Biometrika (1980), 67, 3, *pp.* 725–8
Printed in Great Britain

Quantile estimators of Johnson curve parameters

BY ROBERT E. WHEELER
E. I. du Pont de Nemours & Co., Wilmington, Delaware

SUMMARY

Easy quantile estimators are given for the parameters of Johnson curves. The estimators are at least as good as moment estimators.

Some key words: Johnson curves; Quantile estimators.

1. INTRODUCTION

Johnson's (1949) system of frequency curves for the variable x is defined by

$$z = \gamma + \delta \log f(u), \quad u = (x-\xi)/\lambda,$$

where z is a standard normal variable, and f has three possible forms:

$$\begin{array}{lll} S_L: & f(u) = u, & \text{the log normal;} \\ S_U: & f(u) = u + (1+u^2)^{\frac{1}{2}}, & \text{an unbounded distribution;} \\ S_B: & f(u) = u/(1-u), & \text{a bounded distribution.} \end{array}$$

Estimation of the parameters ξ and λ poses no problems once γ and δ are determined: simple linear regression is usually adequate.

Johnson (1949, 1965) suggested estimation procedures for the parameters based on moments. Fortran algorithms for these have been given by Hill, Hill & Holder (1976). Johnson (1949) discussed also estimation of all four parameters by quantiles in which $z = \gamma + \delta \log[f\{(x-\xi)/\lambda\}]$ is solved for selected pairs (z, x). In general a closed form solution for arbitrary quantiles is not possible. Aitchison & Brown (1957, pp. 58–62) have examined an estimator for the S_L curve whose closed form is obtained by specifying particular pairs (z, x). It is possible to extend this idea to the S_U and S_B curves with the result that very easy estimators of γ and δ are obtained.

The structure of these estimators is such that considerable light is thrown on the shape of the Johnson curves.

2. NOTATION

Let $F(z_n) = (n-\frac{1}{2})/N$ for integers n, N, where F is the standard normal distribution function and $n \leqslant N$. Define $x = \xi + \lambda f^{-1}(w)$, where $w = \exp\{(z-\gamma)/\delta\}$, and choose x_p, x_k, x_0, x_m, x_n, corresponding to $z = -z_n, -\frac{1}{2}z_n, 0, \frac{1}{2}z_n, z_n$. Define for later use $a = e^{-\gamma/\delta}$ and $b = e^{\frac{1}{2}z_n/\delta}$.

Any quantity of the form

$$\frac{x_i - x_j}{x_r - x_s} = \frac{f^{-1}(w_i) - f^{-1}(w_j)}{f^{-1}(w_r) - f^{-1}(w_s)}$$

does not depend on ξ or λ. This fact will be used to express γ and δ in terms of the five quantities x_p, x_k, x_0, x_m, x_n, and then sample order statistics will be substituted to obtain estimates $\hat{\gamma}$ and $\hat{\delta}$. Let $\{\hat{x}_i\}$ be a nondecreasing set of N observations and define $\hat{x}_0$ to be the central $\hat{x}_i$ or the average of the two central $\hat{x}_i$'s if N is even. Define the remaining sample order statistics $\hat{x}_p$, $\hat{x}_k$, $\hat{x}_m$, $\hat{x}_n$ by their subscripts, such that m is that integer in the range $[1, N]$ for which $|F(\frac{1}{2}z_n)-(m-\frac{1}{2})/N|$ is a minimum, and $k = 1+N-m$, $p = 1+N-n$. Except in the case of censored data, $n = N$ is recommended.

3. S_L curves

Only δ is important for S_L curves since $z = \gamma+\delta\log u = \delta\log(u/\theta)$, where $\gamma = -\delta\log\theta$. For S_L curves, $w = e^{z/\delta} = (x-\xi)/\lambda$, and

$$t = \frac{x_n-x_0}{x_0-x_p} = b^2;$$

hence $\delta = z_n/\log t$. A solution exists for $t \neq 1$, and substitution of the sample order statistics yields $\hat{\delta}$.

The usual convention is that δ is positive. If $t<1$ when sample order statistics are substituted, this convention may be observed by redefining the $\hat{x}_i$ as $-\hat{x}_i$; however, there is no essential difficulty with using a negative δ.

4. S_U curves

For S_U curves $\frac{1}{2}(w^2-1)/w = (x-\xi)/\lambda$, and

$$t = \frac{x_n-x_0}{x_0-x_p} = \frac{1+a^2b^2}{a^2+b^2}$$

which may be solved for a^2 to get $a^2 = (1-tb^2)/(t-b^2)$, from which one may find $\gamma = -\delta\log a$. A solution exists for $b^2>t>b^{-2}$.

Consider

$$t_u = \frac{x_n-x_p}{x_m-x_k} = \frac{1+b^2}{b}.$$

Since by convention δ is a positive quantity, only one solution of the quadratic is appropriate:

$$b = \tfrac{1}{2}t_u+\{(\tfrac{1}{2}t_u)^2-1\}^{\frac{1}{2}}$$

and thence $\delta = \frac{1}{2}z_n/\log b$.

A solution exists for $t_u>2$, and substitution of the sample order statistics yields $\hat{\delta}$ and $\hat{\gamma}$.

5. S_B curves

For S_B curves $w/(1+w) = (x-\xi)/\lambda$, and

$$t = \frac{x_n-x_0}{x_0-x_p} = \frac{a+b^2}{1+ab^2}$$

which may be solved for a to get

$$a = (t-b^2)/(1-tb^2)$$

from which one may find $\gamma = -\delta\log a$. A solution exists for $b^2>t>b^{-2}$.

Consider

$$t_b = \frac{(x_m-x_0)(x_n-x_p)}{(x_n-x_m)(x_0-x_p)} = \frac{1+b^2}{b}.$$

Since by convention δ is a positive quantity, only one solution of the quadratic is appropriate:

$$b = \tfrac{1}{2}t_b + \{(\tfrac{1}{2}t_b)^2 - 1\}^{\frac{1}{2}}$$

and thence $\delta = \frac{1}{2}z_n/\log b$. A solution exists for $t_b > 2$, and substitution of the sample order statistics yields $\hat{\delta}$ and $\hat{\gamma}$.

6. The statistics t, t_u, t_b

The five points used by the estimators serve well to describe Johnson's curves, and their relationships are summarized by the statistics t, t_u, t_b.

The statistic t serves to indicate the symmetry of the curve, and t is simply the ratio of the two tail lengths.

The statistic t_u depends only on the relative length of the tails to the central part of the distribution. This indicates that for S_U curves, the 'outer' tails exceed the 'inner' tails by a factor of two.

The statistic t_b depends on a similar sort of ratio for one tail, but only in relation to the length of the other tail. The t_b relationship is symmetric with respect to the tails in that if x_n and x_m are interchanged with x_p and x_k in the expression for t_b, the result is still $(1+b^2)/b$. Let t_b^1 and t_b^2 be the statistics computed from the two tails, then $\frac{1}{2}(t_b^1 + t_b^2)$ may in practice prove a better statistic to use in estimating δ.

In addition, the ratio

$$t_b/t_u = \frac{(x_m - x_0)(x_m - x_k)}{(x_n - x_m)(x_0 - x_p)}$$

serves to differentiate the three types. It is less than unity for S_U, equal to unity for S_L, and greater than unity for S_B. This may be seen by substituting the three forms for f in $f^{-1}(w) = (x - \xi)/\lambda$, and reducing t_b/t_u in a way similar to that in the above sections. Because of the two possible values for t_b, a better discriminator in practice may prove to be $\frac{1}{2}(t_b^1 + t_b^2)/t_u$.

When sample quantiles are substituted, the above symmetries and relationships will, of course, not hold exactly; however, a visual examination of the five sample quantiles on a probability plot in light of the above comments may be very informative.

7. Precision and accuracy

These estimators are of interest because of their simplicity and the light that they shed on the Johnson system. In practice they should provide good starting values for accurate iterative schemes, and they are at least as good as moment estimators.

Aitchison & Brown (1957, pp. 58–62) concluded that the S_L quantile estimator, with z_n the 95th percentile, is superior to the moment estimator.

The same may be concluded about many S_U curves because of the great difficulty in estimating third and fourth moments from modest samples: the moment ratio β_2 for a S_U curve with $\gamma = 0$, $\delta = 1$ is about 36, but several thousand observations are required before the extreme order statistics are large enough to produce a sample moment ratio of such a magnitude. The S_U moment estimations are thus, in general, quite bad for modest sample sizes.

Table 1 compares the quantile and moment estimators on 10 random samples of 30 observations each drawn from two S_B distributions. The moment estimates were calculated by the algorithm of Hill et al. (1976), and the quantile estimates used $n = N$. It may be seen that the quantile estimates are about as good as the moment estimates.

Table 1. *Approximate comparisons of quantile and moment estimators for S_B curve parameters based on* 10 *simulations*

	$\gamma = 0,\ \delta = 0{\cdot}5$				$\gamma = 0,\ \delta = 2$			
	Quantile		Moment		Quantile*		Moment**	
	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\delta}$
Average	0·14	0·55	0·02	0·56	0·29	1·72	0·20	1·59
Standard deviation	0·14	0·13	0·24	0·14	1·39	1·03	0·76	1·27

* No solution was possible in 3 cases.
** Failed to converge in 2 cases.

References

Aitchison, J. & Brown, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press.
Hill, I. D., Hill, R. & Holder, R. L. (1976). Fitting Johnson curves by moments. *Appl. Statist.* **25**, 180–9.
Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–76.
Johnson, N. L. (1965). Tables to facilitate fitting S_U curves. *Biometrika* **52**, 547–58.

[*Received October* 1979. *Revised June* 1980]