points above the line $2\beta_2 - 3\beta_1 - 6 = 0$, *i.e.*, the Type III. curve line, $\epsilon - r + 1$ will be positive, if (β_1, β_2) lie outside the loop of the biquadratic. But within the loop it is negative, or one value of m must be negative, or we reach an infinite ordinate at $x = -a_1$ or a_2, *i.e.*, a J-shaped curve. The other ordinate at $x = a_2$ or $-a_1$ is zero, because the other m must be a finite positive quantity.

If $\epsilon - r + 1 = 0$, *i.e.*, along the biquadratic loop, one value of m is zero, and the other is positive if r be greater than 2, and negative if it be less than 2. But

$$r - 2 = \frac{2(5\beta_2 - 6\beta_1 - 9)}{3\beta_1 - 2\beta_2 + 6}.$$

Accordingly above the line $5\beta_2 - 6\beta_1 - 9 = 0$, and above the line $2\beta_2 - 3\beta_1 - 6 = 0$, $r - 2$ will be negative, but these lines do not meet in the positive quadrant. Hence all

* See "STUDENT," 'Biometrika,' vol. VI., p. 304, and FISHER, 'Biometrika,' vol. X., p. 508.

along the upper boundary of the loop one m is zero and the other negative. Accordingly, from the R-point round the upper boundary of the loop, we have the curve

$$y = y_0 (1+x/a_1)^{-m'_1}.$$

I call this curve Type VIII.

Since $-m'_1/a_1 = m_2/a_2$, and m_2 is zero while m_1 and a_1 are finite, it follows that $a_2 = 0$, and accordingly the range of frequency is from $x = 0$ to $x = -a_1$. The curve is therefore a J-shaped curve with infinite ordinate at one end of the range and a finite ordinate at the other.

Now consider the lower side of the loop. Here $5\beta_2 - 6\beta_1 - 9$ will be positive, for this side is below the R-line and $3\beta_1 - 2\beta_2 + 6$ will also be positive until the point in which the line $2\beta_2 - 3\beta_1 - 6 = 0$ meets the lower side of the loop, *i.e.*, the point $\beta_1 = 4$, $\beta_2 = 9$. Hence from the R-point up to $\beta_1 = 4$, $\beta_2 = 9$, a point practically outside the range of the customary statistical frequencies, $r-2$ will be positive, or m_1 will be positive. Further m_1 and a_1 being finite and m_2 zero, it follows that a_2 is zero, or the curve is

$$y = y_0 (1+x/a_1)^{m_1}.$$

In this case the curve has a zero ordinate at one end and a finite ordinate at the other. I term this curve Type IX.

At the point where the line $2\beta_2 - 3\beta_1 - 6 = 0$ meets the biquadratic, Type IX. agrees with my earlier Type III.

The equation to that type is*

$$y = y_0 (1+x/a)^{\gamma a} e^{-\gamma x},$$

where

$$\gamma a = \frac{4}{\beta_1} - 1 \quad \text{and} \quad \gamma = \frac{2}{\sigma\sqrt{\beta_1}}\cdot$$

Hence for $\beta_1 = 4$, $\gamma a = 0$, and $\gamma = 1/\sigma$. Thus a is zero and the curve becomes

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

the range being from 0 to ∞.

But in Type IX., since r has become infinite, m_1 is infinite and the limit to

$$y = y_0 (1+x/a_1)^{m_1}$$

is accordingly the exponential curve

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

as we shall see shortly λ must equal $1/\sigma$, where σ is the standard deviation.

I propose to call this exponential curve Type X., and the point $\beta_1 = 4$, $\beta_2 = 9$, E or the exponential point.

* 'Phil. Trans.,' A, vol. 186, p. 373.

Beyond the exponential point, our biquadratic branch has entered the area of Type VI. curves,* and m_1 will now again be negative.

Now the equation to Type VI. is

$$y = y_0 \frac{(x-a)^{q_2}}{x^{q_1}},$$

and the range from $x = a$ to ∞. The special case of this along the branch of the biquadratic occurs when $q_2 = 0$, leading to†

$$1 - q_1 = \epsilon_1 = r - 1,$$

or

$$y = y_0/x^{q_1}$$

where

$$q_1 = \frac{2(5\beta_2 - 6\beta_1 - 9)}{2\beta_2 - 3\beta_1 - 6},$$

which is positive, since q_1 is now beneath both the lines

$$5\beta_2 - 6\beta_1 - 9 = 0 \quad \text{and} \quad 2\beta_2 - 3\beta_1 - 6 = 0. \ddagger$$

This curve, which will be more fully considered below, has a range from a certain value a to ∞. It thus starts with a finite ordinate and asymptotes to zero. It is a transition curve extending from the exponential point along the lower limb of the biquadratic loop. I call this curve Type XI. The biquadratic never cuts the cubic along which Type V. lies and no further change occurs in Type XI.

I now pass to the consideration of the R-line or $5\beta_2 - 6\beta_1 - 9 = 0$.

The general differential equation§ to the type of frequency curve under consideration is

$$\frac{1}{y}\frac{dy}{dx} = \frac{-\{\sqrt{\beta_1}(\beta_2+3) + (10\beta_2 - 12\beta_1 - 18)x/\sigma\}}{\sigma\{4\beta_2 - 3\beta_1) + \sqrt{\beta_1}(\beta_2+3)x/\sigma + (2\beta_2 - 3\beta_1 - 6)x^2/\sigma^2\}},$$

the origin being *at the mean.*

Hence if $5\beta_2 - 6\beta_1 - 9 = 0$, the term in x/σ disappears from the numerator, and we can further get rid of β_2 by substituting $\frac{1}{5}(6\beta_1 + 9)$ for it. Making this substitution, we reach

$$\frac{1}{y}\frac{dy}{dx} = \frac{-2\sqrt{\beta_1}}{\sigma(3+\beta_1) - \sigma(\sqrt{\beta_1} - x/\sigma)^2}.$$

* 'Phil. Trans.,' A, vol. 197, p. 449.

† *Loc. cit.*, Equations, bottom of p. 449.

‡ As we pass outwards from the exponential point along the biquadratic q_1 ranges from ∞ to 5, which it reaches at the asymptote to the biquadratic $\beta_1 = 50$, or when $\beta_2 = \infty$, $\beta_1 = 50$.

§ "Mathematical Contributions to the Theory of Evolution, XIV. On the General Theory of Skew Correlation and Non-Linear Regression," p. 6, 'Drapers' Company Research Memoirs,' Cambridge University Press.

This leads on integration to

$$y = y_0 \left(\frac{\sigma(\sqrt{3+\beta_1} + \sqrt{\beta_1}) + x}{\sigma(\sqrt{3+\beta_1} - \sqrt{\beta_1}) - x} \right)^{\sqrt{\frac{\beta_1}{3+\beta_1}}}.$$

I term this Type XII., or the R-line J-curve. The origin is the mean, the range from $x = -\sigma(\sqrt{3+\beta_1} + \sqrt{\beta_1})$ to $\sigma(\sqrt{3+\beta_1} - \sqrt{\beta_1})$. It separates J-curves—so long as we are above the line $2\beta_2 - 3\beta_1 - 6 = 0$—for which $r-2$ is positive from those in which $r-2$ is negative. But $r-2 = m_1 + m_2$. Hence below the R-line the positive m_1 is greater than the negative m_2, but above this line the positive m_1 is less than the negative m_2, *i.e.*, the upright of the J is emphasised at the expense of the horizontal part, while below the R-line this condition is reversed until on the biquadratic the infinite ordinate of the J-upright is replaced by a finite ordinate.

I propose now to consider a little in detail the nature of these new types of frequency and the manner of fitting them to actual data. I have dealt above sufficiently fully with "block-frequency" and its criterion $\beta_2 - \beta_1 - 1 = 0$ and therefore need only consider Types VIII. to XII.

(4) *Frequency Curve. Type VIII.*—

$$y = y_0 (1 + x/a)^{-m}.$$

Range, from $x = 0$ to $x = -a$.*

y_0 is clearly the value of the ordinate at $x = 0$, *i.e.*, the finite ordinate at the tail.

We easily deduce if N be the total frequency $y_0 = \mathrm{N}(1-m)/a$, and taking the origin at $n = -a$,

$$\bar{x} = \mu'_1 = a(1-m)/(2-m), \qquad \mu'_2 = a^2(1-m)/(3-m),$$
$$\mu'_3 = a^3(1-m)/(4-m), \qquad \mu'_4 = a^4(1-m)/(5-m).$$

Hence for the moment-coefficients about the mean

$$\sigma^2 = \mu_2 = a^2(1-m)/\{(3-m)(2-m)^2\},$$

$$\mu_3 = 2a^3 m(1-m)/\{(4-m)(3-m)(2-m)^3\},$$

$$\mu_4 = 3a^4(1-m)(4-5m+3m^2)/\{(5-m)(4-m)(3-m)(2-m)^4\}.$$

These lead to

$$\beta_1 = \frac{4m^2(3-m)}{(1-m)(4-m)^2}, \qquad \beta_2 = \frac{3(3-m)(4-5m+3m^2)}{(1-m)(4-m)(5-m)}.$$

Clearly m could be found from the value of β_1 by solving the cubic equation

$$m^3(4-\beta_1) + m^2(9\beta_1 - 12) - 24\beta_1 m + 16\beta_1 = 0,$$

* Of course, whether a is really positive or negative will depend on the sign given to x, or the direction of the x-axis.

then a is determined from

$$a = \pm\sigma(2-m)\sqrt{\frac{3-m}{1-m}},$$

the sign being determinable from the observed value of μ_3 and y_0 from

$$y_0 = \frac{N(1-m)}{a} = \frac{N}{\sigma}\frac{1-m}{2-m}\sqrt{\frac{1-m}{3-m}},$$

and the placing of the frequency curve on the observations by

$$\mu'_1 = a(1-m)/(2-m).$$

If, however, we find $10\beta_2-12\beta_1-18$ and $3\beta_1-2\beta_2+6$, we have

$$10\beta_2-12\beta_1-18 = +\frac{24m(m-2)^3}{(1-m)(5-m)(4-m)^2},$$

$$3\beta_1-2\beta_2+6 = -\frac{24(m-2)^3}{(1-m)(5-m)(4-m)^2},$$

giving

$$m = -\frac{2(5\beta_2-6\beta_1-9)}{3\beta_1-2\beta_2+6},$$

and thus since m is to be positive, the point (β_1, β_2) must be *above* the line $5\beta_2-6\beta_1-9=0$. The line $2\beta_2-3\beta_1-6=0$ does not meet $5\beta_2-6\beta_1-9=0$ in the positive quadrant, so that a point below both these lines does not exist in real frequency. Clearly

$$1-m = (8\beta_2-9\beta_1-12)/(3\beta_1-2\beta_2+6),$$

$$3-m = (4\beta_2-3\beta_1)/(3\beta_1-2\beta_2+6),$$

$$4-m = 2(\beta_2+3),$$

and thus if these values be substituted in β_1 as given above, we reach

$$\beta_1(\beta_2+3)^2(8\beta_2-9\beta_1-12) = (4\beta_2-3\beta_1)(10\beta_2-12\beta_1-18)^2$$

the equation to the biquadratic, proving that the point associated with the above frequency curve lies on the biquadratic.

Again $1-m$ will always be positive, or m less than unity. For the upper branch of the loop of the biquadratic lies below its asymptote, or $8\beta_2-9\beta_1-12\frac{2}{3}=0$, and accordingly below the line $8\beta_2-9\beta_1-12=0$; thus the numerator of $1-m$ is always positive. So also is the denominator, for the upper branch always lies above the line $2\beta_2-3\beta_1-6=0$.*

* In fact the R-line $(5\beta_2-6\beta_1-9=0)$ the parallel to the asymptote $(8\beta_2-9\beta_1-12=0)$, the limiting frequency line $(\beta_2-\beta_1-1=0)$, and the Type III. line $(2\beta_2-3\beta-6=0)$ meet in the point $\beta_2=-3$, $\beta_1=-4$ of the negative quadrant and the upper branch of the loop lies in the angle between the first two and in the positive quadrant.

As m is positive and less than unity the area and moments of the curve are all real and finite. When the point (β_2, β_1) moves along the loop of the biquadratic towards the R-point $(\beta_2 = 1\cdot8, \beta_1 = 0)$, the value of $1-m$ becomes more and more nearly unity, and ultimately at R we have $m = 0$, or the frequency curve is

$$y = y_0$$

a rectangle, *i.e.*, we reach the rectangle point. If on the other hand we move towards infinity along the upper branch of the biquadratic loop, we find $1-m$ approaches the value $1/\beta_1$ and thus ultimately becomes zero, or $m = 1$. Thus the limiting form of the frequency curve is a rectangular hyperbola, or rather the part of such hyperbola

$$y = y_0/(1+x/a)$$

from the vertical asymptote $x = -a$ to $x = 0$.

But this is clearly only a theoretical limit, for it involves $\beta_1 = \beta_2 = \infty$, and this means that if μ_2 be finite, μ_3 and μ_4 are infinite—results impossible in any actual frequency if the population be finite. It is clear indeed that β_2 must be less than N, for obviously $N\mu_4 < N^2\mu_2^2$. Again, β_1 is $< \beta_2 - 1$, and accordingly $\beta_1 < N-1$.* But these limits are of small service for practical statistics, where even for small samples, say, $N = 20$, they would scarcely ever be approached.† Thus the rectangular hyperbola can only be treated as a limiting form of Type VIII. far beyond the region of actual statistical experience.‡ For practical purposes the point is that m is limited to values between 0 and 1, or Type VIII. ranges from the rectangle to the rectangular hyperbola. The suggestiveness of this is that curves in the I_U and the I_J areas, *i.e.*, above and below the upper branch of the biquadratic loop, must approach these types as they approach the extremes of this branch. Generally a U-curve near the biquadratic will be close to a curve resembling a curtailed hyperbola.

* Mr. G. N. WATSON has given me a nearer limit to β_2, namely, $\beta_2 \leq N - 2 + \frac{1}{N-1}$. But, except as showing that β_2 must be finite, which is otherwise obvious, this is again of no real service.

† The highest *observed* values that I know of for β_2 and β_1 are those given by DUNCKER ('Biometrika,' vol. VIII., p. 238). He gives

'Armzahl,' *Asterina exigua*	N = 600	$\beta_2 = 33\cdot13$,	$\beta_1 = 1\cdot76$,
,, *Archaster typicus*	N = 902	$\beta_2 = 128\cdot48$,	$\beta_1 = 4\cdot76$.

There are only three groups of frequency in each, 4, 5 and 6, and the bulk of the observations are concentrated in 5. The observations do not give, as he suggests, PEARSON'S Type IV. and Type VI. curves respectively; the κ_2 in *both* cases is less than unity, corresponding to Type IV. But both fall into the heterotypic area of Type IV. The attempt to fit with heterotypic curves would hardly be profitable until there was absolute certainty that the group with 4 'Armzahl' was not the result of accident.

‡ *Theoretically* very high values of β_1 and β_2 can easily be found, *i.e.*, for samples of *four*, when the population sampled has, say, a correlation of 0·98; here the frequency curve for the correlation coefficient gives $\beta_1 = 203\cdot325$ and $\beta_2 = 311\cdot731$, but it is the rapidly approaching zero of μ_2 which leads to these results.

In concluding our discussion of this curve we may note that, perhaps, the easiest way of tracing the biquadratic is to calculate β_1 and β_2 from

$$\beta_1 = \frac{4(2\gamma-1)^2(\gamma+1)}{3\gamma-1}, \quad \beta_2 = \frac{3(\gamma+1+\beta_1)}{3-\gamma} = \frac{3(\gamma+1)(16\gamma^2-13\gamma+3)}{(3\gamma-1)(3-\gamma)}$$

by giving a succession of values to γ.

For $\gamma = 0{\cdot}5$ to 3 we get the points on the lower branch of the loop; for $\gamma = 0{\cdot}5$ to $0{\cdot}3$ we obtain the points on the upper branch of the loop. It will be seen that this amounts to taking the origin at $\beta_2 = -3$, $\beta_1 = -4$, and rotating a line through this point round it to intersect the curve. The slope of this line to the β_1 axis is $3/(3-\gamma)$.

The cubic, it may be here noted, which gives the Type V. curve may be traced from

$$\beta_1 = 4(\gamma^2-1), \quad \beta_2 = \frac{3(\gamma+1+\beta_1)}{3-\gamma} = \frac{3(\gamma+1)(4\gamma-3)}{3-\gamma}.$$

Here γ must be given values from 1 to 3.

The Type III. line, which passes through the Gaussian point, also passes through $\beta_2 = -3$ and $\beta_1 = -4$, and the above means of getting at the points on the cubic corresponds to finding the points in which a straight line passing through $(-4, -3)$ and rotating from the position of the Type III. line cuts the cubic—its slope in any position being as before $3/(3-\gamma)$.

Actually if θ be the angle between the above line from $(-4, -3)$ to the cubic, *i.e.*, $\tan\theta = 3/(3-\gamma)$,

$$r = 12(\sec\theta - \operatorname{cosec}\theta),$$

but to use this polar equation has not been found a very ready manner of plotting the cubic.*

(5) *Frequency Curve. Type IX.*—

$$y = y_0\left(1+\frac{x}{a}\right)^m.$$

Range from $x = -a$ to $x = 0$; y is zero at one end of the range and equal to y_0 at the other.

The analysis proceeds precisely as in the case of the curve of Type VIII., except that m is now opposite in sign. We have

$$y_0 = \mathrm{N}(1+m)/a,$$

$$\bar{x}\,(= \text{distance of mean from point } x = -a) = a(m+1)/(m+2),$$

$$\sigma^2 = \mu_2 = a^2(m+1)/\{(m+3)(m+2)^2\},$$

$$\mu_3 = -2a^3m(m+1)/\{(m+4)(m+3)(m+2)^3\},$$

$$\mu_4 = 3a^4(m+1)(3m^2+5m+4)/\{(m+5)(m+4)(m+3)(m+2)^4\},$$

* The parts of the cubic and the quartic lying in the other three quadrants have been plotted by Miss B. C. B. Cave. Geometrically the interrelations of the two curves, their asymptotic and other critical lines are of much interest, but until some interpretation can be put on imaginary values of the moment coefficients, these interrelations have no statistical bearing.

leading to

$$\beta_1 = \frac{4m^2(m+3)}{(m+1)(m+4)^2}, \qquad \beta_2 = \frac{3(m+3)(3m^2+5m+4)}{(m+1)(m+4)(m+5)}.$$

Thus

$$m^3(\beta_1-4)+3m^2(3\beta_1-4)+24m\beta_1+16\beta_1 = 0$$

would give m, and a would be found from

$$a = \pm\sigma(m+2)\sqrt{\frac{m+3}{m+1}},$$

the sign being found from the observed value of μ_3. Lastly

$$y_0 = \frac{N}{\sigma}\frac{m+1}{m+2}\sqrt{\frac{m+1}{m+3}}.$$

Practically it is better to determine m from

$$m = \frac{2(5\beta_2-6\beta_1-9)}{3\beta_1-2\beta_2+6},$$

which value of m substituted in the expression for β_1 gives the biquadratic.

Clearly since the lower branch of the biquadratic lies below the line $5\beta_2-6\beta_1-9=0$, m is positive until the line $2\beta_2-3\beta_1-6=0$ is reached, and in this section of the branch, *i.e.*, from $m=0$ to $m=\infty$, or from $\beta_2 = 1{\cdot}8$, $\beta_1 = 0$ up to $\beta_2 = 9$, $\beta_1 = 4$ (the exponential point, E) occurs an interesting isolated point—the line-point L. When $\beta_2 = 2{\cdot}4$, $\beta_1 = 0{\cdot}32$, then $m = 1$, and Type IX. degenerates into a sloping straight line, $y = y_0(1+x/a)$, or the frequency *line* is

$$y = \frac{\sqrt{2}N}{3\sigma}\left(1+\frac{x}{3\sqrt{2}\sigma}\right).$$

Up to the line-point, Type IX. curve rises at $x=-a$ *perpendicular* to the axis of x, at the line-point it makes a finite angle less than 90 degrees, and after the line-point we start with contact at $x=-a$.

It is interesting to note the sloping line arising as a case of these generalised frequency curves, and we observe that its locus is separated from the rectangle locus by a considerable interval along the biquadratic in which the curve of Type IX. is very trapezoidal in form.

(6) *Frequency Curve of Type X. The Exponential Curve.*—Beyond the line-point, L at $\beta_2 = 2{\cdot}4$, $\beta_1 = 0{\cdot}32$, we reach as m steadily mounts a series of frequency curves which culminate in the exponential curve at E or $\beta_2 = 9$, $\beta_1 = 4$.

Clearly

$$y_0 = N\frac{m+1}{a} = \frac{N}{\sigma}\frac{m+1}{m+2}\sqrt{\frac{m+3}{m+1}} = \frac{N}{\sigma} \text{ when } m \text{ is infinite.}$$

Further

$$y = \frac{N}{\sigma}\left(1 \pm \frac{x}{\sigma(m+2)}\right)^{(m+2)-2},$$

$$= \frac{N}{\sigma} e^{\pm x/\sigma},$$

the range being from $x = 0$ to $-\infty$, if we take the positive sign—and from $x = 0$ to $+\infty$, if we take the negative. It is thus sufficient to consider

$$y = \frac{N}{\sigma} e^{\pm x/\sigma},$$

with range from $x = 0$ to $x = +\infty$. The first two moments of the area about $x = 0$ are $\nu'_1 = \sigma$ and $\nu'_2 = \sigma^2$. Thus $\bar{x} = \sigma$ and $\mu_2 = \sigma^2$, as it should. Lastly, $\mu_3 = 2\sigma^3$ and $\mu_4 = 9\sigma^4$.

The fitting of the exponential curve presents no difficulty.

The exponential point E is a transition point of great interest as being even more than the Gaussian point G—the meeting point of many types. At E, Type IX_2 changes to Type XI., but at E the familiar Type III. passes from a zero ordinate at the limited end of the range to a J-curve with infinite ordinate. Further, E is a point at which the areas of Type I. (Type I_L) as a limited range with zero ordinates at its terminals, and as a limited range with one infinite ordinate at a terminal (Type I_J) meet. Finally, Type VI. area, which lies between Type III. line and Type V. cubic, is divided into two sections by Type XI., which lies along the lower branch of the biquadratic loop below E. Below the biquadratic, Type VI. takes the form

$$y = y_0 (x-a)^{q_2}/x^{q_1},$$

with a range from $x = a$ to ∞, q_1 and q_2 being both positive. In the area, however, below Type III_J and above Type XI., Type VI. takes the form VI_J, or the J-shaped curve

$$y = \frac{y_0}{x^{q_1}(x-a)^{q_2}},$$

with a range from $x = a$ to $x = \infty$. In this case $r = 6(\beta_2-\beta_1-1)/(3\beta_1-2\beta_2+6)$ will be negative, since we are below the line $2\beta_1-3\beta_1-6 = 0$. Further, ϵ is negative since we are above the cubic or Type V. branch

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

Thus our quadratic

$$m'^2 - rm' + \epsilon = 0,$$

corresponds of necessity to real roots, of which one will be negative and the other positive. The positive root will be

$$\tfrac{1}{2}(\sqrt{r^2-4\epsilon}+r),$$

and is therefore numerically the smaller root since r is negative; it will be less than unity, and therefore $m'-1 = m$ will be negative if

$$\tfrac{1}{2}(\sqrt{r^2-4\epsilon}+r) < 1,$$

or

$$\epsilon - r + 1 > 0,$$

but this is the condition for the point β_1, β_2 lying inside the loop of the quadratic. Thus in this case we reach the J-shaped curve of Type VII., or

$$y = \frac{y_0}{x^{q_1}(x-a)^{q_2}}.$$

In order that the area of this curve and its moments should be finite, it is clearly needful that q_2 should be less than unity.

(7) *Frequency Curve. Type XI.*—Beyond the exponential point the lower branch of the biquadratic is below the line $2\beta_2 - 3\beta_1 - 6 = 0$, and consequently m is again negative and the curve takes the form

$$y = y_0 x^{-m},$$

where

$$m = \frac{2(5\beta_2 - 6\beta_1 - 9)}{2\beta_2 - 3\beta_1 - 6}.$$

The range is, however, only limited in one direction, it is from $x = b$ to $x = \infty$, say.

This lower branch of the biquadratic loop tends to become vertical and asymptotic to the line $\beta_1 = 50$. Hence m takes all values from ∞ down to 5.

Clearly, for moments about $x = b$,

$$N\mu'_p = \frac{y_0}{m-p-1}\,\frac{1}{b^{m-p-1}},$$

and these will be real and finite if $p < m-1$, or only the fourth moment would fail at the limit $\beta_2 = \infty$, which indeed cannot in practice be reached. At the same time if we want the probable error of the fourth moment to be finite, it is needful that μ'_8 should be finite or we must have $m > 9$. Thus $m = 9$ must be where the curve passes into the heterotypic region and becomes of doubtful application.

We easily find from the above result for μ'_p

$$\bar{x} = b(m-1)/(m-2), \qquad \mu_2 = \sigma^2 = b^2(m-1)/\{(m-2)^2(m-3)\},$$

$$\mu_3 = 2b^3 m(m-1)/\{(m-2)^3(m-3)(m-4)\},$$

$$\mu_4 = 3b^4(m-1)(3m^2-5m+4)/\{(m-2)^4(m-3)(m-4)(m-5)\},$$

leading to

$$\beta_1 = \frac{4m^2(m-3)}{(m-1)(m-4)^2}, \qquad \beta_2 = \frac{3(m-3)(3m^2-5m+4)}{(m-1)(m-4)(m-5)}.$$

Thus for $m = 9$ we find $\beta_1 = 9{\cdot}72$, $\beta_2 = 22{\cdot}725$, which satisfy the equation $8\beta_2 - 15\beta_1 - 36 = 0$ of the heterotypic line.

m may be found from

$$m = \frac{2(5\beta_2 - 6\beta_1 - 9)}{2\beta_2 - 3\beta_1 - 6},$$

or from β_1 alone by the cubic

$$m^3(4-\beta_1) + m^2(9\beta_1 - 12) - 24\beta_1 m + 16\beta_1 = 0,$$

then

$$b = \pm\sigma(m-2)\sqrt{\frac{m-3}{m-1}},$$

and

$$y_0 = \mathrm{N}b^{m-1}(m-1),$$

while the mean $\bar{x} = b(m-1)/(m-2)$ enables us to place the curve on the observations.

There is no discontinuity in the form of the curve down to $m = 5$, but only discontinuity after $m = 9$ in the probable errors of its moment-coefficients.

The curve starts with a finite ordinate and meets that ordinate at a finite angle; it asymptotes to the x-axis at $x = \infty$, and has no point of inflexion except at infinity.

(8) *Frequency Curve. Type XII.*—

$$y = y_0\left(\frac{\sigma(\sqrt{3+\beta_1} + \sqrt{\beta_1}) + x}{\sigma(\sqrt{3+\beta_1} - \sqrt{\beta_1}) - x}\right)^{\sqrt{\frac{\beta_1}{3+\beta_1}}}.$$

This J-curve arises along the R-line, or $5\beta_2 - 6\beta_1 - 9 = 0$. Its range is from $x = \sigma(\sqrt{3+\beta_1} - \sqrt{\beta_1})$ to $x = -\sigma(\sqrt{3+\beta_1} + \sqrt{\beta_1})$, and then its mean is the origin. When β_1 is zero it degenerates into a rectangle (*i.e.*, at the rectangle point).

In order to illustrate the nature of the curve more fully let us start from the general equation which arises when the denominator of the differential equation has real roots,* *i.e.*,

$$y = y_0(1 + x/a_1)^{m_1}(1 - x/a_2)^{m_2},$$

where

$$y_0 = \frac{\mathrm{N}}{b}\frac{m_1^{m_1}m_2^{m_2}}{(m_1+m_2)^{m_1+m_2}}\frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\Gamma(m_2+1)}$$

and

$$\frac{m_1}{a_1} = \frac{m_2}{a_2} = \frac{m_1+m_2}{b},$$

the origin being the mode and b the range.

Transferring to the mean as origin† this becomes

$$y = \frac{y_0}{a_1^{m_1}a_2^{m_2}}\left(\frac{b(m_1+1)}{m_1+m_2+2} + x\right)^{m_1}\left(\frac{b(m_2+1)}{m_1+m_2+2} - x\right)^{m_2},$$

* 'Phil. Trans.,' A, vol. 186, p. 369.

† *Loc. cit.*, p. 370.

where*

$$b = \tfrac{1}{2}\sigma\,\{\beta_1(m_1+m_2+4)^2+16(m_1+m_2+3)\}^{1/2}.$$

$$\frac{y_0}{a_1^{m_1}a_2^{m_2}} = \frac{\mathrm{N}}{b^{m_1+m_2+1}}\,\frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\,\Gamma(m_2+1)},$$

on substitution for a_1 and a_2 as above.

Now put $m_1+m_2 = 0$, or $m_2 = -m_1 = m$, say.

Then

$$y = \frac{\mathrm{N}\,.\,\Gamma(2)}{b\Gamma(1+m)\,\Gamma(1-m)}\left(\frac{b(m+1)}{2}+x\right)^{m}\left(\frac{b(1-m)}{2}-x\right)^{-m},$$

while

$$b = \tfrac{1}{2}\sigma\,\{16\beta_1+48\}^{1/2} = 2\sigma\,(\beta_1+3)^{1/2}.$$

It remains to find m.

Now m_1 and m_2 are the roots of †

$$m^2-(r-2)\,m+\epsilon-r+1 = 0,$$

where

$$r = \frac{6(\beta_2-\beta_1-1)}{3\beta_1-2\beta_2+6} = m_1+m_2+2,$$

$$\epsilon = \frac{(m_1+m_2+2)^2}{4+\frac{1}{4}\beta_1(m_1+m_2+4)^2/(m_1+m_2+3)}$$

Hence, when $m_1+m_2 = 0$, we have

$$r = 2 \quad \text{or} \quad 5\beta_2-6\beta_1-9 = 0, \text{ the R-line,}$$

and

$$\epsilon = 3/(\beta_1+3).$$

Whence

$$m^2 = 1-\epsilon \quad \text{or} \quad m = \pm\sqrt{\frac{\beta_1}{3+\beta_1}}.$$

But $\Gamma(2) = 1$, and it is well known that

$$\Gamma(1+m)\,\Gamma(1-m) = \frac{m\pi}{\sin m\pi}.$$

Thus

$$y = \frac{\mathrm{N}}{2\pi\sigma}\,\frac{\sin\left\{\sqrt{\frac{\beta_1}{3+\beta_1}}\,\pi\right\}}{\sqrt{\beta_1}}\left(\frac{\sigma(\sqrt{3+\beta_1}+\sqrt{\beta_1})+x}{\sigma(\sqrt{3+\beta_1}-\sqrt{\beta_1})-x}\right)^{\sqrt{\frac{\beta_1}{3+\beta_1}}}$$

This is the full equation to the R-line J-curve, the mean being origin.‡ It requires for its determination only a knowledge of β_1, but we must be also certain that the

* *Loc. cit.*, p. 369.

† *Loc. cit.*, pp. 368–9. Deduced at once from $m'^2-rm'+\epsilon = 0$ by putting $m' = m+1$.

‡ The sign of $\sqrt{\beta_1}$ in $\sigma(\sqrt{3+\beta_1}\pm\sqrt{\beta_1})$ must be determined from that of μ_3.

condition $5\beta_2 - 6\beta_1 - 9 = 0$ is satisfied within the limits of random sampling. Its possibilities extend from $\beta_1 = 0$ to $\beta_1 = \infty$. When $\beta_1 = 0$,

$$y = \frac{N}{2\sqrt{3}\sigma}, \text{ the rectangle.}$$

Now consider what happens for any frequency curve of the limiting character when *both* β_1 and β_2 become infinite, say, in the ratio $\beta_2 = p\beta_1$. Then

$$r = \frac{6(p-1)}{3-2p},$$

and accordingly r will be finite if p is finite, except along the Type III. line. Accordingly for $\beta_1 = \infty$, ϵ will be zero. Thus the ratio of β_2 to β_1 is from their values,

$$\beta_2/\beta_1 = \tfrac{3}{2} \cdot \frac{r+2}{r+3} = p,$$

which agrees with the above result for r.

For the special case when $r = 2$, we have $p = \frac{6}{5}$, which agrees with the limiting ratio of β_2/β_1 along the R-line.

Now when $\epsilon = 0$ we have from

$$m^2 - (r-2)m + \epsilon - r + 1 = 0,$$

$$m = \tfrac{1}{2}\left(r - 2 \pm \sqrt{(r-2)^2 + 4r - 4}\right),$$

$$= \tfrac{1}{2}(r - 2 \pm r) = r - 1 \text{ or } -1.$$

Thus from the equations on page 445,

$$y = \frac{N}{b}\frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\Gamma(m_2+1)}\left(\frac{m_1+1}{m_1+m_2+2} + \frac{x}{b}\right)^{m_1}\left(\frac{m_2+1)}{m_1+m_2+2} - \frac{x}{b}\right)^{m_2},$$

$$= \frac{N(m_2+1)}{b}\frac{1}{\Gamma(m_2+2)}\left(\frac{r}{r} + \frac{x}{b}\right)^{r-1}\left(-\frac{x}{b}\right)^{-1},$$

$$= \frac{N(m_2+1)}{x}\left(1 - \frac{x}{b}\right)^{r-1},$$

if we change the sense of the axis of x and take x from 0 to $+b$.

Now in order that σ should be finite it is needful that b should be infinite when $m_2 = -1$, for

$$\sigma^2 = b^2(m_2+1)/\{r(r+1)\}.$$

But if b be infinite, $y = 0$ owing to the factor m_2+1, for every value of x, except $x = 0$. Hence the frequency is a concentrated lump at $x = 0$, and this involves of itself $\sigma = 0$.

But if $\sigma = 0$, b must be finite or zero, and these both again throw us back on a concentrated frequency at $x = 0$.

Accordingly, when β_1 and β_2 both become infinite, we deal with a concentrated frequency lump. But the ratio of β_1 to β_2 will depend on the manner in which we have reached this limiting case.

For example, if we are dealing with the correlations in samples of two drawn from a population in which the correlation is ρ, the frequency consists of two lumps, but as ρ approaches unity, one lump shrivels up, β_1 and β_2 both become infinite, but their ratio is one of equality, *i.e.*, we approach infinity along the line $\beta_2 - \beta_1 - 1 = 0$.

When we take samples of three from a population of correlation ρ, the frequency curves are U-shaped, but as ρ approaches unity the frequency concentrates in one leg of the U, β_1 and β_2 both become indefinitely larger, but their ultimate ratio β_2/β_1 appears to equal $\frac{5}{4}$.* The U-curve flattens down into an L-curve, of which the horizontal limb extends to infinity and becomes indefinitely thin, while the vertical limb contains all the frequency.

(9) *Scheme of Skew Frequency Curves Represented as a Diagram.*—We are now able to considerably enlarge our diagrammatic representation of frequency curves. (See Diagram, Plate 1.)

Every distribution is represented by its characteristic co-ordinates β_1 and β_2, which must be positive, and therefore we need only deal with the positive β_1, β_2 quadrant. No frequency distribution at all can lie above the line $\beta_2 - \beta_1 - 1 = 0$; this restriction removes more than half the positive quadrant. No frequency distribution can be adequately represented by one of the present system of skew curves, if it falls below the line $8\beta_2 - 15\beta_1 - 36 = 0$. The area below this line is therefore termed *heterotypic*. Heterotypic distributions are to say the least of it very rare, if they be not extremely improbable. We have seen that there is some reason to suppose that bimodal distributions would give rise to such heterotypic distributions, but with our present views as to frequency such distributions when they do not arise from the mere anomalies of random sampling are classed as heterogeneous, and supposed to be due to mixtures.

Having thus limited our area at top and bottom we proceed to consider the various possibilities that arise.

The β_2-axis, where $\beta_1 = 0$, is the axis of *symmetrical* frequency distributions. Possibilities begin at the B-line or the point $\beta_2 = 1$, or we have two equal concentrated frequency blocks at any arbitrary distance b. This is the case of two alternative values, either of which is equally probable. For example, heads or tails in the repeated tossings of a single coin, or positive or negative perfect correlation in samples of two taken from a population of individuals bearing two uncorrelated

* I use the word "appears" advisedly, because the ratio has been obtained by determining the value of β_2/β_1 for high numerical value of ρ. The actual ratio for $\rho = 1$ depends upon approaching a limit in rather complicated elliptic integral expressions, which I have not yet accomplished.

characters. Below the point $\beta_2 = 1$, descending the β_2-axis, the two concentrated frequencies expand into a symmetrical U-curve. This is Type II_{U} with the equation

$$y = y_0 \,(1-x^2/a^2)^{-m}$$

and the criterion $\beta_1 = 0$, $\beta_2 < 1{\cdot}8$.

Here*

$$m = \tfrac{1}{2}\,(9-5\beta_2)/(3-\beta_2),$$

$$a^2 = \sigma^2 \,.\, 2\beta_2/(3-\beta_2),$$

and

$$y_0 = \frac{\mathrm{N}}{\sqrt{2\pi}\sigma}\,\frac{\Gamma\,(\frac{3}{2}-m)}{\Gamma\,(1-m)\,\sqrt{\frac{3}{2}-m}}.$$

When $\beta_2 = 1{\cdot}8$, $m = 0$, and we reach the "rectangle-point" R. Here $y_0 = \mathrm{N}/(2a)$ and $\sigma = a/\sqrt{3}$.

Samples of three individuals from a population whose individuals carry two uncorrelated characters give a symmetrical U-frequency for the coefficients of correlation of those characters in triplets of individuals. In this illustration $\beta_2 = 1{\cdot}5$. Samples of four individuals from the same population give a rectangle for the frequency distribution of the coefficients of correlation. Passing still lower down the axis of symmetrical frequency the type is now Type II_{L}, or the limited range frequency curve

$$y = y_0\,(1-x^2/a^2)^m$$

and the criterion is $\beta_1 = 0$, $\beta_2 > 1{\cdot}8 < 3$.

In this range m increases from 0 to ∞, and

$$m = \tfrac{1}{2}\,(5\,\beta_2-9)/(3-\beta_2)$$

$$a^2 = \sigma^2 \,.\, 2\,\beta_2/(3-\beta_2),$$

$$y_0 = \frac{\mathrm{N}}{\sqrt{2\pi}\sigma}\,\frac{\Gamma\,(\frac{3}{2}+m)}{\Gamma\,(1+m)\,\sqrt{\frac{3}{2}+m}}.$$

We see that the range grows greater as m approaches infinity, or $\beta_2 = 3$, when we reach G the Gaussian point ($\beta_1 = 0$, $\beta_2 = 3$).

If samples of n individuals be taken from an indefinitely large population in which the individuals carry two uncorrelated characters, then if n be 5 or over, all the frequency curves of the correlation coefficients of these samples are of Type II_{L}, only approaching the Gaussian when n is very considerable indeed. For example when $n = 25$, $\beta_2 = 2{\cdot}7692$, and the frequency is still a good way from the Gaussian. When $n = 400$, $\beta_2 = 2{\cdot}9850$, it is thus fairly close to it, but is not coincident.

* It is, perhaps, worth noticing that for $\beta_2 = 15/7$ we obtain the ordinary parabola as a special type of frequency-curve.

After we have passed the Gaussian point we obtain curves of unlimited range of Type VII., of which the equation is

$$y = y_0 (1+x^2/a^2)^{-m},$$

The range of β_2 is from 3 to ∞ and

$$m = \tfrac{1}{2}(5\beta_2-9)/(\beta_2-3),$$

falls from infinity to 2·5; while

$$a^2 = \sigma^2 \,.\, 2\beta_2/(\beta_2-3),$$

$$y_0 = \frac{\mathrm{N}}{\sqrt{2\pi}\sigma}\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})\sqrt{(m-\frac{3}{2})}}.$$

Illustration of curves of Type VII.* are not infrequent in biological statistics. We see that the Gaussian is a mere point in an infinite range of symmetrical frequency curves, and a single point in a doubly infinite series of general frequency distributions.

Now let us consider the asymmetrical frequency curves displayed on the Diagram. If we approach from the "impossible area" we reach on the B-line the first available type of frequency—the alternative concentrated blocks. At one end of the B-line we have two equal isolated frequencies, and at the other a single isolated frequency.

Crossing the B-line we reach the area of limited range U-shaped curves, *i.e.*, Type $\mathrm{I_U}$, which has for its equation :

$$y = y_0 (1+x/a_1)^{-m_1} (1-x/a_2)^{-m_2}.$$

This U-area extends as far as the upper branch of the loop of the biquadratic, the asymptote of which, $24\beta_2-27\beta_1-38 = 0$, is indicated by a broken line. In U-shaped frequency curves both m_1 and m_2 are necessarily less than unity, for their product is $\epsilon-r+1$, which is less than unity and positive above the upper branch of the biquadratic (*i.e.*, $\epsilon-r+1 = 0$). Type $\mathrm{I_U}$ is fitted as Type I. (see 'Phil. Trans.,' A, vol. 186, p. 367), and has been illustrated by me ('Roy. Soc. Proc.,' vol. 62, p. 287), by fitting curves of frequency to cloudiness. The frequency curves for the correlation coefficients of samples of three drawn from a population whose individuals have two characters of any degree of correlation are also skew U-shaped frequency curves, although their algebraic form has not the above simplicity.

* Type $\mathrm{II_L}$ was discussed in my first memoir, 'Phil. Trans.,' vol. 186, p. 372. Type $\mathrm{II_J}$ and Type VII. are briefly referred to in 'Biometrika,' vol. IV., p. 174, but, unfortunately, with some rather disturbing misprints. They are correctly placed on RHIND's diagram, 'Biometrika,' vol. VII., p. 131, but the formulæ for fitting are not given. The formulæ have been given for many years in lecture-notes, and the curves have been frequently used.

On the upper branch of the biquadratic loop we reach curves of Type VIII., *i.e.*,

$$y = y_0\,(1+x/a)^{-m},$$

discussed on p. 444 of the present memoir. Here m is less than unity.

We now pass into the loop of the biquadratic between the upper branch and the R-line. Here we have J-curves, Type I_{J}, of the form

$$y = y_0\,(1+x/a_1)^{-m_1}\,(1+x/a_2)^{-m_2}$$

where m_2 is less than unity, and m_1 is less than m_2.

Coming to the R-line, m_1 becomes equal to m_2 and we have Type XII., or

$$y = y_0\left(\frac{\sigma\,(\sqrt{3+\beta_1}+\sqrt{\beta_1})+x}{\sigma\,(\sqrt{3+\beta}-\sqrt{\beta_1})-x}\right)^{\sqrt{\frac{\beta_1}{3+\beta_1}}}$$

discussed on p. 446 of the present memoir. Below the R-line, we return to Type I_{J}, but m_1 is now greater than m_2.*

We now reach the lower branch of the biquadratic loop. This is divided into three portions by three critical points. The first portion is from the rectangle-point (R) to the line-point L. In this portion we start from R with the curve of Type IX. or,

$$y = y_0\,(1+x/a)^{m}$$

for $m = 0$, or the rectangle, and proceed from that value to $m = 1$, which gives us the line (or triangle); the range is $-a$ to 0. Since m is always < 1, the curve rises perpendicularly at $x = -a$, and approximates to a trapezoidal form. The method of fitting is discussed in this memoir, p. 441. The fitting of the line curve

$$y = y_0\,(1+x/a)$$

is dealt with on p. 442.

Beyond the line-point L we have Type IX_2 which differs in no way from Type IX_1, except that m is now greater than unity, and there is contact of a rapidly increasing order at $x = -a$.

When $m = \infty$ we find Type X. the exponential curve, at the exponential point E. The fitting of this curve

$$y = \frac{\mathrm{N}}{\sigma}\,e^{-x/\sigma}$$

has been discussed on p. 443.

* For example, at the point $\beta_2 = 4$, $\beta_1 = 2$, between the R-line and upper branch,

$$y = y_0\left(1+\frac{x}{a_1}\right)^{0\cdot 2123}\Big/\left(1+\frac{x}{a_2}\right)^{0\cdot 7123},$$

but at $\beta_2 = 8$, $\beta_1 = 4$, between the R-line and the lower branch,

$$y = y_0\left(1+\frac{x}{a_1}\right)^{7\cdot 4011}\Big/\left(1+\frac{x}{a_2}\right)^{0\cdot 4011}$$

Since E is the junction of several types, we turn to consider Type III. which is the curve found along the critical line

$$2\beta_2 - 3\beta_1 - 6 = 0.$$

It passes through the Gaussian point G, and its equation is

$$y = y_0 (1 + x/a)^p e^{-px/a}.$$

It is fully discussed in my first memoir; see 'Phil. Trans.,' A, vol. 186, p. 373, *et seq.*

From G to the exponential point E, p ranges from ∞ to zero, which latter value provides the exponential curve. After the exponential point p becomes negative and we reach Type III_J, a J-curve with range limited in one direction only. This curve separates the doubly limited curves of Type I_J from curves of Type VI_J, which lie below the line $2\beta_2 - 3\beta_1 - 6 = 0$, and above the lower branch of the biquadratic loop. On this lower branch of the loop we have Type XI., or the form

$$y = y_0 x^{-m}$$

the range being from an arbitrary value b to ∞, and m ranging from ∞ to 5. This type is fully discussed in the present memoir; see p. 444. It continues right away along this branch of the biquadratic, but at $\beta_2 = 22{\cdot}725$ and $\beta_1 = 9{\cdot}72$, the eighth moment of the theoretical curve would become infinite, and accordingly the probable error of the fourth moment coefficient would become *theoretically* infinite. Thus since the fitting of the curve depends on the fourth moment its constants would cease to be reliable measures of the distribution. We enter at this point the "heterotypic area," for this type of curve.* We have now two further areas to clear off, namely those between the Type III. line and the lower branch of the biquadratic loop. Above the former and below the latter we have the range of double limited frequency curves, *i.e.*, Type I_L, or

$$y = y_0 (1 + x/a_1)^{m_1} (1 - x/a_2)^{m_2}.$$

This curve was fully discussed in my first memoir ('Phil. Trans.,' A, vol. 186, p. 376, *et seq.*) m_1 and m_2 are both positive, and experience has shown that probably the bulk of all frequency distributions cluster into this area.

Above the biquadratic loop and below the line $2\beta_2 - 3\beta_1 - 6 = 0$, we have curves of Type VI_J, or

$$y = \frac{y_0}{x^{q_1}(x-a)^{q_2}}$$

with range from $x = a$ to $x = \infty$.

* Of course, by using the actual eighth moment of the data, instead of the eighth moment of the theoretical curve, the standard deviation of the fourth moment would be finite, but this procedure would really indicate that, as far as the high moments are concerned, curve and data were discordant, and that we should not really be finding the probable error of a constant of our theoretical frequency curve.

They have been considered on p. 443 of the present memoir. Their full theory is precisely that of curves of Type VI. in general, discussed in the first supplement to my memoir on skew variation ('Phil. Trans.,' A, vol. 197, p. 448, *et seq.*). The only point to be emphasised is that the q_2 of Equation XIX. of that memoir in this area is negative and less than unity. The treatment is identical.

Below both the Type III. line and the biquadratic, we have a space bounded by the cubic

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

This is the area of Type VI. proper, *i.e.*,

$$y = y_0(x-a)^{q_2}/x^{q_1}$$

with range from $x = a$ to $x = \infty$, $q_2 < q_1$ being positive, and is fully discussed in the memoir just cited.

The area of Type VI. is limited by the above cubic along which Type V., or,

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

from $x = 0$ to $x = \infty$, describes the frequency. Its full consideration will be found in 'Phil. Trans.,' A, vol. 197, p. 446, *et seq.* Below the Type V. cubic we reach the area of Type IV. curve, or

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

This has unlimited range in both directions and its treatment is fully discussed in my first memoir ('Phil. Trans.,' A, vol. 186, p. 376, *et seq.*). Theoretically, Types IV. and VI. describe all types lying below the line $2\beta_2-3\beta_1-6 = 0$. The objection to their use lies in the increasing probable errors of their constants, however good their general fit may be. To warn the statisticians of this, the line $8\beta_2-15\beta_1-36 = 0$, is drawn on the diagram and the area below it is marked "heterotypic area." I use this term to signify that it is doubtful whether my skew-frequency curves, depending only on the first *four* moments, can adequately describe distributions of types falling below this line; they require the use of the fifth and higher moment coefficients. Their occurrence in practice, however, must be rare.

It will be noticed that the line $\beta_2-\beta_1-3 = 0$ is drawn through the Gaussian point. This is the relation which must be satisfied in the case of POISSON's exponential limit to the binomial. Hence, in the case of a distribution with β_1, β_2, near this line, it is worth while investigating whether the "law of small numbers" is appropriate. Above this line every real binomial distribution, *i.e.*, cases of p and q both positive and less than unity, and n positive (taking the binomial as $(p+q)^n$) must lie, for

$$\frac{\beta_2-3}{\beta_1} = \frac{1-6pq}{1-4pq},$$

and the right-hand side is clearly less than unity. This limited area covered by the real binomial explains its relative infrequency as a descriptive series in practical statistics. If, however, we take the negative binomial as admissible, *i.e.*, allow forms of the type

$$(p-q)^{-n}, \quad \text{where} \quad p-q=1$$

we extend the possible area of a binomial down to the line $2\beta_2-3\beta_1-6=0$.

Such a type of binomial is by no means of infrequent occurrence and can be more or less justified on *a priori* grounds.* Below Type III. line, the values of p and q become in the mathematical sense unreal, *i.e.*, imaginary. It is by no means certain, however, that such imaginary binomials with real moment coefficients may not, like imaginary hypergeometricals, give statistically good fits and be ultimately provided with physical interpretations.

(10) *Concluding Remarks.*—It is very difficult to assert finality for any scientific investigation, but I trust this second supplement to my original memoir on skew variation of 1894 has garnered the last harvest of possible types within the limits proposed in that investigation. The object was the discovery of a system of frequency curves providing for every possible variation of the first four moment coefficients of a distribution and provision for their rapid treatment and calculation. Since 1894 much has been done by the provision of tables of the new functions and improved tables of old functions necessary to carry this out.† Diagrams like that accompanying this memoir, enable the statistician who has calculated the characteristic β_1 and β_2, to select at once the appropriate type, from the position of the point β_1, β_2 in the β_1, β_2 plane. The first diagram, prepared by Mr. A. J. Rhind at my suggestion, has been long in use.‡ For the present very carefully prepared and much extended diagram I have to thank my colleague, Miss Adelaide G. Davin, whose labours cannot fail to be appreciated by those having to handle practically statistical data.

Since the publication of my original memoir on skew variation, many attempts have been made to express the nature of skew distributions by other systems of curves or by expansions in series. I have given careful attention to these competing systems and have discussed some of them elsewhere ('Biometrika,' vol. IV., pp. 169 to 212). My chief objections to them arise from the fact that they either (i.) cover far less than the necessary area; or (ii.) involve constants the probable errors of which can be indefinitely great; or (iii.) involve constants the probable errors of which have not been or possibly cannot be calculated. In no case that I know of have they systematically been applied to extensive ranges of data, and the goodness of fit compared with that of other systems. The existence of such competing systems is at any rate

* See 'Biometrika,' vol. IV., p. 209, and vol. XI., p. 139.

† Now collected in "Tables for Statisticians and Biometricians," issued by the Cambridge University Press.

‡ 'Biometrika,' vol. VII., p. 131.

noteworthy evidence that to attempt to describe frequency by the Gaussian curve is hopelessly inadequate. It is strange how long it takes to uproot a prejudice of that character! If the reader will turn again to the present diagram, he will see that the Gaussian frequency occupies a *single point* in an indefinitely extended area. Those who support the Gaussian theory have to prove that no distribution occurs at a distance from the point G of our diagram greater than could be accounted for by the probable errors of sampling of β_1 and β_2. These errors are known and have been tabled* and that position is quite untenable. Frequency distributions occur every day which by no manner of means can be described by Gaussian systems.

It has been said that my skew curves suddenly change their algebraic type and that the statistician is puzzled by a slight change in the constants β_1 and β_2 involving such radical changes in the equation to the type. But if the reader examines the present diagram, he will see that the main Types I_U, I_J, I_L, IV., VI. and VI_J occur in *areas*, while the remaining types occur in the critical curved or straight lines which bound these areas. Special cases like the Gaussian, the exponential or the rectangular distributions occur where critical lines intersect. Now all these critical lines are really critical in the sense that a change of important physical significance occurs in this neighbourhood, and it is very unlikely that physical changes will be unaccompanied by sharp algebraical changes of form, such as are directly obvious in my curves, but are disguised by discontinuities in some of the proposed alternative expressions in series.†

Any *one* illustration that the frequencies which occur in actual statistical data can practically cover the whole possible area of the β_1, β_2 planes, and can present frequency distributions which change abruptly in type, will suffice to confute both the argument that frequency is concentrated in or near the Gaussian point, and the argument that it is undesirable that skew-frequency curves should be so manifold in form, although how they are to change from U to J, to "cocked hat," to rectangle and to exponential forms without this abrupt change will be a puzzling problem to solve for the professed mathematician. An illustration of this character has been several times referred to in the course of this paper. Let us suppose there exists an indefinitely large population, each individual of which carries any number of characteristics which are correlated together, for simplicity we will say according to the normal law. We may suppose that there are enough pairs of characters to give all values of the correlation ρ from $+1$ to -1.

* 'Tables for Statisticians and Biometricians,' pp. 68–71.

† An analogy might be given in the case of the expression of a "cocked-hat" shape of finite range and a U-shaped distribution by a *single* FOURIER'S series. Here the trigonometrical expression by the FOURIER'S series would be superficially the same if kept in symbolic form, while the algebraic form of the U-curve would require two vertical asymptotes and its equation would be wholly different from that of the "cocked-hat" form. The Fourier expression would only disguise the real discontinuity. In the same manner real discontinuity of form is disguised in the series which express skew frequency in terms of a long series of moment coefficients.

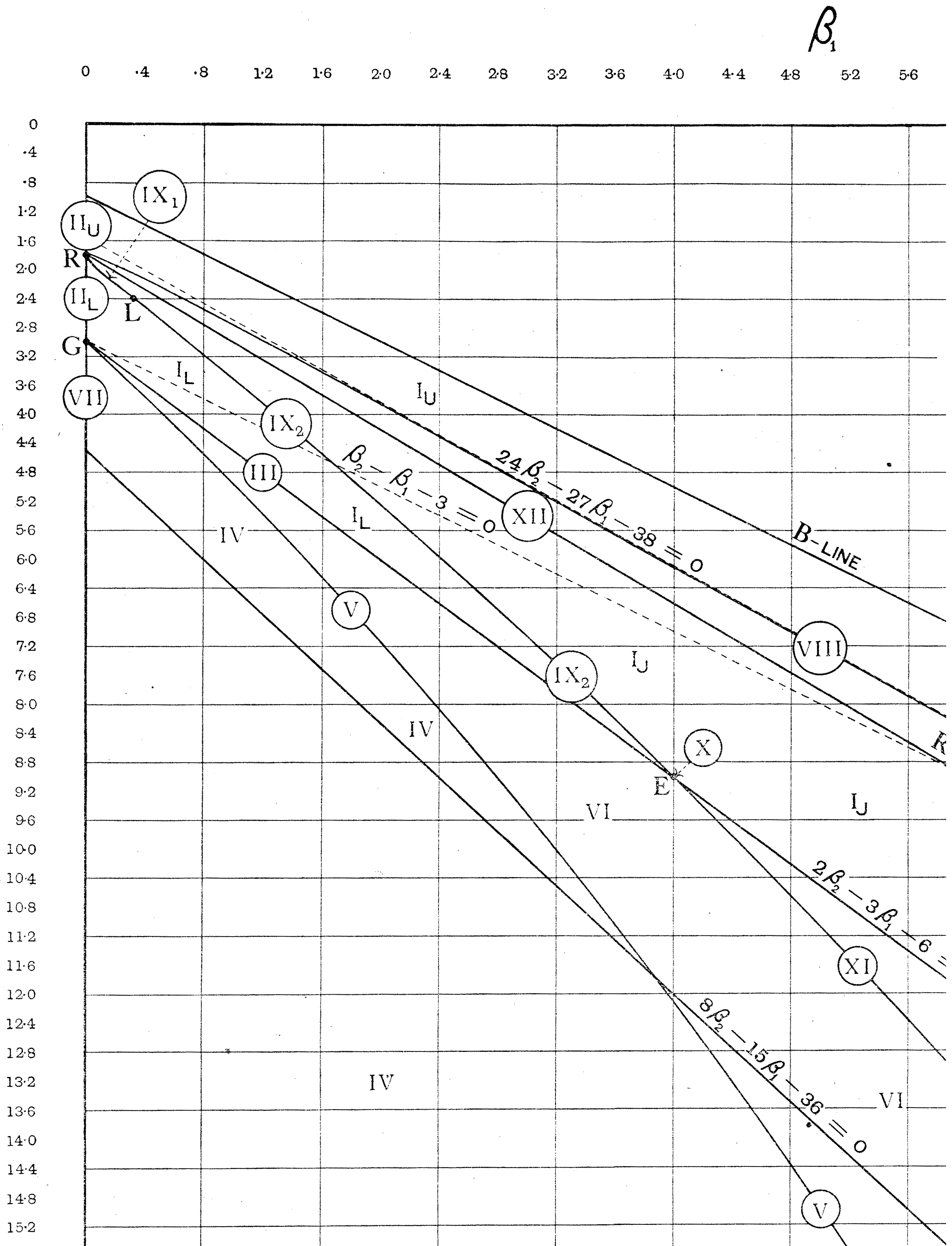
Now from this population we will take a large number m of samples of n individuals. If in each one of these samples we calculate the correlation, r, between two variates, then r will not be equal to the value of ρ in the sampled population, but the m samples will give a frequency curve for r, which is limited in range between $+1$ and -1 and is determined by n the number of individuals in the sample and by ρ the correlation of the characters in the indefinitely large population sampled. We thus obtain a doubly infinite series of frequency distributions. The general theory of such distributions has been worked out by "STUDENT" ('Biometrika' vol. VI., p. 302, *et seq.*), Mr. H. E. SOPER (*Ibid.*, vol. IX., p. 91, *et seq.*), and Mr. R. A. FISHER (*Ibid.*, vol. X., p. 507, *et seq.*). The actual forms of the frequency curves are not usually expressible by simple single functions, but the ordinates and the β_1, β_2 admit of numerical determination. The calculations are extremely laborious, but up to the present the members of my laboratory staff have calculated some 270 frequency curves with nearly 40 ordinates each for values of ρ ranging from 0 to 1, and of n from 2 to 400. The great bulk of these curves show no approach to normality. The values of β_1, β_2 range from points on the B-line down to infinity, the distributions contain concentrated blocks, U-shaped curves, J-shaped curves, rectangles, trapezoid-like forms and every variety of skewness in doubly limited range curves. Only in cases where n is very considerable and ρ is neither a positive nor a negative high correlation is there an approximation to the Gaussian. For a series of curves in which β_1 can be 5 and $\beta_2 = 9$,—or both, if we will—ten times these amounts, it is idle to talk about the value of the Gaussian curve ($\beta_1 = 0$, $\beta_2 = 3$) in describing variation. These frequency curves can be actually obtained by experimental sampling, although the process is laborious, and indeed were so obtained in the first place.* They arise from observation and experiment. The remarkable point about them is that they illustrate all the types we have been discussing and justify sharp transitions in algebraic forms by showing that such transitions correspond to actual physical facts arising from experimental statistical data. The whole illustration, details of which will shortly be published, indicates the evil of implicit reliance on a classical theory.

The Gaussian theory of error has, with great weight of authority, been applied to determine significant differences in statistical constants. The theory of the "probable error" must be justified in the case of each statistical constant to which it is applied. Psychologists have been busy discussing the differences found in mental correlations deduced from small samples on the basis of significance judged by the Gaussian theory of probable error. That theory has practically no application, as the "probable error" has really no meaning in the case of the bulk of the samples dealt with. Applications of the theory of probable error in other sciences than psychology to experimental results based on small samples will readily occur to the reader. The conclusions may be correct or incorrect, but they are unquestionably based on an

* 'Biometrika,' vol. VI., pp. 305–7.

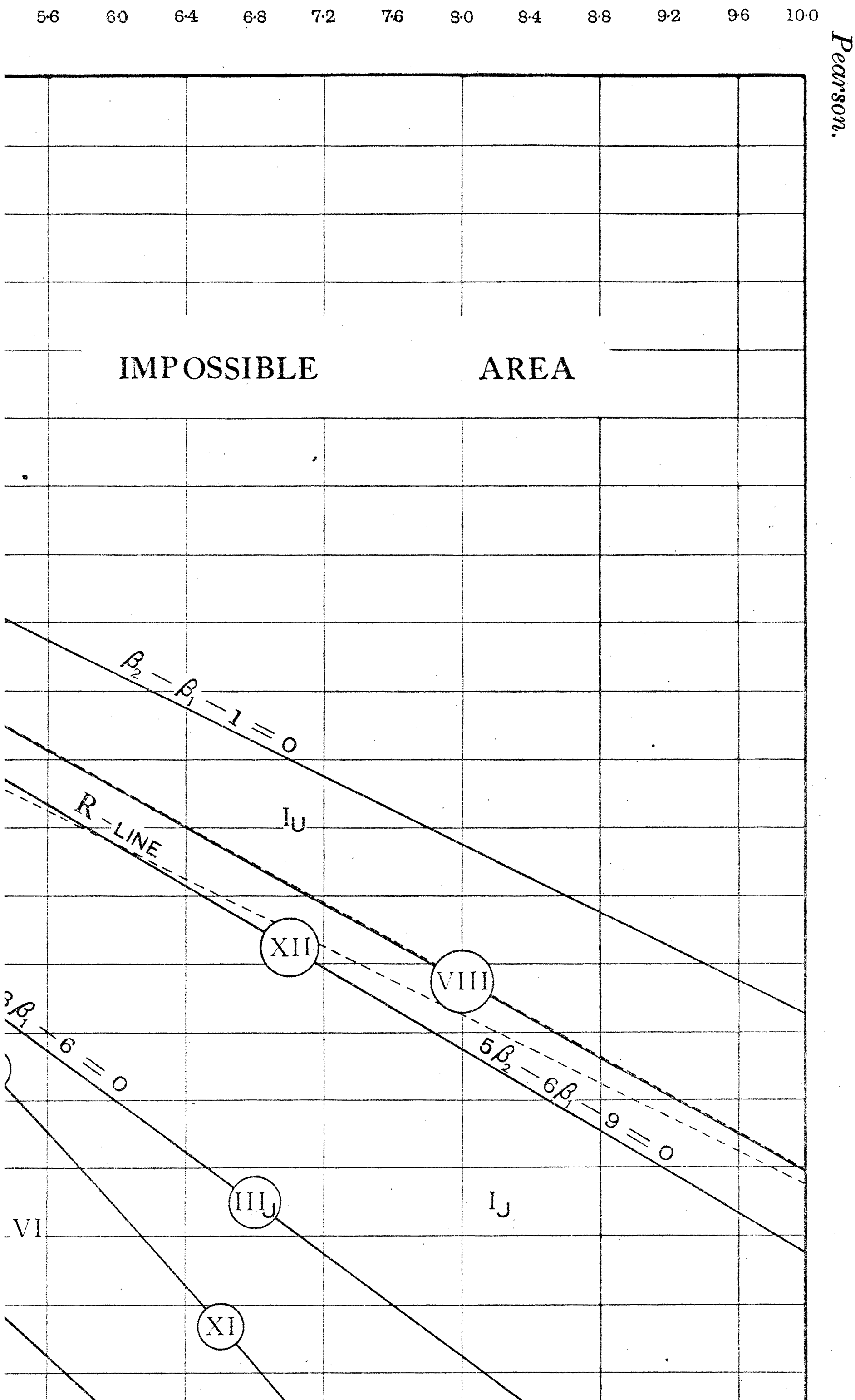
TYPES OF SKEW FREQUE

OF β_1 AND

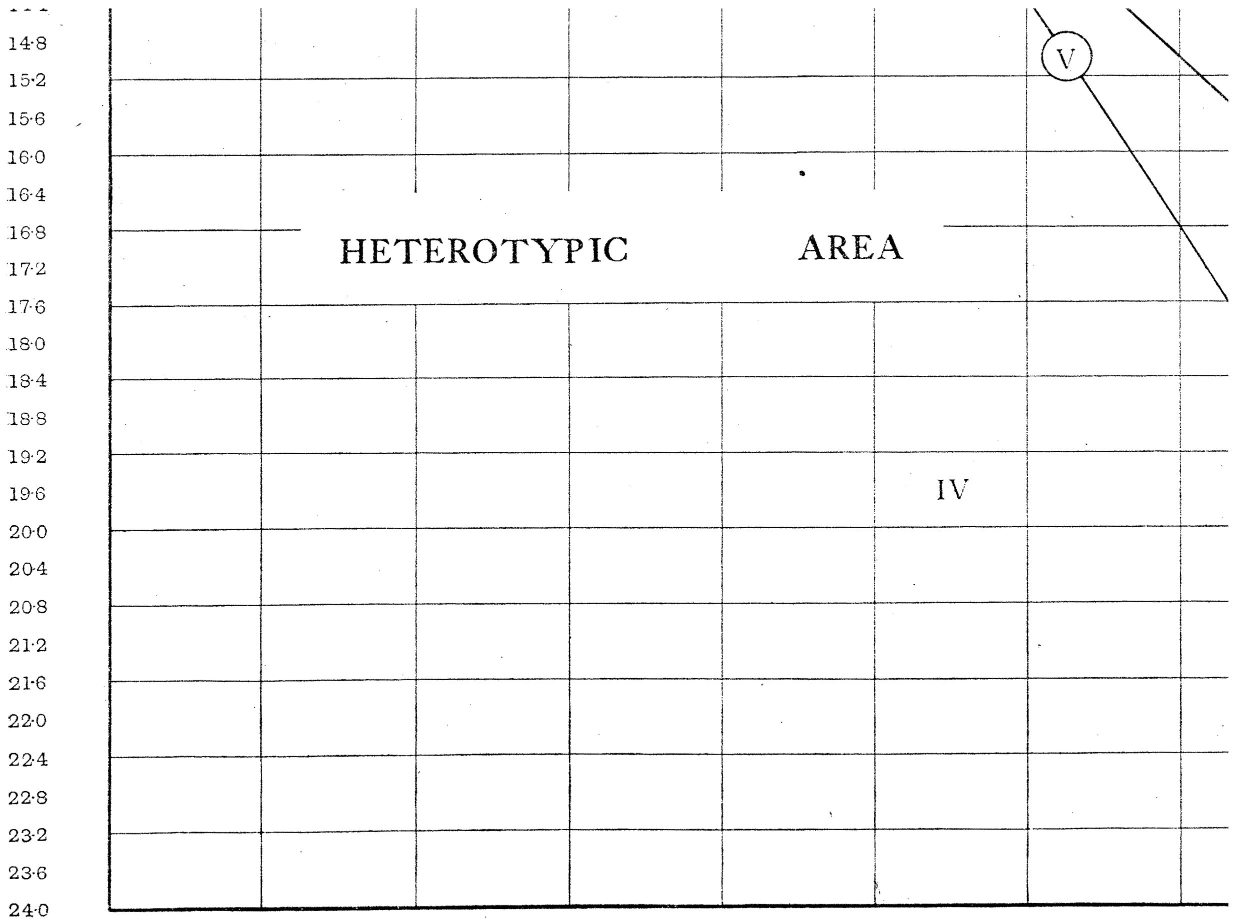


JENCY FOR VALUES

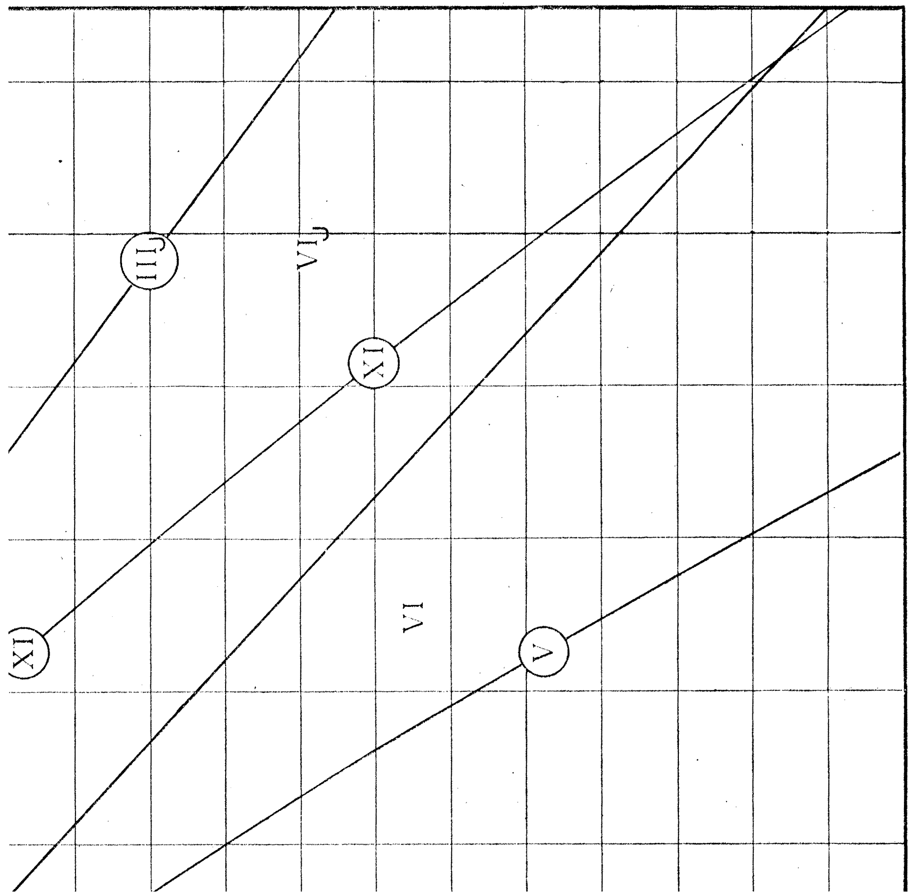
D β_2



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES



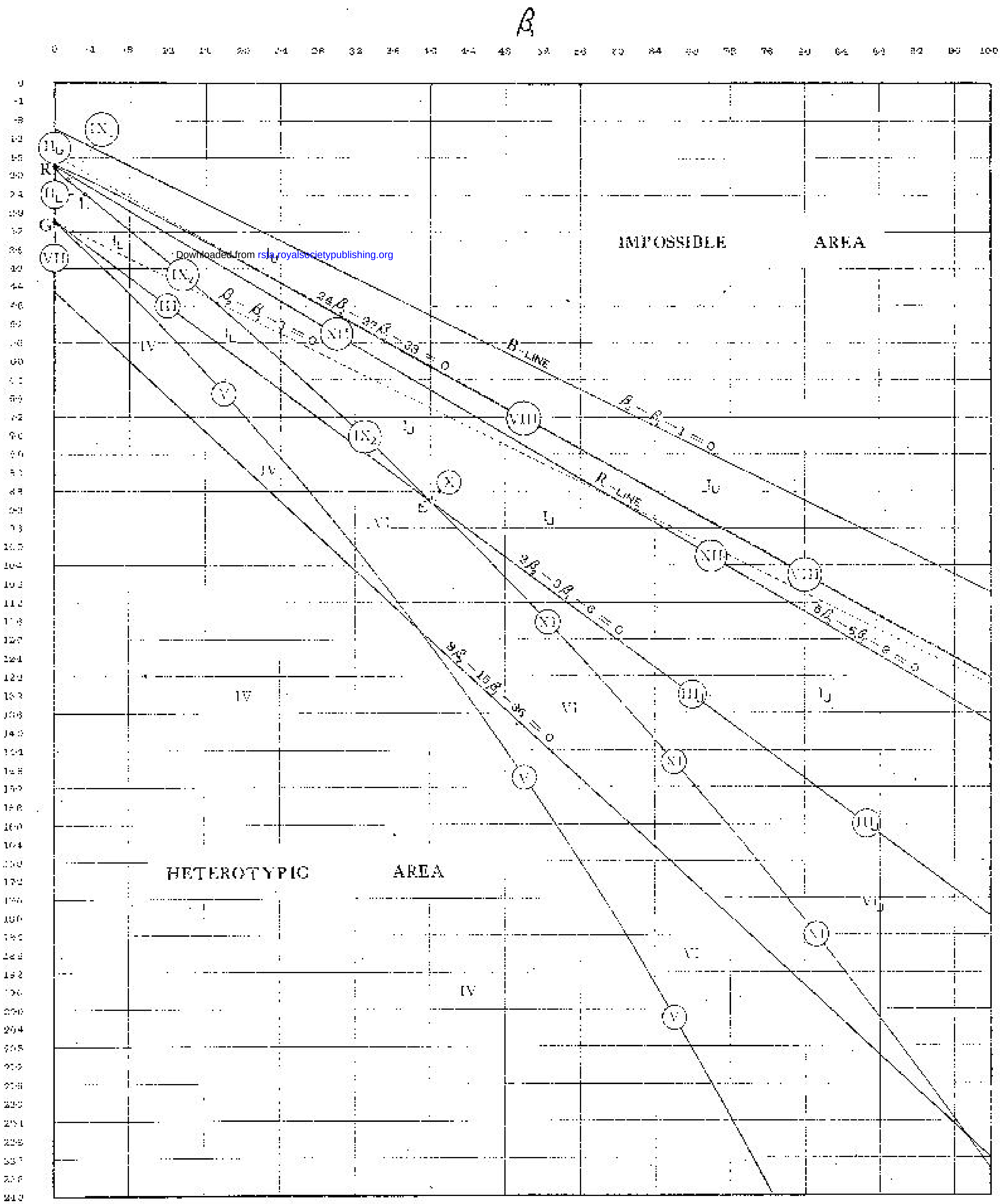
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES


XI
III J
VI J
XI
VI
V

inflation of the Gaussian point, G, to cover all that may be happening in the whole area of possible β_1, β_2 points in our diagram. It cannot at present be too often emphasised that such inflation is illegitimate, and that, as Dr. ISSERLIS has recently indicated,* the assumption that the distribution curves of statistical constants follow the Gaussian curve is not legitimate, especially in the case of "small samples," which not only for many commercial purposes, *e.g.*, experimental brewing, but in numerous branches of science, *e.g.*, psychology, astronomy, and even physics, are all that economy of money or time permits of being recorded.

* 'Roy. Soc. Proc.,' A, vol. 92, p. 23.

TYPES OF SKEW FREQUENCY FOR VALUES OF β_1 AND β_2



Pearson.