SELECTING AMONG WEIBULL, LOGNORMAL AND GAMMA DISTRIBUTIONS USING COMPLETE AND CENSORED SAMPLES

Siswadi and C. P. Quesenberry

North Carolina State University
Raleigh, NC 27650-5457

Institute of Statistics Mimeograph Series #1609

ABSTRACT

In a recent paper, Kent and Quesenberry [19] considered using certain optimal invariant statistics to select the best fitting member of a collection of probability distributions using complete samples of life data. In the present work, extension of this approach in two directions are given. First, selection for complete samples based on scale and shape invariant statistics is considered. Next, the selection problem for type I censored samples is considered, and both scale invariant and maximum likelihood selection procedures are studied. The two-parameter (scale and shape) Weibull, lognormal, and gamma distributions are considered and applications to real data are given. Results from a (small) comparative simulation study are presented.

1. INTRODUCTION

The two-parameter (scale and shape) Weibull, lognormal and gamma distributions are all commonly used in reliability and life testing problems. The problem of selecting one of these three distributions for a particular sample, either complete or censored, is a difficult one. In this work we consider basing the selection on the values of certain selection statistics computed from the sample. Although we consider selection based only on sample information, it should be noted that in some practical problems further information may be available which should also be weighed in the final selection of a distribution. Such information could be derived, for example, from known physical characteristics of a failure mechanism and its failure rate.

Also, it should be mentioned that throughout this paper we treat the selection of one of the distributions as a forced selection problem. That is, we formally select exactly one of the competing distributions. In practice, we may sometimes wish to use a partial selection procedure which does not necessarily always select one distribution over the others when they are close together, as indicated by selection statistics that are nearly equal. The selection statistics posed here may be readily used in partial selection schemes, however, we do not explicitly consider these applications in the present paper.

Kent and Quesenberry [19], KQ, proposed a forced selection procedure based upon statistics that are invariant under scale transformations. Other relevant literature includes a paper by Dumonceaux, Antle and Haas [11] who examined maximum likelihood ratio (MLR) tests for discriminating between two models with unknown location and scale parameters, and compared empirically the power of MLR tests with that of uniformly most powerful invariant (UMPI) tests for discriminating between normal and Cauchy distributions. They actually recommend the MLR test over the UMPI test on the basis of relatively good power and ease of computation. Dumonceaux and Antle [10] gave an MLR procedure for discriminating between Weibull and lognormal distributions that is based on the fact that the logarithms of both Weibull and lognormal random variables have location-scale parameter distributions. In a recent paper, Bain and Engelhardt [2] considered a likelihood ratio selection statistic for selecting between gamma and Weibull distributions.

Some graphical procedures for the selection problem have been given by Nelson [21], and by Barlow and Campo [3]. Other papers that are related to the present work include Hogg, Uthoff, Randles and Davenport [18], who discuss a number of selection procedures, including one based on location-scale invariant statistics; Dyer ([12], [13]) who considers a number of selection procedures for discriminating between pairs of classes of location-scale distributions; and Uthoff ([24], [25]) who considers some particular invariant statistics. As general references for invariant tests see Hajek and Sidak [15] and Lehmann [20], and for MLR tests see Cox [8]. Volodin [26] considers a generalized three-parameter gamma distribution and discriminates between two-parameter gamma and Weibull distributions by making scale invariant tests on the other parameters.

As mentioned above, KQ considered selecting among the gamma, lognormal and Weibull families for the complete sample problem. The selection statistic posed was formed by first deriving a scale invariant statistic that is optimal in the sense that it minimizes the sum of the two probabilities of selecting the incorrect distribution for two conformable (cf. Quesenberry and Starbuck[22]) distributions: and then replacing the shape parameter by its ML estimator. Such procedures were called suboptimal, and the selection statistics for the three families were set out in simple closed form formulas in that paper.

In the present work we consider two major changes in the approach and problem considered in KQ. First, for the lognormal and Weibull families we use optimal scale and shape invariant selection statistics. Also, we consider the selection problem for type I censored samples as well as for complete samples. For these cases the selection statistics are generally expressed as definite integrals whose evaluation requires numerical integration. Thus, a substantial part of this work has necessarily been concerned with the development of computer algorithms to evaluate these integrals.

2. DENSITIES AND SELECTION FUNCTIONS

In many applied problems it is reasonable to assume that the location parameters of life distributions are known. Thus, we consider distributions with only scale and shape parameters unknown. The densities of the gamma, lognormal and Weibull distributions to be studied are given in Table 1. The symbol $I_{(a,b)}(x)$ in table 1, and elsewhere, is the indicator function of the interval (a,b), *i.e.*, $I_{(a,b)}(x) = 1$ if $a < x < b$, and is zero otherwise.

We consider type I censored samples, which are obtained when a number of items are put on life test and observed for a previously specified fixed time T. Thus the parent density for the observed lives is a truncated version of the complete samples density given in Table 1.

Table 1. Densities of Weibull, Lognormal and Gamma Distributions

Name	Symbol	Density	
Weibull	$W(\theta,\beta)$	$f_1 = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} \exp[-(x/\theta)^{\beta}] \cdot I_{(0,\infty)}(x)$;	$\theta,\beta > 0$
Lognormal	$LN(\theta,\sigma)$	$f_2 = \frac{1}{\sqrt{2\pi}\,\sigma x} \cdot \exp\{-[\ell n(x/\theta)]^2/2\sigma^2\} \cdot I_{(0,\infty)}(x)$;	$\theta,\sigma > 0$
Gamma	$G(\theta,\alpha)$	$f_3 = \theta^{-\alpha}[\Gamma(\alpha)]^{-1} x^{\alpha-1} \exp(-x/\theta) \cdot I_{(0,\infty)}(x)$;	$\theta,\alpha > 0$

The approach used here is of the same general form as that in KQ. A selection statistic, S, is defined for each of the three parametric classes, and the class with the largest selection statistic is chosen as the best fitting family for a given sample.

We consider some transformation properties of these distributions before defining selection statistics for them. If X is a random variable with either a $G(\theta,\alpha)$, $LN(\theta,\sigma)$ or $W(\theta,\beta)$ distribution, then consider the transformation

$$Y = aX^b, \; a > 0, \; b > 0 . \qquad (2.1)$$

If X is a $W(\theta,\beta)$ random variable, then Y is a $W(a\theta^b, \beta/b)$ random variable; and if X is a $LN(\theta,\sigma)$ random variable, then Y is a $LN(a\theta^b, b\sigma)$ random variable. Thus Weibull random variables are transformed to Weibull random variables by (2.1), and lognormal random variables are transformed to lognormal random variables by this transformation. Unfortunately, for present purposes at least, the gamma distributions do not share this property since a gamma random variable is not always transformed to another gamma random variable by (2.1). That is to say, the $G(\theta,\alpha)$ class is not a scale-shape class that is conformable with the lognormal and Weibull classes as defined by Quesenberry and Starbuck [22]. Nevertheless, we use here a selection statistic for the complete samples problem which is essentially the value of the density function of a maximal invariant when each of the three parents is assumed.

For $x_1,\ldots,x_n$ an observed sample, we define the selection statistic for a density function f_i, $(i=1,2,3)$ by

$$S_i = \int_0^\infty \int_0^\infty f_i(\gamma x_1^\lambda, \ldots, \gamma x_n^\lambda)\gamma^{n-1}\lambda^{n-2}(x_1\ldots x_n)^\lambda d\gamma d\lambda \ . \qquad (2.2)$$

Due to the property of the $G(\theta,\alpha)$ distribution discussed above, the selection statistic S_3 of (2.2) is a function of the parameter α. We obtain a selection statistic by replacing α by its maximum likelihood estimator, $\hat{\alpha}$, in this function. The selection statistics used in this work are given in Table 2.

Table 2. Scale-shape Invariant Selection Statistics for Complete Samples.

Family	i	S_i
$W(\theta,\beta)$	1	$\Gamma(n) \int_0^\infty (\Pi x_i)^{\lambda}(\Sigma x_i^{\lambda})^{-n} \lambda^{n-2}\, d\lambda$
$LN(\theta,\sigma)$	2	$\frac{1}{2} n^{-\frac{1}{2}} \pi^{-\frac{1}{2}(n-1)} \Gamma[\frac{1}{2}(n-1)][\Sigma \ell n^2 x_i - (\Sigma \ell n x_i)^2/n]^{-\frac{1}{2}(n-1)}$
$G(\theta,\hat{\alpha})$	3	$\Gamma(n\hat{\alpha})\Gamma^{-n}(\hat{\alpha}) \int_0^\infty (\Pi x_i)^{\hat{\alpha}\lambda}(\Sigma x_i^{\lambda})^{-n\hat{\alpha}} \lambda^{n-2}\, d\lambda$

The evaluation of these functions requires numerical integration, except for the lognormal selection function. It is often computationally easier to compute and compare the logarithms of the selection statistics than the statistics themselves. To estimate the parameter α of the gamma distribution, we use the ML estimator $\hat{\alpha}$ of Greenwood and Durand [14] which has been studied further by Bowman and Shenton [6], and was recently used and given in detail by KQ.

The selection procedures proposed here are closely related to uniformly most powerful invariant (UMPI) tests for separate families testing problems. For the particular case of selecting between log-normal and Weibull distributions, using the selection statistics of Table 2 is equivalent to using the UMPI test statistic for classifying a sample into one of these two distributions. If α_1 denotes the probability that a sample from a lognormal parent will be classified as a Weibull sample, and α_2 the probability that a Weibull sample will be classified as a lognormal sample, then the above selection procedure will minimize $\alpha_1 + \alpha_2$ among all procedures invariant with respect to the transformations of (2.1). Or, if the probability of selecting each distribution is ½, then this procedure minimizes the total probability of misclassification, viz., $(\alpha_1 + \alpha_2)/2$ (see KQ, section 3).

3. SIMULATION RESULTS FOR COMPLETE SAMPLES

In this section we report results of a simulation study of the selection rules proposed above.

In order to obtain results that can be compared with those of KQ, samples were generated from nine parent distributions: W(1,½), W(1,2), W(1,4), LN(1,0.4), LN(1,1), LN(1,2.5), G(1,½), G(1,2) and G(1,5) for $n = 10, 20, 30$. One thousand samples were generated from each of these distributions. The pairwise selection error rates are given in Table 3 and the observed rates of correct classification in the 3-way procedure are given in Table 5.

The entries in Table 3 are read as follows, using the first set of results as an example. The selection is to be made between W and LN families. One thousand samples of size 10 were generated from a W distribution, and 28 percent of these samples were classified as being from a LN distribution. One thousand samples of size 10 were generated from a LN distribution and 36 percent of these samples were classified as being from a W distribution. Note that since W and LN are conformable scale-shape families, these results do not depend upon which particular members of the families that are involved.

This procedure has total error probabilities for the case of a lognormal vs a Weibull that are the smallest possible for a scale-shape invariant procedure. Comparison of the results for this case given in Table 3 and those in Table 4 of KQ shows that the present optimal procedure has very little advantage over the suboptimal procedure of KQ For the other two-way selection problems, gamma vs lognormal and gamma vs Weibull, the comparisons of the procedures of this paper with those of KQ are not clearcut. This is because the gamma distribution does not admit a UMPI statistic with respect to the transformations of (2.1).

In view of these observations we recommend the selection procedures set out in KQ on the grounds that (i) the selection statistics in

Table 2 of KQ have convenient formulas that are readily evaluated, and (ii) the error rates achieved by those procedures appear to be about as favorable as for those achieved by the much more computationally difficult scale-shape invariant procedures.

Also, Bain and Engelhardt [2] give in their Table 2 some probabilities of correct selection between gamma and Weibull

Table 3. Selection Error Rates for Pairwise Procedures*

n	X ~ W	X ~ LN	Total
10	.36	.28	.32
20	.21	.23	.22
30	.15	.18	.16

n	X ~ G(½)	X ~ W(½)	Total	X ~ G(½)	X ~ W(2)	Total
10	.40	.42	.41	.40	.44	.42
20	.37	.38	.38	.37	.36	.37
30	.34	.35	.34	.34	.33	.34
	X ~ G(½)	X ~ W(4)		X ~ G(2)	X ~ W(½)	
10	.40	.40	.40	.44	.42	.43
20	.37	.29	.33	.42	.38	.40
30	.34	.23	.29	.40	.35	.37
	X ~ G(2)	X ~ W(2)		X ~ G(2)	X ~ W(4)	
10	.44	.44	.44	.44	.40	.42
20	.42	.36	.39	.42	.29	.36
30	.40	.33	.36	.40	.23	.31
	X ~ G(5)	X ~ W(½)		X ~ G(5)	X ~ W(2)	
10	.39	.42	.41	.39	.44	.41
20	.31	.38	.35	.31	.36	.34
30	.30	.35	.32	.30	.33	.31
	X ~ G(5)	X ~ W(4)				
10	.39	.40	.39			
20	.31	.29	.30			
30	.30	.23	.26			

Table 3 (continued).

n	X ∼ G(½)	X ∼ LN(0.4)		X ∼ G(½)	X ∼ LN(1)	
10	.25	.41	.33	.25	.33	.29
20	.15	.38	.27	.15	.27	.21
30	.09	.35	.22	.09	.21	.15
	X ∼ G(½)	X ∼ LN(2.5)		X ∼ G(2)	X ∼ LN(0.4)	
10	.25	.23	.24	.37	.41	.39
20	.15	.14	.15	.31	.38	.34
30	.09	.08	.08	.24	.35	.29
	X ∼ G(2)	X ∼ LN(1)		X ∼ G(2)	X ∼ LN(2.5)	
10	.37	.33	.35	.37	.23	.30
20	.31	.27	.29	.31	.14	.22
30	.24	.21	.22	.24	.08	.16
	X ∼ G(5)	X ∼ LN(0.4)		X ∼ G(5)	X ∼ LN(1)	
10	.44	.41	.42	.44	.33	.38
20	.36	.38	.37	.36	.27	.32
30	.33	.35	.34	.33	.21	.27
	X ∼ G(5)	X ∼ LN(2.5)				
10	.44	.23	.33			
20	.36	.14	.25			
30	.33	.08	.20			

* See section 3 for explanation of table entries and discussion of these results.

distributions using a likelihood ratio test statistic. Their results can be used to construct total error rates comparable to those of Table 3, for a few selected values of the gamma and Weibull shape parameters. We have computed these values and give them in Table 4.

Table 4. Total Error Rates for Gamma vs Weibull for Likelihood Ratio Procedure of Bain and Engelhardt.

α	n	β : .5	2	4
.5	10	.435	.405	.385
	20	.375	.345	.320
2	10	.470	.440	.420
	20	.415	.385	.360

Comparison of the total error rates of Tables 3 and 4 shows no trend in favor of either procedure.

Table 5 gives the selection rates in our simulation study for the three-way scale-shape invariant selection procedure.

The entries in Table 5 are read as follows, using the first set of results as an example. The selection is to be made among the gamma, Weibull and lognormal distributions. One thousand samples of size 10 were generated from a $G(1,\frac{1}{2})$ distribution, of which 57 percent are classified as G, 21 percent are classified as W and 23 percent are classified LN. The results in Table 5 can be compared with the results in Table 5 of KQ. The comparisons do not show that either of these procedures has a clear advantage, however, the selection procedure of KQ may have a slight edge. Thus, as for the two-way selection procedures above, we favor the computationally simpler scale invariant, SI_3, procedure of KQ.

Table 5. Selection Rates for Three-way Procedure

n	$X \sim G(\frac{1}{2})$			$X \sim G(2)$			$X \sim G(5)$		
	G	W	LN	G	W	LN	G	W	LN
10	.57	.21	.23	.20	.44	.37	.18	.39	.44
20	.62	.26	.12	.28	.43	.31	.32	.31	.36
30	.66	.29	.05	.37	.40	.24	.38	.30	.33
	$X \sim W(\frac{1}{2})$			$X \sim W(2)$			$X \sim W(4)$		
	G	W	LN	G	W	LN	G	W	LN
10	.42	.23	.35	.15	.57	.29	.16	.60	.24
20	.38	.41	.21	.24	.64	.12	.19	.71	.10
30	.35	.51	.15	.26	.67	.07	.18	.77	.05
	$X \sim LN(0.4)$			$X \sim LN(1)$			$X \sim LN(2.5)$		
	G	W	LN	G	W	LN	G	W	LN
10	.17	.24	.59	.14	.19	.67	.14	.15	.72
20	.23	.15	.62	.18	.09	.73	.04	.19	.77
30	.26	.10	.65	.16	.05	.79	.02	.16	.82

4. SELECTION WITH CENSORING

Suppose that from a random sample of size n on a parent random variable with density and distribution functions f and F, respectively, only the values less than a prespecified time T are observed. If r is the number of values less than T, then r is a binomial rv with probability function $b(r;F(T),n)$. Let $x_1,\ldots,x_r$ be the observed values, indexed in the same order as the original sample, and $x_{(1)},\ldots,x_{(r)}$ be the corresponding order statistics. We require selection procedures based on the values $x_1,\ldots,x_r$, and r. We have studied procedures based on scale-shape invariance, as considered above for complete samples, scale invariance as in KQ, and maximum likelihood ratio procedures. Of these procedures, only the SI and maximum likelihood, ML, procedures will be described now, since these procedures will be recommended for reasons given below.

When f and F are functions of parameters θ, say f_θ and F_θ, the likelihood function corresponding to $x_1,\ldots,x_r$ and r is

$$L_\theta(x_{(1)},\ldots,x_{(r)},r;T) =$$
$$\{n!/(n-r)!\}\{1-F_\theta(T)\}^{n-r}\prod_{i=1}^{r} f_\theta(x_{(i)})I_{(0,T)}(x_{(i)}) \qquad (4.1)$$

The scale invariant selection statistic is defined by

$$S = \int_0^\infty L(\lambda x_{(1)},\ldots,\lambda x_{(r)},\ r;\lambda T)\lambda^{r-1}d\lambda, \qquad (4.2)$$

where the scale parameter in L has been set equal to one, and S depends upon a shape parameter. The shape parameter in S for each of the three families considered here will be replaced by its maximum likelihood estimator, obtained by maximizing the likelihood in (4.1).

We also consider selection for the censored case using, essentially, a likelihood ratio procedure. In this approach we use the maximum value of the likelihood function in (4.1) as the selection statistic. Formally, the selection statistic is

$$\begin{aligned} S &= \sup_\theta\{L_\theta(x_{(1)},\ldots,x_{(r)},\ r;T)\}, \\ &= L_{\hat\theta}(x_{(1)},\ldots,x_{(r)};\ r,T)\ , \end{aligned} \qquad (4.3)$$

for $\hat\theta$ the ML estimator(s) of θ.

The selection functions for these two methods for the three families are given in Table 6.

Table 6. Selection Statistics for Censored Samples.

Family	i	S_i
Scale Invariant, SI		
$W(\theta,\beta)$	1	$\frac{n!}{(n-r)!}\Gamma(r)\hat{\beta}^{r-1}\left(\prod_{i=1}^{r} x_{(i)}\right)^{\hat{\beta}-1}\left(\sum_{i=1}^{r} x_{(i)}^{\hat{\beta}} + (n-r)T^{\hat{\beta}}\right)^{-r}$
$LN(\theta,\sigma)$	2	$\frac{n!}{(n-r)!}(\sqrt{2\pi}\,\hat{\sigma})^{1-r}r^{-\frac{1}{2}}\left(\prod_{i=1}^{r} x_{(i)}\right)^{-1}\exp\left\{-\frac{1}{2\hat{\sigma}^2}\left(\sum_{i=1}^{r}(\ell n\, x_{(i)}^2) - \frac{1}{r}\left(\sum_{i=1}^{r}\ell n\, x_{(i)}\right)^2\right)\right\}\cdot E\left(1-\Phi\left(\frac{u+\ell nT}{\hat{\sigma}}\right)\right)^{n-r}$ where u is a $N\left(-\sum_{i=1}^{r}\ell n\, x_{(i)}/r\ ,\ \hat{\sigma}^2/r\right)$ r.v.
$G(\theta,\alpha)$	3	$\frac{n!}{(n-r)!}\Gamma(r\hat{\alpha})\Gamma^{-r}(\hat{\alpha})\left(\prod_{i=1}^{r} x(i)\right)^{\hat{\alpha}-1}\left(\sum_{i=1}^{r} x(i)\right)^{-r\hat{\alpha}}$ $E\left(1-\Gamma\left(Tu/\sum_{i=1}^{r} x_{(i)},\hat{\alpha}\right)/\Gamma(\hat{\alpha})\right)^{n-r}$ where u is a $G(1,r\hat{\alpha})$ r.v.
Maximum Likelihood, ML		
$W(\theta,\beta)$	1	$\frac{n!}{(n-r)!}\hat{\beta}^{r}\hat{\theta}^{-r\hat{\beta}}\left(\prod_{i=1}^{r} x_{(i)}\right)^{\hat{\beta}-1}\exp\left[-\hat{\theta}^{-\hat{\beta}}\left(\sum_{i=1}^{r} x_{(i)}^{\hat{\beta}} + (n-r)T^{\hat{\beta}}\right)\right]$
$LN(\theta,\sigma)$	2	$\frac{n!}{(n-r)!}\left[1-\Phi\left(\frac{\ell n\, T-\ell n\hat{\theta}}{\hat{\sigma}}\right)\right]^{n-r}\left(\prod_{i=1}^{r} x_{(i)}\right)^{\alpha-1}(\sigma\sqrt{2\pi})^{-r}$ $\exp\left\{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{r}(\ell n\, x_{(i)} - \ell n\hat{\theta})^2\right\}$.

Table 6 (continued.

Family	i	S_i
$G(\theta,\alpha)$	3	$\frac{n!}{(n-r)!}\Gamma^{-n}(\hat{\alpha})\left[\Gamma(\hat{\alpha}) - \Gamma\left(\frac{T}{\hat{\theta}},\hat{\alpha}\right)\right]^{n-r}\hat{\theta}^{-r\hat{\alpha}} \left(\prod_{i=1}^{r} x_{(i)}\right)^{\hat{\alpha}-1} \exp -\left(\frac{1}{\hat{\theta}}\sum_{i=1}^{r} x_{(i)}\right)$

Note: $\Gamma(a,b) = \int_0^a s^{b-1}\exp(-s)ds$

We have written programs to evaluate the selection statistics of Table 6. A brief description of this work follows in the remainder of this section. For more detail see Siswadi [23].

Maximum likelihood estimates for the scale and shape parameters are required for the ML selection functions, and for the shape parameter for the SI selection functions. For the Weibull class, these estimates were obtained as solutions of the ML equations in Cohen [7]. For the lognormal class solutions for the ML equations were obtained using results of Harter and Moore [17], adjusted for type I censoring. Another procedure for lognormal type I censored samples is given by Aitchison and Brown [1]. Solutions of the ML equations for the gamma class were obtained from the results of Harter and Moore [16], adjusted for type I censoring. After the ML estimates of the scale and shape parameters are obtained, the evaluations of the ML selection functions of Table 6 are straightforward.

After the ML estimate $\hat{\beta}$ of the shape parameter of the Weibull distribution is obtained, the SI selection function is readily evaluated. However, the selection functions for both the lognormal and gamma scale invariant procedures are difficult to evaluate, and we have used Monte Carlo and importance sampling from the normal distribution and gamma distribution, respectively, (see Davis and Rabinowitz, [9]) to evaluate them.

5. SIMULATION RESULTS FOR CENSORED SAMPLES

We have conducted a small Monte Carlo simulation study of the two selection methods discussed above for censored samples to provide some information on the error rates for these procedures. These empirical error rates allow comparisons of the two procedures with each other as well as with complete sample rates given in section 3 and in KQ. Comparison with complete sample rates gives a measure of the loss of information due to censoring.

The families of distributions considered were W(½), W(4), G(½), G(2), LN(0.4), and LN(2.5); and the sample size was $n = 30$ in all cases. The truncation point T was chosen so that the df $F(T) = 0.90$, i.e., for a mean rate of 10% censoring. One hundred samples were generated for each of the above distributions except W(4), for which 16 samples were generated. The running time for some cases was very long and this limited the number of samples that could be generated.

The misclassification rates for pairwise selection are given in Table 7, which is comparable to Table 3 for the complete samples case. Note in Table 7 that the ML and SI procedures give similar error rates for W vs LN and, in fact, both give the same total error rate of 0.29. Comparison of these results with those of the SI procedure for complete samples (see Table 4 of KQ) shows that there is a rather large loss of information due to censoring since the W, LN and total error rates are 0.19, 0.15, and 0.17, respectively.

For the two-way selection error rates in Table 7 that involve a gamma distribution, neither the ML nor the SI procedure appears to have an overall advantage. Also, by comparing these cases with the

Table 7. Misclassification Rates for Pairwise Selection Procedures - Censored Sample (n = 30).

Procedure	$X \sim W$	$X \sim LN$	Total
ML	.40	.18	.29
SI	.45	.12	.29

Procedure	$X \sim G(\frac{1}{2})$	$X \sim W(\frac{1}{2})$	Total	$X \sim G(\frac{1}{2})$	$X \sim W(4)$	Total
ML	.38	.32	.35	.38	.44	.41
SI	.52	.16	.34	.52	.50	.50
	$X \sim G(2)$	$X \sim W(\frac{1}{2})$		$X \sim G(2)$	$X \sim W(4)$	
ML	.36	.32	.34	.36	.44	.40
SI	.39	.16	.28	.39	.50	.45
	$X \sim G(\frac{1}{2})$	$X \sim LN(.4)$		$X \sim G(\frac{1}{2})$	$X \sim LN(2.5)$	
ML	.14	.32	.23	.14	.08	.11
SI	.20	.31	.26	.20	.09	.15
	$X \sim G(2)$	$X \sim LN(.4)$		$X \sim G(2)$	$X \sim LN(2.5)$	
ML	.35	.32	.34	.35	.08	.22
SI	.42	.31	.37	.42	.09	.21

Table 8. Classification Rates for Three-way Procedures - Censored Sample (n = 30)

Procedure	$X \sim G(\frac{1}{2})$			$X \sim G(2)$			$X \sim W(\frac{1}{2})$		
	G	W	LN	G	W	LN	G	W	LN
ML	.60	.29	.11	.29	.36	.35	.27	.35	.38
SI	.47	.35	.18	.28	.32	.40	.16	.40	.44
	$X \sim W(4)$			$X \sim LN(.4)$			$X \sim LN(2.5)$		
	G	W	LN	G	W	LN	G	W	LN
ML	.13	.56	.31	.23	.09	.68	.01	.17	.82
SI	.25	.50	.25	.23	.08	.69	.01	.13	.86

same cases in Table 4 of KQ, we feel that the loss of information due to censoring is not so large as for the W vs LN case commented on above.

The classification rates for three-way selection procedures are given in Table 8. Again, neither the ML nor the SI procedure appear to have any overall advantage, and both perform quite well. Also, by comparison with Table 5 of KQ it appears that ten percent censoring has little effect on the probability of correctly classifying a lognormal sample, but the probabilities of correctly classifying either Weibull or gamma samples are reduced somewhat.

6. A USER PROGRAM AND EXAMPLES FOR CENSORED SAMPLES

The selection procedures for the three families of distributions have been programmed in FORTRAN. A listing of this program can be obtained from the authors. The program computes the selection statistics for complete and censored samples according to the formulas given in Tables 2 and 6, respectively.

For the scale invariant procedure, the selection statistics are computed by the Monte Carlo method given in Davis and Rabinowitz [9]. The program was tested on several examples and on many samples produced through simulation. In general, the selection statistics estimated did not appear reliable for heavily censored samples. Therefore, in the user program for the scale invariant procedure, the selection results are not printed if the coefficient of variation of the replicated values in the Monte Carlo method is larger than 35%

<u>Example 1</u>.

Birnbaum and Saunders [5] considered a set of data of lifetimes, in thousands of cycles, of aluminum sheeting under periodic loading, to illustrate the gamma family. If we assume that the experiment was terminated at a prespecified time, say T = 1900, then the censored observations and the results of the selection procedure are presented in Table 9. For these data, the Weibull family is selected by both the ML and SI procedures, however, the selection statistics for the gamma family are only slightly smaller. It is also to be noted, although the details are not given here, for the complete sample the selection procedure based on the selection statistics given in Table 2 yields the same results.

Example 2.

Bartholomew [4], p. 370 gave the failure times of 15 items that failed during a specified period of testing from an original sample of size n = 20. He states that the items have an exponential life distribution, and uses the exponential distribution to perform analyses of the data. We have used these data in the selection program,and the results are given in Table 10. Both the maximum likelihood and scale invariant procedures prefer the lognormal distributions, which casts some doubt on the assumption of an exponential parent distribution.

Table 9. Results of Selection Procedure.

Lifetimes of Aluminum Sheeting under Periodic Loading					
370	706	716	746	785	797
844	855	858	886	886	930
960	988	990	1000	1010	1016
1018	1020	1055	1085	1102	1102
1108	1115	1120	1134	1140	1199
1200	1200	1203	1222	1235	1238
1252	1258	1262	1269	1270	1290
1293	1300	1310	1313	1315	1330
1355	1390	1416	1419	1420	1420
1450	1452	1475	1478	1481	1485
1502	1505	1513	1522	1522	1530
1540	1560	1567	1578	1594	1602
1604	1608	1630	1642	1674	1730
1750	1750	1763	1768	1781	1782
1792	1820	1868	1881	1890	1893
1895					

Sample Size = 101
Sample Observed = 91
Truncation point = 1900

Maximum Likelihood Estimates		
Family	Scale	Shape
Weibull	0.154149D+04	0.404114D+01
Gamma	0.125214D+03	0.112550D+02
Lognormal	0.135159D+04	0.317034D+00

Family	Selection Statistic
Maximum Likelihood Procedure	
Weibull	0.147164D+03
Gamma	0.146470D+03
Lognormal	0.144092D+03

The family selected is Weibull

Scale Invariant Procedure	
Weibull	0.144432D+03
Gamma	0.144103D+03
Lognormal	0.141762D+03

The family selected is Weibull

Table 10. Results of Selection Procedure

Bartholomew Data					
3	19	23	26	27	37
38	41	45	58	84	90
99	109	138			

Sample Size = 20
Sample Observed = 15
Truncation point = 150

Maximum Likelihood Estimates		
Family	Scale	Shape
Weibull	0.105498D 03	0.108289D 01
Gamma	0.876146D 02	0.116892D 01
Lognormal	0.682517D 02	0.122585D 01

Family	Selection Statistic
Maximum Likelihood Procedure	
Weibull	-0.669185D 01
Gamma	-0.664251D 01
Lognormal	-0.660273D 01
The family selected is lognormal	
Scale Invariant Procedure	
Weibull	-0.720101D 01
Gamma	-0.699896D 01
Lognormal	-0.667347D 01
The family selected is lognormal	

REFERENCES

[1] Aitchison, J. and J. A. C. Brown, The Lognormal Distribution, Cambridge University Press, London (1957).

[2] Bain, L. J. and M. Engelhardt, "Probability of Correct Selection of Weibull versus Gamma based on Likelihood Ratio," Communications in Statistics A9:375-381(1980).

[3] Barlow, R. E. and R. Campo, "Total Time on Test Processes and Applications to Failure Data Analysis," in Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assessment: 451-431, SIAM, Philadelphia (1975).

[4] Bartholomew, D. J., "The Sampling Distribution of an Estimate Arising in Life Testing," Technometrics 5:361-374 (1963).

[5] Birnbaum, Z. W. and S. C. Saunders, "A Statistical Model for Life-length of Materials," Journal of the American Statistical Association 53:151-160 (1958).

[6] Bowman, K. O. and L. R. Shenton, Properties of Estimator for the Gamma Distribution, Report CTC-1, Oak Ridge, Union Carbide Corporation, Nuclear Division, TN (1968).

[7] Cohen, A. C., "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples," Technometrics 7:579-588 (1965).

[8] Cox, D. R., "Tests of Separate Families of Hypotheses," in Proceedings of the Fourth Berkeley Sumposium 1: 105-123, University of California Press, Berkeley (1961).

[9] Davis, P. J. and P. Rabinowitz, Methods of Numerical Integration, Academic Press, New York (1975).

[10] Dumonceaux, R. and C. E. Antle, "Discrimination between the Log-normal and the Weibull Distributions," Technometrics 15: 923-926 (1973).

[11] Dumonceaux, R., C. E. Antle and G. Haas, "Likelihood Ratio Test for Discrimination between Two Models with Unknown Location and Scale Parameters," Technometrics 15:19-27 (1973).

[12] Dyer, A. R., "Discrimination Procedures for Separate Families of Hypotheses," Journal of the American Statistical Association 68:970-974 (1973).

[13] Dyer, A. R., "Hypotheses Testing Procedures for Separate Families of Hypotheses," Journal of the American Statistical Association 69:140-145 (1974).

[14] Greenwood, J. A. and D. Durand, "Aids for Fitting the Gamma Distribution by Maximum Likelihood," Technometrics 2:55-65 (1960).

[15] Hajek, J. and Z. Sidak, Theory of Rank Tests, Academic Press, New York (1967).

[16] Harter, H. L. and A. H. Moore, "Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples," Technometrics 7:639-643 (1965).

[17] Harter, H. L. and A. H. Moore, "Local-maximum-likelihood Estimation of the Parameters of Three-parameter Lognormal Populations from Complete and Censored Samples," Journal of the American Statistical Association 61:841-842 (1966).

[18] Hogg, R. V., V. A. Uthoff, R. H. Randles and A. S. Davenport, "On the Selection of the Underlying Distribution and Adaptive Estimation," Journal of the American Statistical Association 67:597-600 (1972).

[19] Kent, J. and C. P. Quesenberry, "Selecting Among Probability Distributions Used in Reliability," Technometrics, 24, 59-65.

[20] Lehmann, E. L., Testing Statistical Hypotheses, Wiley, New York (1959)

[21] Nelson, W., "Theory and Applications of Hazard Plotting for Censored Failure Data," Technometrics 14:945-966 (1972).

[22] Quesenberry, C. P. and R. R. Starbuck, "On Optimal Tests for Separate Hypotheses and Conditional Probability Integral Transformations," Communications in Statistics: Theory and Methods. Part A, 5:507-524 (1976).

[23] Siswadi, "Selecting Among Weibull, Lognormal and Gamma Distributions for Complete and Censored Samples," Unpublished North Carolina State University thesis (1981).

[24] Uthoff, V. A., "An Optimum Test Property of Two Well-known Statistics,", Journal of the American Statistical Association 65:1597-1600 (1970).

[25] Uthoff, V. A., "The Most Powerful Scale and Location Invariant Test of the Normal versus the Double Exponential," Annals of Statistics 1:170-174 (1973).

[26] Volodin, I. N., "On the Discrimination of Gamma and Weibull Distributions," Theory of Probability and its Applications 19:383-393 (1974).