TECHNOMETRICS©, VOL. 22, NO. 2, MAY 1980

The Johnson System: Selection and Parameter Estimation

James F. Slifker and Samuel S. Shapiro
Department of Mathematical Sciences
Florida International University
Miami, FL 33199

This paper presents simple criteria which can be used to select which of the three members of the Johnson System of distributions should be used for fitting a set of data. The paper also presents elementary formulas for estimating the parameters for each of the members of the family. Thus, many obstacles to the use of the Johnson System are resolved.

KEY WORDS

Empirical distributions
Johnson distributions
Approximating distributions
Parameter estimation

1. INTRODUCTION

Statisticians are often faced with the problem of summarizing a set of data by means of a mathematical function which will fit the data and also allow them to obtain estimates of percentiles. A common practice is to use a flexible family of distributions to accomplish this, often a family with four parameters being chosen. The generalized lambda distribution which was described by Ramberg et al. (1979), the Pearson system for which Johnson, Nixon and Amos (1963) provided tables, and the Johnson (1949) system are examples of such empirical distributions.

The system proposed by Johnson contains three families of distributions which are generated by transformations of the form

$$z = \gamma + \eta\, k_i(x;\lambda,\epsilon), \tag{1}$$

where z is a standard normal variable and the $k_i\,(x;\lambda,\epsilon)$ are chosen to cover a wide range of possible shapes.

Johnson suggested the following functions:

$$k_1(x;\lambda,\epsilon) = \sinh^{-1}\left(\frac{x-\epsilon}{\lambda}\right),$$

Received December 1978; revised July 1977

denoted the S_U distribution,

$$k_2(x;\lambda,\epsilon) = \ln\left(\frac{x-\epsilon}{\lambda+\epsilon-x}\right),$$

denoted the S_B distribution, and

$$k_3(x;\lambda,\epsilon) = \ln\left(\frac{x-\epsilon}{\lambda}\right),$$

denoted the S_L distribution (lognormal).

The S_L is in essence a three-parameter lognormal distribution since the parameter λ can be eliminated by letting $\gamma^* = \gamma - \eta\ln\lambda$ so that $z = \gamma^* + \eta\ln(x-\epsilon)$. The S_B is a distribution bounded on $(\epsilon,\epsilon+\lambda)$ and the S_U is an unbounded distribution. See Hahn and Shapiro (1967) for further description of these distributions. The chosen functions are such that, in a plot of the third and fourth standardized moments $\sqrt{\beta_1}$ and β_2, the S_L distributions form a curve which divides the $(\sqrt{\beta_1},\beta_2)$ plane into two regions. The S_B distributions lie in one of the regions and the S_U lie in the other.

In using the Johnson system, the first step is to determine which of the three families should be used. The usual procedure is to compute the sample estimates of the standardized moments and choose the distribution according to which of the two regions the computed point falls into. Major shortcomings of this procedure are:

1. the variances of the estimates of the third and fourth moments are quite high;
2. the estimates of these moments are highly biased for small samples; see Johnson and Lowe (1979);
3. the moment estimators are greatly affected by outliers.

Once the family is selected, the next step is to estimate the parameters. Procedures for estimating the parameters of the S_L were given by Aitchison and Brown (1957). Johnson (1949) suggested the technique of matching percentiles for the S_B distribution and gave formulas for the cases in which one or both endpoints are known. If neither endpoint is known, his procedure required the solution of simultaneous nonlinear equations. Johnson suggested the technique of matching moments for the S_U distribution which required a two-way lookup table to determine the parameters γ and η from the sample estimates of $\sqrt{\beta_1}$ and β_2. Procedures for estimating parameters of the S_B distribution have been given by Bukac (1972) and Mage (1980) based on symmetrical points.

The objectives of this paper are to introduce a simple selection rule which is a function of four percentiles for selecting one of the three families and to give estimates of the parameters for all the families without requiring a solution of simultaneous equations or special tables. The results are summarized in Section 2. The procedures for selecting a specific family of distributions are described and, for each family, explicit formulas are given for obtaining estimates of the parameters. Applications of these techniques to two examples are presented in Section 3. The derivations of the results are left to Section 4.

2. SELECTION PROCEDURES AND PARAMETER ESTIMATION

This study was motivated by an attempt to find a property of the transformation (1) which could be used to select an appropriate member of the Johnson family to approximate a set of data which did not require use of high order moments. It was heuristically felt that there must be some relationship concerning distances in the tails vs. distances in the central portion of the distribution which could be used to distinguish bounded from unbounded cases. This led to the following formalization.

Consider any of the transformations described by equation (1). Choose any fixed value $z > 0$ of a standard normal variate. Then the four points $\pm z$ and $\pm 3z$ determine three intervals of equal length. The transformation (1) yields four values of x which are no longer equally spaced. It was hypothesized that, for a bounded symmetrical Johnson distribution, the distances between each of the outer and inner points would be smaller than the distance between the two inner points, and that the converse would be true for the unbounded case. This is easy enough to demonstrate and led to the following more general result. Let x_{3z}, x_z, x_{-z} and x_{-3z} be the values corresponding to $3z$, z, $-z$ and $-3z$ under the transformation (1). Let

$$\begin{aligned} m &= x_{3z} - x_z \\ n &= x_{-z} - x_{-3z} \\ p &= x_z - x_{-z}\,. \end{aligned} \tag{2}$$

In Section 4, the following will be proved.

$$\begin{aligned} &\text{(i)} \quad \frac{mn}{p^2} > 1 \text{ for any } S_U \text{ distribution;} \\ &\text{(ii)} \quad \frac{mn}{p^2} < 1 \text{ for any } S_B \text{ distribution;} \\ &\text{(iii)} \quad \frac{mn}{p^2} = 1 \text{ for any } S_L \text{ distribution.} \end{aligned} \tag{3}$$

Hence, this property can be used to discriminate among the three families.

Selection Procedure

The selection procedure thus consists of the following steps.

1. Choose a value of $z > 0$. This choice should be motivated by the number of data points. In general, for moderate-sized data sets, a value of z less than 1.0 would be chosen. A choice of z of 1.0 or higher would make it difficult to estimate the percentile points corresponding to $\pm 3z$. A more "typical" choice would be to use a value of z near 0.5 such as $z = 0.524$. This would dictate the use of $3z = 1.572$ and these 2 points would require estimating the 70th and 94.2th percentiles. However, the larger the number of observations, the larger the value of z that can be selected. It will be subsequently shown that z and the estimates for p, m and n can be used to obtain estimates of the distribution parameters. Thus, the choice of z can be motivated by choosing a value which assumes a close match of data and empirical distribution in the areas of greatest interest.

2. Determine from a table of areas for the normal distribution the percentages P_ζ corresponding to $\zeta = 3z$, z, $-z$ and $-3z$ respectively. For example, if $z = 0.2$ then $P_{0.2} = 0.5793$.

3. For each such ζ, obtain from the data the percentile $x^{(i)}$ corresponding to P_ζ using the relationship $(i - 1/2)/n = P_\zeta$ and set $x_\zeta = x^{(i)}$. Here n is the number of data points. Thus, x_ζ is the i^{th} ordered observation where $i = nP_\zeta + 1/2$. Since i in general will not be an integer it will be necessary to interpolate.

4. From the values obtained in the previous step, compute the sample values of m, n and p and use the criteria of (3) to select the appropriate member of the family.

It should be noted that since the $x^{(i)}$'s are random variables, the probability that $mn/p^2 = 1$ will be nil. If one wishes to use the S_L distribution it will be necessary to allow a tolerance interval around 1.

After the selection process is completed, the next problem is to estimate the parameters of the chosen distribution. As noted earlier, various parameter estimation techniques for the Johnson system have been suggested by several authors. In this paper, a uniform approach of matching percentiles is taken. Explicit formulas are given in each of the three cases which allow estimation of the parameters to be easily done on a hand calculator without the use of tables. The estimates are given in terms of the chosen value of z and the values of m, n and p computed in the selection process from the data. It is assumed in all instances that none of the parameters are known.

For each of the three families, the formulas are obtained by starting with a given Johnson distribution and fixed $z > 0$, and then solving explicitly for the parameters in terms of z and the population values of m, n and p.

It should be noted that the following formulas express the parameter values as functions of m, n, and p which, in turn, are functions of the population values of x_{3z}, x_z, x_{-z}, *and* x_{-3z}. In practice, one would compute the corresponding parameter estimates based on the sample x values.

(i) Johnson S_U Distribution

$$z = \gamma + \eta \sinh^{-1}\left(\frac{x-\epsilon}{\lambda}\right). \qquad (4)$$

The values of the parameters are presented in such a way as to emphasize their dependence on the ratios m/p and n/p.

Parameter Estimates for Johnson S_U Distribution

$$\eta = \frac{2z}{\cosh^{-1}\left[\frac{1}{2}\left(\frac{m}{p} + \frac{n}{p}\right)\right]}; \qquad (\eta > 0)$$

$$\gamma = \eta \sinh^{-1}\left[\frac{\frac{n}{p} - \frac{m}{p}}{2\left(\frac{m}{p}\frac{n}{p} - 1\right)^{1/2}}\right]; \qquad (5)$$

$$\lambda = \frac{2p\left(\frac{m}{p}\frac{n}{p} - 1\right)^{1/2}}{\left(\frac{m}{p} + \frac{n}{p} - 2\right)\left(\frac{m}{p} + \frac{n}{p} + 2\right)^{1/2}}; \qquad (\lambda > 0)$$

$$\epsilon = \frac{x_z + x_{-z}}{2} + \frac{p\left(\frac{n}{p} - \frac{m}{p}\right)}{2\left(\frac{m}{p} + \frac{n}{p} - 2\right)}.$$

(ii) Johnson S_B Distribution

$$z = \gamma + \eta \ln\left(\frac{x-\epsilon}{\lambda + \epsilon - x}\right). \qquad (6)$$

The solutions for the S_B parameters turn out to depend on the ratios p/m and p/n (as opposed to m/p and n/p for S_U).

Parameter Estimates for Johnson S_B Distribution

$$\eta = \frac{z}{\cosh^{-1}\left(\frac{1}{2}\left[\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right)\right]^{1/2}\right)}; \qquad (\eta > 0)$$

$$\gamma = \eta \sinh^{-1}\left[\frac{\left(\frac{p}{n} - \frac{p}{m}\right)\left\{\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right) - 4\right\}^{1/2}}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)}\right]; \qquad (7)$$

$$\lambda = \frac{p\left[\left\{\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right) - 2\right\}^2 - 4\right]^{1/2}}{\frac{p}{m}\frac{p}{n} - 1}; \qquad (\lambda > 0)$$

$$\epsilon = \frac{x_z + x_{-z}}{2} - \frac{\lambda}{2} + \frac{p\left(\frac{p}{n} - \frac{p}{m}\right)}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)}.$$

(iii) Johnson S_L Distribution (lognormal)

$$z = \gamma^* + \eta \ln(x - \epsilon). \qquad (8)$$

Estimates for the parameters in this case are well known (e.g. Aitchison and Brown, 1957). However, for completeness, formulas are included which express these parameters in terms of the interval lengths m, n and p. These latter are no longer independent since, by (3),

$$\left(\frac{n}{p}\right) = \left(\frac{m}{p}\right)^{-1}.$$

Parameter Estimates for Johnson S_L Distribution

$$\eta = \frac{2z}{\ln\left(\frac{m}{p}\right)}$$

$$\gamma^* = \eta \ln\left[\frac{\frac{m}{p} - 1}{p\left(\frac{m}{p}\right)^{1/2}}\right] \qquad (9)$$

$$\epsilon = \frac{x_z + x_{-z}}{2} - \frac{p}{2}\,\frac{\frac{m}{p} + 1}{\frac{m}{p} - 1}.$$

Generalization

The four points $\pm z$ and $\pm 3z$ which are symmetric about 0 would be an appropriate choice if the entire set of data is to be fitted into a Johnson distribution.

At times, however, it may be desirable to approximate only a portion of the data. The same selection technique and essentially the same formulas (with one minor modification) may still be used for any four equally spaced normal values. We wish to thank David T. Mage for this observation (see Mage, 1980).

Four points which are equally spaced and symmetric about a point c may be expressed as

$$c - 3z, c - z, c + z, c + 3z.$$

A Johnson transformation

$$z = \gamma + \eta k_i(x; \lambda, \epsilon) \tag{1}$$

applied to these points gives four values of x from which, similar to (2), corresponding values of m, n and p can be obtained. But the given four points transform by (1) to the same x values that the points $-3z$, $-z$, z and $3z$ do by

$$z = (\gamma - c) + \eta k_i(x; \lambda, \epsilon). \tag{1'}$$

The results for this latter situation have already been stated.

The following conclusions can then be drawn for any four equally spaced normal values.

(i) Since (1′) is another transformation of the same type as (1), the criteria stated in (3) regarding mn/p^2 remain valid.

(ii) With the exception that γ is replaced by $\gamma - c$, the parameters in (1) are the same as the parameters in (1′). The formulas in (5), (7) and (9) give the parameters in the case described by (1′). It follows that to obtain the parameters in (1), the same formulas apply except that the expression yielding γ (or γ^* in (9)) will now give $\gamma - c$. Thus, the only modification to the formulas in (5), (7) and (9) is that c, the center of symmetry of the four chosen points, should be added to the expression for γ.

3. EXAMPLES

1. Johnson (1949) used a set of data considered by Pretorius (1930) on the length of beans to illustrate the fitting of an S_U distribution. The data and the fit obtained by Johnson using matching of moments are shown in Table 1. Using z values of ± 1 and ± 3 and interpolating in the table, the percentiles are estimated as the $(i - 1/2)/n$ ordered observations. The percentiles are:

$$x_3 = x_{(0.9986)} = 16.689$$

$$x_1 = x_{(0.8413)} = 15.242$$

$$x_{-1} = x_{(0.1587)} = 13.581$$

$$x_{-3} = x_{(0.0014)} = 10.409.$$

For example, the percentage 0.9986 corresponding to $z = 3$ is read from a table of normal values. The i^{th} ordered observation corresponding to this percentage is

$$i = (0.9986)(9440) + 0.5 = 9427.3.$$

TABLE 1—*Length of 9440 beans in mm.*

Central Values	Observed Frequencies	Johnson moment fit	Percentile fit
less than 9.25	-	2.6	2.1
9.5	1	2.7	2.4
10.0	7	5.8	5.3
10.5	18	12.1	11.4
11.0	36	25.7	23.2
11.5	70	55.2	49.1
12.0	115	118.0	107.6
12.5	199	249.3	232.2
13.0	437	508.7	491.8
13.5	929	970.6	975.2
14.0	1787	1642.5	1688.8
14.5	2294	2240.6	2294.9
15.0	2082	2130.3	2086.2
15.5	1129	1151.5	1087.5
16.0	275	290.1	315.3
16.5	55	32.2	56.6
17.0	6	2.0	9.4
more than 17.25	-	0.1	1.0
TOTAL	9440	9440	9440
Value of χ^2		87.1	48.0

Interpolating between the values 16.25 and 16.75 gives the value of x_3. Then

$$p = x_1 - x_{-1} = 1.661$$

$$m = x_3 - x_1 = 1.447$$

$$n = x_{-1} - x_{-3} = 3.172$$

yielding

$$\frac{mn}{p^2} = 1.664.$$

Since the discriminant is greater than 1, an S_U distribution is appropriate. The formulas given in (5) are used to obtain

$$\eta = \frac{2(1)}{\cosh^{-1}[\frac{1}{2}(0.871 + 1.910)]} = 2.333$$

$$\gamma = 2.333\sinh^{-1}\left[\frac{1.910 - 0.871}{2\sqrt{(1.910)(0.871) - 1}}\right] = 1.402$$

$$\lambda = \frac{2(1.661)\sqrt{(1.910)(0.871) - 1}}{(0.871 + 1.910 - 2)\sqrt{0.871 + 1.910 + 2}} = 1.585$$

$$\epsilon = \frac{15.242 + 13.581}{2} + \frac{1.661(1.910 - 0.871)}{2(1.910 + 0.871 - 2)} = 15.516.$$

The fit is shown in Table 1.

A comparison to the moment fit given by Johnson shows that the percentile fit is superior. The chi-squared value for the percentile fit was 48.0 as compared to 87.1 for the moment fit and thus, the percentile fit represents a considerable improvement. (The smallest three values and the largest two values were combined for the χ^2 computation.)

2. Hahn and Shapiro (1967, p. 207) presented a tabulation of the resistance of five hundred 1/2 ohm resistors. See Table 2 for data. Choosing a match at the 5% and 95% points corresponding to $z = \pm 1.645$ requires the other two percentiles to correspond to $z = \pm 0.5483$. These are associated with the 29.2% and 70.8% points of the data. Interpolation in the tabulation yields

$$x_{1.645} = x_{(0.9500)} = 0.786$$

$$x_{0.5483} = x_{(0.7082)} = 0.635$$

$$x_{-0.5483} = x_{(0.2918)} = 0.516$$

$$x_{-1.645} = x_{(0.0500)} = 0.432$$

and

$$\frac{mn}{p^2} = 0.896.$$

Since the discriminant is less than 1, an S_B distribution should be used and the formulas in (7) apply.

$$\eta = \frac{0.5483}{\cosh^{-1}[\frac{1}{2}\sqrt{(1.788)(2.417)}]} = 1.959$$

$$\gamma = 1.959\sinh^{-1}\left[\frac{(1.417 - 0.788)\sqrt{(1.788)(2.417) - 4}}{2[(0.788)(1.417) - 1]}\right] = 2.373$$

$$\lambda = \frac{(0.119)\sqrt{[(1.788)(2.417) - 2]^2 - 4}}{(0.788)(1.417) - 1} = 1.203$$

$$\epsilon = \frac{0.635 + 0.516}{2} - \frac{1.203}{2} + \frac{(0.119)(1.417 - 0.788)}{2[(0.788)(1.417) - 1]} = 0.295.$$

The fit is shown in Table 2.

TABLE 2—*Resistance of 500 1/2 ohm resistors.*

Central Value	Actual observations	S_B
less than 0.4	4	6.5
0.425	33	36.1
0.475	78	74.1
0.525	99	93.8
0.575	87	90.4
0.625	76	73.0
0.675	51	52.0
0.725	32	33.7
0.775	21	19.9
0.825	7	10.9
0.875	5	5.5
more than 0.9	7	4.1
TOTAL	500	500
Value of χ^2		3.64

The resulting fit is quite good, the chi-squared value being 3.64 which is not significant at the 5% level. (The first and last cells were combined with their neighbors in computing the chi-squared value.)

4. ALGEBRAIC DERIVATIONS

In this section, the criteria stated in (3) are established and the parameter estimates given in (5), (7) and (9) are developed.

(i) S_U Distributions

Solving (4) for x in terms of z yields

$$x = \epsilon - \lambda\sinh\left(\frac{\gamma - z}{\eta}\right).$$

With a fixed positive value of z being chosen, the values of m, n and p as defined in (2) can be expressed as

$$m = \lambda\left[\sinh\left(\frac{\gamma - 2z + z}{\eta}\right) - \sinh\left(\frac{\gamma - 2z - z}{\eta}\right)\right]$$

$$n = \lambda\left[\sinh\left(\frac{\gamma + 2z + z}{\eta}\right) - \sinh\left(\frac{\gamma + 2z - z}{\eta}\right)\right] \quad (10)$$

$$p = \lambda\left[\sinh\left(\frac{\gamma + z}{\eta}\right) - \sinh\left(\frac{\gamma - z}{\eta}\right)\right].$$

Using the standard formula

$$\sinh(A + B) - \sinh(A - B) = 2\cosh A \sinh B,$$

it is easy to obtain

$$\frac{m}{p} = \frac{\cosh\left(\frac{\gamma - 2z}{\eta}\right)}{\cosh\left(\frac{\gamma}{\eta}\right)}$$

$$\frac{n}{p} = \frac{\cosh\left(\frac{\gamma + 2z}{\eta}\right)}{\cosh\left(\frac{\gamma}{\eta}\right)} . \tag{11}$$

On the one hand, the identity

$$\cosh(A + B) \cdot \cosh(A - B) = \cosh^2 A + \cosh^2 B - 1$$

implies that

$$\frac{m}{p} \cdot \frac{n}{p} = \frac{\cosh^2\left(\frac{\gamma}{\eta}\right) + \cosh^2\left(\frac{2z}{\eta}\right) - 1}{\cosh^2\left(\frac{\gamma}{\eta}\right)} . \tag{12}$$

Since $\cosh^2(2z/\eta) > 1$, it is clear that $m/p \cdot n/p > 1$ which establishes (i) in (3).

On the other hand, the identity

$$\cosh(A + B) + \cosh(A - B) = 2\cosh A \cosh B$$

can be applied to (11) to give

$$\frac{m}{p} + \frac{n}{p} = 2\cosh\left(\frac{2z}{\eta}\right) . \tag{13}$$

This immediately gives the indicated value of η in (5). Since it is always assumed that $\eta > 0$, the positive value of the inverse cosh must be chosen.

The expression for $\cosh(2z/\eta)$ in (13) can be substituted into (12) and the result is readily solvable for $\cosh^2(\gamma/\eta)$.

$$\cosh^2\left(\frac{\gamma}{\eta}\right) = \frac{(m + n)^2 - 4p^2}{4(mn - p^2)} . \tag{14}$$

This leads to

$$\sinh\left(\frac{\gamma}{\eta}\right) = \frac{[(m - n)^2]^{1/2}}{2(mn - p^2)^{1/2}} .$$

Unlike η, the parameter γ may be either positive or negative, so a determination of the sign of the numerator in this expression must be made. Another hyperbolic identity applied to (11) gives

$$\frac{n}{p} - \frac{m}{p} = \frac{2\sinh\left(\frac{\gamma}{\eta}\right)\sinh\left(\frac{2z}{\eta}\right)}{\cosh\left(\frac{\gamma}{\eta}\right)} .$$

Both $\cosh(\gamma/\eta) > 0$ and $\sinh(2z/\eta) > 0$ (since $z > 0$) and hence the sign of $\sinh(\gamma/\eta)$ is the same as that of $n/p - m/p$, or equivalently, of $n - m$. It follows that

$$\sinh\left(\frac{\gamma}{\eta}\right) = \frac{n - m}{2(mn - p^2)^{1/2}} \tag{15}$$

and γ as given in (5) is obtained.

It is now straightforward but tedious to deduce the remaining parameters. As indicated earlier, the expression for p in (10) can be rewritten as

$$p = 2\lambda \cosh\left(\frac{\gamma}{\eta}\right)\sinh\left(\frac{z}{\eta}\right) .$$

$\operatorname{Cosh}(\gamma/\eta)$ is known from (14) and $\sinh(z/\eta)$ can be obtained from (13). This leads to

$$\lambda = \frac{2p[p(mn - p^2)]^{1/2}}{(m + n - 2p)(m + n + 2p)^{1/2}} \tag{16}$$

which is equivalent to the result given in (5).

Finally, to get the location parameter ϵ, consider the sum

$$x_z + x_{-z} = 2\epsilon - \lambda\left[\sinh\left(\frac{\gamma - z}{\eta}\right) + \sinh\left(\frac{\gamma + z}{\eta}\right)\right] ;$$

that is

$$x_z + x_{-z} = 2\epsilon - 2\lambda\sinh\left(\frac{\gamma}{\eta}\right)\cosh\left(\frac{z}{\eta}\right) .$$

Using (15), (16) and the value of $\cosh(z/\eta)$ obtained from (13), this simplifies to

$$x_z + x_{-z} = 2\epsilon - \frac{p(n - m)}{m + n - 2p}$$

which yields ϵ as given in (5).

(ii) S_B Distribution

In the case of the S_B distribution, the solution for x in (6) is

$$x = \epsilon + \frac{\lambda}{1 + \exp\left(\frac{\gamma - z}{\eta}\right)} .$$

Proceeding as in the previous case, it is a straightforward algebraic exercise to express m, n and p in the following forms.

$$m = \frac{\lambda \sinh\left(\frac{z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma - 2z}{\eta}\right)}$$

$$n = \frac{\lambda \sinh\left(\frac{z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma + 2z}{\eta}\right)} \tag{17}$$

$$p = \frac{\lambda \sinh\left(\frac{z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)} .$$

Hence,

$$\frac{p}{m} = \frac{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma - 2z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)}$$

$$\frac{p}{n} = \frac{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma + 2z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)} . \tag{18}$$

With the aid of basic identities for the hyperbolic functions, it is easy to obtain

$$\frac{p}{m} \cdot \frac{p}{n} = \frac{A}{B}$$

where

$$A = \cosh^2\left(\frac{z}{\eta}\right) + \cosh^2\left(\frac{\gamma}{\eta}\right) + 2\cosh\left(\frac{z}{\eta}\right)\cosh\left(\frac{\gamma}{\eta}\right)\cosh\left(\frac{2z}{\eta}\right) + \sinh^2\left(\frac{2z}{\eta}\right)$$

and

$$B = \cosh^2\left(\frac{z}{\eta}\right) + \cosh^2\left(\frac{\gamma}{\eta}\right) + 2\cosh\left(\frac{z}{\eta}\right)\cosh\left(\frac{\gamma}{\eta}\right) \cdot$$

B is obviously smaller than A since the cosh function assumes only positive values and since $\cosh(2z/\eta) > 1$ and $\sinh^2(2z/\eta) > 0$. This gives $p/m \cdot p/n > 1$, or equivalently, $mn/p^2 < 1$ which establishes (ii) in (3).

From (18)

$$1 + \frac{p}{m} = \frac{2\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right) + \cosh\left(\frac{\gamma - 2z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)}$$

$$1 + \frac{p}{n} = \frac{2\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right) + \cosh\left(\frac{\gamma + 2z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)} .$$

In these expressions, both $\cosh((\gamma - 2z)/\eta)$ and $\cosh((\gamma + 2z)/\eta)$ are expanded in terms of γ/η and z/η. It is then a tedious exercise to verify that

$$\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right) = 4\cosh^2\left(\frac{z}{\eta}\right) \tag{19}$$

which yields the value of η given in (7).

Consider now the sum of the terms in (18).

$$\frac{p}{m} + \frac{p}{n} = \frac{2\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma + 2z}{\eta}\right) + \cosh\left(\frac{\gamma - 2z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)} .$$

After the terms in the numerator are expanded, this can be solved for $\cosh(\gamma/\eta)$ in terms of z/η.

$$\cosh\left(\frac{\gamma}{\eta}\right) = \frac{\left(\frac{p}{m} + \frac{p}{n} - 2\right)\cosh\left(\frac{z}{\eta}\right)}{4\sinh^2\left(\frac{z}{\eta}\right) - \left(\frac{p}{m} + \frac{p}{n} - 2\right)} .$$

From (19), $\cosh(z/\eta)$ and $\sinh^2(z/\eta)$ are obtained and substituted into this equation. The result is

$$\cosh\left(\frac{\gamma}{\eta}\right) = \frac{\left(\frac{p}{m} + \frac{p}{n} - 2\right)\left[\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right)\right]^{1/2}}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)} . \tag{20}$$

Rather than solve this for γ, it appears preferable to derive $\sinh(\gamma/\eta)$ because $\sinh^{-1}$, being single-valued, will yield the correct sign of γ. After some algebraic manipulations,

$$\sinh\left(\frac{\gamma}{\eta}\right) = \frac{\left[\left(\frac{p}{m} - \frac{p}{n}\right)^2\right]^{1/2}\left[\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right) - 4\right]^{1/2}}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)} \cdot \tag{21}$$

To see which root should be taken in the first factor in the numerator, observe that, by (18),

$$\frac{p}{n} - \frac{p}{m} = \frac{2\sinh\left(\frac{\gamma}{\eta}\right)\sinh\left(\frac{2z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)} ,$$

where the appropriate hyperbolic identity has been used. The denominator on the right is always positive and $\sinh(2z/\eta) > 0$ since $z > 0$. Hence, it follows that the sign of $\sinh(\gamma/\eta)$ is the same as that of $p/n - p/m$. Equation (21) then becomes

$$\sinh\left(\frac{\gamma}{\eta}\right) = \frac{\left(\frac{p}{n} - \frac{p}{m}\right)\left[\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right) - 4\right]^{1/2}}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)}$$

and gives the value of γ in (7).

The parameter λ can be evaluated using the expression for p in (17).

$$\lambda = \frac{p\left[\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)\right]}{\sinh\left(\frac{z}{\eta}\right)} . \tag{22}$$

Cosh (z/η), $\sinh(z/\eta)$ (by (19)) and $\cosh(\gamma/\eta)$ (by (20)) are already known in terms of m, n and p. After some algebraic simplification, the formula for λ in (7) is obtained.

For ϵ, consider the sum

$$x_z + x_{-z} = 2\epsilon + \lambda \left[\frac{1}{1 + \exp\left(\frac{\gamma - z}{\eta}\right)} + \frac{1}{1 + \exp\left(\frac{\gamma + z}{\eta}\right)} \right].$$

This is easily shown to be equivalent to

$$x_z + x_{-z} = 2\epsilon + \lambda \left[1 - \frac{\sinh\left(\frac{\gamma}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)} \right].$$

From (22), this reduces to

$$x_z + x_{-z} = 2\epsilon + \lambda - \frac{p \sinh\left(\frac{\gamma}{\eta}\right)}{\sinh\left(\frac{z}{\eta}\right)}$$

Substituting the previously determined values of $\sinh(\gamma/\eta)$ and $\sinh(z/\eta)$ in terms of p, m and n quickly yields the desired result for ϵ in (7).

(iii) S_L Distribution

By (8)

$$x = \epsilon + \exp\left(\frac{z - \gamma^*}{\eta}\right).$$

Then

$$m = \exp\left(\frac{3z - \gamma^*}{\eta}\right) - \exp\left(\frac{z - \gamma^*}{\eta}\right)$$
$$n = \exp\left(\frac{-z - \gamma^*}{\eta}\right) - \exp\left(\frac{-3z - \gamma^*}{\eta}\right) \qquad (23)$$
$$p = \exp\left(\frac{z - \gamma^*}{\eta}\right) - \exp\left(\frac{-z - \gamma^*}{\eta}\right).$$

It follows that

$$\frac{m}{p} = \exp\left(\frac{2z}{\eta}\right) \quad \text{and} \quad \frac{n}{p} = \exp\left(-\frac{2z}{\eta}\right) \qquad (24)$$

and therefore $mn/p^2 = 1$.

Moreover, the value of η is obtained from (24).

$$\eta = \frac{2z}{\ln\left(\frac{m}{p}\right)}$$

From (23) and (24),

$$p = \exp\left(-\frac{\gamma^*}{\eta}\right)\left[\left(\frac{m}{p}\right)^{1/2} - \left(\frac{m}{p}\right)^{-1/2}\right].$$

which yields γ^* as given in (9).

Finally,

$$x_z + x_{-z} = 2\epsilon + \exp\left(-\frac{\gamma^*}{\eta}\right)\left[\exp\left(\frac{z}{\eta}\right) + \exp\left(-\frac{z}{\eta}\right)\right].$$

Substituting the known expressions for $\exp(-\gamma^*/\eta)$ and $\exp(z/\eta)$ give the desired result for ϵ.

5. CONCLUSION

This paper has presented a procedure for determining by a simple means a method of choosing which of the three members of the Johnson System of distributions should be used to approximate a given set of data. The procedure chooses four percentiles which satisfy the relationship that their corresponding standard normal values are equally spaced with intervals $2z$, i.e., choose value of $3z$, z, $-z$ and $-3z$. In addition, techniques are given for estimating the parameters in the case where none are assumed known which do not require the use of tables or the solution of simultaneous nonlinear equations. All computations can be done on a hand calculator having the square root and exponential functions.

6. ACKNOWLEDGMENTS

The authors wish to thank the referees for their helpful comments.

REFERENCES

AITCHISON, J. and BROWN, J. A. C. (1957). *The Lognormal Distribution.* Cambridge, England: Cambridge University Press.

BUKAC, J. (1972). Fitting S_B curves using symmetrical percentile points. *Biometrika, 59,* 688–690.

HAHN, G. J. and SHAPIRO, S. S. (1967). *Statistical Models in Engineering.* New York: John Wiley & Sons, Inc.

JOHNSON, M. E. and LOWE, V. W. (1979). Bounds on the sample skewness and kurtosis. *Technometrics, 21,* 377–378.

JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika, 36,* 149–176.

JOHNSON, N. L., NIXON, E., and AMOS, D. E. (1963). Tables of percentage points of Pearson curves for given $\sqrt{\beta_1}$, β_2, expressed in standard measure. *Biometrika, 50,* 459.

MAGE, D. T. (1980). An explicit solution for S_B parameters using four percentile points. *Technometrics, 22,* 247–251.

RAMBERG, J. S., TADIKAMALLA, P. R., DUDEWICZ, E. J., and MYKYTKA, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics, 21,* 201–214.