Proceedings of the 1983
Winter Simulation Conference
S. Roberts, J. Banks, B. Schmeiser (eds.)

FITTING JOHNSON CURVES TO UNIVARIATE AND MULTIVARIATE DATA

James R. Wilson
Mechanical Engineering Department
The University of Texas
Austin, TX 78712

RESEARCH SUMMARY

Moment matching and percentile matching are the standard methods for fitting a distribution from Johnson's translation system (the S_L, S_U, S_B, and S_N families) to a univariate data set (Johnson and Kotz 1970). One method for fitting a multivariate Johnson distribution to a vector-valued data set is to: (a) fit each marginal distribution separately with a univariate Johnson distribution; and then (b) fit a multivariate normal distribution to the transformed vectors using the sample correlation coefficients between the transformed coordinates. To make this approach numerically feasible, Wilson (1983) implemented an interactive percentile matching algorithm based on a modified Newton-Raphson procedure. When the algorithm was applied to 80 trivariate data sets arising in a large-scale policy analysis project, excessive manual intervention was required in some situations to obtain acceptable fits (Martin 1983).

As an alternative method for fitting a Johnson distribution to a univariate data set

$$\{ x_j : 1 \leqslant j \leqslant n \} \tag{1}$$

that has been sorted in ascending order, it is proposed that the "best" curve of the form

$$y = y(x) = \Phi\{\gamma + \delta\cdot f[(x-\xi)/\lambda]\} \tag{2}$$

should be fitted by nonlinear weighted least squares regression (Jennrich and Sampson 1968) to the set of sample points

$$\{ (x_j, y_j) : 1 \leqslant j \leqslant n \}, \tag{3}$$

where

$$y_j = j/(n+1), \qquad 1 \leqslant j \leqslant n, \tag{4}$$

$$\Phi(z) \equiv (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{z} \exp(-\tfrac{1}{2}t^2)\, dt, \quad -\infty < z < \infty, \tag{5}$$

and where $f(\cdot)$ is selected from among the alternative forms

$$f(u) = \begin{cases} \ln(u) & \text{for the } S_L \text{ family} \\ \sinh^{-1}(u) & \text{for the } S_U \text{ family} \\ \ln[u/(1-u)] & \text{for the } S_B \text{ family} \\ u & \text{for the } S_N \text{ family} \end{cases} \tag{6}$$

When this method was applied to a univariate data set arising in animal science, the resulting fit was significantly better than that obtained by conventional methods (Vaughan 1983).

REFERENCES

Jennrich RI, Sampson PF (1968), Application of stepwise regression to non-linear estimation, Technometrics, Vol. 10, No. 1, pp. 63 - 72.

Johnson NL, Kotz S (1970), Distributions in Statistics: Continuous Univariate Distributions - 1, John Wiley & Sons, New York.

Martin OJ (1983), ADSSIM: Aged and Disabled Services Simulation, Master's Report, Mechanical Engineering Department, University of Texas, Austin, Texas, 246 pp.

Vaughan DK (1983), Predictive Models of Postweaning Cattle Growth, Master's Report, Mechanical Engineering Department, University of Texas, Austin, Texas, 148 pp.

Wilson JR (1983), Modeling multiattribute populations with Johnson's translation system, Technical Report, Mechanical Engineering Department, University of Texas, Austin, Texas, 30 pp.

CH1953-9/83/0000-0115 $01.00 © 1983 IEEE