† This work was supported in part by the U.S. Army Research Office-Durham under Contract No. DAHC04-71-C-0042.

1. University of North Carolina at Chapel Hill.

2. Temple University, Philadelphia, and University of North Carolina at Chapel Hill.

EXTENDED AND MULTIVARIATE TUKEY LAMBDA DISTRIBUTIONS†

N. L. Johnson[1] and S. Kotz[2]

Institute of Statistics Mimeo Series No. 831

Department of Statistics
University of North Carolina at Chapel Hill

June, 1972

EXTENDED AND MULTIVARIATE TUKEY LAMBDA DISTRIBUTIONS†

N. L. Johnson[1] and S. Kotz[2]

SUMMARY

The symmetrical Tukey lambda distributions are the distributions of $[X^{\lambda}\ 1\ (1-X)^{\lambda}] \times \lambda^{-1}$ where X has a standard uniform distribution (range 0 to 1). Systems of multivariate distributions can be formed by applying transformations

$$Y_i = [X_i^{\lambda_i} - (1-X_i)^{\lambda_i}] \times \lambda_i^{-1}$$

when X_i $(i = 1,\ldots,m)$ have a joint Dirichlet distribution (with $0 \le X_i \le \sum_{j=1}^{m} X_j \le 1$). Since no more than one of the X's can have a uniform distribution, though all have beta distributions, we are led to study distributions of $[X^{\lambda} - (1-X)^{\lambda}]$ when X has a general beta distribution on $[0,1]$. These distributions are termed *extended* Tukey lambda distributions. Properties of these distributions are studied. Properties of the multivariate ones are also described and a numerical illustration is presented.

1. INTRODUCTION

The Tukey lambda distributions [2,7,8] have been found to be useful in a number of univariate sampling problems. These are distributions of

$$(1) \qquad Y = [aT^{\lambda} - (1-T)^{\lambda}]/\lambda$$

where $a > 0$, and T has a standard uniform distribution, with density

$$p_T(t) = 1 \quad (0 < t < 1).$$

The transformation (1) is monotonic increasing if $\lambda > 0$, and decreasing if $\lambda \le 0$. If $a = 1$, the distribution of Y is symmetrical. Some properties of such *symmetrical lambda* distributions have been described by Joiner and Rosenblatt [5].

If it is desired to form a multivariate distribution with variables which can be obtained by some such transformation as (1), it is natural to consider a joint Dirichlet distribution for $T_1,\ldots,T_m$, with density

$$(2)\quad p_{T_1,\ldots,T_m}(t_1,\ldots,t_m) = \frac{\Gamma\left(\sum_{j=0}^{m}\theta_j\right)}{\prod_{j=0}^{m}\Gamma(\theta_j)}\left(\prod_{j=1}^{m}t_j^{\theta_j-1}\right)\left(1-\sum_{j=1}^{m}t_j\right)^{\theta_0-1}$$

$$(\theta_0,\ldots,\theta_m > 0;\ t_1,\ldots,t_m > 0;\ \sum_{j=1}^{m}t_j < 1).$$

The marginal distribution of T_j is standard beta, with parameters θ_j, $\sum_{i=0}^{m}\theta_i - \theta_j$. Clearly, no more than one of the m variables $T_1,\ldots,T_m$ can have a uniform or indeed a symmetrical distribution.

We are thus led to consider the distribution of Y in (1) when T has a standard beta distribution with general parameters θ (> 0) and ϕ (> 0).

We shall refer to r.v. Y in (1) when T has a standard beta distribution as an *extended lambda variable*.

Since we can now obtain skew distributions even with $a = 1$, we will restrict ourselves to this case. In the next section, we will describe some properties of this family of distributions, and in the final section we will describe multivariate distributions with members of this family as marginals.

2. EXTENDED TUKEY LAMBDA DISTRIBUTIONS

If $\lambda > 0$ the range of variation of

$$Y = [T^\lambda - (1-T)^\lambda]\lambda^{-1}$$

is from λ^{-1} to λ^{+1}. If $\lambda < 0$ it is unlimited. Corresponding to $\lambda = 0$, we take $Y = \log[T/(1-T)]$ which also has unlimited range of variation (see [3]).

The density function of Y is

$$(3)\quad p_Y(y) = [B(\theta,\phi)]^{-1}\, t^{\theta-1}(1-t)^{\phi-1}\,[t^{\lambda-1} + (1-t)^{\lambda-1}]^{-1}$$

where each t has to be expressed in terms of y to satisfy the relation

$$y = [t^{\lambda} - (1-t)^{\lambda}]\lambda^{-1}.$$

(A simple explicit form for $p_Y(y)$ is not available.)

The r-th moment about zero is

$$(4) \qquad [B(\theta,\phi)]^{-1} \lambda^{-r} \sum_{j=0}^{r} \binom{r}{j} (-1)^j B(\theta + (r-j)\lambda, \phi + j\lambda).$$

In the case $\lambda = 0$, the j-th cumulant of Y is $K_j(Y) = \psi^{(j-1)}(\theta) + (-1)^j \psi^{(j-1)}(\phi)$, (see [3]). From (4) it is possible to compute the mean, variance, $\sqrt{\beta_1}$ and β_2 of Y. Some values of $\sqrt{\beta_1}$ and β_2 are given in Table 1.

Consider now the variation in shape of the distribution of Y with changes in λ, the parameters θ and ϕ having fixed values. Since

$$Y - (1-Y) = 2Y - 1 = Y^2 - (1-Y)^2$$

we see that the values of $\sqrt{\beta_1}$ and β_2 must be the same for $\lambda = 1$ and $\lambda = 2$. In each one of these two cases we have a beta distribution with the same parameters (though different range) as that of the original variables. In virtue of the continuity of the function involved, β_2 must take a maximum or minimum value for some value of λ between 1 and 2. It is, in fact, a minimum when $\theta = \phi$, but can be a maximum when $\theta \neq \phi$, and occurs for a value of λ in a remarkably narrow range about $\lambda = 1.45$ -- see Table 2 for a few examples. By considering neighboring loci of $(\sqrt{\beta_1}, \beta_2)$ (for λ varying) with slightly different values of θ and ϕ it can be seen that, at least for some values of $\sqrt{\beta_1}$ and β_2 there will be more than one set of values (θ,ϕ,λ) giving these values for the shape parameters.

Table 1   Moment-ratios for extended Tukey lambda distributions

		$\lambda=0.5$		$\lambda=1,2$		$\lambda=4$	
θ	ϕ	$\sqrt{\beta_1}$	β_2	$\sqrt{\beta_1}$	β_2	$\sqrt{\beta_1}$	β_2
0.25	0.25	0	1.42	0	1.29	0	1.44
0.25	0.5	0.59	1.98	0.68	1.91	0.59	2.03
0.25	1.0	1.09	3.23	1.38	3.79	1.03	3.12
0.25	5.0	1.67	5.99	2.93	13.59	1.90	6.23
0.25	10.0	1.72	6.35	3.38	18.36	2.45	9.63
0.5	0.5	0	1.72	0	1.50	0	1.83
0.5	1.0	0.48	2.20	0.64	2.14	0.40	2.27
0.5	5.0	1.08	3.96	1.93	7.25	1.05	3.28
0.5	10.0	1.14	4.29	2.31	9.94	1.52	5.00
1.0	1.0	0	2.08	0	1.80	0	2.45
1.0	5.0	0.64	3.08	1.18	4.20	0.35	2.24
1.0	10.0	0.73	3.38	1.52	5.78	0.77	2.91
5.0	5.0	0	2.73	0	2.54	0	3.70
5.0	10.0	0.16	2.84	0.33	2.82	-0.39	3.12
10.0	10.0	0	2.86	0	2.74	0	3.61

Table 2   Values of λ giving maximum or minimum β_2

θ	ϕ	λ	$\sqrt{\beta_1}$	β_2	θ	ϕ	λ	$\sqrt{\beta_1}$	β_2
0.25	0.25	1.43	0	1.268*	1	1	1.45	0	1.753
0.25	0.5	1.45	0.691	1.897*	1	5	1.45	1.299	4.549
0.25	1.0	1.43	1.425	3.868	1	10	1.43	1.688	6.573
0.25	5.0	1.42	3.179	15.64	5	5	1.48	0	2.489*
0.25	10.0	1.30	3.718	22.00	5	10	1.51	0.380	2.836
0.5	0.5	1.44	0	1.469*	10	10	1.49	0	2.705*
0.5	1.0	1.47	0.662	2.134*					
0.5	5.0	1.43	2.109	8.171					
0.5	10.0	1.41	2.549	11.69					

(* = minimum)

This might be expected since there are three parameters available to give two specified values ($\sqrt{\beta_1}$ and β_2). Another way of looking at the situation, is to consider what happens when λ is fixed, but θ and ϕ vary. It is known that for $\lambda = 1$, the region between the boundary $\beta_2 - (\sqrt{\beta_1})^2 - 1 = 0$ and the Type III line $2\beta_2 - 3(\sqrt{\beta_1})^2 - 6 = 0$ is covered. For $\lambda > 0$, generally, the region covered is that between the line $\beta_2 - (\sqrt{\beta_1})^2 - 1 = 0$ and the line (Johnson and Kotz [4]) corresponding to (Type III variable)$^\lambda$. The latter is approached as $\phi \to \infty$ with θ remaining constant.

Clearly, $p_Y(y) \to 0$ or ∞ at the extremes of the range of variation of y, according as $p_T(t) \to 0$ or ∞, for $\lambda > 1$.

Other modal values (if any) will correspond to solutions of the equation

$$(5) \quad \frac{dp_Y(y)}{dt} = 0 \quad \left(\text{note that } \frac{dp_Y(y)}{dy} = \frac{dp_Y(y)}{dt} \circ \frac{dt}{dy} \text{ and } \frac{dt}{dy} \neq 0\right).$$

Since

$$\frac{1}{p_Y(y)} \frac{dp_Y(y)}{dt} = \frac{\theta - 1}{t} - \frac{\phi - 1}{1 - t} - \frac{(\lambda-1)(t^{\lambda-2} - (1-t)^{\lambda-2})}{t^{\lambda-1} + (1-t)^{\lambda-1}},$$

(5) is equivalent to

$$(5)' \quad \theta - 1 - (\phi-1)u - (\lambda-1)u(u^{\lambda-2} - 1)(u^{\lambda-1} + 1)^{-1} = 0$$

with $u = t(1-t)$. Note that u increases from 0 to ∞ as t increases from 0 to 1.

From (5)', we have

$$(5)'' \quad u^{\lambda-1} = \frac{(\phi-\lambda)u - (\theta-1)}{\theta - \lambda - (\phi-1)u} = \frac{1 - \lambda u + (\phi u-\theta)}{u - \lambda - (\phi u-\theta)} = g(u), \text{ say.}$$

We consider a few special cases.

For $\lambda < 1 < \min(\theta,\phi)$, $g(u)$ increases from 0 to ∞ as u increases from $(\theta-1)/(\phi-\lambda)$ to $(\theta-\lambda)/(\phi-1)$, while $u^{\lambda-1}$ decreases as u increases.

Equation (5)" thus has just one root (since for u outside this range, $g(u)$ is negative).

A similar situation (in reverse) holds if $1 < \lambda < \min(\theta,\phi)$. If $\phi < \min(1,\lambda)$ and $\theta > \max(1,\lambda)$ then $g(u)$ is always negative (for $u > 0$) and (5)" has no solution. In this case, since $\theta > 1$ and $\phi < 1$, $p_Y(y) \to 0$ as $y \to -1$ and $\to \infty$ as $y \to 1$. The density function of Y is J-shaped, as is that of T. A similar situation (in reverse) holds if $\phi > \max(1,\lambda)$ and $\theta < \min(1,\lambda)$. For $\phi > 1$ and $\theta > 1$, $p_Y(y) = 0$ at $t = 0$ or 1, hence the slope of $p_Y(y)$ is also zero at these points.

An interesting special case corresponds to $\lambda = \theta+\phi-1$. Then (5)" becomes

$$u = \left(\frac{\theta-1}{\phi-1}\right)^{(\theta+\phi-2)^{-1}} .$$

We also note that in the symmetrical case $\theta = \phi$, (5)" is satisfied by

$$u^{\lambda-1} = 1$$

whence $t = \frac{1}{2}$ and the corresponding modal value for Y is 0, as is to be expected. The distribution can, however, be bimodal (with antimode at 0). This is so, for example, if $\lambda > 1$ and $\theta = \phi < \frac{1}{2}(\lambda^2-\lambda+2)$.

3. METHODS OF FITTING

Although the values of $\sqrt{\beta_1}$ and β_2 do not determine the values of θ, ϕ and λ uniquely, it is possible to fit the extended lambda distributions by moments, using the first, second and third sample moments. (If the range of variation is finite, it needs to be known, also.) The procedure is not, however, very convenient, at any rate at present. An iterative procedure using a trial value of λ to produce values of T corresponding to observed Y's (either using tables or a computing machine) can be employed. From the "observed T's" values of θ and ϕ can be obtained by equating sample mean and variance (of T) to the corresponding values ($\theta(\theta+\phi)^{-1}$, $\theta\phi(\theta+\phi)^{-2}(\theta+\phi+1)^{-1}$ respectively) for a standard beta distribution. Using these values of θ and ϕ, a new value for λ can be obtained by solving the equation

$$\text{Sample mean} \quad Y = \frac{\lambda\Gamma(\theta+\phi)}{\Gamma(\theta+\phi+\lambda)}\left[\frac{\Gamma(\theta+\lambda)}{\Gamma(\theta)} - \frac{\Gamma(\theta+\lambda)}{\Gamma(\phi)}\right].$$

The process is then repeated.

Alternatively a form of fitting to specified percentile points might be employed.

Whatever method is used, tables enabling the transformation $y = t^{\lambda} - (1-t)^{\lambda}$ to be easily inverted are useful.

4. MULTIVARIATE DISTRIBUTION

Consider the joint distribution of $Y_1,\ldots,Y_m$ where

$$Y_i = [T_i^{\lambda_i} - (1-T_i)^{\lambda_i}]\lambda_i^{-1} \qquad (i = 1,\ldots,m)$$

and the joint density of $T_1,T_2,\ldots,T_m$ is given by (2). The conditional distribution of T_j, given any subset $T_{a_1},\ldots,T_{a_k}$ of the remaining T's, is beta with parameters θ_j, $\sum_{i=0}^{m}\theta_i - \sum_{i=1}^{k}\theta_{a_i} - \theta_j$ and range of variation 0 to $1 - \sum_{h=1}^{k}T_{a_h}$.

The corresponding conditional distribution of Y_j is not an extended lambda distribution. It is, one might say, a "generalized extended lambda distribution", being the distribution of

$$[(cT')^{\lambda_j} - (1-cT')^{\lambda_j}]\lambda_j^{-1}$$

when $c = 1 - \sum_{h=1}^{k}T_{a_h}$, and T' has a standard beta distribution with parameters θ_j, $\sum_{i=0}^{m}\theta_i - \sum_{i=1}^{k}\theta_{a_i} - \theta_j$ and range of variation 0 to λ_j^{-1}.

The range of variation of Y_j is thus from -1 to $\lambda_j^{-1}[c^{\lambda_j}-(1-c)^{\lambda_j}]$. Although this distribution is undoubtedly of somewhat complicated form--simple expressions, even for moments, not being available -- we can get quite a good idea of the joint distribution by considering percentile points of the array distributions.

The *median regression* of Y_j on $Y_{a_1},\ldots,Y_{a_k}$, for example is

$$\text{Med}\,(Y_j | Y_{a_1},\ldots,Y_{a_k}) = (cT'_{\frac{1}{2}})^{\lambda_j} - (1-cT'_{\frac{1}{2}})^{\lambda_j}$$

where

$$I_{T'_{\frac{1}{2}}}(\theta_j, \sum_{i=0}^{m}\theta_i - \sum_{i=1}^{k}\theta_{a_i} - \theta_j) = \frac{1}{2}$$

$$c = 1 - \sum_{h=1}^{k} T_{a_h}; \quad Y_{a_h} = [T_{a_h}^{\lambda_h} - (1-T_{a_h})^{\lambda_h}]\lambda_h^{-1}.$$

(Note that $T'_{\frac{1}{2}}$ does not depend on $Y_{a_1},\ldots,Y_{a_k}$.) An *approximate* formula for $T'_{\frac{1}{2}}$ (based on (Mean-Mode) $\doteqdot$ 3(Mean-Median)) is

$$\left\{\theta_j + \frac{1}{3}(\theta_j-\theta_0-\sum_{i=1}^{k}\theta_{a_i})(\theta_j+\theta_0+\sum_{i=1}^{k}\theta_{a_i}-2)^{-1}\right\}(\theta_j + \theta_0 + \sum_{i=1}^{k}\theta_{a_i})^{-1}.$$

Other percentile points can be obtained in a similar way, but special care must be taken to allow for the sign of λ_j.

The effect of a change in the values of $Y_{a_1},\ldots,Y_{a_k}$ from $y_{a_1},\ldots,y_{a_k}$ to $y'_{a_1},\ldots,y'_{a_k}$ is easily assessed. Since

$$\Pr[Y_j < y | y_{a_1},\ldots,y_{a_k}] = \begin{cases} I_t(\theta_j, \sum_{i=0}^{m}\theta_i - \sum_{i=1}^{k}\theta_{a_i} - \theta_j) & \text{if } \lambda_j > 0 \\ 1 - I_t(\theta_j, \sum_{i=0}^{m}\theta_i - \sum_{i=1}^{k}\theta_{a_i} - \theta_j) & \text{if } \lambda_j < 0 \end{cases}$$

where

$$y = [(ct)^{\lambda_j} - (1-ct)^{\lambda_j}]\lambda_j^{-1}$$

with

$$c = 1 - \sum_{i=1}^{k} t_{a_i} \qquad (\lambda_h^{-1}[t_h^{\lambda_h} - (1-t_h)^{\lambda_h}] = y_h)$$

we have

$$\Pr[Y_j < y | y'_{a_1},\ldots,y'_{a_k}] = \begin{cases} I_{ct/c'}(\theta_j, \sum_{i=0}^{m}\theta_i - \sum_{i=1}^{k}\theta_{a_i} - \theta_j) & \text{if } \lambda_j > 0 \\ 1 - I_{ct/c'}(\theta_j, \sum_{i=0}^{m}\theta_i - \sum_{i=1}^{k}\theta_{a_i} - \theta_j) & \text{if } \lambda_j < 0, \end{cases}$$

where $c' = 1-\sum_{i=1}^{k} t'_{a_i}$, provided of course that y lies within the conditional limits of variation of Y_j in each case. If $\lambda_j > 0$, these are

$$-1 \text{ to } [c^{\lambda_j} - (1-c)^{\lambda_j}]\lambda_j^{-1}$$

$$-1 \text{ to } [c'^{\lambda_j} - (1-c')^{\lambda_j}]\lambda_j^{-1}$$

respectively, if $\lambda_j < 0$, they are

$$\lambda_j^{-1}[c^{\lambda_j} - (1-c)^{\lambda_j}] \text{ to } \infty,$$

$$\lambda_j^{-1}[c'^{\lambda_j} - (1-c')^{\lambda_j}] \text{ to } \infty,$$

respectively.

Tables of percentile points of the beta distribution [1, 6] can be used to construct regions of various kinds containing specified proportions of the joint distribution.

5. NUMERICAL ILLUSTRATION

As a simple example, we shall consider the case

$$\theta_0 = 2, \quad \theta_1 = \theta_2 = 1, \quad \lambda_1 = 0.5 \text{ and } \lambda_2 = 0.5.$$

Then

$$Y_j = 2[\sqrt{T_j} - \sqrt{(1-T_j)}] \qquad (j = 1,2),$$

where T_j has a beta distribution with parameters 1 and 3. The moments and moment ratios of Y_j are

$$E[Y_j] = -\frac{4}{45} = -0.089; \qquad \text{var}(Y_j) = 0.1093;$$

$$\sqrt{\beta_1}(Y_j) \;=\; -3.391 \quad \text{and} \quad \beta_2 \;=\; 14.903.$$

(The (β_1,β_2) point is in the Pearson Type I region.) The conditional distribution of Y_2, given Y_1, has cumulative distribution function

$$\Pr[Y_2 < y|Y_1 = y_1] \;=\; I_t(1,2) \quad (= \; 1 - (1-t)^2),$$

where $y = 2[(ct)^{\frac{1}{2}}-(1-ct)^{\frac{1}{2}}]$ with $c = 1-t_1$ and $y_1 = 2[t_1^{\frac{1}{2}}-(1-t_1)^{\frac{1}{2}}]$ (so that according as $y_1 \gtrless 0$

$$t_1 \;=\; \tfrac{1}{2}[1 \pm \sqrt{1-(1-\tfrac{1}{4}y_1^{\,2})^2}]).$$

Constants needed for calculation of a few percentiles of this conditional distribution are set out below:

$P \;=\; \Pr[Y_2 < y\|Y_1 = y_1]$	$t \;=\; 1 - \sqrt{(1-P)}$
0.01	0.00513
0.05	0.02532
0.25	0.13397
0.50	0.29289
0.75	0.50000
0.95	0.77639
0.99	0.90000.

Then $y = 2[\{t(1-t_1)\}^{\frac{1}{2}} - \{1-(1-t_1)t\}^{\frac{1}{2}}]$ where t_1 is given in the following table

y_1	t_1	y_1	t_1
-0.5	0.32600	0.1	0.53533
-0.4	0.36000	0.2	0.57053
-0.3	0.39453	0.3	0.60547
-0.2	0.42947	0.4	0.64000
-0.1	0.46467	0.5	0.67400
0	0.50000.		

Thus for example, the median of Y_2, when $Y_1 = 0.3$ is

$$2[\{0.29289 \times 0.39453\}^{\frac{1}{2}} - \{1 - 0.29289 \times 0.39453\}^{\frac{1}{2}}] \;=\; -\,1.201$$

For $Y_1 = -0.3$, the median is

$$2[\{0.28289 \times 0.60547\}^{\frac{1}{2}} - \{1 - 0.29289 \times 0.60547\}^{\frac{1}{2}}] = -0.986.$$

Note that since $t = 0.29289$ for $P = 0.5$, the conditional median is not greater than

$$2[(0.29289)^{\frac{1}{2}} - (0.70711)^{\frac{1}{2}}] = -0.547$$

whatever be the value of Y_1.

6. CONCLUDING REMARKS

Although the preceeding results are mathematically not very elegant, the distributions do have some practical appeal. In particular, computation of probabilities, is relatively straightforward. The original Dirichlet distribution determines the mathematical structure of the results. We have not so far been able to find a natural parent distribution giving formulae which are any simpler.

By comparison with Dirichlet distributions, the extra parameters $\lambda_1, \lambda_2, \ldots, \lambda_m$ introduce added flexibility. Evidence is lacking on the general value of this feature. It is likely that in some, perhaps most, cases only a few of the variables will need transformation (i.e., most λ's will be 1 (or, equivalently, 2)). In such cases, the untransformed variates will, of course, still have conditional beta distributions (with nonstandard range) just as in Dirichlet distributions.

REFERENCES

[1] Harter, H.L. (1964) *New Tables of the Incomplete Gamma-Function Ratio and of Percentage Points of the Chi-square and Beta Distributions*, Aerospace Res. Lab., Wright-Patterson AFB, Ohio.

[2] Hastings, C., Mosteller, F., Tukey, J.W. and Winsor, C.P. (1947) Low moments for small samples: a comparative study of order statistics, *Ann. Math. Statist.*, 18, 413-426.

[3] Johnson, N.L. (1949) Systems of frequency curves generated by methods of translation, *Biometrika*, 36, 149-176.

[4] Johnson, N.L. and Kotz, S. (1972) Power transformations of gamma variables, *Biometrika*, 59, 226-229.

[5] Joiner, B.L. and Rosenblatt, Joan R. (1971) Some properties of the range in samples from Tukey's symmetric lambda distributions, *J. Amer. Statist. Assoc.*, 66, 394-399.

[6] Pearson, E.S. and Hartley, H.O. (1964) *Biometrika Tables for Statisticians*, 1, Cambridge University Press.

[7] Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples), *Biometrika*, 52, 591-611.

[8] Tukey, J.W. (1962) The future of data analysis, *Ann. Math. Statist.*, 33, 1-67.

FOOTNOTES

Some Key Words: Tukey lambda distributions; Beta distributions; Dirichlet distributions; Transformations; Median regression.

† This work was supported in part by the U.S. Army Research Office-Durham under Contract No. DAHC04-71-C-0042.

1. University of North Carolina at Chapel Hill.

2. Temple University, Philadelphia, and University of North Carolina at Chapel Hill.