Systems of Frequency Curves Generated by Methods of Translation
Author(s): N. L. Johnson
Source: *Biometrika*, Vol. 36, No. 1/2 (Jun., 1949), pp. 149-176
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2332539
Accessed: 05-01-2025 12:21 UTC

SYSTEMS OF FREQUENCY CURVES GENERATED BY METHODS OF TRANSLATION

By N. L. JOHNSON

1. Introduction

1·1. *Preliminary remarks*

This paper is concerned with the discussion of some of the uses which may be made of transformations of variables such that the transformed variables may be considered to have a normal distribution. The concept of such transformations was put forward by Edgeworth (1898) and termed by him the Method of Translation. Edgeworth considered, in fact, only transformations which could be represented by polynomials, as did Kapteyn (1903). Later, however, Kapteyn & Van Uven (1916), Wicksell (1917) and Rietz (1922) extended the method to more general kinds of transformation. As we shall see later, the particular case of the logarithmic transformation, which is given some prominence in each of the last two references, had been anticipated by other authors who had not, however, considered the transformation as more than a special device applicable to particular cases.

1·2. *Historical development*

It is of interest to consider the reasons why the need for such transformations should have arisen. There is no doubt that in the earlier phases of their development the primary object was that of graduating observed frequency distributions. The normal distribution had played a dominant role in both theoretical and applied statistics since the time of Laplace. It was, however, apparent that the normal curve could not provide an adequate representation of many of the distributions encountered in statistical practice. Towards the end of the nineteenth century attempts were made to construct systems of frequency curves which should be capable of representing a wider variety of distributions than those for which a normal curve would suffice. It may be noted that the most obvious departure from normality was that which is described as skewness, and that much of the work at this time was described as the construction of systems of 'skew frequency curves'. The most successful of the systems then proposed have been those of K. Pearson (1895) and Charlier (1905). The work of Edgeworth and others, referred to in § 1·1, constitutes a third line of approach which, though not so widely used as those of Pearson and Charlier, has certain advantages of its own. In view of the important position occupied by the normal curve, it was, of course, natural to consider the possibility of relating observed distributions to the standard form. The fact that functions associated with the normal curve were well tabulated must also have been a strong contributory factor. An important reason for the lack of general acceptance of the method of translation is the fact that it became apparent that compared with the Pearson system the curves proposed covered only a very limited variety of shapes. A similar criticism might, of course, be directed at the Charlier system, though the latter system possesses advantages in respect of the aid which its analytic form offers to theoretical investigations.

The main purpose of this paper is to propose certain systems of curves derived by the method of translation, which, it is hoped, retain most of the advantages while eliminating some of the drawbacks of the systems first based on this method.

1·3. *Transformation to normality*

Subsequently to the construction of the systems of curves described in § 1·2, the normal distribution has gained added importance as a result of developments in statistical theory. In particular, the theory of significance tests and the associated probability distributions have been worked out much more thoroughly for normal populations than for other cases. Originally this may have been a consequence of the theoretical importance of the normal distribution, based on Laplace's theorem and the central limit theorems, but a factor of considerable importance is the simplicity of results based on normal populations. 'Normal theory', as it may be termed, is so much simpler than theory based on any general system of curves that it is of great importance to be able to use it if possible. To this end, two lines of inquiry have been put forward. E. S. Pearson (1931) and R. C. Geary (1947), *inter alia*, have considered the problem of how far normal theory may be invalidated by various kinds of departure from normality in the original distributions. The other approach is in effect an application of the method of translation. A function of the observed variable is sought which shall be, with sufficient approximation, a normal variable. Normal theory, with its simplicity and convenience, is then applied to the transformed variables. Curtiss (1943) gives a good critical summary of many of these methods. It may be noted that the interest, in these applications, lies in the significance tests to be applied and not in the creation of systems of frequency curves.

A further application of the method of translation is found in the approximate normalization of certain test criteria. In this case it is implicitly assumed that the original distribution is normal, and the method of translation is used to simplify certain parts of 'normal theory'. Examples are the Wilson-Hilferty (1931) transformation of χ^2, and the transformations proposed by Hotelling & Frankel (1938) and by Cornish & Fisher (1937).

1·4. *General theoretical background*

Pretorius (1930), in the course of a long paper dealing with non-normal distributions, remarks: 'The superiority of one frequency function over another depends rather on the success with which that function can be applied to graduate data than on the manner in which it originated.' This point of view has much to recommend it and, if accepted, absolves us from the necessity of providing a plausible probability theory basis for any proposed system of frequency curves. On the other hand, it must be remembered that the normal curve was first reached from probability theory rather than from the graduation of data. While, therefore, from a utilitarian point of view a probability theory basis is unnecessary, it is useful to keep the theory in mind when constructing new systems. For example, Pearson's fundamental differential equation was based on certain considerations of probability, though it was applied in cases where these considerations could hardly be presumed valid. Rather similarly the method of translation can be related to probability theory in a general and somewhat tentative manner. The argument, as described below, is due to Kapteyn (1903) and Wicksell (1917).

The normal distribution can be considered as arising from the summation of a large number of small independent effects which have occurred in a specified order. If it be now supposed that the magnitude of an effect be proportional to some function of the value of the variable before the addition of the effect, it can be shown that a certain function of the final variables should be normally distributed.

Suppose $x_1, x_2, \ldots$ to be independent random variables, each capable of taking only a small range of values near zero. The first sentence of the preceding paragraph can be interpreted as meaning that

$$X_n = x_1 + x_2 + \ldots + x_n \tag{1}$$

will be approximately normal if n is large. The second part of the paragraph means that if

$$Y_n = x_1 + x_2\, G(Y_1) + \ldots + x_n\, G(Y_{n-1}), \tag{2}$$

where G is some function, then it is possible to determine a function $f(Y_n)$ which is approximately normally distributed. This would be the case if

$$f(Y_n) = x_1 + x_2 + \ldots + x_n.$$

Now if this be so,

$$f(Y_n) - f(Y_{n-1}) = x_n. \tag{3}$$

From (2),

$$Y_n - Y_{n-1} = x_n\, G(Y_{n-1}). \tag{4}$$

Hence

$$\frac{f(Y_n) - f(Y_{n-1})}{Y_n - Y_{n-1}} = \frac{1}{G(Y_{n-1})}.$$

Since x_n is supposed small, it follows that

$$f'(Y) \doteqdot 1/G(Y). \tag{5}$$

Van Uven, in an Appendix to Kapteyn & van Uven (1916), and Baker (1934) have pointed out that it is always possible in theory to transform any continuous distribution into a normal distribution. Van Uven gives a graphic method of doing so, while Baker proposes an approximation based on the method of moments. In both cases, however, practical difficulties are considerable.

The parallelism of equation (5) and the equation for a function which shall 'equalize variances' (when the standard deviation is proportional to the function G of the expected value) is notable, and, in fact, equalization of variance and approximate normalization often go together. Curtiss (1943) gives a full discussion of these two aspects of certain transformations used in the analysis of variance.

Recently, in connexion with problems concerning the distribution of particle sizes, Kolmogoroff (1941), Halmos (1944) and Epstein (1947) have developed another theoretical basis leading, under certain conditions, to the most common form of the distributions which we shall consider (see § 3·1 below).

1·5. *Order of discussion*

In this paper we shall not consider in any further detail theoretical arguments for the use of transformations to normality. We shall be concerned, rather, with the study of the properties of distributions for which simple transformations to normality are possible.

In §§ 2·1–2·4 the problem will be considered in a general manner. Certain properties which are valid for wide classes of transformation will be described and a basis will be developed for the discussion of any special system. Three such special systems will be put forward, and their properties discussed, in §§ 3·1–3·6; bivariate distributions based on these systems will be considered in a later paper.

2. General theory

2·1. *Translation as a method of generating systems of frequency curves*

Any curve of the Pearson system of frequency curves is a solution of the differential equation

$$\frac{1}{y}\frac{dy}{dx} = -\frac{a+x}{c_0 + c_1 x + c_2 x^2},$$

and is defined by the values of the parameters a, c_0, c_1 and c_2 in that equation. Somewhat similarly, a curve in the Charlier A system is defined by the values of coefficients in the well-known expansion of derivatives of the normal function.

If we write a transformation of a variable x to normality in the formal manner

$$z = f(x),$$

where z is a unit normal variable, we have, clearly, defined a multiply infinite system of frequency curves, corresponding to the possible functions $f(x)$ which might be chosen. In order to obtain a system of curves analogous to the Pearson or Charlier systems, $f(x)$ must be specialized, preferably in a simple form, and made to depend on a certain number of parameters. The values of these parameters will then determine which curve of the system represents the distribution of x.

It is convenient to introduce four parameters (as in the case of the Pearson system) and to write

$$z = \gamma + \delta f\left(\frac{x-\xi}{\lambda}\right). \tag{6}$$

Here f is, preferably, a function of simple form, depending on no variable parameters. $f\{(x-\xi)/\lambda\}$ should also be a monotonic function of x. Without loss of generality it will be supposed that $f\{(x-\xi)/\lambda\}$ is a non-decreasing function of x and that δ and λ are positive.

From quite general considerations, it is possible to appreciate the roles played by certain of the parameters. If we write

$$y = (x-\xi)/\lambda,$$

then

$$z = \gamma + \delta f(y), \tag{7}$$

whence

$$p(y) = \delta f'(y)\, p(z)\,|_{z=\gamma+\delta f(y)} \tag{8·1}$$

$$= \frac{\delta}{\sqrt{(2\pi)}} f'(y) \exp\{-\tfrac{1}{2}[\gamma + \delta f(y)]^2\}. \tag{8·2}$$

Equation (8·1) is, of course, of general validity and does not depend on the definition of z as a unit normal variable. Since $x = \xi + \lambda y$, it follows that the distribution of x will be of the same shape as that of y, which is given, in general, by (8·1). The standard deviation of x will be λ times that of y, while changes in ξ will affect only the expected value (or other central measure) of the distribution of x.

It follows that the parameters γ and δ determine the shape of the distribution of x, that λ is a scale factor and ξ a location factor. It follows also that attention should be concentrated on the relation between the values of γ and δ and the distribution of x, since the parameters ξ and λ affect the distribution only in a simple manner. It will therefore be convenient to take as our standard form of transformation

$$z = \gamma + \delta f(y), \tag{7 bis}$$

rather than the more complicated expression (6), and to investigate the relation between γ, δ and the shape of the distribution of y.

2·2. *Requirements of a translation system*

The system of frequency curves obtained depends on the function $f(y)$ which is chosen. For practical convenience this function should possess the following properties:

(1) It should be a monotonic function of y.

(2) Apart from being simple in form it should be easy to calculate. Preferably, tables of the function should be in existence.

(3) The range of values of $f(y)$ corresponding to the actual range of possible values of y should be from $-\infty$ to $+\infty$. Although good approximation may sometimes be obtained even when this requirement is ignored, it is highly desirable that it should be satisfied, since z, being a normal variable, is supposed to vary from $-\infty$ to $+\infty$.

(4) The resulting system of distributions of y (and so of x) should include distributions of most, if not all, of the kinds encountered in collected data.

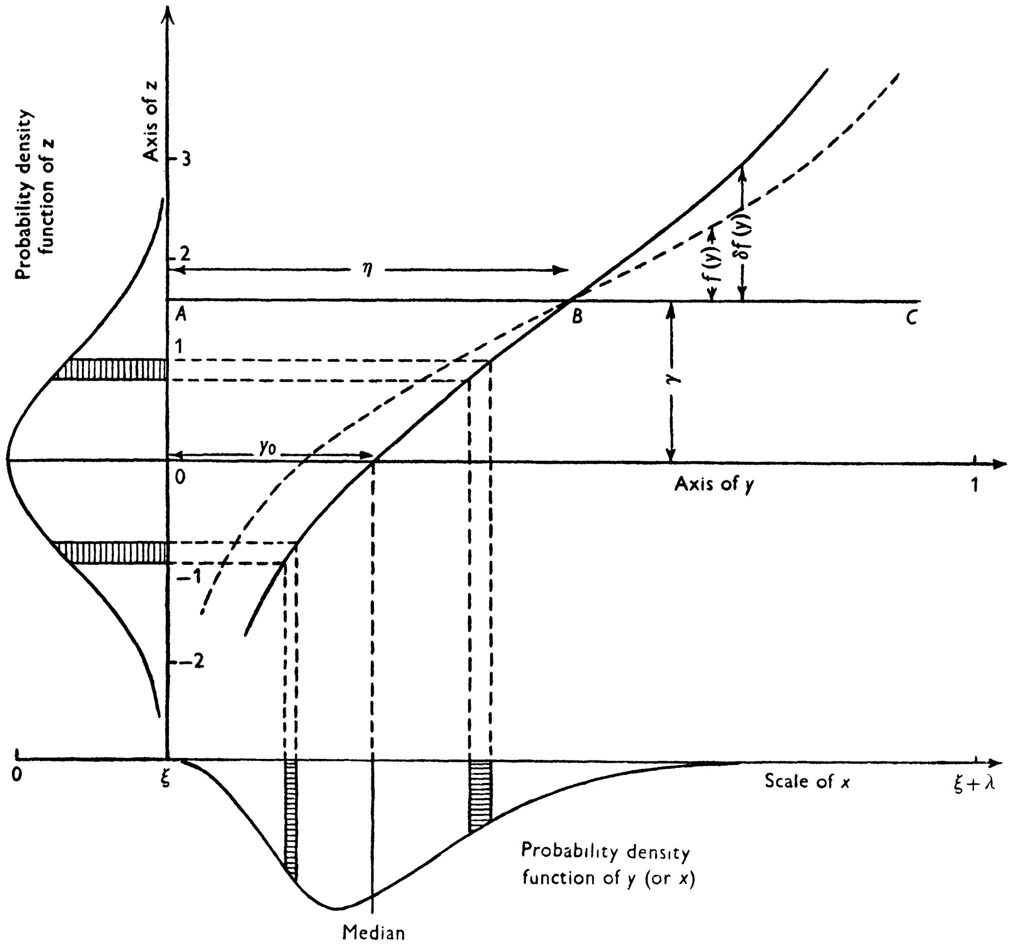

Fig. 1

2·3. *General properties of translation systems*

In this section we shall study the transformation $z = \gamma + \delta f(y)$, remembering that x is related to y by the linear equation $x = \xi + \lambda y$. We shall suppose in the first place that z is a standardized variable with a symmetrical distribution. The general properties of the relationship can be most easily appreciated with the help of Fig. 1.* This diagram actually represents the case where z is a standardized normal variable and the function f has the form $\log\{y/(1-y)\}$, but it illustrates the general properties of the transformation.

Relatively to the base-line ABC, which is parallel to and at a distance γ from the axis of y, the dotted curve has been plotted with ordinates $f(y)$ and abscissae y. For the solid-line

* I am indebted to Prof. E. S. Pearson for suggesting the use of this diagram.

curve the ordinates, measured from ABC, have been multiplied by δ. As a result, when referred to the axis of y and z, the solid-line curve represents the functional relation between y and z.

The effect of the distortion of the z-scale, due to this relationship, on the distribution of y (or of x) is also illustrated. The shaded columns, equal in area, under the two distribution curves, represent the probabilities of z and y (or x) falling in corresponding small intervals δz and δy (or δx). Clearly where $f'(y)$ has a high value, the contraction on the y-scale due to the transformation is greater than where $f'(y)$ is smaller. The values of γ and δ affect the distribution of y in so far as they determine over what parts of the total range of y these augmentations and diminutions of probability density shall occur, and to which parts of the distribution of z they shall correspond.

As δ is increased, it is seen that the range within which observations are likely to be found (e.g. corresponding to $-3 < z < 3$) will correspond to a smaller and smaller length of the dotted curve representing $f(y)$, which in the limit may be regarded as linear. Thus if y_0 be defined by

$$0 = \gamma + \delta f(y_0), \tag{9}$$

it is seen that y_0 will be the median of the y distribution and, further, if δ be sufficiently large, we shall have to a close degree of approximation

$$z \doteqdot \delta(y - y_0) f'(y_0) \tag{10}$$

for the bulk of the distribution of y. (10) may be written

$$y \doteqdot y_0 + z/\delta f'(y_0). \tag{11}$$

Hence if δ is large, y will have a distribution of approximately the same shape as z. We also note from (11) that an increase in δ may be expected to decrease the standard deviation of y.

We shall now restrict ourselves to a special class of transformation functions. A transformation will be called *symmetrical* if there is a unique number η such that

$$f(\eta + y') = -f(\eta - y')$$

for all y'.* It follows that $f(\eta) = 0$. For symmetrical transformations, therefore, $y_0 = \eta$ if $\gamma = 0$; further, if $\gamma = 0$, the distribution of y is symmetrical about η since the changes in probability density are symmetrical about the centre of the distribution of z. If γ is not zero, the distribution of y is skew. The parameter γ is thus particularly associated with skewness. In general, however, δ also affects skewness, and γ affects the kurtosis. As suggested by Fig. 1, other relations may be traced between (a) the shape of the distribution of y, and (b) the form of $f(y)$ and the magnitude and sign of γ.

2·4. *Fitting and errors in fitting*

The methods of fitting curves in general use are

(i) The method of percentile points.
(ii) The method of moments.
(iii) The method of maximum likelihood.

Method (i) is peculiarly suitable for fitting curves of a translation system. The percentile points of the distribution of y can easily be expressed in terms of the corresponding points of the distribution of z, and these latter will usually be tabled.

* The transformation shown in Fig. 1, $f(y) = \log\{y/(1-y)\}$, is symmetrical about $\eta = \frac{1}{2}$.

Should the moments of y be of fairly simple form, method (ii) may be used. If all four parameters γ, δ, ξ and λ are to be estimated, γ and δ are first determined from $\beta_1(x)$ and $\beta_2(x)$; then ξ and λ are determined so that agreement in mean and standard deviation is obtained.

Method (iii) is rather difficult to apply to translation systems. However, a method of successive approximation can be worked out which, though tedious, is straightforward and applicable to all cases of transformation to normality.

Although the process of fitting reduces to the estimation of γ, δ, ξ and λ, the accuracy of these estimates is usually of less intrinsic interest than the accuracy of probabilities (or expected frequencies) calculated from the fitted curve. For a given form of distribution of z, it is a simple matter to investigate the variation in computed probabilities associated with variation in the values assigned to the parameters, assuming that the correct form of function $f\{(x-\xi)/\lambda\}$ has been chosen. A brief study of the effect of an incorrect choice of this function in certain special cases will be given in § 3·6.

3. Special systems

3·1. *The log-normal system*

The most common transformation of type (6) is that termed by Gaddum (1945) the log-normal transformation.

If
$$z = \gamma + \delta \log\left(\frac{x-\xi}{\lambda}\right) \tag{12·1}$$
or
$$z = \gamma + \delta \log y, \tag{12·2}$$
z being a unit normal variable, the distribution of x (or of y) is said to be log-normal.

The transformation was proposed by Galton (1879), anticipating the form of argument used by Kapteyn, and some properties of the distribution were obtained by McAlister (1879). Fechner (1897) also used the transformation in a special application, but the idea was not further pursued by these authors. Kapteyn & Van Uven (1916) gave a graphical method of fitting the distribution and investigated its shape. Wicksell (1917) dealt rather more fully with the subject. He pointed out that the log-normal transformation is obtained by putting $G(Y) = Y$ in (5); that is to say, by assuming random increments proportional to the variable to which they apply. Wicksell also obtained the moments of the distribution of y. We have

$$\begin{aligned}\mu_r'(y) &= \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{\{r(z-\gamma)\}/\delta}\, e^{-\frac{1}{2}z^2}\,dz \\ &= e^{\frac{1}{2}r^2\delta^{-2}-r\gamma\delta^{-1}}.\end{aligned} \tag{13}$$

It follows that
$$\left.\begin{aligned}\beta_1 &= (\omega-1)(\omega+2)^2 \quad (\sqrt{\beta_1}>0),\\ \beta_2 &= \omega^4+2\omega^3+3\omega^2-3,\end{aligned}\right\} \tag{14}$$

where $\omega = e^{\delta^{-2}}$. The (β_1, β_2) points for log-normal distributions therefore lie on a curve defined by the parametric equations (14). This curve is shown in Fig. 2. This restriction of the locus of (β_1, β_2) is to be expected since (12·1) can be written

$$z = (\gamma - \delta\log\lambda) + \delta\log(x-\xi),$$

so that there are only three independent parameters and, without any loss of generality, (12·1) may be rewritten
$$z = \gamma + \delta\log(x-\xi). \tag{15}$$

Wicksell also proposed a method of fitting log-normal distributions based on the observed moments m_1', m_2 and m_3 of the distribution of x. The positive root of the equation

$$t^3 + 3t - \sqrt{b_1} = 0 \tag{16}$$

(where $\sqrt{b_1} = m_3/m_2^{\frac{3}{2}}$) is found. ξ, in (15), is then estimated by means of the formula $\xi = m_1' - \sqrt{m_2}/t$. Estimates of γ and δ are then obtained quite straightforwardly from (13). Yuan (1933) gave tables to facilitate the solution of (16). Quensel (1945) has given expressions for the standard deviations of estimates obtained by this method. Finney (1941) pointed out that the mean and variance of the transformed variable $\log(x-\xi)$ should be used if efficient estimates of γ and δ are to be obtained, and obtained expressions for such estimates. These could not of course be applied directly if ξ is unknown.

If y is log-normal
$$p(y) = \frac{\delta}{\sqrt{(2\pi)}\,y} e^{-\frac{1}{2}(\gamma+\delta\log y)^2} \quad (0 < y). \tag{17}$$

Yuan pointed out that this distribution has infinitely high contact at either end of its range of variation, since

$$\lim_{y\to 0} y^{-n}p(y) = 0 = \lim_{y\to\infty} y^n p(y)$$

for all values of n.

The log-normal system has proved useful in a number of applications. We may mention its use in dosage-mortality problems (e.g. Gaddum, 1945) in the graduation of economic data (Gibrat, 1931; Frechet, 1945) and in agriculture (Cochran, 1938). Williams (1937, 1940) has applied the system to a varied collection of problems.

3·2. *Extension of the logarithmic type of transformation*

Despite its successful application in a number of cases, the log-normal system is restricted in flexibility, just as is Pearson's Type III distribution, because the associated (β_1, β_2) point must lie on the curve defined by equation (14). It seems reasonable to suppose that useful extensions of the system might be obtained by using different functions $f(y)$ in (7) (or $f\{(x-\xi)/\lambda\}$ in (6)). We shall now consider the construction of such new systems, and will start by laying down certain properties which it appears desirable that they should possess.

(i) In order to avoid a restricted locus of variation for (β_1, β_2), the function f should be such that in equation (6) there are four truly independent parameters.

(ii) The new systems should fit in naturally with the log-normal system which could be regarded as a transition form, lying between two systems of distributions, one with a range of variation bounded at both extremities, the other unbounded at either extremity. By analogy with the Pearson system, it is to be expected that in the (β_1, β_2) plane the system with a bounded range of variation will cover the region between the log-normal line, and the limiting line $\beta_2 - \beta_1 - 1 = 0$; while the other system will cover the remainder of the (β_1, β_2) plane.

These regions are indicated in Fig. 2, wherein are also introduced the symbols S_L for 'log-normal system', S_B for 'bounded system' and S_U for 'unbounded system', which will be used in the remainder of this paper. It may be noted that the scheme of curves sketched above is not strictly analogous to the Pearson system, as in the latter there is a *region* in the (β_1, β_2) plane corresponding to range of variation bounded at one end only (Type VI).

(iii) Finally, in the choice of the function $f(y)$ the considerations detailed in § 2·2 must be kept in mind, and the requirements therein scheduled satisfied as far as is possible.

3·3. *Choice of new transformation functions*

Consider the log-normal variable x, defined by

$$z = \gamma + \delta \log (x - \xi) \quad (\xi < x).$$

Putting $y = 1 - \xi/x$, we have

$$z = (\gamma + \delta \log \xi) + \delta \log \frac{y}{1-y} \quad (0 < y < 1). \tag{18}$$

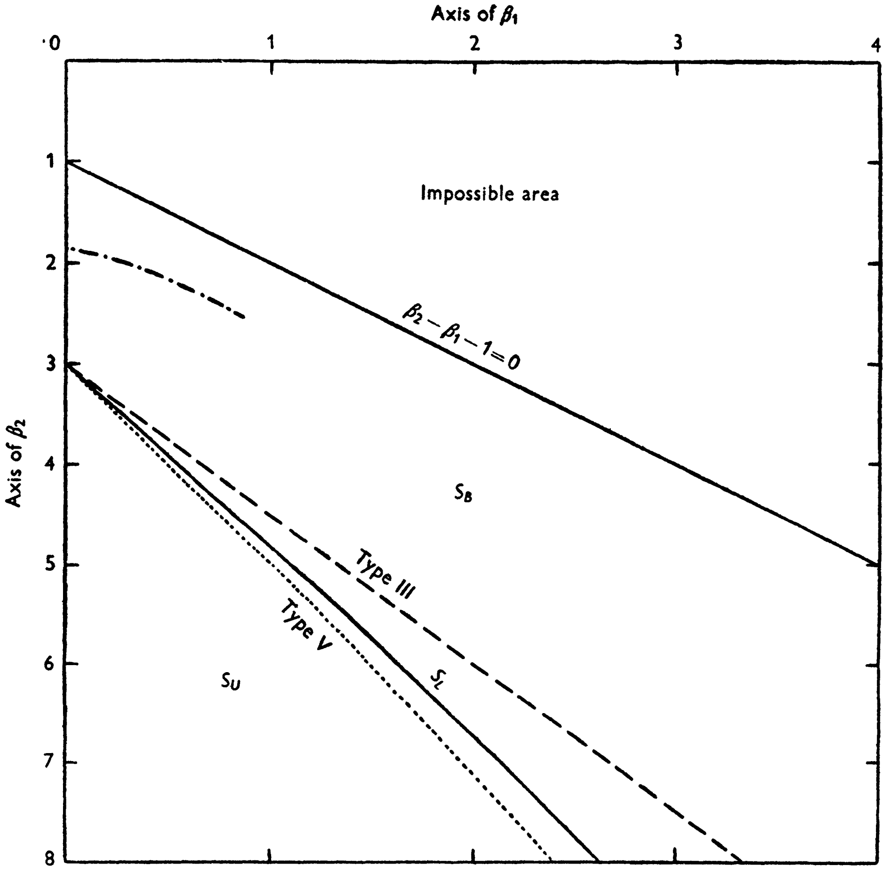

Fig. 2

The log-normal line is marked S_L

— — — — — — — Pearson Type III.
. . . . Pearson Type V.
.—.—.—.—.— Boundary of bimodal curves of system S_B.

A transformation of this type is also obtained from the general formula (7) by putting $f(y) = \log\{y/(1-y)\}$. Putting $y = (x-\xi)/\lambda$, we have as a particular case of (6)

$$z = \gamma + \delta \log \frac{x - \xi}{\xi + \lambda - x} \quad (\xi < x < \xi + \lambda). \tag{19}$$

The system of curves generated by (18) or (19), z being a unit normal variable, will be our system S_B.

We may note that the proposed function $f(y)$ satisfies the conditions laid down in § 2·2, that it should be simple and calculable without undue difficulty. In fact

$$f(y) = \log\frac{y}{1-y} = 2\tanh^{-1}(2y-1). \tag{20}$$

Tables of inverse hyperbolic tangents (Milne-Thompson & Comrie, 1931) may therefore be used to evaluate $f(y)$. We further note that $f(y)$ has the desirable property that it increases from $-\infty$ to $+\infty$ as y increases from 0 to 1.

The transformation (19) was suggested by Wicksell (1917) as possibly being worthy of study. Bartlett (1937) has stated that (19) proves useful in certain analysis of variance problems. It may also be noted that Fisher's z' transformation for the correlation coefficient is of form (19) with

$$\xi = -1, \quad \lambda = 2, \quad \gamma \doteqdot -\tfrac{1}{2}\sqrt{(n-3)}\log\frac{1+\rho}{1-\rho}, \quad \delta \doteqdot \tfrac{1}{2}\sqrt{(n-3)},$$

where n = sample size, ρ = population correlation coefficient and $x = r$, the sample correlation.

The construction of our system S_U is rather more arbitrary, though suggested by analogy with S_L and S_B. The transformation by which we shall define S_U is

$$z = \gamma + \delta\sinh^{-1}\left(\frac{x-\xi}{\lambda}\right) = \gamma + \delta\log\left\{\frac{x-\xi}{\lambda} + \sqrt{\left[\left(\frac{x-\xi}{\lambda}\right)^2 + 1\right]}\right\} \tag{21}$$

or

$$z = \gamma + \delta\sinh^{-1}y = \gamma + \delta\log\left[y + \sqrt{(y^2+1)}\right]. \tag{22}$$

Milne-Thompson & Comrie's tables of inverse hyperbolic sines may be used to evaluate $f(y)$ in this case. As required in § 2·2, $f(y)$ increases from $-\infty$ to $+\infty$ as y increases from $-\infty$ to $+\infty$. Beall (1942) and Bartlett (1947) have suggested the use of the function $\sinh^{-1}\sqrt{y}$ or $\sinh^{-1}\sqrt{(y+\frac{1}{2})}$, especially with reference to negative binomial variables.

3·4. *The system S_B*

From (19) we have immediately

$$p(y) = \frac{\delta}{\sqrt{(2\pi)}}\frac{1}{y(1-y)}\exp\left[-\frac{1}{2}\left\{\gamma + \delta\log\frac{y}{1-y}\right\}^2\right] \quad (0<y<1). \tag{23}$$

Hence
$$p(y) = \frac{\delta e^{-\frac{1}{2}\gamma^2}}{\sqrt{(2\pi)}}y^{-\frac{1}{2}\delta^2\log y-(\gamma\delta+1)}(1-y)^{-\frac{1}{2}\delta^2\log(1-y)+\gamma\delta-1}e^{-\delta^2\log y\log(1-y)},$$

so that
$$\lim_{y\to 0} y^{-n}p(y) = 0 = \lim_{y\to 1}(1-y)^{-n}p(y)$$

for any value of n. The distribution curve of y therefore has 'high contact' at either end of its finite range of variation.

Inverting (19) we have
$$y = (1+e^{-(z-\gamma)/\delta})^{-1}. \tag{24}$$

Hence the median value of y is $(1+e^{\gamma/\delta})^{-1}$. The equation to be satisfied by any modal value of y, other than the extremities of the range of variation, is

$$2y - 1 = \delta\left(\gamma + \delta\log\frac{y}{1-y}\right).$$

Putting
$$y = \tfrac{1}{2}(y'+1),$$

$$y' - \gamma\delta = \delta^2\log\frac{1+y'}{1-y'}. \tag{25}$$

The number of intersections of the straight line $u = y' - \gamma\delta$ and the curve

$$u = \delta^2 \log \frac{1+y'}{1-y'} \tag{25·1}$$

determines whether the distribution of y is or is not bimodal. If there is only one intersection, the distribution is unimodal, if there are three intersections it is bimodal. Supposing, for the moment, that $\gamma > 0$, there is clearly one intersection in the interval $-1 < y' < 0$. There may be two other intersections. These must be in the interval $0 < y' < 1$ if they exist. In the limiting case, the straight line in (25·1) will touch the curve at some point in the interval $0 < y' < 1$. At this point the slopes of line and curve must be equal, so that

$$1 = 2\delta^2(1-y'^2)^{-1}, \quad \text{i.e.} \quad y' = \sqrt{(1-2\delta^2)}.$$

Hence the line will touch the curve at this value of y' if

$$\sqrt{(1-2\delta^2)} - \gamma\delta = 2\delta^2 \tanh^{-1}\sqrt{(1-2\delta^2)},$$

i.e.

$$\gamma = \delta^{-1}[\sqrt{(1-2\delta^2)} - 2\delta^2 \tanh^{-1}\sqrt{(1-2\delta^2)}].$$

It follows that the necessary and sufficient conditions for bimodality (whatever the sign of γ) are

$$\delta < 1/\sqrt{2}, \quad |\gamma| < \delta^{-1}\sqrt{(1-2\delta^2)} - 2\delta \tanh^{-1}\sqrt{(1-2\delta^2)}. \tag{26}$$

Table 1 shows the limiting values of $|\gamma|$ for various values of δ.

Table 1

δ	Maximum $\lvert\gamma\rvert$	δ	Maximum $\lvert\gamma\rvert$
0·7	0·0027	0·3	2·12
0·6	0·175	0·2	4·02
0·5	0·533	0·1	9·37
0·4	1·12		

Figs. 3 and 4 show the limiting curves $\gamma = 0$, $\delta = 1/\sqrt{2}$ and $\gamma = 0·533$, $\delta = 0·5$. Clearly these limiting curves will have a nearly flat horizontal portion with an inflexion at $y' = \sqrt{(1-2\delta^2)}$; i.e. at $y = \frac{1}{2}[1+\sqrt{(1-2\delta^2)}]$. We also note that, if δ be fixed, as γ increases, the 'permanent' mode is always below $y = \frac{1}{2}[1-\sqrt{(1-2\delta^2)}]$, the anti-mode, when present, is between $\frac{1}{2}[1-\sqrt{(1-2\delta^2)}]$ and $\frac{1}{2}[1+\sqrt{(1-2\delta^2)}]$ and the secondary mode is above $\frac{1}{2}[1+\sqrt{(1-2\delta^2)}]$. Fig. 5 shows a symmetrical bimodal distribution, while Figs. 6, 7 and 8 show typical unimodal curves of S_B. The boundary above which (β_1, β_2) points correspond to bimodal curves has been shown in Fig. 2, as far as it has been explored.

The moments of the distribution of y are complicated in form. They are discussed in the Appendix, where some numerical values are given. It is also shown in the Appendix that the (β_1, β_2) points of curves of the system S_B cover the area between the log-normal line and the straight line $\beta_2 - \beta_1 - 1 = 0$. Until sufficiently comprehensive tables are available, it is clear that it will not be possible to fit curves of types S_B by the method of moments. For practical purposes the method of percentiles is the most convenient to use, though in certain special cases it will be possible to apply the method of maximum likelihood in quite a simple manner.

The process of fitting is considerably simplified if one or both end-points of the distribution of x are known. We shall deal with the three cases of (*a*) both, (*b*) one, (*c*) neither of the end-points known.

(*a*) *Both end-points known*

In this case, both ξ and λ are known. Hence, given the value of x, the value of the transformed variable $\log(x-\xi)/(\xi+\lambda-x)$ can be obtained directly. Corresponding to observed values $x_1, x_2, \ldots, x_n$ there will be transformed values $f_1, f_2, \ldots, f_n$, where

$$f_i = \log\frac{x_i-\xi}{\xi+\lambda-x_i} \quad (i = 1, \ldots, n). \tag{27}$$

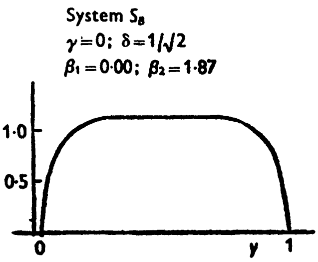

Fig. 3

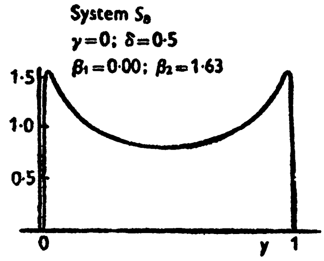

Fig. 5

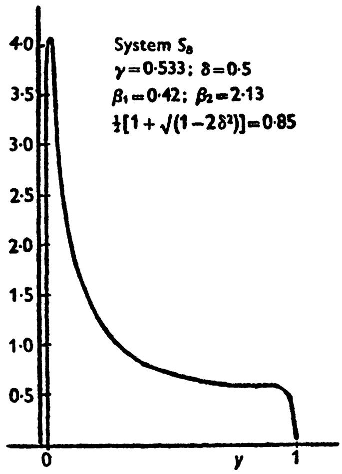

Fig. 4

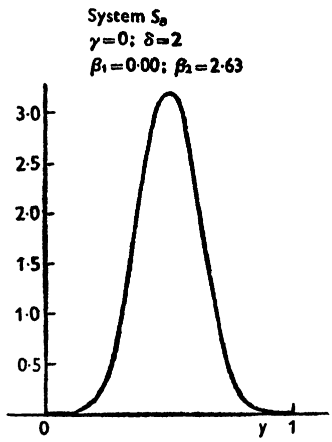

Fig. 6

The problem then reduces to that of fitting a normal curve to the observed f_i's. Fitting this curve by moments we have

$$\hat{\gamma} = -\bar{f}/s_f, \quad \hat{\delta} = 1/s_f,$$

where
$$\bar{f} = n^{-1}\sum_{i=1}^{n} f_i, \quad s_f^2 = n^{-1}\sum_{i=1}^{n}(f_i-\bar{f})^2.$$

This will give the maximum likelihood estimates for γ and δ.

However, a difficulty arises if the original data are not given *in extenso* but as a grouped distribution. If the original groups (for the variable x) are of equal length, the transformed groups (for the variable f) will be of unequal length and there will be groups of infinite length at either end of the distribution. Moments calculated from such data would require corrections which would be difficult to ascertain. The method of percentiles is very simple to apply in this case. An application of the method is described in Example 1 (pp. 168, 169 below).

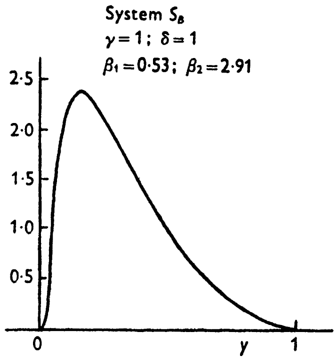

Fig. 7

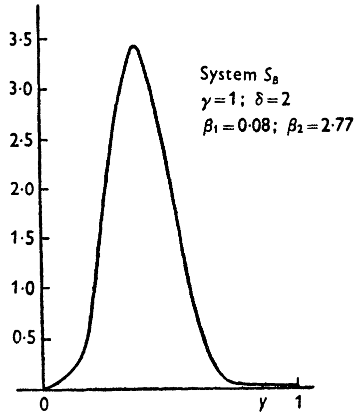

Fig. 8

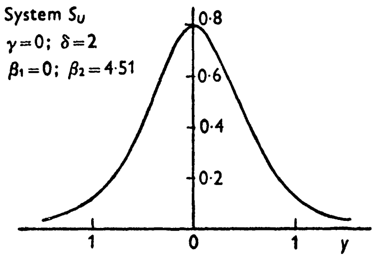

Fig. 9

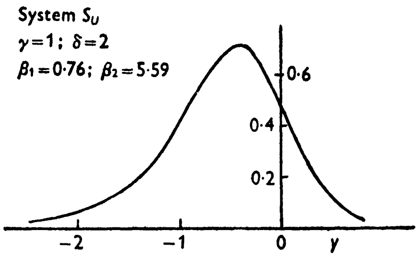

Fig. 10

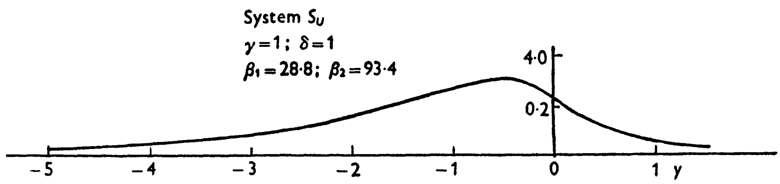

Fig. 11

(*b*) *One end-point known*

Suppose the lower end-point, i.e. ξ, to be known. In this case a convenient application of the method of percentiles to estimate λ, γ and δ is as follows:

From the data we estimate the median $\hat{x}_0$, and the lower and upper $100P\,\%$ points $\hat{x}_1$ and $\hat{x}_2$. Then we have to estimate λ, γ and δ from the equations

$$\left.\begin{aligned} -z_P &= \hat{\gamma}+\hat{\delta}\log\frac{\hat{x}_1-\xi}{\xi+\hat{\lambda}-\hat{x}_1},\\ 0 &= \hat{\gamma}+\hat{\delta}\log\frac{\hat{x}_0-\xi}{\xi+\hat{\lambda}-\hat{x}_0},\\ z_P &= \hat{\gamma}+\hat{\delta}\log\frac{\hat{x}_2-\xi}{\xi+\hat{\lambda}-\hat{x}_2},\end{aligned}\right\} \tag{28}$$

where
$$\frac{1}{\sqrt{(2\pi)}}\int_{z_P}^{\infty} e^{-\frac{1}{2}t^2}\,dt = P.$$

From equations (28) we obtain

$$\frac{(\hat{x}_0-\xi)^2}{(\xi+\hat{\lambda}-\hat{x}_0)^2}=\frac{(\hat{x}_1-\xi)(\hat{x}_2-\xi)}{(\xi+\hat{\lambda}-\hat{x}_1)(\xi+\hat{\lambda}-\hat{x}_2)}, \tag{29}$$

whence

$$\hat{\lambda}=\frac{X_0(X_0X_1+X_0X_2-2X_1X_2)}{X_0^2-X_1X_2}, \tag{30}$$

where

$$X_i=\hat{x}_i-\xi \quad (i=0,1,2).$$

$\hat{\gamma}$ and $\hat{\delta}$ may then be found from (28). Alternatively, using the value of $\hat{\lambda}$ obtained from (30), γ and δ may be estimated from the observed mean and standard deviation of

$$\log(x-\xi)/(\xi+\hat{\lambda}-x),$$

as in case (*a*).

(*c*) *Neither end-point known*

In this case all four parameters ξ, λ, γ and δ have to be estimated. The method of percentile points in this case requires that estimates be obtained of four values x_A, x_B, x_C, x_D say, such that certain fixed proportions P_A, P_B, P_C, P_D, respectively, of the distribution of x fall below these values. $\hat{\xi}$, $\hat{\lambda}$, $\hat{\gamma}$ and $\hat{\delta}$ have then to be found from the equations

$$z_k=\hat{\gamma}+\hat{\delta}\log\frac{x_k-\hat{\xi}}{\hat{\xi}+\hat{\lambda}-x_k}, \tag{31}$$

where

$$\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{z_k}e^{-\frac{1}{2}t^2}dt=P_k. \tag{32}$$

These equations may be solved by successive approximation. In Example 2 (pp. 169–171 below) only an approximate solution has been obtained. The values shown could be improved by the standard method based on Taylor's expansion.

3·5. *The system S_U*

From (22) we have immediately

$$p(y)=\frac{\delta}{\sqrt{(2\pi)}}\frac{1}{\sqrt{(y^2+1)}}\exp\left[-\tfrac{1}{2}\{\gamma+\delta\log[y+\sqrt{(y^2+1)}]\}^2\right]. \tag{33}$$

Evidently $y^n p(y)\to 0$ as $y\to-\infty$ or $y\to+\infty$, so that there is 'high contact' at either end of the infinite range of variation of y.

Inverting (22) we have $y=\tfrac{1}{2}(e^{(z-\gamma)/\delta}-e^{-(z-\gamma)/\delta})=\sinh\left(\dfrac{z-\gamma}{\delta}\right)$. (34)

Hence the median value of y is $-\sinh(\gamma/\delta)$.

From (33) the equation to be satisfied by any modal value of y, other than the extremities of the range of variation, is

$$y/(1+y^2)=-\delta\{\gamma+\delta\log[y+\sqrt{(y^2+1)}]\}. \tag{35}$$

From graphical considerations it is evident that there is only one solution of (35), and that this solution is between the median and zero. Hence when γ is positive the mode is greater than the median, implying negative skewness, and vice versa. Since the transformation generating system S_U is symmetrical about $y=0$, and $f'(y)=(y^2+1)^{-\frac{1}{2}}$ is a decreasing function of $|y|$, this is to be expected.

Figs. 9–11 show typical curves of system S_U.

The moments of the system S_U are determined with much greater facility than are those of S_B. We have
$$\mu'_r = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\, 2^{-r}(e^{(z-\gamma)/\delta} - e^{-(z-\gamma)/\delta})^r\, dz.$$

Hence if r is even
$$\mu'_r = 2^{-(r-1)}\left\{\sum_{s=0}^{\frac{1}{2}r-1}(-1)^s\binom{r}{s}e^{\frac{1}{2}(r-2s)^2\delta^{-2}}\cosh\left[(r-2s)(\gamma/\delta)\right] + (-1)^{\frac{1}{2}r}\frac{1}{2}\binom{r}{\frac{1}{2}r}\right\}, \qquad (36\cdot1)$$
and if r is odd
$$\mu'_r = 2^{-(r-1)}\sum_{s=0}^{\frac{1}{2}(r-1)}(-1)^{s+1}\binom{r}{s}e^{\frac{1}{2}(r-2s)^2\delta^{-2}}\sinh\left[(r-2s)(\gamma/\delta)\right]. \qquad (36\cdot2)$$

From equations (36) it follows that
$$\left.\begin{aligned}
\mu'_1 &= -\omega^{\frac{1}{2}}\sinh\Omega,\\
\mu_2 &= \tfrac{1}{2}(\omega-1)(\omega\cosh 2\Omega+1),\\
\mu_3 &= -\tfrac{1}{4}\omega^{\frac{1}{2}}(\omega-1)^2\{\omega(\omega+2)\sinh 3\Omega+3\sinh\Omega\},\\
\mu_4 &= \tfrac{1}{8}(\omega-1)^2\{\omega^2(\omega^4+2\omega^3+3\omega^2-3)\cosh 4\Omega+4\omega^2(\omega+2)\cosh 2\Omega+3(2\omega+1)\},
\end{aligned}\right\} \qquad (37)$$
where
$$\omega = e^{\delta^{-2}}, \quad \Omega = \gamma/\delta.$$

From (37) we see that if γ is positive the inequalities mean < median < mode hold, while if γ is negative the direction of the inequalities is reversed. Also when $\gamma = 0$, $\beta_1 = 0$ (as should be the case) and $\beta_2 = \frac{1}{2}(\omega^4+2\omega^2+3)$. As γ tends to infinity, δ remaining fixed, we have
$$\lim_{\gamma\to\infty}\beta_1 = (\omega-1)(\omega+2)^2, \quad \lim_{\gamma\to\infty}\beta_2 = \omega^4+2\omega^3+3\omega^2-3. \qquad (38)$$
As γ increases from zero to infinity, therefore, the (β_1,β_2) point varies from $(0, \frac{1}{2}(\omega^4+2\omega^2+3))$ to a point on the S_L line.

As δ decreases from infinity to zero, ω increases from zero to infinity. Hence the (β_1,β_2) points for system S_U cover the region of the (β_1,β_2) plane 'below' the S_L line, as sketched in Fig. 2.

The calculations involved not being too lengthy, fairly extensive series of values of the mean, standard deviation, β_1 and β_2 for distributions of systems S_U were computed. They are not reproduced here, but Fig. 12 is an abac based on these calculations. Using this abac the parameters γ and δ can be estimated, and ξ and λ then determined to give the required mean and standard deviation (cf. Burr (1942)). Given γ and δ, it is not difficult to calculate ξ and λ from (37). No further tables or abacs are, therefore, given. In the case of system S_B, however, where μ'_1 and σ are not easily calculated, a second abac, giving μ'_1 and σ as functions of γ and δ, would be required.

It may be noted that it is in the system S_U that the necessity for estimation of all four parameters is likely to be of most frequent occurrence. In the case of S_L and S_B, ξ (also λ in S_B) often has an obvious and simple meaning. ξ (and λ) may often be fixed in advance in such cases. In general there is no such simple interpretation of ξ and λ in the case of S_U. There is, indeed, the particular result that for the symmetrical curves of S_U, $x = \xi$ is the axis of symmetry. Otherwise, the relation of ξ and λ to the position and size of the curves is not simple.

Examples 3 and 4 (pp. 171, 172 below) describe the fitting of curves of system S_U to observational data. It appears that the curves give a good approximation to Pearson Type IV curves. As S_U is much easier to deal with than Type IV, especially with regard to the computation of probabilities, it seems that S_U might be used as an approximation to Type IV even when the latter is considered the more reasonable curve to fit.

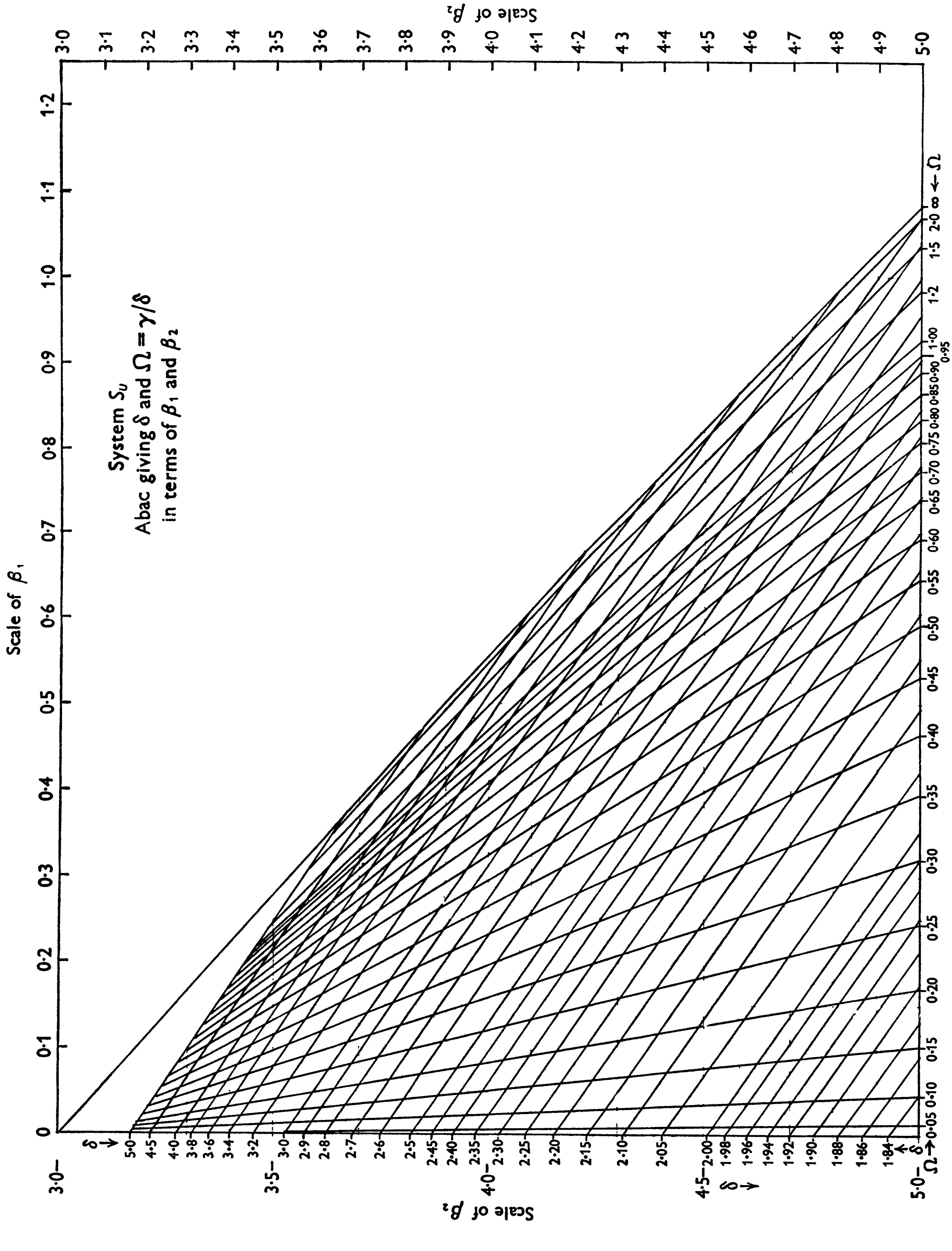
System S_U
Abac giving δ and $\Omega = \gamma/\delta$
in terms of β_1 and β_2
Scale of β_1
Scale of β_2
Scale of β_2

Fig. 12

3·6. *The transformation applied to certain Pearson curves*

Since experience has shown that curves of the Pearson system are representative of a wide range of frequency distributions met in practice, it is of interest to ask how far the application of the S_L, S_B and S_U transformations to variables following distributions of this system will result in a transformed variable following approximately the normal law. We shall suppose, then, that y follows a probability law of the Pearson system and compare $\beta_1(y), \beta_2(y)$ and $\beta_1(z), \beta_2(z)$ with the normal values 0, 3 respectively.

Some of the results obtained in this section coincide with those which Aroian (1941) and Wishart (1947) obtained in the course of work on the distribution of statistics employed in analysis of variance.

(i) S_L *applied to Type III*

If
$$p(y) = \frac{1}{\Gamma(\nu)} y^{\nu-1} e^{-y} \quad (0 < y < \infty),$$

then the cumulants of $z = \gamma + \delta \log y$ are

$$\kappa_1(z) = \gamma + \delta \Psi(\nu), \quad \kappa_r(z) = \delta^r \Psi^{(r-1)}(\nu) \quad (r \geqslant 2), \tag{39}$$

where $\Psi^{(s)}(\nu) = \dfrac{d^{s+1} \log \Gamma(\nu)}{d\nu^{s+1}}$ is the $(s+2)$-gamma function and we write $\Psi^{(0)}(\nu)$ as $\Psi(\nu)$. We have

$$\beta_1(z) = [\Psi^{(2)}(\nu)]^2 [\Psi^{(1)}(\nu)]^{-3}, \quad \beta_2(z) = 3 + \Psi^{(3)}(\nu) [\Psi^{(1)}(\nu)]^{-2}. \tag{40}$$

Using the asymptotic expansions for $\Psi^{(3)}(\nu)$ we obtain

$$\beta_1(z) \doteqdot \nu^{-1}, \quad \beta_2(z) \doteqdot 3 + 2\nu^{-1} \tag{41}$$

(valid for ν not too small).

Since $\beta_1(y) = 8\nu^{-1}$ and $\beta_2(y) = 3 + 12\nu^{-1}$, we see that S_L does produce a variable with shape coefficients nearer the normal values than those of the original distribution, when applied to Type III variables.

(ii) S_L *applied to Type VI*

If
$$p(y) = \frac{\Gamma(\tau)}{\Gamma(\tau - \nu)\,\Gamma(\nu)} y^{\nu-1} (y+1)^{-\tau} \quad (0 < y < \infty),$$

then
$$\left.\begin{aligned} \beta_1(z) &= [\Psi^{(2)}(\tau-\nu) - \Psi^{(2)}(\nu)]^2 [\Psi^{(1)}(\tau-\nu) + \Psi^{(1)}(\nu)]^{-3}, \\ \beta_2(z) &= 3 + [\Psi^{(3)}(\tau-\nu) + \Psi^{(3)}(\nu)] [\Psi^{(1)}(\tau-\nu) + \Psi^{(1)}(\nu)]^{-2}. \end{aligned}\right\} \tag{42}$$

If both $(\tau - \nu)$ and ν be sufficiently large we have

$$\left.\begin{aligned} \beta_1(z) &\doteqdot \nu^{-1} - 4\tau^{-1} + (\tau-\nu)^{-1}, \\ \beta_2(z) &\doteqdot 3 + 2\nu^{-1} + 2(\tau-\nu)^{-1} - 6\tau^{-1}, \end{aligned}\right\} \tag{43}$$

which compare with
$$\left.\begin{aligned} \beta_1(y) &\doteqdot 4\nu^{-1} - 4\tau^{-1} + 16(\tau-\nu)^{-1}, \\ \beta_2(y) &\doteqdot 3 + 6\nu^{-1} - 6\tau^{-1} + 30(\tau-\nu)^{-1}. \end{aligned}\right\} \tag{44}$$

(iii) S_L *applied to Type V*

Here
$$p(y) = \frac{1}{\Gamma(\nu)} y^{-(\nu+1)} e^{-1/y} \quad (0 < y < \infty).$$

This case is similar to (ii), as is to be expected, since a Type V variable may be regarded as the reciprocal of a Type III variable, while S_L remains of the same form if y be replaced by its reciprocal.

(iv) *S_L applied to Type I (and II)*

If
$$p(y) = \frac{\Gamma(\nu+\tau)}{\Gamma(\nu)\,\Gamma(\tau)} y^{\nu-1}(1-y)^{\tau-1} \quad (0<y<1),$$

then
$$\left.\begin{aligned} \beta_1(z) &= [\Psi^{(2)}(\nu)-\Psi^{(2)}(\nu+\tau)]^2\,[\Psi^{(1)}(\nu)-\Psi^{(1)}(\nu+\tau)]^{-3}, \\ \beta_2(z) &= 3+[\Psi^{(3)}(\nu)-\Psi^{(3)}(\nu+\tau)]\,[\Psi^{(1)}(\nu)-\Psi^{(1)}(\nu+\tau)]^{-2}. \end{aligned}\right\} \tag{45}$$

If ν and τ be sufficiently large then

$$\left.\begin{aligned} \beta_1(z) &\doteqdot 4\tau^{-1}+\nu^{-1}-(\nu+\tau)^{-1}, \\ \beta_2(z) &\doteqdot 3+6\tau^{-1}+2\nu^{-1}-2(\nu+\tau)^{-1}, \end{aligned}\right\} \tag{46}$$

which may be compared with

$$\left.\begin{aligned} \beta_1(y) &\doteqdot 4\nu^{-1}+4\tau^{-1}-16(\nu+\tau)^{-1}, \\ \beta_2(y) &\doteqdot 3+6\tau^{-1}+6\nu^{-1}-30(\nu+\tau)^{-1}. \end{aligned}\right\} \tag{47}$$

(v) *S_B applied to Type I (and Type II)*

If, as before,
$$p(y) = \frac{\Gamma(\nu+\tau)}{\Gamma(\nu)\,\Gamma(\tau)} y^{\nu-1}(1-y)^{\tau-1} \quad (0<y<1),$$

then
$$\left.\begin{aligned} \beta_1(z) &= [\Psi^{(2)}(\nu)-\Psi^{(2)}(\tau)]^2\,[\Psi^{(1)}(\nu)+\Psi^{(1)}(\tau)]^{-3}, \\ \beta_2(z) &= 3+[\Psi^{(3)}(\nu)+\Psi^{(3)}(\tau)]\,[\Psi^{(1)}(\nu)+\Psi^{(1)}(\tau)]^{-2}. \end{aligned}\right\} \tag{48}$$

If ν and τ both be sufficiently large, then

$$\left.\begin{aligned} \beta_1(z) &\doteqdot \nu^{-1}+\tau^{-1}-4(\nu+\tau)^{-1}, \\ \beta_2(z) &\doteqdot 3+2\nu^{-1}+2\tau^{-1}-6(\nu+\tau)^{-1}. \end{aligned}\right\} \tag{49}$$

These formulae may be compared with (46) and (47).

As would be expected, it appears that the S_B transformation generally produces a closer approach to normality than does the S_L transformation applied to the same Type I (or Type II) variable. In particular, if the original variable be symmetrically distributed, S_B preserves the symmetry while S_L does not. Table 2 below provides numerical comparisons in a number

Table 2

ν	τ	y		z in S_L		z in S_B	
		β_1	β_2	β_1	β_2	β_1	β_2
2	2	0·000	2·143	2·233	6·444	0·000	3·594
4	4	0·000	2·455	1·136	4·769	0·000	3·278
6	6	0·000	2·600	0·757	4·180	0·000	3·180
2	4	0·219	2·625	1·384	5·243	0·131	3·625
2	6	0·480	3·109	1·114	4·868	0·244	3·741
4	6	0·051	2·608	0·835	4·333	0·022	3·262

of special cases. The fifth and sixth lines of this table indicate that, as is to be expected, the relative superiority of S_B diminishes as the (β_1, β_2) points of the Pearson curves approach the Type III line (and are hence nearer the log-normal line). Fisher's z' transformation for the correlation coefficient, in the case $\rho = 0$, provides an example of the application of S_B to a Type I variable. We have, putting $r = 2R-1$,

$$p(R) = \frac{\{\Gamma(\frac{1}{2}n-1)\}^2}{\Gamma(n-2)} R^{\frac{1}{2}(n-4)}(1-R)^{\frac{1}{2}(n-4)} \quad (0<R<1),$$

$$z = \tfrac{1}{2}\sqrt{(n-3)}\log\frac{1+r}{1-r} = \tfrac{1}{2}\sqrt{(n-3)}\log\frac{R}{1-R}.$$

Hence
$$\left.\begin{aligned} \sigma^2(z) &= \tfrac{1}{2}(n-3)\,\Psi'^{(1)}(\tfrac{1}{2}n-1),\\ \beta_1(z) &= 0,\\ \beta_2(z) &= 3+\tfrac{1}{2}\Psi'^{(3)}(\tfrac{1}{2}n-1)\,[\Psi'^{(1)}(\tfrac{1}{2}n-1)]^{-2}.\end{aligned}\right\} \tag{50}$$

(vi) S_U *applied to Type VII*

If
$$p(y) = \frac{\Gamma(\nu)}{\sqrt{\pi}\,\Gamma(\nu-\frac{1}{2})}(1+y^2)^{-\nu} \quad (-\infty<y<+\infty),$$

then
$$\beta_1(z) = 0, \quad \beta_2(z) = 3+\tfrac{1}{2}\Psi'^{(3)}(\nu-\tfrac{1}{2})\,[\Psi'^{(1)}(\nu-\tfrac{1}{2})]^{-2}. \tag{51}$$

Table 3 compares the β_2's of corresponding distributions of y and z for various values of ν.

Table 3

ν	$\beta_2(y)$	$\beta_2(z)$	ν	$\beta_2(y)$	$\beta_2(z)$
1	∞	5·000	4	5·000	3·322
2	∞	3·806	5	4·200	3·245
3	9·000	3·466	6	3·857	3·107

The above discussion has been concerned only with the shape of the distribution of the transformed variable, as judged by the values of $\beta_1(z)$ and $\beta_2(z)$. Even where these quantities differ appreciably from the normal values of 0 and 3, it is, however, possible that the transformed probability integral could be regarded as normal for practical purposes. Further investigation on this point would be of interest.

4. Conclusion

4·1. *Critical summary*

The following comments may prove helpful in assessing the value of the systems of curves described in the foregoing pages:

(i) The systems S_L, S_B, S_U, together with the normal curve combine to give a variety of shapes of curve as wide as that provided by the systems of frequency curves in general use.

(ii) The fundamental property that each of these curves may be transformed to a normal curve by a simple transformation may be regarded either as a practical convenience, or as a desirable property based on the arguments of Kapteyn and Wicksell. The first of these reasons is of considerable importance, but it should be noted that simple exact tests of significance are, even theoretically, obtainable only for a very restricted range of problems (Bartlett, 1947).

(iii) S_L is, of course, a well-established system. Of the systems S_B and S_U, S_B is based on the simpler transformation, but S_U has the simpler expressions for its moments. The fitting of S_U seems to be more straightforward than that of S_B, except when the limits of variation of the latter are definitely known.

(iv) Curves of S_L, S_B and S_U all have high contact at the extremes of their range of variation. This may sometimes prove a drawback at the finite limits of variation for systems S_L and S_B.

(v) Except for discrepancies at the ends, which may be associated with (iv), curves of S_L, S_B and S_U agree generally with Pearson curves having the same (or nearly the same) first four moments. The use of the former curves may sometimes be considered simply as a convenient aid in calculating rough figures for subrange frequencies of the latter curve. In a note added to the paper by Pretorius (1930), K. Pearson suggests the use of S_L in this capacity relative to certain Type VI curves.

4·2. *Other translation systems*

A further point of interest arises in the fact that all the moments of all curves in the systems S_L, S_B and S_U are finite. By comparison, it is known that the higher moments of certain of the Pearson curves can be infinite. While finiteness of moments is in many respects a desirable property, it may be argued that such finiteness might restrict the systems relatively to curves with very long tails. It may be noted that such curves might be covered by choosing a different distribution for z. In particular, we might suppose z to be distributed according to the first law of Laplace

$$p(z) = \tfrac{1}{2}e^{-|z|} \quad (-\infty < z < \infty). \tag{52}$$

Frechet (1928, 1939) has suggested that more use might be made of this law; his arguments, combined with those of Kapteyn, would lead to systems of curves defined by

$$z = \gamma + \delta f\left(\frac{x-\xi}{\lambda}\right), \tag{53}$$

with $p(z)$ given by (52). Inserting the particular forms for $f\{(x-\xi)/\lambda\}$, we would obtain systems S'_L, S'_B, S'_U corresponding to the systems S_L, S_B, S_U. It is easy to show that the system S'_U can have infinite moments.

Clearly, by choosing fresh forms for $p(z)$ a great variety of systems of curves could be constructed, but practical considerations naturally limit those cases which it is worth while to study. Mention may be made of the work of Olshen (1938), who has investigated certain transformations of the Pearson Type III distribution.

5. Numerical examples

Example 1

For this example data used by Pearse (1928) were employed. The data gave the distribution of cloudiness at Greenwich for the period 1890–1904, excluding 1901. The last column, headed Type I, gives the frequencies obtained from a Pearson Type I curve fitted to the data by Pearse. Three moments were used in fitting this curve, the length of the range of variation being fixed in advance.

Curve $S_B(1)$, the frequencies for which are shown in the third column, was fitted to the observed data on the assumption that the degrees of cloudiness stated as $0, 1, 2, \ldots, 10$ could be regarded as corresponding to groups $-0{\cdot}5$ to $0{\cdot}5$, $0{\cdot}5$ to $1{\cdot}5$, ..., $9{\cdot}5$ to $10{\cdot}5$. ξ was thus fixed at $-0\cdot5$ and λ at $11\cdot0$. γ and δ were then chosen to give exact agreement in the two extreme groups. These values were

$$\gamma = -0\cdot3110, \quad \delta = 0\cdot25166.$$

Table 4

Degree of cloudiness	Observed frequencies	$S_B(1)$	$S_B(2)$	Type I
0	320	320·0	320·0	321·7
1	129	100·9	120·9	121·5
2	74	73·9	72·0	75·1
3	68	63·8	57·5	61·4
4	45	59·8	52·1	56·0
5	45	59·9	51·6	55·2
6	55	63·4	54·9	57·8
7	65	72·0	63·9	65·5
8	90	90·0	85·5	83·2
9	148	135·4	160·7	139·6
10	676	676·0	676·0	678·0
	1715	1715·1	1714·9	1715·0
Value of χ^2	—	18·44	5·76	6·52

Curve $S_B(2)$ was fitted in the same way, except that it was assumed that the successive groups were 0 to 0·5, 0·5 to 1·5, ..., 9·5 to 10. ξ was therefore put equal to 0 and λ put equal to 10. The values of γ and δ obtained were

$$\gamma = -0\cdot3110 \text{ (as before)}, \quad \delta = 0\cdot19681.$$

$S_B(2)$ gives a much better fit then $S_B(1)$ and, on the whole, a better fit than the Type I curve. All three curves fail to give sufficiently small frequencies in the trough in the centre of the distribution.

Example 2

The data used in this example relate to the age of Australian mothers at birth of a child (single births only) in the period 1922–6. Pretorius (1930) fitted a Type I curve to these data. The values of the moment ratios ($\beta_1 = 0\cdot101$, $\beta_2 = 2\cdot430$) indicate that a curve of system S_B might be fitted. In this case, however, there are no obvious values to assign to the parameters ξ and λ. The method actually adopted (like that of Pretorius) was based on trial and error. It was decided to attempt a fit which would give nearly correct values for the 5, 30, 70 and 95 % points of the distribution. Values were assigned to ξ and λ and then values of γ and δ obtained, so that the specified percentage points were, as far as possible, unaltered. The process was repeated to improve the fit. Details of the working are now given in respect of the curve finally fitted.

For this curve $\xi = 15\cdot0, \quad \lambda = 36\cdot5.$

From a cumulative diagram the following percentage points of the observed distribution were estimated:

5 % point	20·3 years	70 % point	32·8 years
30 % point	25·6 years	95 % point	40·4 years

With $\xi = 15{\cdot}0$, $\lambda = 36{\cdot}5$, the values of $\log\{(x-\xi)/(\xi+\lambda-x)\}$ at these points are:

$$-0{\cdot}7096, \quad -0{\cdot}4192, \quad -0{\cdot}0247, \quad 0{\cdot}3918 \quad \text{respectively.}$$

The normal equivalent deviates for 5, 30, 70 and 95 % are:

$$-1{\cdot}6449, \quad -0{\cdot}5244, \quad 0{\cdot}5244, \quad 1{\cdot}6449 \quad \text{respectively.}$$

Hence γ and δ should satisfy the four equations

$$-1{\cdot}6449 = \gamma - 0{\cdot}7096\delta, \tag{i}$$

$$-0{\cdot}5244 = \gamma - 0{\cdot}4192\delta, \tag{ii}$$

$$0{\cdot}5244 = \gamma - 0{\cdot}0247\delta, \tag{iii}$$

$$1{\cdot}6449 = \gamma + 0{\cdot}3918\delta. \tag{iv}$$

From (i) and (iv) we obtain $\gamma = 0{\cdot}5978$, $\delta = 1{\cdot}2649$; from (ii) and (iii) we obtain $\gamma = 0{\cdot}5857$, $\delta = 1{\cdot}2424$. For the curve to be fitted, we took the values

$$\gamma = 0{\cdot}5918, \quad \delta = 1{\cdot}2536.$$

The table below compares the observed distribution, the distribution corresponding to the fitted S_B curve and the Type I distribution fitted by Pretorius.

Table 5

Age of mother (central values in years)	Observed frequencies	S_B	Type I
13	3	—	—
15	191	33	46
17	4,573	4,697	6,105
19	21,322	22,510	22,871
21	42,758	44,048	41,996
23	62,620	60,994	58,455
25	73,423	71,123	69,796
27	74,834	74,662	75,176
29	72,640	73,063	74,822
31	65,182	67,476	69,637
33	58,407	59,236	60,909
35	48,834	49,380	50,071
37	39,932	38,837	38,524
39	31,050	28,491	27,485
41	18,975	19,530	17,878
43	11,283	10,594	10,359
45	4,365	5,179	5,088
47	1,072	1,619	1,943
49	199	208	476
51	13	2	44
53	4	—	—
55	2	—	—
Total	631,682	631,682	631,682

Neither the S_B curve nor the Type I curve fit well at the ends of the distribution. The S_B curve appears to give the closer fit in the central part of the distribution. For the range with central values 17–47 inclusive, we have

$$S_B \text{ curve:}\quad \chi^2 = 718; \qquad \text{Type I curve:}\quad \chi^2 = 1759.$$

Excluding both groups 17 and 47 (central values), we have

$$S_B \text{ curve:}\quad \chi^2 = 530; \qquad \text{Type I curve:}\quad \chi^2 = 1375.$$

As is almost invariably found to be the case when dealing with very large samples, exceedingly high values of χ^2 are obtained. Differences between observation and theory, which may be practically unimportant from the point of view of graduation and which might not be picked out in samples of more usual size, are statistically significant having regard to the large numbers involved.

Example 3

The data on length and breadth of beans used in this example and in Example 4 were used by Pretorius (1930). In this case we shall fit a curve of system S_U to the distribution of lengths of beans. The mean, standard deviation, β_1 and β_2 of the observed distribution are

$$\begin{aligned} &\text{Mean} &&= 14{\cdot}399\,\text{mm.}; &&\beta_1 = 0{\cdot}829;\\ &\text{Standard deviation} &&= 0{\cdot}9036\,\text{mm.}; &&\beta_2 = 4{\cdot}863.\end{aligned}$$

Using the abac of Fig. 12 we find

$$\delta = 2{\cdot}64; \quad \Omega = \gamma/\delta = 0{\cdot}90,$$

whence $$\gamma = 2{\cdot}38.$$

From (37) we calculate $$\mathscr{E}\left(\frac{x-\xi}{\lambda}\right) = 1{\cdot}1029, \quad \sigma\left(\frac{x}{\lambda}\right) = 0{\cdot}5948.$$

Hence $$\lambda = \frac{0{\cdot}9036}{0{\cdot}5948} = 1{\cdot}5192,$$

$$\xi = 14{\cdot}399 + 1{\cdot}1029\lambda = 16{\cdot}0745.$$

There is necessarily some uncertainty in the determination of γ and δ from the chart, but investigation indicated that the degree of uncertainty should not seriously affect the fitted frequencies. Table 6 shows the observed frequencies, the frequencies calculated from the fitted S_U curve, and the frequencies calculated for the Type IV curve fitted by Pretorius.*

Example 4

A curve of system S_U is fitted to the distribution of breadth of beans referred to in Example 3. For the observed distribution

$$\begin{aligned} &\text{Mean} &&= 7{\cdot}9755\,\text{mm.}; &&\beta_1 = 0{\cdot}1943;\\ &\text{Standard deviation} &&= 0{\cdot}3399\,\text{mm.}; &&\beta_2 = 3{\cdot}6544.\end{aligned}$$

Following the same procedure as in Example 3 the following values of the parameters of the S_B curve were obtained:

$$\delta = 3{\cdot}55, \quad \gamma = 2{\cdot}13, \quad \lambda = 0{\cdot}9721, \quad \xi = 8{\cdot}6195.$$

Table 7 compares observed frequencies, the fitted frequencies calculated from the S_U curve and those calculated from the Type IV curve fitted by Pretorius.*

* The groupings in the tails used in calculating χ^2 are shown by the braces to the right of the table.

Table 6

Length (central values in mm.)	Observed frequencies	S_U	Type IV
< 9·25	—	2·6	1·9
9·5	1	2·7	2·6
10·0	7	5·8	5·4
10·5	18	12·1	11·3
11·0	36	25·7	24·2
11·5	70	55·2	52·5
12·0	115	118·0	113·8
12·5	199	249·3	243·7
13·0	437	508·7	503·6
13·5	929	970·6	968·9
14·0	1787	1642·5	1638·9
14·5	2294	2240·6	2229·8
15·0	2082	2130·3	2132·6
15·5	1129	1151·5	1181·6
16·0	275	290·1	299·3
16·5	55	32·2	28·5
17·0	6	2·0	1·4
> 17·25	—	0·1	
Total	9440	9440·0	9440·0
Value of χ^2	—	87·1*	102·5†

* Excluding the 'over 16·25' group, $\chi^2 = 66·3$.
† Excluding the 'over 16·25' group, $\chi^2 = 70·1$.

Table 7

Breadth (central values in mm.)	Observed frequencies	S_U	Type IV
< 6·25		1·3	4·8
6·375	4	3·7	
6·625	10	13·8	13·3
6·875	72	53·2	49·9
7·125	170	182·2	177·2
7·375	530	557·8	557·9
7·625	1397	1394·2	1413·3
7·875	2579	2507·0	2530·5
8·125	2742	2757·0	2732·5
8·375	1483	1544·4	1515·4
8·625	400	381·5	393·6
8·875	48	41·5	48·6
9·125	5	2·3	3·0
> 9·25		0·1	
Total	9440	9440·0	9440·0
Value of χ^2	—	17·47	14·36

APPENDIX

Moments of distributions in system S_B

If $z = \gamma + \delta \log\{y/(1-y)\}$ and z is a unit normal variable, then the rth moment of y about zero is

$$\mu_r'(y) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} (1 + e^{-(z-\gamma)/\delta})^{-r}\, dz. \tag{54}$$

This integral is not easy to evaluate directly, and values of μ_1', μ_2', μ_3' and μ_4' were obtained by the following steps.

(i) For the case $r = 1$, the expected value

$$\mu_1' = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} (1 + e^{-(z-\gamma)/\delta})^{-1}\, dz \tag{55}$$

can be evaluated directly using a result due to Mordell (1920, 1933). To throw (55) into a form to which Mordell's result applies, we make the transformation $t - \gamma = -2\pi\delta z$ leading to

$$\mu_1' = \sqrt{(2\pi)}\,\delta e^{-\frac{1}{2}\gamma^2} \int_{-\infty}^{\infty} e^{\pi i \psi t^2 - 2\pi\nu t} (1 + e^{2\pi t})^{-1}\, dt,$$

where $\nu = -\gamma\delta, \quad \psi = 2\pi\delta^2 i.$

By Mordell's formula

$$\mu_1' = \sqrt{(2\pi)}\,\delta e^{-\frac{1}{2}\gamma^2} \frac{1}{\theta_{00}(\nu, \psi)} \left\{ \frac{i}{\psi} \sum_{-\infty}^{\infty} \frac{q_1^{n^2} e^{2n\pi i \nu/\psi}}{1 + q_1^{2n}} - i \sum_{-\infty}^{\infty} \frac{q^{n^2 - \frac{1}{4}} e^{(2n-1)\pi i \nu}}{1 - q^{2n-1}} \right\},$$

where $q = e^{\pi i \psi}, \quad q_1 = e^{-\pi i/\psi}$

and $\theta_{00}(\nu, \psi) = \sum_{-\infty}^{\infty} e^{n^2 \pi i \psi} e^{2n\pi i \nu} = 1 + 2 \sum_{n=1}^{\infty} e^{n^2 \pi i \psi} \cos 2n\pi\nu.$

After some algebraic simplification, this leads to the result

$$\mu_1' = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}\gamma^2} \frac{\frac{1}{2}\delta^{-1} + \delta^{-1} \sum_{n=1}^{\infty} e^{-n^2/2\delta^2} \cosh \frac{n(1-2\gamma\delta)}{2\delta^2} \operatorname{sech} \frac{n}{2\delta^2} - 2\pi\delta \sum_{n=1}^{\infty} e^{-\frac{1}{2}(2n-1)^2 \pi^2 \delta^2} \sin(2n-1)\pi\gamma\delta \operatorname{cosech}(2n-1)\pi^2\delta^2}{1 + 2 \sum_{n=1}^{\infty} e^{-2n^2\pi^2\delta^2} \cos 2n\pi\gamma\delta}. \tag{56}$$

(ii) The higher moments can be expressed in terms of the partial derivatives of the first moment with respect to γ. From (54)

$$\begin{aligned}
\frac{\partial \mu_r'}{\partial \gamma} &= -\frac{r}{\delta} \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \frac{e^{-(z-\gamma)/\delta} e^{-\frac{1}{2}z^2}}{(1 + e^{-(z-\gamma)/\delta})^{r+1}}\, dz \\
&= -\frac{r}{\delta} \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \frac{\{(1 + e^{-(z-\gamma)/\delta}) - 1\} e^{-\frac{1}{2}z^2}}{(1 + e^{-(z-\gamma)/\delta})^{r+1}}\, dz \\
&= -\frac{r}{\delta} (\mu_r' - \mu_{r+1}').
\end{aligned}$$

Hence

$$\mu_{r+1}' = \mu_r' + \frac{\delta}{r} \frac{\partial \mu_r'}{\partial \gamma}. \tag{57}$$

Applying this formula with $r = 1, 2, 3$ successively we obtain

$$\left.\begin{aligned}
\mu_2' &= \mu_1' + \delta \frac{\partial \mu_1'}{\partial \gamma}, \\
\mu_3' &= \mu_1' + \tfrac{3}{2}\delta \frac{\partial \mu_1'}{\partial \gamma} + \tfrac{1}{2}\delta^2 \frac{\partial^2 \mu_1'}{\partial \gamma^2}, \\
\mu_4' &= \mu_1' + \tfrac{11}{6}\delta \frac{\partial \mu_1'}{\partial \gamma} + \delta^2 \frac{\partial^2 \mu_1'}{\partial \gamma^2} + \tfrac{1}{6}\delta^3 \frac{\partial^3 \mu_1'}{\partial \gamma^3}.
\end{aligned}\right\} \tag{58}$$

Formulae (58), together with (56), make it possible to calculate μ_2', μ_3' and μ_4'. Although the analytical expressions for these moments must be very complicated, their numerical computation is straightforward, though tedious.

(iii) The computational labour may be reduced by using the recurrence formula developed below. To emphasize the dependence of μ_r' on γ and δ, we shall write

$$\mu_r'(\gamma,\delta) = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}(1+e^{-(z-\gamma)/\delta})^{-r}\,dz.$$

Then
$$\mu_r'(\gamma,\delta) = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\{(1+e^{-(z-\gamma)/\delta})-e^{-(z-\gamma)/\delta}\}(1+e^{-(z-\gamma)/\delta})^{-r}\,dz$$
$$= \mu_{r-1}'(\gamma,\delta) - e^{\frac{1}{2}\delta^{-2}+\gamma\delta^{-1}}\mu_r'(\gamma+\delta^{-1},\delta).$$

Hence
$$\mu_r'(\gamma+\delta^{-1},\delta) = e^{-(\frac{1}{2}\delta^{-2}+\gamma\delta^{-1})}[\mu_{r-1}'(\gamma,\delta)-\mu_r'(\gamma,\delta)]. \tag{59}$$

Remembering that $\mu_0' = 1$ for all γ and δ, (59) makes it possible for the first four moments to be calculated with fair rapidity for series of values of γ at intervals of δ^{-1}.

A few of the calculated values of β_1 and β_2 are shown in Table 8. Generally the moments were calculated by methods (i) and (ii) for values of γ between 0 and δ^{-1}; further results were then obtained by means of (iii). It was necessary to take care to avoid accumulation of errors in applying (iii). The first sets of moments were calculated to eleven decimal places, and the values of β_1 and β_2 obtained should be accurate to the five places of decimals shown.

Table 8

γ	$\delta = 0{\cdot}5$				$\delta = 1{\cdot}0$			
	μ_1'	σ	β_1	β_2	μ_1'	σ	β_1	β_2
0·0	0·50000	0·31396	0·00000	1·62731	0·50000	0·20829	0·00000	2·13828
0·5	0·35227	0·29610	0·36363	2·07024	0·39797	0·20151	0·12803	2·32409
1·0	0·22480	0·24873	1·64723	3·65177	0·30327	0·18262	0·52856	2·90911
1·5	0·12959	0·18679	4·59189	7·36733	0·22147	0·15541	1·24787	3·98260
2·0	0·06767	0·12615	11·15751	15·97162	0·15546	0·12465	2·36420	5·71101
2·5	0·03225	0·07717	26·40534	37·17352	0·10536	0·09468	3·98326	8·34077

γ	$\delta = 2{\cdot}0$			
	μ_1'	σ	β_1	β_2
0·0	0·50000	0·11813	0·00000	2·63131
0·5	0·44125	0·11665	0·02084	2·66720
1·0	0·38402	0·11235	0·08279	2·77419
1·5	0·32971	0·10560	0·18409	2·95062
2·0	0·27942	0·09697	0·32168	3·19309
2·5	0·23394	0·08710	0·49116	3·49645

The transformation generating the system S_B is symmetrical about $y = \frac{1}{2}$, so that when $\gamma = 0$ the distribution of y is symmetrical. Positive values of γ correspond to positive skewness.

Since the curve is symmetrical when $\gamma = 0$, it follows that $\mu_1'(0,\delta) = \frac{1}{2}$, whatever be δ. This can be verified from (56). Putting $\gamma = 0$, we obtain

$$\mu_1'(0,\delta) = \frac{1}{2}\frac{1+2\sum_{n=1}^{\infty} e^{-n^2/2\delta^2}}{\sqrt{(2\pi)}\,\delta\left[1+2\sum_{n=1}^{\infty} e^{-2n^2\pi^2\delta^2}\right]}.$$

This expression can be shown to be equal to $\frac{1}{2}$ by putting $t = 0$, $\tau = 2\pi\delta_t^2$ in Jacobi's imaginary transformation of the theta function

$$\theta_3(t\,|\,\tau) = (-i\tau)^{-\frac{1}{2}} e^{t^2/\pi i\tau}\,\theta_3\left(\frac{t}{\tau}\,\middle|\, -\frac{1}{\tau}\right)$$

(Whittaker & Watson, 1946, p. 475).

Also, since $\mu_0'(0,\delta) = 1$, it follows from (59) that

$$\mu_1'(\delta^{-1},\delta) = e^{-\frac{1}{2}\delta^{-2}}(1-\tfrac{1}{2}) = \tfrac{1}{2}e^{-\frac{1}{2}\delta^{-2}}.$$

Again $$\mu_1'(2\delta^{-1},\delta) = e^{-\frac{3}{2}\delta^{-2}}(1-\tfrac{1}{2}e^{-\frac{1}{2}\delta^{-2}}) = e^{-\frac{3}{2}\delta^{-2}} - \tfrac{1}{2}e^{-\frac{4}{2}\delta^{-2}}.$$

Similarly $$\mu_1(3\delta^{-1},\delta) = e^{-\frac{5}{2}\delta^{-2}} - e^{-\frac{8}{2}\delta^{-2}} + \tfrac{1}{2}e^{-\frac{9}{2}\delta^{-2}},$$

and generally, if k be a positive integer greater than unity,

$$\mu_1'(k\delta^{-1},\delta) = (-1)^{k+1}e^{-\frac{1}{2}k^2\delta^{-2}}\left[\tfrac{1}{2} + \sum_{s=1}^{k}(-1)^s e^{\frac{1}{2}s^2\delta^{-2}}\right]. \tag{60}$$

(60) is useful as a check formula.

Further interesting formulae may be obtained starting from the equation

$$0 = \mu_3(0,\delta) = \mu_3'(0,\delta) - \tfrac{3}{2}\mu_2'(0,\delta) + \tfrac{1}{4}. \tag{61}$$

Such formulae are useful in checking calculations, but do not lead to simple formulae for $\mu_3'(k\delta^{-1},\delta)$, since $\mu_2'(0,\delta)$ is not a sufficiently simple function of δ.

We now proceed to consider limiting values of β_1 and β_2 for S_B. We can write (59) in the form

$$\mu_r'(k\delta^{-1},\delta) = -e^{-\frac{1}{2}(2k-1)\delta^{-2}}\Delta\mu_{r-1}'(\overline{k-1}\,\delta^{-1},\delta), \tag{62}$$

the forward difference, Δ, applying to the subscript of μ'. Applying (62) repeatedly and noting that (62) holds for negative as well as positive values of r, we have

$$\mu_r'(k\delta^{-1},\delta) = (-1)^k e^{-\frac{1}{2}k^2\delta^{-2}}\Delta^k\mu_{r-k}'(0,\delta). \tag{63}$$

The sth negative moment of y about zero is easily found to be

$$\mu_{-s}'(\gamma,\delta) = \sum_{t=0}^{s}\binom{s}{t}e^{\frac{1}{2}t^2\delta^{-2}+t\gamma\delta^{-1}},$$

whence, if $k > r$,

$$\mu_r'(k\delta^{-1},\delta) = (-1)^k e^{-\frac{1}{2}k^2\delta^{-2}}\left[\sum_{t=0}^{k-r}(-1)^{t-r}\binom{k-\overline{t+1}}{r-1}e^{\frac{1}{2}t^2\delta^{-2}} + \sum_{t=0}^{r-1}(-1)^{t-1}\binom{k}{t}\mu_{r-t}'(0,\delta)\right]. \tag{64}$$

Hence if k is large

$$\begin{aligned}\mu_r'(k\delta^{-1},\delta) &\doteqdot e^{-\frac{1}{2}k^2\delta^{-2}}e^{\frac{1}{2}(k-r)^2\delta^{-2}} \\ &\doteqdot e^{-\frac{1}{2}[2r(k/\delta)\delta - r^2]\delta^{-2}}.\end{aligned} \tag{65}$$

$\mu_r'(\gamma,\delta)$ is a continuous decreasing function of γ for sufficiently large values of γ. Hence

$$\mu_r'(\gamma,\delta) \doteqdot e^{-r\gamma\delta^{-1}+\frac{1}{2}r^2\delta^{-2}} \quad \text{when } \gamma \text{ is large,}$$

and so

$$\left.\begin{aligned}\lim_{\gamma\to\infty}\beta_1(\gamma,\delta) &= (\omega-1)(\omega+2)^2, \\ \lim_{\gamma\to\infty}\beta_2(\gamma,\delta) &= \omega^4 + 2\omega^3 + 3\omega^2 - 3, \\ \text{where}\quad \omega &= e^{\delta^{-2}}.\end{aligned}\right\} \tag{66}$$

As γ increases from zero to infinity, therefore, the (β_1,β_2) point moves from a point on the axis of β_2 to a point on the S_L line (cf. (14)).

Now consider the behaviour of $\mu_r'(\gamma,\delta)$ as δ decreases, γ remaining fixed. When δ is small we have

$$\begin{aligned}(1+e^{-(z-\gamma)/\delta})^{-r} &\doteqdot 0 \quad \text{for} \quad z < \gamma \\ &\doteqdot 1 \quad \text{for} \quad z > \gamma,\end{aligned}$$

so that

$$\mu_1' \doteqdot \mu_2' \doteqdot \mu_3' \doteqdot \mu_4' \doteqdot \frac{1}{\sqrt{(2\pi)}}\int_\gamma^\infty e^{-\frac{1}{2}z^2}\,dz. \tag{67}$$

As δ decreases the (β_1,β_2) point therefore approaches a point on the boundary line $\beta_2 - \beta_1 - 1 = 0$. As γ varies all points on the boundary are covered.

REFERENCES

Aroian, L. A. (1941). *Ann. Math. Statist.* **12**, 429.
Baker, G. A. (1934). *Ann. Math. Statist.* **5**, 113.
Bartlett, M. S. (1937). *Suppl. J. R. Statist. Soc.* **4**, 137.
Bartlett, M. S. (1947). *Biometrics*, **3**, 39.
Beall, G. (1942). *Biometrika*, **32**, 243.
Burr, I. W. (1942). *Ann. Math. Statist.* **13**, 215.
Charlier, C. V. L. (1905). *Ark. Mat. Astr. Fys.* **2**, nos. 8 and 20.
Cochran, W. G. (1938). *Empire J. Exp. Agric.* **6**, 157.
Cornish, E. A. & Fisher, R. A. (1937). *Rev. Inst. Int. Statist.* **5**, 307.
Curtiss, J. H. (1943). *Ann. Math. Statist.* **14**, 107.
Edgeworth, F. Y. (1898). *J. R. Statist. Soc.* **61**, 670.
Epstein, B. (1947). *J. Franklin Inst.* **244**, 471.
Fechner, G. T. (1897). *Kollektivmasslehre.* Leipzig: Engelmann.
Finney, D. J. (1941). *Suppl. J. R. Statist. Soc.* **7**, 155.
Frechet, M. (1928). *Bull. Sci. Math.* **63**, 203.
Frechet, M. (1937). *Recherches Theoriques Modernes*, **1**. Paris: Gauthier Villars.
Frechet, M. (1939). *Rev. Inst. Int. Statist.* **7**, 32.
Frechet, M. (1945). *Rew. Inst. Int. Statist.* **13**, 16.
Gaddum, J. H. (1945). *Nature, Lond.*, **156**, 463.
Galton, F. (1879). *Proc. Roy. Soc.* **29**, 365.
Geary, R. C. (1947). *Biometrika*, **34**, 209.
Gibrat, R. (1931). *Les Inégalités Économiques.* Paris: Libraire du Recueil Sirey.
Halmos, P. R. (1944). *Ann. Math. Statist.* **15**, 182.
Hotelling, H. & Frankel, L. R. (1938). *Ann. Math. Statist.* **9**, 87.
Kapteyn, J. C. (1903). *Skew Frequency Curves in Biology and Statistics.* Groningen: Astronomical Laboratory.
Kapteyn, J. C. & van Uven, M. J. (1916). Title as above.
Kolmogoroff, A. N. (1941). *C.R. Acad. Sci. U.R.S.S.* **31**, no. 2, 99.
McAlister, D. (1879). *Proc. Roy. Soc.* **29**, 367.
Milne-Thompson, L. & Comrie, L. J. (1931). *Standard Four-Figure Mathematical Tables.* London: Macmillan.
Mordell, L. J. (1920). *Quart. J. Math., Oxford*, **48**, 329.
Mordell, L. J. (1933). *Acta Math.* **61**, 323.
Olshen, C. A. (1938). *Ann. Math. Statist.* **9**, 176.
Pearse, G. E. (1928). *Biometrika*, **20** A, 314.
Pearson, E. S. (1931). *Biometrika*, **23**, 114.
Pearson, K. (1895). *Philos. Trans.* A, **186**, 343.
Pretorius, S. J. (1930). *Biometrika*, **22**, 109.
Quensel, C. E. (1945). *Skand. Aktuar.* **28**, 141.
Rietz, H. L. (1922). *Ann. Math.* **23**, 291.
Whittaker, E. T. & Watson, G. N. (1946). *A Course of Modern Analysis*, 4th ed. Cambridge University Press.
Wicksell, S. D. (1917). *Ark. Mat. Astr. Fys.* **12**, no. 20.
Williams, C. B. (1937). *Ann. Appl. Biol.* **24**, 404.
Williams, C. B. (1940). *Biometrika*, **31**, 356.
Wilson, E. B. & Hilferty, M. M. (1931). *Proc. Nat. Acad. Sci., Wash.*, **17**, 694.
Wishart, J. (1947). *Biometrika*, **34**, 170.
Yuan, P. T. (1933). *Ann. Math. Statist.* **4**, 30.