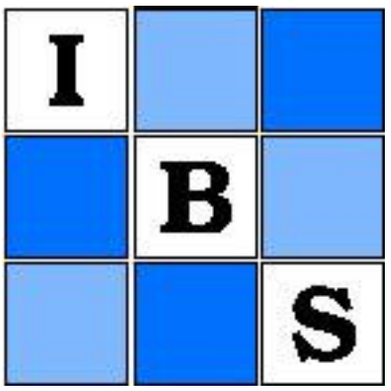

A GENERALISED LOGIT-NORMAL DISTRIBUTION

R. MEAD

National Vegetable Research Station, Wellesbourne, Warwick, England.

INTRODUCTION

The use of a transformation of a variate, such that the transformed variate is approximately normally distributed, has been discussed by several authors, in particular by Johnson [1949], and his paper should be consulted for the historical and general theoretical background. In this paper a generalisation of one of the transformations proposed by Johnson is considered.

The practical use of any particular transformation and the corresponding family of distributions depends on the versatility of the distribution, i.e. the possible variation in its shape, and also on the ease with which the distribution can be fitted. The possible shapes of the distribution obtainable with the transformation, together with some limiting forms, are considered in the first part of this paper. The second part is concerned with methods of fitting the distribution to grouped data.

DERIVATION OF THE TRANSFORMATION

The transformation, $z = f(x)$ where z is a standardised normal variate will be referred to as the transformation of x, the observed variate, to normality (equivalently $f(x)$ can be considered as being normally distributed).

One of the transformations discussed by Johnson is

$$z = \gamma + \delta \log \frac{x - \xi}{\xi + \lambda - x}$$

or, equivalently,

$$z = \left[\left(\log \frac{x - \xi}{\xi + \lambda - x}\right) - \mu\right] \Big/ \sigma$$

where ξ and $(\xi + \lambda)$ are the lower and upper limits of x. In many practical situations the lower limit of x is zero, i.e. $\xi = 0$, giving the logit transformation,

$$z = \gamma + \delta \log \frac{x}{A - x}$$

where x takes values between 0 and A. Following Nelder [1961] I have generalised the transformation by including a power, θ, for reasons which will be discussed later in the paper, so that the full transformation is

$$z = \gamma + \delta \log \frac{x^\theta}{A^\theta - x^\theta}$$

or (1)

$$z = \gamma + \delta \operatorname{logit} \left(\frac{x}{A}\right)^\theta .$$

The distribution of x corresponding to this transformation is reasonably called the generalised logit-normal distribution.

APPLICATION

The interest in the distribution arose from the search for a mathematical form to fit the distribution of individual root diameters in a population of carrots grown in competition in a row. The form of the empirical distribution was found to vary considerably with density, ratio of inter-row distance to mean intra-row distance, and maturity of crop, and it seemed likely that a distribution with three or four parameters would be needed to describe the data.

The major points arising from the observed distributions which affected the choice of a theoretical distribution were:—

(i) The observed distribution was bounded above by the maximum size attainable under the particular, non-competitive conditions of growth and maturity,
(ii) the skewness of the distribution varied considerably, either to right or left,
(iii) the distribution was generally more platykurtic than the normal distribution, and
(iv) the position of the mode relative to the upper end of the distribution, and also the shapes of the tails, varied even among distributions of similar skewness.

The logit-normal distribution with $\theta = 1$ satisfies the first three criteria quite well. Alteration of θ produces some additional variation in the shape of the distribution, principally in the length of the comparatively 'straight' part of the transformation curve and in its position relative to the point of inflexion of the curve.

It is possible for the distribution to be bimodal, and the conditions for this when $\theta = 1$ are set out by Johnson in his paper. In the general case, the condition for a modal value of x, other than A or zero is

$$(\theta + 1)\left(\frac{x}{A}\right)^{\theta} - 1 = \theta\delta\left(\gamma + \delta \log \frac{x^{\theta}}{A^{\theta} - x^{\theta}}\right)$$

and other required results follow as in the case, $\theta = 1$.

Histograms with fitted curves for four sets of carrot data are contained in the Appendix of this paper.

THEORETICAL PROPERTIES OF THE TRANSFORMATION

The probability densities $g_1(x)$ of x and $g_2(z)$ of z are related by the equation

$$g_1(x) = g_2(z)\frac{dz}{dx}. \tag{2}$$

It is therefore the gradient of the transformation, and particularly the variation of the gradient for varying x and fixed parameter values, that determine the shape of the x-distribution. Further it is the way the variation of the gradient changes as the values of the parameters, μ, σ, θ and A, alter that determines the range of shapes of the distribution obtainable with the transformation. In other words, the important aspect of a family of transformations is the variation (for varying parameter values) of the variation (for varying x) of the gradient of the transformation.

The gradient of (1) is

$$\theta\delta\left[x\left(1 - \left(\frac{x}{A}\right)^{\theta}\right)\right]^{-1}$$

and the point of inflexion is given by $x = A/(1 + \theta)^{1/\theta}$,

In order to examine the changes in the variation of the gradient for varying θ it is convenient to consider the ratio of the gradient at x to the gradient at the point of inflexion. The relative gradient thus obtained is

$$g = \frac{\theta}{(1 + \theta)^{1+1/\theta}\frac{x}{A}\left(1 - \left(\frac{x}{A}\right)^{\theta}\right)}.$$

To find the limiting forms of the transformation as θ tends to zero or infinity the limiting forms of the relative gradient have been found and the appropriate forms of the transformation have been obtained by integration (this is the most convenient method, as the generalised logit tends to infinity for all x as $\theta \to 0$, and its gradient tends to infinity for all x as $\theta \to \infty$).

As θ tends to zero the relative gradient tends to

$$g_0 = -\frac{A}{ex \log \frac{x}{A}}$$

and the appropriate form of the transformation is

$$z = \gamma' + \delta'\left[-\log\left(-\log \frac{x}{A}\right)\right].$$

This is the direct log log transformation; the complementary log log transformation, $z = \log[-\log(1 - x/A)]$, has been tabulated by Fisher and Yates [1957] who comment on its similarity to the logit distribution when $(1 - x/A)$ is small, i.e. when x is near A.

As θ tends to infinity the relative gradient tends to $g_\infty = A/x$ and the appropriate form of the transformation is

$$z = \gamma' + \delta'\left[A \log \frac{x}{A}\right].$$

Thus the log normal distribution can be considered as a limiting form of the generalised logit-normal distribution.

FORM OF THE DISTRIBUTION

If the probability density of z is

$$g_2(z) = (\sqrt{2\pi})^{-1} \exp(-z^2/2)$$

then the corresponding probability density of x is

$$g_1(x) = \theta\left[\sigma\sqrt{2\pi}\, x\left(1 - \left(\frac{x}{A}\right)^\theta\right)\right]^{-1} \exp\left[-\left(\text{logit}\left(\frac{x}{A}\right)^\theta - \mu\right)^2/2\sigma^2\right].$$

Each of the transformation curves consists of an approximately straight section about the point of inflexion (z_0, x_0) with tails of varying degrees of curvature. As the gradient increases in the tails of the curve the ratio of $g_1(x)/g_2(z)$ increases; consequently the frequency distributions of x are generally platykurtic.

To exemplify the various shapes of the frequency distribution for x we begin by keeping σ^2 small and considering the effects of varying μ and θ. If $\mu = z_0$ then the frequency distribution of x is symmetrical close to the median value of x and has a longer tail to the right than to the left if $\theta < 1$ and a longer tail to the left if $\theta > 1$. If $\mu < z_0$ the mode of the frequency distribution for x is less than the median, and the central portion of the frequency distribution of x is skew to the right; at the same time the right hand tail increases in length and the left hand tail decreases as μ decreases. If $\mu > z_0$ the effects are reversed. Thus the skewness of the central portion of the frequency distribution of x

(about the median) is determined by the position of μ relative to z_0, while variation of θ shifts the distribution along the x-axis and alters the relative lengths of the tails.

In general, increasing σ^2 makes the frequency distribution of x more platykurtic. For large σ the skewness of the frequency distribution of x depends on the value of μ relative to $z_1 = \text{logit } (\frac{1}{2})^\theta$. Thus if $\mu - z_0$ and $\mu - z_1$ have the same sign the direction of the skewness remains the same as σ increases whereas if they have opposite signs the direction of the skewness is eventually altered as σ increases.

METHODS OF FITTING

The maximum likelihood equations for the four parameters lead to a complicated iterative procedure when ungrouped data are being used; when grouped data are used, as is frequently the case for size distribution, the procedure becomes extremely difficult. Consequently an alternative procedure is desirable, and the general method adopted has been to estimate μ and σ to give the best fit, as measured by minimum χ^2, for given A and θ and then to minimise this minimum χ^2 for variations in A and θ. Apart from the relative simplicity of such a two-step procedure this method makes it easier to investigate possible invariance of A or θ over different sets of data.

For this general method to be successful we need a quick and efficient method of estimating μ and σ for given A and θ. Two methods have been tried. The first is the ordinary method of moments, which has a major drawback in that the interval lengths on the transformed scale are unequal, which means that the estimated values of μ and σ contain biases which depend on the true values of μ and σ and therefore admit of no simple, 'Sheppard's-like', corrections. In addition, as the intervals with end-points zero or A on the x-scale become open-ended on the z-scale, a procedure for dealing with these has to be defined.

The second method is an iterative procedure which involves fitting a weighted regression, on the generalised logit of x_i, of the normal equivalent deviate of the cumulative frequency less than x_i, the weights being the normal 'probit' weights estimated from the fitted values of the previous iteration (the method is thus equivalent to fitting a probit curve to the cumulative frequencies, ignoring the correlations between these frequencies).

The equations used are

$$\begin{aligned} \gamma \sum_i w_i + \delta \sum_i w_i h_i &= \sum_i w_i d_i \\ \gamma \sum_i w_i h_i + \delta \sum_i w_i h_i^2 &= \sum_i w_i d_i h_i \end{aligned} \tag{3}$$

where

$\gamma = -\mu/\sigma$, $\delta = 1/\sigma$,
h_i = generalised logit of the upper end, x_i , of the group interval i,
$d_i = \Phi^{-1}$ (observed proportion $< x_i$), where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\tfrac{1}{2}t^2)\, dt$$

and

$$w_i = \frac{\exp(-c_i^2)}{g_i(1 - g_i)}$$

where

g_i = fitted proportion $< x_i$ (for the initial iteration use observed proportions)

and

$c_i = \Phi^{-1}(g_i)$ (initial iteration, d_i)

In practice it has been found that 2 iterations, including the initial fit with empirical weights, give fits sufficiently close to the correct values.

An objection to this method is that the correlations between the various cumulative frequencies are ignored. Results will be quoted in the next two sections to show that the existence of these correlations does not prevent the method from giving satisfactory fits.

COMPARISON OF METHODS

In order to demonstrate the effectiveness of the iterative method we need to show first that the use of χ^2 as a measure of the goodness-of-fit gives results similar to those obtained using Maximum Likelihood, and secondly that the iteration method gives a value of χ^2 which is close to the best possible χ^2 and is better than that obtained using the method of moments.

The effectiveness of χ^2 as a criterion of the goodness-of-fit depends on the χ^2 surface for variation of parameters being similar to the corresponding likelihood surface; i.e. it depends on $(-2 \log L) - \chi^2$ being relatively constant at least in the neighbourhood of the minimum of the surface. Using Stirling's approximation we obtain the relationship

$$-2 \log L = (k - 1) \log 2\pi n + \sum_i \log \frac{n_i}{n} + \sum_i \frac{(n_i - np_i)^2}{np_i} + Q(p_i) \quad (4)$$

where

n_i = observed number in group i, $n = \sum_i n_i$,
p_i = actual probability density of group i,
k = number of groups and
Q = consists of terms $O(n^{-1})$ depending on p_i .

To investigate the actual values of Q at the minimum χ^2 and the variation of Q about the minimum, values of χ^2 and $(-2 \log L)$ were calculated for constant A and sets of values of (θ, μ, σ) arranged in cuboctahedronal grids about an approximate minimum position. Two grids were used for each of 16 sets of the carrot root data, one grid having a very short step length, such that the maximum of χ^2 over the grid was about 2 greater than the minimum, and the other having rather longer steps.

Q was calculated for each point from (4) and δQ was defined as $|Q - Q_0|$ where Q_0 is the value of Q corresponding to χ_0^2 , the minimum observed value of χ^2. Within a data set δQ was fairly constant for small values of $\delta\chi^2 (= \chi^2 - \chi_0^2)$ but for larger values of $\delta\chi^2$, δQ tended to increase with $\delta\chi^2$ with $C = 100\ \delta Q/\delta\chi^2$ relatively constant. The Root Mean Square (RMS) of δQ for points with $\delta\chi^2 < 2$ was calculated for each set and for points with $\delta\chi^2 \geq 2$ the mean value of C was found.

The values of Q_0 were fairly constant for 14 of the data sets, giving a mean value of 0.05, but for the two worst fitting data sets Q_0 was approximately 0.25. Apart from the worst-fitting set the RMS of δQ varied between 0.01 and 0.22 with a mean value of 0.10 and $\bar{C}$ varied between 4% and 12% with a mean value of 7%. In general the higher values of both criteria were associated with the worse fits. For the data set giving the worst fit the values calculated were RMS $(\delta Q) = 0.48$ and $\bar{C} = 36\%$. However the χ^2 surface for this fit is extremely flat and it is therefore relatively unimportant to obtain the exact minimum.

It therefore appears that for these data sets, in which the numbers of roots were quite high (150–700), differences between the two surfaces are unimportant for the purpose of finding the minimum position of either surface (only in the very bad fit did the minimum positions differ) and the use of χ^2 is justified. Further investigations into the differences between the two surfaces might be useful if the distribution was being fitted to data with smaller numbers in the groups, or with markedly different values of the parameters.

The iterative regression method and the method of moments were compared for a number of grids of (A, θ) for various sets of data. The χ^2 obtained were also compared with an estimated minimum χ^2 for some (A, θ) found by fitting the data for a grid of μ and σ values. The problem

of open-ended grades on the transformed scale in the method of moments was dealt with by finding mean weights of roots in the various grades and estimating the mean diameter for the open-ended grades by assuming a linear relationship between log mean weight and log mean diameter. On the direct comparison it was found that the regression method gave a lower χ^2 than the method of moments in 133 out of the 144 cases (consisting of 3×3 (A, θ) grids for each of 16 sets of data). Furthermore the χ^2 from the regression method compared well with the minimum χ^2 as estimated from the (μ, σ) grid. The values of χ^2 obtained by the three methods are given in Table 1 for 9 pairs of (A, θ) values for one set of the carrot root data.

TABLE 1

VALUES OF χ^2 FOR THREE METHODS OF ESTIMATING THE MEAN AND VARIANCE OF THE TRANSFORMED VARIATE, z

log d	A	Method of moments	2 Iterations of the Regression Method	Estimated Minimum χ^2
−0.5	9.0	0.859	0.557	0.548
−0.5	9.5	0.308	0.102	0.090
−0.5	10.0	0.256	0.111	0.098
−0.3	9.0	0.370	0.192	0.183
−0.3	9.5	0.232	0.118	0.109
−0.3	10.0	0.498	0.427	0.417
−0.1	9.0	0.236	0.155	0.148
−0.1	9.5	0.542	0.500	0.486
−0.1	10.0	1.145	1.135	1.102

EFFECT OF USING THE APPROXIMATE LEAST-SQUARES EQUATIONS

The correct least-square estimates of the parameters γ and δ are given by

$$(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})\hat{\psi} = \mathbf{A}'\mathbf{V}^{-1}\mathbf{d} \tag{5}$$

where

$$\boldsymbol{\Psi}' = (\gamma, \delta)$$

$$\mathbf{V} = \{v_{ij}\} \text{ is the variance—covariance matrix of the } d_i \text{ ,}$$

$$\mathbf{d}' = (d_1 , d_2 , \cdots d_k)$$

$$\mathbf{A}' = \begin{bmatrix} 1, & 1, & \cdots 1 \\ h_1 , & h_2 , & \cdots h_k \end{bmatrix}$$

and

k = number of points used in the regression.

The elements of $\mathbf{V}$ and of $\mathbf{U} = \mathbf{V}^{-1}$ can easily be calculated as

$$\begin{aligned} v_{ij} &= nP_iQ_j\phi_i^{-1}\phi_j^{-1} && (i \leq j) \\ u_{ii} &= n^{-1}\phi_i^2(p_i^{-1} + p_{i-1}^{-1}) && i = 1, 2, \cdots k \\ u_{i-1,i} &= n^{-1}\phi_{i-1}\phi_i p_i^{-1} && i = 2, 3, \cdots k \\ u_{i,i+1} &= n^{-1}\phi_i\phi_{i+1}p_{i+1}^{-1} && i = 1, 2, \cdots (k-1) \end{aligned} \tag{6}$$

where

$$P_i = 1 - Q_i = \Phi(\gamma + \delta h_i);\ \phi_i = \phi(\gamma + \delta h_i);$$

and

$$p_i = P_i - P_{i-1} (i = 2, 3, \cdots k);\ p_1 = P_1\ ;\ p_{k+1} = Q_k\ .$$

Thence

$$\mathbf{A}'\mathbf{V}^{-1}\mathbf{A} = \begin{bmatrix} \sum_{i=1}^{k} \phi_i f_i(\phi) & \sum_{i=1}^{k} t_i f_i(\phi) \\ \sum_{i=1}^{k} t_i f_i(\phi) & \sum_{i=1}^{k} t_i f_i(t) \end{bmatrix}$$

$$\mathbf{A}'\mathbf{V}^{-1}\mathbf{d} = \begin{bmatrix} \sum_{i=1}^{k} \phi_i d_i f_i(\phi) \\ \sum_{i=1}^{k} \phi_i d_i f_i(t) \end{bmatrix}$$

where $t_i = \phi_i h_i$, $\phi_0 = \phi_k = 0$ and

$$f_i(x) = x_{i-1}p_i^{-1} + x_i p_i^{-1} + x_i p_{i+1}^{-1} + x_{i+1}p_{i+1}^{-1}\ .$$

Four sets of carrot root data were fitted for 9 pairs of (A, θ) values arranged in a 3×3 grid about an approximate minimum position, using both the 'exact' method (equation (5)) and the 'approximate' method (equation (3)) with the Pearsonian χ^2,

$$\left[\sum \frac{(\text{expected number}—\text{observed number})^2}{\text{expected number}}\right]$$

as a measure of the goodness-of-fit. It was found the approximate method gave better fits than the exact method in most cases. In fact, of the 4 sets of data, on only one did the exact method show an improvement for any of the 9 points, and on this set the greatest difference was 1.3% of the exact value. On the other 3 sets 19 of the 27 fittings gave

values less than 5% different, and only 2 pairs of fittings were more than 10% different.

In these sets of data the numbers in the groupings were fairly large (an average of about 100 per group) but from (6) it can be seen that the correlations between the cumulative frequencies do not depend on the overall number. The evidence therefore suggests that the approximate method of fitting is all that is required.

DISCUSSION

In the two previous sections of this paper we have shown that, at least for certain forms of data, the iterative regression method provides a satisfactory method of fitting the normal distribution to grouped data where the groupings are unequally spaced. While it is of course true that the greatest advantage of this method lies in its ability to deal with irregular groupings there is no reason to suppose that it will not prove equally effective in fitting the normal distribution to equally spaced groups.

Earlier in the paper we have described briefly the effects of including θ in the transformation on the versatility of the family of distributions. To see whether the inclusion of θ is necessary for the satisfactory fitting of the original carrot data (which consisted of 2 harvest dates $\times 2$ mean plant densities $\times 4$ inter-row distances $= 16$ sets of data) the best fits, using minimum χ^2 as the criterion, were obtained under the following three hypotheses:—

(i) $\theta = 1$ and A is constant at each harvest;
(ii) A is constant at each harvest and θ is allowed to vary;
(iii) $\theta = 1$ and A is allowed to vary.

TABLE 2

BEST ESTIMATES OF θ FOR 16 DATA SETS: A HELD CONSTANT

	Harvest 1 ($A = 2\frac{1}{8}''$)		Harvest 2 ($A = 2\frac{5}{8}''$)	
Mean plant density (plants/sq. ft.)	16	8	16	8
Inter-row spacing				
4″	0.28	0.49	0.03	0.67
8″	0.77	0.04	2.72	0.53
12″	1.27	1.18	1.82	0.92
16″	1.05	1.35	3.22	3.19

For (i) 7 of the 16 data sets gave fits singificantly bad at the 5% level and using (ii) only 2 data sets gave bad fits. Thus if we wish to consider A as an upper limit to the size attainable at a given harvest we must obviously allow θ to vary. The way in which θ varies is shown in Table 2 in which the best estimates of θ for the 16 data sets are tabulated.

When the best fits using (iii) were found, 4 of the data sets gave bad fits (including the 2 giving bad fits for (ii)); thus, even if we are prepared to abandon the idea of A as an asymptotic upper limit to the size of the carrot root at a given harvest, we still obtain a better fit by allowing θ to vary, keeping A constant.

APPENDIX

EXAMPLES OF FITTING THE GENERALISED LOGIT-NORMAL DISTRIBUTION

Histograms from four sets of data of carrot root diameters are shown in Figure 1 together with the curves fitted by the iterative regression method.

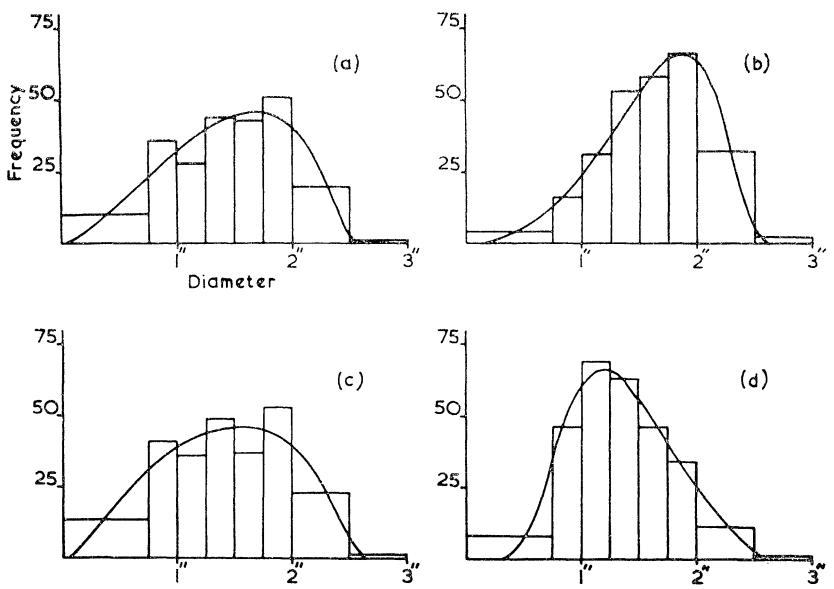

FIGURE 1

FITTED CURVES AND HISTOGRAMS FOR FOUR SETS OF CARROT ROOT DIAMETERS
(a) inter-row distance 4″ (b) inter-row distance 8″
(c) inter-row distance 12″ (d) inter-row distance 16″

The carrots were sown at an approximate overall mean density of 8 plants per square foot with inter-row distances of 4″, 8″, 12″ and 16″. At harvest the roots were graded through a series of circular holes of increasing diameter. For these sets of data it was found that A could be kept constant at $2\frac{5}{8}$″ without making the fits significantly worse. The estimated values of the other parameters are tabulated in Table 3, together with the corresponding χ^2.

TABLE 3

VALUES OF THE FITTED PARAMETERS AND χ^2 FOR THE FITTING OF FIGURE 1

	(a)	(b)	(c)	(d)
Inter row distance	4″	8″	12″	16″
θ	0.670	0.527	0.923	3.190
μ	0.778	1.349	0.294	−2.099
σ	0.871	0.729	0.991	1.300
χ^2	6.57	1.78	6.64	0.81

All χ^2 based on 3 d.f.

A programme in Mercury Autocode for the fitting of the distribution is available from the author.

ACKNOWLEDGEMENT

I am very grateful to Mr. J. A. Nelder for his considerable help and encouragement during the course of this work, in particular for his suggestion of the iterative regression method for fitting the distribution, and also to Mr. G. Berry for several helpful suggestions.

REFERENCES

Fisher, R. A. and Yates, F. [1957]. *Statistical tables for biological, agricultural and medical research*, 5th ed. Edinburgh: Oliver and Boyd.

Johnson, N. L. [1949]. Systems of frequency curves generated by methods of translation. *Biometrika 36*, 149–76.

Nelder, J. A. [1961]. The fitting of a generalization of the logistic curve. *Biometrics 17*, 89–110.