SELECTION AND RANKING PROCEDURES FOR TUKEY'S GENERALIZED LAMBDA DISTRIBUTIONS: SYMMETRIC CASE*

by

Shanti S. Gupta
Purdue University
West Lafayette, IN

Joong K. Sohn
Kyungpook National University
Taegu, Republic of Korea

Technical Report #85-29

Department of Statistics
Purdue University

Revised June 1988

*This research was supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

SELECTION AND RANKING PROCEDURES FOR TUKEY'S GENERALIZED LAMBDA DISTRIBUTIONS: SYMMETRIC CASE*

Shanti S. Gupta
Purdue University
West Lafayette, Indiana 47907
U.S.A.

Joong K. Sohn
Kyungbook National University
Taegu, Republic of Korea

SYNOPTIC ABSTRACT

A selection and ranking procedure based on sample medians for Tukey's generalized symmetric lambda populations is considered. Also,the properties of the proposed procedure such as asymptotic relative efficiencies (ARE) are studied. Tables of constants necessary to carry out the procedures,along with ARE's of the proposed procedure,are computed and tabulated. The performance of the procedure is studied under both a slippage-configuration and an equi-spaced configuration . An application of the lambda distribution for approximating some constants used in selection and ranking problems for some well-known symmetric distributions is discussed.

Key words and phrases: Subset Selection; Tukey's Generalized Lambda Distributions; Sample Medians, Asymptotic Relative Efficiency.

1. INTRODUCTION

Tukey's generalized lambda distribution (hereafter called lambda distribution) was suggested by Tukey (1960) as a wide class of symmetric distributions. It has been generalized by Ramberg and Schmeiser (1972, 1974) so as to include both symmetric and asymmetric distributions. Let $F(\cdot)$ denote the cdf of a lambda distribution and let $F^{-1}(\cdot)$ be its inverse. Then for $0 \leq p \leq 1$ and $x \in \mathbb{R}^1$,

$$x = F^{-1}(p) = \theta + \frac{1}{\beta}\{p^{\gamma_1} - (1-p)^{\gamma_2}\},$$

where θ and β are location and scale parameters, respectively, and γ_1 and γ_2 are shape parameters. If $\gamma_1 = \gamma_2$, then the lambda distribution is symmetric. Originally, Ramberg and Schmeiser (1972, 1974) generalized and used the lambda distribution for the purpose of generation of continuous unimodal symmetric and asymmetric random variates. Moberg, Ramberg and Randles (1978) have used the lambda distribution for Monte Carlo studies to check the robustness of the adaptive M-estimator for the selection problem under the indifference zone approach formulation. However, as can be seen from Table 5 in Section 5, the lambda distribution can be used to approximate many continuous theoretical distributions and empirical distributions. Also, compared with other families of distributions, the lambda family of distributions is known to be simple, flexible and easy to use as well as it is quite broad and general. Ramberg et al. (1979), Mykytka and Ramberg (1979), among others, have considered fitting an appropriate lambda distribution to a set of data. Hogg, Fisher and Randles (1975) have studied the (empirical) power of the adaptive distribution-free test by using the lambda distribution for various combinations of skewness and kurtosis. Joiner and Rosenblatt (1971) have studied the problem of the distribution of ranges of samples from the lambda distribution. It has also been shown by Sohn (1985) that the lambda distribution can be used to approximate the distribution of the sample mean of some symmetric continuous distributions which are not infinitely divisible. For example, Goel (1974) has derived the distribution of the sample mean from a logistic population as a series by using the method of characteristic functions and has provided tables for the cdf for n = 2(1)12 at points 0.00(0.01)3.89 and n = 13(1)15 at points 1.2(0.01)3.89. By using the lambda distribution, the cdf of the logistic sample mean can be computed easily. This was done by Sohn (1985). Comparison of the two sets of tabulated values shows that the maximum difference is less than 0.00155 for all values of n. This maximum difference occurs at the point x = 0.6 for all values of n. For $x \geq 1.0$, the difference decreases as x increases and for $x \in [1.2, 3.89]$, the maximum difference is less than 0.0007 for all n. The above discussion shows that the distribution of the sample mean of a logistic population can be approximated very well by using the lambda distribution.

For statistical selection and ranking problems, the use of the lambda distribution as a model provides results applicable to several parametric distributions, at least to get good approximate results and, of course, for the lambda distribution itself. Furthermore, by changing the values of the parameters of the lambda distribution, we can examine the performance of the selection procedures and thus can investigate the robustness of the statistical selection procedures.

It is well-known that for a symmetric distribution the sample median is an unbiased estimate of the location parameter and is robust in the presence of contamination from heavy-tailed distributions. Hence selection procedures based on sample medians, under the formulation of the subset selection approach, have been developed for several distributions. Gupta and Leong (1979) have considered a procedure for selecting the largest location parameter for the case of Laplace distributions. Gupta and Singh (1980) have studied the same problem for normal distributions. Lorenzen and McDonald (1981) have proposed and studied procedures based on sample medians for the case of logistic distributions.

In this paper, we are concerned with developing statistical selection procedures based on sample medians for selecting the population associated with the largest unknown location parameter among k symmetric lambda distributions with different (unknown) location parameters. In Sections 2 and 3, statistical selection procedures are proposed and their properties and performance are investigated. A numerical example is given in Section 4. Applications of the lambda distribution to the statistical selection and ranking problems for various parametric models are made in Section 5.

2. FORMULATION OF THE PROBLEM

Let $\pi_1,\pi_2,\ldots,\pi_k$ be $k(\geq 2)$ independent populations which are characterized by observable random variables $X_1,X_2,\ldots,X_k$, respectively. Let X_i follow a symmetric lambda distribution $F(x|\theta_i, \beta, \gamma)$, where θ_i is a location parameter, β and γ are known common scale and shape parameters. Without loss of generality, we assume that $\mathrm{Var}(X_i) = 1$, $i = 1,2,\ldots,k$, where

$$\mathrm{Var}(X_i) = \frac{2}{\beta^2}\left\{\frac{1}{2\gamma+1} - \frac{[\Gamma(\gamma+1)]^2}{\Gamma(2\gamma+2)}\right\}. \tag{1}$$

Let X_{ij}, $j = 1,2,\ldots,n$ be n independent observations from π_i, $i = 1,2,\ldots,k$, respectively. Let $\Omega = \{\underline{\theta}|\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \mathbb{R}^k\}$ be the parameter space and let $\Omega_0 = \{\underline{\theta} \in \Omega|\theta_1 =\ldots= \theta_k\}$. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ denote the ordered θ_i's. The population associated with $\theta_{[k]}$ is called the best population. Also let $\pi_{(i)}$ denote the population corresponding to $\theta_{[i]}$. It is assumed that no prior knowledge is available for the correct pairing between the sets, θ_i and $\pi_{(i)}$, $i = 1,2,\ldots,k$. Our goal is to define a procedure to select a nontrivial (nonempty) subset including the best population so as to satisfy the basic P*-condition, i.e., $\inf_{\theta\in\Omega} P_{\underline{\theta}}(CS|R) \geq P^*$, where CS stands for a correct selection (which means selection of any subset which includes the best). For convenience, let $n = 2m+1$ $(m \geq 1)$, and let $X_{i:\,m+1}$ be the sample median of π_i. Let $X_{[1]:\,m+1} \leq X_{[2]:\,m+1} \leq \cdots \leq X_{[k]:\,m+1}$ be ordered $X_{i:\,m+1}$'s. Also let $X_{(i):\,m+1}$ be the sample median corresponding to $\theta_{[i]}$. Now we propose the following selection rule R_T:

R_T: Select π_i if and only if $X_{i:\,m+1} \geq X_{[k]:\,m+1} - d_0$,

where $d_0(\geq 0)$ is chosen so as to satisfy the P*-condition.

Let $f(\cdot)$ and $F(\cdot)$ denote the pdf and cdf of X_i under $\underline{\theta} \in \Omega_0$. Then the following theorem holds.

Theorem 1. For the rule R_T,

$$\inf_{\underline{\theta}\in\Omega} P_{\underline{\theta}}(CS|R_T) = \inf_{\underline{\theta}\in\Omega_0} P_{\underline{\theta}}(CS|R_T)$$

$$= \frac{(2m+1)!}{(m!)^2} \int_{-\infty}^{\infty} I^{k-1}_{F(x+d_0)}(m+1,m+1)[F(x)]^m$$

$$\cdot[1-F(x)]^m f(x)dx, \qquad (2)$$

where $I_x(a,b)$ is an incomplete beta function with parameters a and b.

Proof. The proof follows immediately by using the general theory developed in Gupta and Panchapakesan (1979).

Corollary 1. By equating the right-hand-side of (2) to P*, one can obtain values of $d_0 \equiv d_0(k,m,P^*)$ for various values of k, m and P*.

Values of $d_0 \equiv d_0(k, m, P^*)$ are given in Table 1 for m = 1(1)5, k = 2,3(2)9,10,11, P* = 0.90, 0.95 and for (β, γ) = (-0.0466, -0.0246), (-0.0870, -0.0443), (-0.1389, -0.0667) and (-0.2306, -0.1045), where the corresponding values of kurtosis are 4.6, 5.0, 5.6 and 7.0, respectively and the common variance is 1.

\- - - - - - - - - - - - - - - - -

Table 1 approximately

\- - - - - - - - - - - - - - - - -

3. PROPERTIES AND PERFORMANCE OF THE PROPOSED PROCEDURE R_T

For any rule R, let $\phi_i(y_1,\ldots,y_k)$ be the probability that $\pi_{(i)}$ is selected given the observations y_i of the statistics Y_i, $i = 1,\ldots,k$. The ϕ_i are called the individual selection probabilities. Let $p_i = E[\phi_i(Y_1,\ldots,Y_k)]$, where the expectation is with respect to the distribution of $\underline{Y} = (Y_1,\ldots,Y_k)$. Then p_i denotes the probability that $\pi_{(i)}$ is selected by the rule R.

Definition 1.

(a) The rule R is strongly monotone in $\pi_{(i)}$ if p_i is nondecreasing in $\theta_{[i]}$ when all other components but $\theta_{[i]}$ of $\underline{\theta}$ are kept fixed and p_i is nonincreasing in $\theta_{[j]}$ for each $j \neq i$ when all components of $\underline{\theta}$ other than $\theta_{[j]}$ are kept fixed.

(b) For $\underline{\theta} \in \Omega$, R is said to be monotone if $p_i \leq p_j$ for $1 \leq i < j \leq k$.

(c) For $\underline{\theta} \in \Omega$ and $1 \leq i < k$, R is said to be unbiased if $p_i \leq p_k$.

Note that strong monotonicity for all i $\Rightarrow$ monotonicity $\Rightarrow$ unbiasedness.

(d) R is said to be invariant (symmetric) if

$$\phi_i(y_1,\ldots,y_i,\ldots,y_j,\ldots,y_k) = \phi_j(y_1,\ldots,y_j,\ldots,y_i,\ldots,y_k). \tag{3}$$

Now we have the following theorem.

Theorem 2. The rule R_T has the following properties.

(a) The proposed selection procedure R_T is strongly monotone in $\pi_{(i)}$, for all $i = 1,2,\ldots,k$; hence R_T is monotone and unbiased.

(c) The procedure R_T is invariant.

Let S denote the size of the subset selected by the rule R_T. Then the expected subset size, $E_{\underline{\theta}}(S|R_T)$, is given by

$$E_{\underline{\theta}}(S|R_T) = \sum_{i=1}^{k} \Pr\{\pi_{(i)} \text{ is selected}\}$$

$$= \sum_{i=1}^{k} \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j\neq i}}^{k} G(x+d_0+\theta_{[i]}-\theta_{[j]})dG(x) , \tag{4}$$

where $G(\cdot)$ is a cdf of $X_{i:\,m+1}$ under $\underline{\theta} \in \Omega_0$. We have the following result.

Theorem 3. For given k and $P^*(1/k < P^* < 1)$,

$$\sup_{\underline{\theta}\in\Omega} E_{\underline{\theta}}(S|R_T) \geq \sup_{\underline{\theta}\in\Omega_0} E_{\underline{\theta}}(S|R_T) = k\int_{-\infty}^{\infty} G^{k-1}(x+d_0)dG(x) = kP^*. \quad (5)$$

Now under a slippage configuration, that is, $\theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]}-\delta$, where $\delta > 0$, the asymptotic relative efficiency (ARE) of the proposed rule R_T relative to the means procedure R_G, which will be defined later, will be discussed.

Definition 2.

Under any given configuration of $\underline{\theta}$ let S' be the number of non-best populations selected. For $0 < \varepsilon < 1$, let $n_1(\varepsilon)$ and $n_2(\varepsilon)$ be minimum numbers of observations so that

$$E_{\underline{\theta}}(S'|R_i) \leq \varepsilon, \qquad i = 1,2, \quad (6)$$

for procedures R_1 and R_2. Then the ARE of the rule R_2 relative to R_1 under the configuration $\underline{\theta}$ is defined by

$$\text{ARE}(R_2, R_1 | \underline{\theta}) = \lim_{\varepsilon \downarrow 0} \frac{n_1(\varepsilon)}{n_2(\varepsilon)}, \tag{7}$$

provided that both procedures R_1 and R_2 satisfy the P*-condition. In the sequel, without loss of generality it will be assumed that $\theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]} - \delta = 0$. Also, the procedure R_G is defined by

R_G: Select π_i if and only if $\bar{X}_i \geq \max_j \bar{X}_j - d_G$,

where $\bar{X}_i$'s are sample means and d_G is a nonnegative constant chosen so as to meet the P*-condition. Then the following theorem holds.

Theorem 4. Under the slippage configuration as defined above,

$$\begin{aligned} \text{ARE}(R_T, R_G | \underline{\theta}) &= 4f^2(0) \\ &= 2^{2(\gamma-1)}\left(\frac{\beta}{\gamma}\right)^2. \end{aligned} \tag{8}$$

Proof. Proof is along the lines of the proof in Lorenzen and McDonald (1981) and hence is omitted.

Table 2 provides $\text{ARE}(R_T, R_G | \underline{\theta})$ for various values of β and γ for the following values of kurtosis $\mu_4/\mu_2^2 = 1.8, 3.0, 4.2, 5.0(1.0)\ 9.0$, with $\mu_2 = 1$, where μ_i is the i-th central moment.

Table 2. Values of $ARE(R_T, R_G|\underline{\theta})$

μ_4/μ_2^2	β	γ	$ARE(R_T, R_G\|\underline{\theta})$
1.8	.5744	1.0000	.3299
3.0	.1974	.1349	.6454
4.2	$-.0659\times10^{-2}$	$-.0363\times10^{-2}$	.8235
5.0	-.0870	-.0443	.9068
6.0	-.1686	-.0802	.9886
7.0	-.2306	-.1045	1.0532
8.0	-.2800	-.1233	1.0867
9.0	-.3203	-.1359	1.1503

It is known that (see Gupta and Singh (1980), Lorenzen and McDonald (1981), and Gupta and Leong (1979)) for the slippage configuration, ARE's of the median selection rules for the normal, logistic and double exponential distributions are 0.6366, 0.8225 and 1.0000, respectively. On the other hand, for values of kurtosis 3.0, 4.2, and 6.0 for the lambda distribution, the corresponding values of $ARE(R_T, R_G|\underline{\theta})$ are 0.6454, 0.8235 and 0.9886, respectively. These differences are mainly due to the approximation by lambda distributions with parameters β and γ for the corresponding distributions. Also one can see that when the tail of the underlying distribution becomes heavier, $ARE(R_T, R_G|\underline{\theta})$ increases and thus the rule R_T becomes as efficient as the procedure R_G and the rule R_T is more efficient than the rule R_G for very heavy-tailed distributions.

Now the performance of the rule R_T will be discussed in terms of $P_{\underline{\theta}}(CS|R_T)$, $E_{\underline{\theta}}(S'|R_T)$ and $P_{\underline{\theta}}(CS|R_T)/E_{\underline{\theta}}(S'|R_T)$. One can easily verify the following:

For $\underline{\theta} \in \Omega$,

$$P_{\underline{\theta}}(CS|R_T) = \frac{(2m+1)!}{(m!)^2} \int_0^1 \prod_{j=1}^{k-1} I_{F[\frac{1}{\beta}\{t^\gamma-(1-t)^\gamma\}+d_0+\theta_{[k]}-\theta_{[j]}]}(m+1,m+1) \cdot [t(1-t)]^m dt, \qquad (9)$$

$$E_{\underline{\theta}}(S|R_T) = \sum_{i=1}^{k} P_{\underline{\theta}}\{\pi_{(i)} \text{ is selected}|R_T\}$$

$$= P_{\underline{\theta}}(CS|R_T) + E_{\underline{\theta}}(S'|R_T), \qquad (10)$$

and

$$E(S'|R_T) = \sum_{i=1}^{k-1} \frac{(2m+1)!}{(m!)^2} \int_0^1 \prod_{\substack{j=1\\j\neq i}}^{k} I_{F[\frac{1}{\beta}\{t^\gamma-(1-t)^\gamma\}+d_0+\theta_{[i]}-\theta_{[j]}]} (m+1,m+1)[t(1-t)]^m dt. \qquad (11)$$

Here two configurations are considered, i.e., a slippage configuration $\theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]}-\delta$ and an equi-spaced configuration $\theta_{[1]} = \theta_{[2]}-\delta = \ldots = \theta_{[i]}-(i-1)\delta = \theta_{[k]}-(k-1)\delta$, where $\delta > 0$. Under a slippage configuration, equations (9) and (11) can be rewritten as

$$P_{\underline{\theta}}(CS|R_T) = \frac{(2m+1)!}{(m!)^2} \int_0^1 I^{k-1}_{F[\frac{1}{\beta}\{t^\gamma-(1-t)^\gamma\}+\delta+d_0]}(m+1,m+1)[t(1-t)]^m dt \qquad (12)$$

and

$$E_{\underline{\theta}}(S'|R_T) = (k-1) \frac{(2m+1)!}{(m!)^2} \int_0^1 I^{k-2}_{F[\frac{1}{\beta}\{t^\gamma-(1-t)^\gamma\}+d_0]}(m+1,m+1) \cdot I_{F[\frac{1}{\beta}\{t^\gamma-(1-t)^\gamma\}+d_0-\delta]}(m+1,m+1) \cdot [t(1-t)]^m dt. \qquad (13)$$

Values of $P_{\underline{\theta}}(CS|R_T)$, $E_{\underline{\theta}}(S'|R_T)$ and $P_{\underline{\theta}}(CS|R_T)/E_{\underline{\theta}}(S'|R_T)$ under both a slippage configuration and an equi-spaced configuration are computed for $\delta = 0.1(0.2)0.5$, $m = 3,5$, $k = 5,7$, $P^* = 0.90, 0.95$ and $(\beta,\gamma) = (-0.0466, -0.0246), (-0.2306, -0.1045)$. These are given in Tables 3 and 4. From these tables, the following conclusions can be drawn:

(1) As the value of kurtosis increases, the value of $P_{\underline{\theta}}(CS|R_T)/E_{\underline{\theta}}(S'|R_T)$ increases and hence the proposed rule R_T can be more effective (efficient) as the peakedness of the distribution becomes larger.

(2) Values of $P_{\underline{\theta}}(CS|R_T)/E_{\underline{\theta}}(S'|R_T)$ for $P^* = 0.90$ are uniformly larger than those for $P^* = 0.95$ for all combinations of values of k, m and δ for slippage configurations and also for equi-spaced configurations. One could anticipate this because an increase in the value of P* would tend to cause R_T to select more non-best populations compared with the improvement on $P_{\underline{\theta}}(CS|R_T)$.

These tabulated values may give some idea on the optimal choice of the value of P* in the sense of (approximately) maximizing the value of $P_{\underline{\theta}}(CS|R_T)$ and (approximately) minimizing the values of $E_{\underline{\theta}}(S'|R_T)$, simultaneously.

(3) An increase in either δ or m results in a decrease in $E_{\underline{\theta}}(S'|R_T)$; however, the expected value of S' is more sensitive to a change in δ. As δ becomes larger, it decreases substantially for both configurations.

Tables 3 and 4 approximately here

---

4. AN EXAMPLE

To illustrate the use of the proposed procedure we consider an example with real data. The data are taken from Edwards and Hsu (1983). Seven brands of water filters were tested for their ability to filter microorganisms from the river water. A high count of microorganism cultures grown for 24 hours on a given filter defines it as good. Three observations were made on each of the seven brands. The data are as follows.

Filter	Sample Observations	Median
5	139,133,124	133
7	127,125,114	125
1	117, 97, 70	97
3	112, 95, 87	95
2	108, 87, 72	87
6	66, 65, 61	65
4	68, 54, 53	54

We assume that the data belong to the generalized symmetric lambda distribution family with common values of the scale and shape parameters β, γ, respectively. In order to have the variance in each brand equal to unity, we assume the variance of each brand is 176 (note the usual pooled estimator of the common variance leads to this number approximately), and hence each observation is divided by 13.27. By changing values of β and γ, the results of using the selection procedure R_T are summarized as follows:

P*	(β, γ)	d_0-value	Selected Subset
0.90	(-0.0466, -0.0246)	1.7380	{5,7}
	(-0.1389, -0.0667)	1.6905	{5,7}
	(-0.2306, -0.1045)	1.6448	{5,7}
0.95	(-0.0466, -0.0246)	2.0462	{5,7}
	(-0.01389,-0.0667)	2.0000	{5,7}
	(-0.2306,-0.1045)	1.9540	{5,7}

Here with (β, γ) = (-0.0466, -0.0246), (-0.1389, -0.0667) and (-0.2306, -0.1045), lambda distribution has values of kurtosis 4.6, 5.6, 7.0, respectively. From the above table, one finds that the procedure R_T selects the same subset for reasonable perturbations in the values of (β,γ) thereby affecting the nature of the tails. This points to the 'robust' behavior of the selection rule R_T against the underlying assumptions of various symmetric distributions.

5. APPLICATIONS OF THE LAMBDA DISTRIBUTIONS

In this section, some applications of the lambda distribution for the evaluation of the d-values of subset selection rules are illustrated.

Here we restrict our attention to the symmetric case. As mentioned in the introduction, rules are illustrated. The lambda distribution can be used to approximate theoretical continuous symmetric distributions if values of location, scale and shape parameters are chosen properly.

Table 5 shows values of scale and shape parameters β and γ, respectively, with which the lambda distribution can be used to approximate some well-known symmetric distributions with $\mu_2 = 1$.

Table 5. Values of β and γ for some well-known distributions

distribution	μ_4/μ_2^2	β	γ
uniform	1.80	.5774	1.0000
normal	3.00	.1975	.1349
logistic	4.20	$-.0659\times10^{-2}$	$-.0363\times10^{-2}$
Laplace	6.00	-.1686	-.0802
t with 5 df	9.00	-.3202	-.1359
t with 10 df	4.00	.0261	.0148
t with 34 df	3.20	.1563	.1046

Now we consider an approximation of values of d_G of the procedure R_G defined in Section 2 for the normal model. If one wants to use the selection rule R_G, one needs values of d_G and these values are provided by many authors (for example, Gupta (1956), Gupta (1963), Gupta, Nagel and Panchapakesan (1973), among others). But by using the lambda distribution one can obtain approximate values of d_G, denoted by d_G', by solving the equation

$$\int_{-\infty}^{\infty} F^{k-1}(x+d_G')dF(x) = P^*, \tag{14}$$

where $F(\cdot)$ is a cdf of the lambda distribution with a scale parameter $\beta = 0.1975$ and a shape parameter $\gamma = 0.1349$. In the following table, values of d_G come from Gupta, Nagel and Panchapakesan (1973) and values of d_G' are evaluated from the equation (14).

Table 6. Values of d_G and $d_{G'}$

P*	k	d_G	d'_G
0.90	2	1.8125	1.8126
	5	2.5997	2.6024
	9	2.9301	2.9339
0.95	2	2.3262	2.3279
	5	3.0551	3.0596
	9	3.3678	3.3728
0.99	2	3.2899	3.2931
	5	3.9196	3.9227
	9	4.1999	4.2015

From Table 6, we see that the values of d'_G are fairly close to those of d_G. These agree to at least two decimal places. Furthermore, values of d'_G are conservative (larger than values of d_G); hence the P*-condition will not be violated if one uses d'_G-values in place of d_G-values.

Now we consider another approximation of the d-values of the subset selection procedures based on sample medians for the logistic distribution and compare those values with values from tables of Lorenzen and McDonald (1981). We know that a logistic distribution can be approximated by a lambda distribution with a scale parameter $\beta = -0.0659x10^{-2}$ and a shape parameter $\gamma = -0.0363x10^{-2}$. In the following table values of d_t come from the table of Lorenzen and McDonald (1981) and values of d_a are based on the approximation by using the lambda distribution.

Table 7. Values of d_t and d_a

m	P* / k	0.90		0.95	
		d_t	d_a	d_t	d_a
2	2	0.879	0.879	1.137	1.137
	5	1.274	1.273	1.510	1.510
	7	1.377	1.376	1.609	1.609
5	2	0.599	0.598	0.771	0.771
	5	0.863	0.863	1.019	1.018
	7	0.931	0.930	1.083	1.083
7	2	0.514	0.513	0.661	0.661
	5	0.740	0.739	0.872	0.872
	7	0.797	0.797	0.927	0.926
9	2	0.457	0.457	0.588	0.587
	5	0.657	0.657	0.775	0.774
	7	0.708	0.708	0.823	0.822

From Table 7, we can see that the approximation by using the lambda distribution works fairly well. The values agree with each other at least to two decimal places and for many cases they agree up to three decimal places.

Based on the comparisons made so far it can be concluded that approximations based on the lambda distribution with proper values of scale and shape parameters work very well and we may not need tables for selection procedures for different distributions. More generally, for any (parametric) statistical inference problem, one may use the lambda distribution model to get approximate good results. This advantage may be useful for developing package programs for selection and ranking problems, since tables of constants for selection procedures can be prepared for ranges of values of (β,γ) of practical interest.

ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research Contracts N00014-84-C-0167 and N00014-88-K-0170 at Purdue University, West Lafayette, Indiana. Reproduction in whole or in part is permitted for any purpose of the United States Government. The authors are thankful to Professors S. Panchapakesan and TaChen Liang for helpful comments.

REFERENCES

Edwards, D. G. and Hsu, J. C. (1983), "Multiple comparisons with the best treatment," Journal of the American Statistical Association, 78, 965-971.

Goel, P. K. (1974), "On the distribution of standardized mean of samples from logistic population," Technical Report No. 380, Dept. of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. (1956), "On a decision rule for a problem in ranking means," Ph.D. Thesis (Mimeo. Ser. No. 150), Institute of Statistics, Univ. of North Carolina, Chapel Hill.

Gupta, S. S. (1963), "Probability integrals of the multivariate normal and multivariate t," Annals of Mathematical Statistics, 34, 792-828.

Gupta, S. S. (1965), "On some multiple decision (selection and ranking) rules," Technometrics, 7, 225-245.

Gupta, S. S. and Leong, Y.-K. (1979), "Some results on subset selection procedures for double exponential populations," Decision Information (C. P. Tsokos and R. M. Thrall, eds.), Academic Press, New York, 277-305.

Gupta, S. S., Nagel, K., and Panchapakesan, S. (1973), "On the order statistics from equally correlated normal random variables," Biometrika, 60, 403-413.

Gupta, S. S. and Panchapakesan, S. (1979), "Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations," John Wiley, New York.

Gupta, S. S. and Singh, A. K. (1980), "On rules based on sample medians for selection of the largest location parameter," Communications in Statistics - Theory and Methods, A9, 1277-1298.

Hogg, R. V., Fisher, D. M. and Randles, R. H. (1975), "A two-sample adaptive distribution-free test," Journal of the American Statistical Association, 70, 656-661.

Joiner, B. L. and Rosenblatt, J. R. (1971), "Some properties of the range in samples from Tukey's symmetric lambda distribution," Journal of the American Statistical Association, 66, 394-399.

Lorenzen, T. J. and McDonald, G. C. (1981), "Selecting logistic populations using the sample medians," Communications in Statistics - Theory and Methods, A10(2), 101-124.

Moberg, T. F., Ramberg, J. S. and Randles, R. H. (1978), "An adaptive M-estimator and its application to a selection problem," Technometrics, 20, 255-263.

Mykytka, E. F. and Ramberg, J. S. (1979), "Fitting a distribution to data using an alternative to moments," Proceedings of the IEEE 1979 Winter Simulation Conference, 361-374.

Ramberg, J. S. and Schmeiser, B. W. (1972), "An approximate method for generating symmetric random variables," Communications in ACM, 15, 987-990.

Ramberg, J. S. and Schmeiser, B. W. (1974), "An approximate method for generating asymmetric random variables," Communications in ACM, 17, 78-82.

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.

Sohn, J. (1985), "Multiple decision procedures for Tukey's generalized lambda distributions," Technical Report No. 85-20, Dept. of Statistics, Purdue University, West Lafayette, Indiana.

Tukey, J. W. (1960), "The practical relationship between the common transformations of percentages of counts and amounts," Technical Report 36, Statistical Techniques Research Group, Princeton University.

TABLE 1: Values of d_0 for the Procedure R_T

$(\beta, \gamma) = (-0.0466, -0.0246)$, kurtosis = 4.6

m	P* k	2	3	5	7	9	10	11
1	0.90	1.0970	1.3599	1.6026	1.7380	1.8317	1.8696	1.9033
	0.95	1.4282	1.6788	1.9139	2.0462	2.1382	2.1755	2.2088
2	0.90	0.8606	1.0640	1.2492	1.3511	1.4210	1.4491	1.4740
	0.95	1.1148	1.3064	1.4836	1.5821	1.6500	1.6774	1.7017
3	0.90	0.7305	0.9021	1.0571	1.1417	1.1996	1.2227	1.2433
	0.95	0.9440	1.1046	1.2520	1.3334	1.3893	1.4117	1.4316
4	0.90	0.6455	0.7966	0.9325	1.0064	1.0567	1.0768	1.0946
	0.95	0.8330	0.9739	1.1027	1.1734	1.2219	1.2413	1.2585
5	0.90	0.5846	0.7210	0.8434	0.9098	0.9549	0.9729	0.9883
	0.95	0.7537	0.8806	0.9963	1.0597	1.1030	1.1204	1.1357

$(\beta, \gamma) = (-0.0870, -0.0443)$, kurtosis = 5.0

m	P* k	2	3	5	7	9	10	11
1	0.90	1.0798	1.3399	1.5813	1.7166	1.8107	1.8488	1.8827
	0.95	1.4085	1.6575	1.8924	2.0252	2.1180	2.1557	2.1893
2	0.90	0.8451	1.0455	1.2285	1.3295	1.3990	1.4270	1.4518
	0.95	1.0960	1.2853	1.4609	1.5589	1.6266	1.6539	1.6782
3	0.90	0.7165	0.8852	1.0380	1.1216	1.1788	1.2018	1.2221
	0.95	0.9267	1.0849	1.2305	1.3111	1.3665	1.3887	1.4085
4	0.90	0.6328	0.7811	0.9148	0.9876	1.2373	1.0572	1.0748
		0.8171	0.9557	1.0825	1.1524	1.2003	1.2195	1.2365
5	0.95	0.5728	0.7067	0.8270	0.8923	0.9367	0.9545	0.9702
		0.7389	0.8636	0.9774	1.0400	1.0826	1.0998	1.1150

Table 1 (continued)

$(\beta, \gamma) = (-0.1389, -0.0667)$, kurtosis = 5.6

m	P* k	2	3	5	7	9	10	11
1	0.90	1.0589	1.3156	1.5553	1.6905	1.7849	1.8233	1.8575
	0.95	1.3845	1.6315	1.8661	2.0000	2.0934	2.1315	2.1656
2	0.90	0.8264	1.0231	1.2035	1.2828	1.3506	1.4001	1.4023
	0.95	1.0732	1.2597	1.4334	1.5064	1.5727	1.5996	1.6234
3	0.90	0.6997	0.8649	1.0149	1.0973	1.1537	1.1764	1.1965
	0.95	0.9059	1.0611	1.2045	1.2840	1.3388	1.3609	1.3805
4	0.90	0.6175	0.7625	0.8135	0.9500	0.9980	1.0335	1.0344
	0.95	0.7979	0.9336	1.0582	1.1093	1.1558	1.1745	1.1910
5	0.90	0.5586	0.6894	0.8071	0.8712	0.9148	0.9323	0.9477
	0.95	0.7210	0.8430	0.9546	1.0160	1.0580	1.0749	1.0900

$(\beta, \gamma) = (-0.2306, -0.1045)$, kurtosis = 7.0

m	P* k	2	3	5	7	9	10	11
1	0.90	1.0231	1.2736	1.5101	1.6448	1.7395	1.7782	1.8127
	0.95	1.3427	1.5861	1.8196	1.9540	2.0489	2.0877	2.1225
2	0.90	0.7947	0.9851	1.1608	1.2587	1.3266	1.3541	1.3785
	0.95	1.0345	1.2159	1.3862	1.4820	1.5488	1.5759	1.6000
3	0.90	0.6714	0.8306	0.9759	1.0560	1.1111	1.1334	1.1531
	0.95	0.8706	1.0209	1.1604	1.2380	1.2917	1.3134	1.3327
4	0.90	0.5917	0.7312	0.8576	0.9270	0.9744	0.9935	1.0104
	0.95	0.7656	0.8965	1.0172	1.0840	1.1300	1.1486	1.1650
5	0.90	0.5349	0.6605	0.7739	0.8357	0.8780	0.8949	0.9099
	0.95	0.6911	0.8086	0.9164	0.9758	1.0166	1.0330	1.0475

Table 3. Performance of the Rule R_T under the slippage configuration, $\underline{\theta} = (\theta,\theta,\ldots,\theta+\delta)$, where $\delta > 0$.

$(\beta, \gamma) = (-0.0466, -0.0246)$, kurtosis = 4.6

P*			0.90			0.95		
δ	k	m	P(CS)	E(S')	P(CS)/E(S')	P(CS)	E(S')	P(CS)/E(S')
0.1	5	3	.9291	3.5597	.2610	.9661	3.7766	.2558
		5	.9357	3.5462	.2639	.9698	3.7684	.2573
	7	3	.9295	5.3580	.1735	.9663	5.6758	.1703
		5	.9363	5.3438	.1752	.9701	5.6670	.1712
0.3	5	3	.9668	3.4251	.2823	.9854	3.6937	.2668
		5	.9765	3.3466	.2918	.9901	3.6408	.2720
	7	3	.9674	5.2103	.1857	.9856	5.5859	.1765
		5	.9772	5.1191	.1909	.9904	5.5253	.1793
0.5	5	3	.9857	3.1912	.3089	.9941	3.5364	.2811
		5	.9926	2.9772	.3334	.9972	3.3759	.2954
	7	3	.9862	4.9349	.1998	.9943	5.4048	.1840
		5	.9929	4.6655	.2128	.9923	5.2078	.1915

$(\beta, \gamma) = (-0.2306, -0.1045)$, kurtosis = 7.0

P*			0.90			0.95		
δ	k	m	P(CS)	E(S')	P(CS)/E(S')	P(CS)	E(S')	P(CS)/E(S')
0.1	5	3	.9306	3.5566	.2617	.9668	3.7752	.2561
		5	.9380	3.5411	.2649	.9708	3.7656	.2578
	7	3	.9310	5.3551	.1738	.9669	5.6744	.1704
		5	.9385	5.3382	.1758	.9711	5.6642	.1714
0.3	5	3	.9688	3.4053	.2845	.9861	3.6830	.2677
		5	.9790	3.3107	.2957	.9911	3.6184	.2739
	7	3	.9693	5.1892	.1868	.9863	5.5748	.1769
		5	.9795	5.0777	.1929	.9913	5.5004	.1802
0.5	5	3	.9871	3.1332	.3150	.9946	3.5009	.2841
		5	.9938	2.8719	.3460	.9976	3.3005	.3023
	7	3	.9874	4.8677	.2028	.9947	5.3651	.1854
		5	.9941	4.5331	.2193	.9977	5.1165	.1950

Table 4. Performance of the Rule R_T under the equally-spaced configuration, $\underline{\theta} = (\theta, \theta+\delta, \ldots, \theta+(k-1)\delta)$, where $\delta > 0$

$(\beta, \gamma) = (-0.0466, -0.0246)$, kurtosis = 4.6

		P*	0.90			0.95		
δ	k	m	P(CS)	E(S')	P(CS)/E(S')	P(CS)	E(S')	P(CS)/E(S')
0.1	5	3	.9550	3.4127	.2798	.9794	3.6812	.2661
		5	.9633	3.3275	.2895	.9837	3.6205	.2717
	7	3	.9653	4.9594	.1946	.9844	5.4077	.1821
		5	.9723	4.7363	.2054	.9882	5.2354	.1887
0.3	5	3	.9875	2.4713	.3996	.9948	2.9066	.3423
		5	.9921	2.0643	.4806	.9969	2.4834	.4014
	7	3	.9914	2.8956	.3424	.9964	3.4703	.2871
		5	.9947	2.3105	.4305	.9979	2.7799	.3590
0.5	5	3	.9956	1.5375	.6476	.9983	1.8930	.5273
		5	.9979	1.1687	.8538	.9992	1.4495	.6894
	7	3	.9970	1.6943	.5884	.9988	2.0555	.4859
		5	.9986	1.2889	.7748	.9995	1.5700	.6366

$(\beta, \gamma) = (-0.2306, -0.1045)$, kurtosis = 7.0

		P*	0.90			0.95		
δ	k	m	P(CS)	E(S')	P(CS)/E(S')	P(CS)	E(S')	P(CS)/E(S')
0.1	5	3	.9568	3.3907	.2822	.9801	3.6687	.2672
		5	.9656	3.2891	.2936	.9846	3.5952	.2739
	7	3	.9668	4.9027	.1972	.9849	5.3726	.1833
		5	.9747	4.6358	.2103	.9889	5.1623	.1916
0.3	5	3	.9884	2.2362	.4231	.9950	2.7849	.3573
		5	.9931	1.9029	.5219	.9972	2.3131	.4311
	7	3	.9919	2.6909	.3686	.9966	3.2497	.3067
		5	.9953	2.1131	.4710	.9981	2.5533	.3909
0.5	5	3	.9960	1.4010	.7109	.9984	1.7399	.5738
		5	.9982	1.0443	.9559	.9993	1.3055	.7655
	7	3	.9973	1.5479	.6443	.9989	1.8907	.5283
		5	.9988	1.1558	.8642	.9996	1.4181	.7048

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Technical Report #85-29	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) SELECTION AND RANKING PROCEDURES FOR TUKEY'S GENERALIZED LAMBDA DISTRIBUTIONS: SYMMETRIC CASE	5. TYPE OF REPORT & PERIOD COVERED Technical
	6. PERFORMING ORG. REPORT NUMBER Technical Report #85-29
7. AUTHOR(s) Shanti S. Gupta and Joong K. Sohn	8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0167
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK, AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC	12. REPORT DATE November 1985
	13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION, DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Subset Selection; Tukey's Generalized Lambda Distributions; Sample Medians; Asymptotic Relative Efficiencies.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A selection and ranking procedure based on sample medians for Tukey's generalized symmetric lambda populations is considered. Also the properties of the proposed procedure such as asymptotic relative efficiencies (ARE) are studied. Tables of constants necessary to carry out the procedures along with ARE's of the proposed procedure are computed and tabulated. Also performance of the procedure is studied under both a slippage-configuration and an equi-spaced configurations. An application of the lambda distribution for approximating some constants used in selection and ranking problems for some well-known symmetric distributions is studied.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)