Journal of Applied Statistical Science
Volume 16, Number 1, pp. 115–126
ISSN 1067-5817

APPLICATIONS OF LIKELIHOOD BASED-METHODS FOR THE RELIABILITY PARAMETER OF THE LOCATION AND SCALE EXPONENTIAL DISTRIBUTION

***Shaul K. Bar-Lev*[1] *and Frank A. Van der Duyn Schouten*[2]**
[1]Department of Statistics, University of Haifa,
Haifa 31905, Israel
[2]Center for Economic Research, University of Tilburg,
5000 LE Tilburg, The Netherlands

Abstract

Based on a type-2 censored sample we consider a likelihood-based inference for the reliability parameter $R(t)$ of the location and scale exponential distribution. More specifically, we derive the profile and marginal likelihoods of $R(t)$. A numerical example is presented demonstrating the flavor of results that can be obtained by likelihood-based methods.

Key Words: Life testing; Likelihood interval; Location and scale exponential distribution; Marginal likelihood; Plausibility; Profile likelihood; Reliability; Type-2 censored sample.

AMS 2000 Subject Classifications: Primary 62A10; Secondary 62N05.

1. Introduction

In this note we apply the likelihood approach to draw likelihood-based inference on the reliability parameter associated with a scale and location exponential distribution. The likelihood-based approach for inference has been thoroughly developed for about two decades from mid-sixties to mid-eighties and has been widely applied to various areas, such as time series, linear models and psychological stochastic learning. The likelihood approach was first suggested by Fisher (1934) and later developed by many authors and applied in various contexts. A good survey of likelihood-based methods can be found in Severini (2000), Pace and Salvan (1997, Chapter 4), Royall (1997) and Kalbfleisch (1985). Applications of likelihood-based inference to some problems in life testing can be found in the above cited references. A recent reference is Bar-Lev (2003) in which likelihood-based methods were employed for inference on the shape parameter of the scale and shape Weibull distribution.

Basically, this approach embraces the likelihood principle stating that the likelihood function contains all available information on the unknown parameters that can be extracted

from the sample. Those parameter values, for which there is a relatively large probability of obtaining the observed sample, are considered as being supported by the data and are therefore regarded more plausible; and vice versa. The most plausible value of an unknown parameter is obviously its related maximum likelihood estimate (MLE). If $L(\omega)$ is the likelihood function of ω (possibly a vector), based on a given sample, and $\hat{\omega}$ is the MLE of ω, then the relative likelihood function of ω is the ratio $R(\omega) \doteq L(\omega)/L(\hat{\omega})$ which ranges between 0 to 1. Values of ω for which $R(\omega)$ is "small" can be regarded as implausible, whereas values of ω making $R(\omega)$ "large" can be viewed as plausible. The set $\{\omega : R(\omega) \geq \alpha\}$ is called a $100\alpha\%$ likelihood interval for ω. Accordingly, one might consider values of ω within a 90% or a 95% likelihood intervals as highly plausible whereas values of ω ranging outside a 5% or a 10% likelihood intervals as being highly implausible. Several comments regarding the use of the likelihood principle for inference are presented in the concluding section.

Consider now a random sample $\underline{x}$ drawn from a two-parameter $\omega = (\omega_1, \omega_2)$ distribution. Let $f(\underline{x}; \omega)$ and $L(\omega) = L(\omega : \underline{x})$ denote, respectively, the probability density function (p.d.f.) of $\underline{x}$ and the likelihood function of ω based on the sample $\underline{x}$. In various inferential situations, as in the present note, it is required to draw inference on a sub-parameter of ω, say ω_1, only. In such situations the sub-parameter of interest is called the structural parameter whereas ω_2 is regarded as the nuisance parameter. Inferences on the structural parameter ω_1 can be deduced by eliminating the nuisance parameter ω_2 from the model and constructing a likelihood which depends on ω_1 only. Several likelihood-based methods have been suggested in the literature for such an elimination, all resulting in likelihoods depending on ω_1 only. Resulting likelihoods are called profile, marginal, conditional and integrated likelihoods. The first two, which are utilized in this paper, are briefly outlined in Section 2. In Section 3 we treat the location-scale exponential distribution. Based on a type-2 censored sample we derive the profile and marginal likelihoods of the associated reliability parameter. Section 4 briefly outlines some frequency-based and fiduicial results obtained in the literature concerning the reliability parameter. A numerical example is provided in Section 5. Some concluding remarks regarding the use of likelihood-based approach are presented in Section 6.

2. Profile and Marginal Likelihoods

We first briefly outline the concept of a profile likelihood and then that of a marginal likelihood. Relevant references in this context are Sprott and Kalbfleisch (1969), Kalbfleisch and Sprott (1973), Barndorff-Nielsen (1978), Kalbfleisch (1987), Pace and Salvan(1997), Royall (1997) and Severini (2000).

The Profile likelihood of ω_1 eliminates ω_2 by simply replacing it with $\hat{\omega}_2(\omega_1)$, the MLE of ω_2 when ω_1 is held fixed. The profile and relative profile likelihoods of ω_1 are then defined, respectively, by

$$P(\omega_1) = \sup_{\omega_2} L(\omega_1, \omega_2) = L(\omega_1, \hat{\omega}_2(\omega_1))$$

and

$$RP(\omega_1) \doteq P(\omega_1) / \sup_{\omega_1} P(\omega_1). \tag{2.1}$$

The main disadvantage of the use of $RP(\omega_1)$ for likelihood inference on ω_1 is that it assumes that for any fixed ω_1 the nuisance parameter ω_2 attains its most likely value. This may lead to a loss of accuracy concerning inferential statements on ω_1, especially when the sample size is small.

The marginal likelihood of ω_1 eliminates ω_2 in a more "sophisticated" way as follows. Consider a minimal sufficient statistic $y = y(\underline{x})$ for (ω_1, ω_2). Assume that y can be partitioned as $y = (y_1, y_2)$ such that y_1 is an ancillary statistic for ω_1 in the presence of ω_2; i.e., the p.d.f. of (y_1, y_2) can be decomposed as

$$f(y_1, y_2 : \omega_1, \omega_2) = g(y_1 : \omega_1) h(y_2 : \omega_1, \omega_2 \mid y_1), \tag{2.2}$$

where g and h denote, respectively, the marginal p.d.f. of y_1 and the conditional p.d.f. of y_2 given y_1. In this case, inference on ω_1 can be based on the marginal submodel $g(y_1 : \omega_1)$. The marginal and relative marginal likelihoods of ω_1 are therefore defined, respectively, by

$$M(\omega_1) \doteq g(y_1 : \omega_1), \tag{2.3}$$

and

$$RM(\omega_1) \doteq \frac{M(\omega_1)}{\sup_{\omega_1} M(\omega_1)}. \tag{2.4}$$

One drawback of the marginal procedure is that there should exist an ancillary statistic y_1 allowing the decomposition of the form given in (2.2). In case that more than one ancillary statistic exists, the problem arises which one to choose. However, a more substantial drawback of this procedure is, that even in case that (2.2) holds, the information on ω_1 that might be contained in the conditional submodel $h(y_2 : \omega_1, \omega_2 \mid y_1)$ is ignored. This potential loss of information has motivated numerous authors to define the notion of a *nonformative* conditional submodel with respect to ω_1 in the presence of a nuisance parameter ω_2, i.e., a submodel which contains no available information on ω_1 in the absence of knowledge of ω_2. Indeed, various definitions have been proposed for this notion implying that a marginal likelihood is not unique. A good description of this problem, i.e., whether there is a universal definition for a conditional submodel to be nonformative for a structural parameter in the presence of a nuisance parameter, as well as additional relevant references can be found in Jorgensen (1993).

3. An Application to the Reliablity Pararmeter of the Location-Scale Exponential Distribution

The location-scale exponential distribution has, respectively, a p.d.f. and a cumulative distribution function (c.d.f.) of the form

$$f(x : \theta, \delta) = \theta^{-1} \exp\{-(x-\delta)/\theta\} I_{(\delta,\infty)}(x) \tag{3.1}$$

and

$$F(t : \theta, \delta) = [1 - \exp\{-(t-\delta)/\theta\}] I_{(\delta,\infty)}(t), \tag{3.2}$$

where both parameters $\theta \in \mathbb{R}^+, \delta \in \mathbb{R}$ are unknown and $I_A(x)$ is the indicator function of a set A. This distribution is designated henceforth by $\exp(\theta, \delta)$.

The reliability function at the point t associated with (3.2) is

$$R(t) = R \doteq 1 - F(t : \theta, \delta) = \begin{cases} \exp\{-(t-\delta)/\theta\}, t > \delta \\ 1, \qquad\qquad t < \delta. \end{cases} \tag{3.3}$$

An inference on $R(t)$ is considered to be based on a type-2 censored sample stemming from (3.2). More specifically, n items with survival density (3.1) are placed on a test. The test is stopped once a predetermined $r-th$ failure time, $1 \leq r \leq n$, occurs. Let $x_1 \leq x_2 \leq ... \leq x_r$ denote the r failure times, then their respective joint p.d.f. is

$$f(x_1, ..., x_r : \theta, \delta) = C_{r,n} \theta^{-r} \exp\{-T(x_r - \delta)/\theta\} I_{(\delta,\infty)}(x_1), \tag{3.4}$$

where $C_{r,n} \doteq n!/(n-r)!$ and

$$T(x_r - \delta) \doteq \sum_{i=1}^{r} (x_i - \delta) + (n-r)(x_r - \delta). \tag{3.5}$$

In the next two subsections we derive the profile and marginal likelihood of $R(t)$.

3.1. The Profile Likelihood of R(t)

The likelihood function of (θ, δ) is proportional to (3.4) up to a constant which does not depend on (θ, δ). For deriving the profile likelihood of $R = R(t)$ we shall consider here a reparameterization of the location-scale exponential distribution by (R, δ) rather than by (θ, δ), i.e., in terms of the general setting of Section 2, $(R, \delta) = (\omega_1, \omega_2)$ with R and δ being the structural and nuisance parameters, respectively. Indeed, by using (3.3), we obtain that for $t > \delta$, $\theta = (t-\delta)/(\ln R^{-1})$. Hence the joint likelihood function of (R, δ), denoted by $L(R, \delta : x_1, ..., x_r) = L(R, \delta)$, is given by

$$L(R, \delta) = \left[\frac{\ln R^{-1}}{t-\delta}\right]^r \exp\left\{-(\frac{\ln R^{-1}}{t-\delta}) T(x_r - \delta)\right\}, \ \delta < \min(t, x_1), 0 < R < 1. \tag{3.6}$$

Employing (2.1), the relative profile likelihood of the structural parameter R is defined by

$$RP(R) = \frac{L(R, \hat{\delta}(R))}{\sup_R L(R, \hat{\delta}(R))} = \frac{L(R, \hat{\delta}(R))}{L(\hat{R}, \hat{\delta})}, \tag{3.7}$$

where $L(R, \hat{\delta}(R)) \doteq \sup_\delta L(R, \delta)$. To find the supremum in (3.7), one should distinguish between two cases: (i) $\min(t, x_1) = x_1$, and (ii) $\min(t, x_1) = t$. For case (i), $L(R, \delta))$ is increasing in $\delta < x_1 < t$ for any given R. Hence $L(R, \hat{\delta}(R)) = L(R, x_1)$ and $\sup_R L(R, x_1)$ is obtained at

$$\hat{R} = \exp\left[-r(t - x_1)/T(x_r - x_1)\right].$$

Substituting this in (3.7) we obtain

$$RP(R) = \left[\frac{\ln R^{-1}}{(t-x_1)} \frac{T(x_r - x_1)}{r}\right]^r \exp\left\{-\frac{\ln R^{-1}}{t-x_1} T(x_r - x_1) + r\right\}, x_1 < t. \tag{3.8}$$

Table 1. 10% **profile likelihood intervals for** $R(t)$ based on (3.9)

n	2	5	20	100
10% likelihood intervals for $R(t)$	(.32,1)	(.63,1)	(.89,1)	(.98,1)

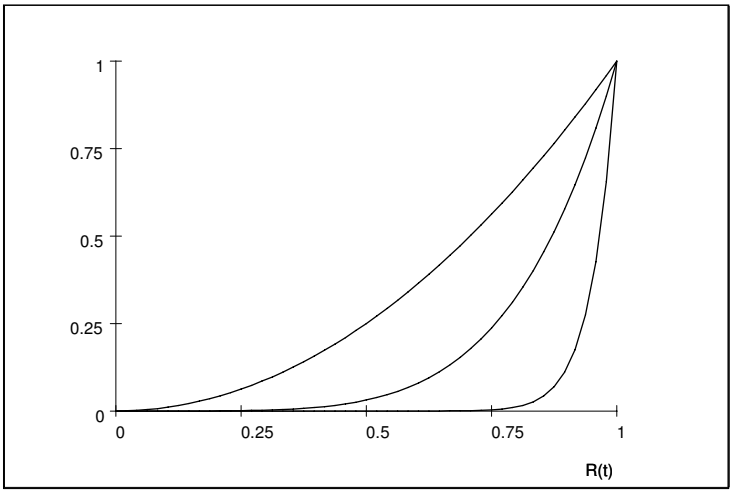

Figure 1. $RP(R)$ in (3.9) for n=2,5,20.

For case (ii), the quantities in (3.7) are obtained by a straightforward differentiation, yielding $\hat{\delta}(R) = t + r^{-1}T(x_r - t)\ln R$ and $\sup_R L(R, \hat{\delta}(R)) = L(1, \hat{\delta}(1))$. Hence

$$RP(R) = R^n,\ t < x_1. \tag{3.9}$$

Remark 1. Note that $RP(R)$ in (3.9) depends on n but remains constant in t regardless of the value of $t < x_1$. Such a result is, however, not too surprising. Indeed, since δ is unknown, then x_1, a strongly consistent estimate for δ, solely provides the only information on the "location" of $\delta < x_1$. For an arbitrary choice of $t\ (< x_1)$, it is not feasible to determine whether or not t is still larger than δ. Hence, the constancy of $RP(R)$ in t, for any $t < x_1$, is reasonable as it serves as a measure of our ignorance regarding the location of δ. Moreover, the dependence of $RP(R)$ on n is reasonable too since the larger the sample size n is the closer x_1 gets to δ, and, consequently, more information on R is gained. This can be seen in both Table 1 and Figure 1. Table 1 displays, for increasing n, 10% relative profile likelihood intervals for R, while Figure 1 plots the relative profile likelihood of $R(t)$ for $n = 2, 5$, and 20.

Combining the two cases (i) and (ii) in (3.8) and (3.9), respectively, we obtain that the relative profile likelihood interval for $R(t)$ is

$$RP(R) = \begin{cases} \left[\frac{\ln R^{-1}}{(t-x_1)}\frac{T(x_r-x_1)}{r}\right]^r \exp\left\{-\frac{\ln R^{-1}}{t-x_1}T(x_r-x_1)+r\right\}, x_1 < t \\ \\ R^n,\ t < x_1. \end{cases} \tag{3.10}$$

3.2. A Marginal Likelihood of $R(t)$

We use (2.4) to obtain the marginal likelihood of $R(t)$. It can be readily seen that the statistic $(x_1, T(x_r - x_1))$ is minimal sufficient for (θ, δ). Consequently, the statistic $(n(x_1 - t)/T(x_r - x_1), T(x_r - x_1))$ is minimal sufficient for (R, θ). Here, we shall use the parameterization $(\omega_1, \omega_2) = (R, \theta)$, i.e., with R and θ being the structural and nuisance parameters, respectively.

In order to derive the marginal likelihood of $R(t)$ we use the decomposition in (2.2) as applied to joint p.d.f. of the minimal sufficient statistic. Indeed, letting

$$(y_1, y_2) \doteq (n(x_1 - t)/T(x_r - x_1), T(x_r - x_1)), \tag{3.11}$$

we will show that the joint density of (y_1, y_2) can be decomposed as

$$f(y_1, y_2 : R, \theta) = g(y_1 : R) h(y_2 : R, \theta \mid y_1), \tag{3.12}$$

where the marginal density g of y_1 depends on R only, meaning that y_1 is ancillary for θ for any given R; whereas the conditional density of y_2 given y_1 depends on both R and θ but contains no information on R in the absence of knowledge of θ (in the sense of Sprott and Kalbfleisch, 1969, and Sprott, 1975). To observe this we use the following lemma.

Lemma 1. For $r > 1$, the joint density of (y_1, y_2) defined by (3.11) and the marginal density of y_1 in (3.12) are given, respectively, by

$$f(y_1, y_2 \quad : \quad R, \theta) = \frac{1}{(r-2)!\theta^r} R^n \exp\left\{\frac{-y_2(y_1+1)}{\theta}\right\} y_2^{r-1} I_{(0,\infty)}(y_2) \times I_{(n\theta \ln R, \infty)}(y_1 y_2) \tag{3.13}$$

and

$$g(y_1 : R) = \begin{cases} \frac{(r-1)R^n}{(y_1+1)^r}\left[1 - \sum_{i=0}^{r-1} \frac{1}{i!}\left(\frac{y_1+1}{y_1}\ln R^n\right)^i \exp\left(-\frac{y_1+1}{y_1}\ln R^n\right)\right], & y_1 < 0 \\ \frac{(r-1)R^n}{(y_1+1)^r}, & y_1 > 0. \end{cases} \tag{3.14}$$

The expression (3.13) is simply obtained by transforming $(u_1, u_2) \doteq (n(X_1 - \delta), T(X_r - X_1)) \to (y_1, y_2)$ and noting that $u_1 \perp u_2$ with $u_1 \sim \exp(\theta)$ and $u_2 \sim gamma(r-1, \theta)$. (3.14) then follows by integrating (3.13) with respect to y_2.

Note that the two conditions $y_1 < 0$ and $y_1 > 0$ are equivalent to the conditions $t > x_1$ and $t < x_1$, respectively. Hence, by using (3.14) we obtain that up to a constant not depending on R the marginal likelihood of R is

$$M(R) = \begin{cases} R^n\left[1 - \sum_{i=0}^{r-1} \frac{1}{i!}\left(\frac{T(x_r - t)}{n(x_1 - t)}\right)^i (\ln R^n)^i \exp\left(-\frac{T(x_r-t)}{n(x_1-t)}\ln R^n\right)\right], \\ \qquad\qquad\qquad\qquad\qquad\qquad if \;\; x_1 < t \\ R^n, \quad if \;\; t < x_1, \end{cases} \tag{3.15}$$

To show that $M(R)$ contains all available information on R in the absence of knowledge of θ, we use the fact that the conditional distribution of y_2/θ given y_1 does not depend on θ. Hence, the r.v. y_2/θ conditional on y_1 is a pivotal quantity for θ. This result satisfies one of

the criteria given in Sprott and Kalbfleisch (1969) (see also Sprott, 1975) required for the conditional model $h(y_2 : R,\theta \mid y_1)$ to be nonformative with respect to R in the absence of knowledge of θ. Hence, by that criterion, $M(R)$ contains all of the available information on R that can be extracted from the sample.

Note also that for $t < x_1$, the supremum of the second term in (3.15) is obtained at $\hat{R} = 1$ and hence $RM(R) = R^n$. This term coincides with that of $RP(R)$ in (3.10). For $t > x_1$, the supremum of the first term of (3.15) cannot be expressed explicitly and should be solved numerically for specific sample observations. Accordingly, the resulting form of $RM(R)$ is

$$RM(R) = \begin{cases} \left\{ \sup_R R^n \left[1 - \sum_{i=0}^{r-1} \frac{1}{i!} \left(\frac{T(x_r - t)}{n(x_1 - t)}\right)^i (\ln R^n)^i \exp\left(-\frac{T(x_r-t)}{n(x_1-t)} \ln R^n\right) \right] \right\}^{-1} \\ \qquad \times R^n \left[1 - \sum_{i=0}^{r-1} \frac{1}{i!} \left(\frac{T(x_r - t)}{n(x_1 - t)}\right)^i (\ln R^n)^i \exp\left(-\frac{T(x_r-t)}{n(x_1-t)} \ln R^n\right) \right], \\ \qquad\qquad\qquad\qquad\qquad\qquad if \;\; t > x_1, \\ R^n, \quad if \;\; t < x_1. \end{cases} \tag{3.16}$$

4. Some Frequency-Based and Fiduicial Approach Results

The MLE of (θ, δ) is $(T(x_r - x_1)/r, x_1)$. Hence, using (3.3), the MLE for $R = R(t)$ is

$$\hat{R} = \exp\left\{ -\frac{(t - x_1)}{T(x_r - x_1)/r} \right\}, t > x_1. \tag{4.1}$$

Pugh (1963) derived an expression for the minimum variance unbiased estimate for $R(t)$ for the non-censored sample case and $\delta = 0$. Balasubramanian and Balakrishnan (1992) dealt with parameter estimation for the location and scale exponential distribution under multiple type-2 censoring. Additional references can be found in Johnson, Kotz and Balakrishnan (1994, pp. 507-509). Engelhardt and Bain (1978) derived the distribution (3.14) of the ancillary statistic $y_1 = n(x_1 - t)/T(x_r - x_1)$ and used it to construct confidence limits for $R(t)$. The resulting limits can be calculated only numerically due to the rather cumbersome expression of the p.d.f. of y_1. More specifically, if $t_p = \delta + \theta[-\log(1-p)]$ denotes the p-th quantile of (3.1), Engelhardt and Bain used the pivotal quantity $Z_p = r(x_1 - t_p)/T(x_r - x_1)$ to base confidence intervals for t_p by utilizing the relation

$$\gamma = P(Z_p \le \zeta_{p,\gamma}) = P(t_p \ge x_1 - \zeta_{p,\gamma} T(x_r - x_1)/r),$$

where $\zeta_{p,\gamma}$ designates the γ-th quantile of Z_p. Such confidence intervals can be converted to confidence intervals for $R(t)$ by employing the relation $P(t_p \ge t) = P(R(t) \ge 1 - p)$. Their calculations require though an intensive simulation work

A fiducial approach was carried out by Pierce (1973) and Grubbs (1971). Pierce derived a version of fiducial distribution of $R(t)$ from the joint fiducial distribution of (θ, δ). For obtaining the latter joint distribution, he used a prior distribution for (θ, δ) being proportional to θ^{-1} multiplied by the likelihood function of (θ, δ). Grubbs (1971) used a fiducial procedure to obtain an approximate one-sided confidence interval for $R(t)$. By holding x_1 and $\tilde{\theta} \doteq T(x_r - x_1)/(r-1)$ fixed and letting $w = (t - x_1)/(r-1)\tilde{\theta}$, Grubbs presented the quantity

$$Q \doteq \ln\left(\frac{1}{R(\tau)}\right) = \frac{t - \delta}{\theta} = \frac{x_1 - \delta}{\theta} + w\frac{(r-1)\tilde{\theta}}{\theta}.$$

Table 2. R(t),$\hat{R}(t)$ and $\hat{R}_M(t)$

t	$R(t)$	$\hat{R}(t)$	$\hat{R}_M(t)$
4	.7408	.8940	.8580
8	.4966	.7129	.7065

He then used the fact that $2n(x_1-\delta)/\theta \sim \chi^2(2)$ and $2(r-1)\tilde{\theta}/\theta \sim \chi^2(2r-2)$ as well as some known approximations to represent Q, properly normalized, as a standard normal variate. These lead to

$$P\left(R(t) > \exp\left\{-m\left[1-\frac{v}{3m^2}+\frac{z_{1-\alpha}}{3m}v^{1/2}\right]^3\right\}\right) \cong 1-\alpha, \tag{4.2}$$

as an approximated $(1-\alpha)$ one-sided confidence interval for $R(t)$, where $m = 1/n+(r-1)w$, $v = 1/n^2+(r-1)w^2$ and z_p designates the p-th quantile of standard normal variate. Grubbs' expression in (4.2) does not distinguish, however, between the two cases $t > x_1$ and $t < x_1$ that separate the two terms of $RP(R)$ and $RM(R)$ in (3.10) and (3.16), respectively. For some values of $t < x_1$, such an oversight may lead to unacceptable values of $R(t)$ such as $P(R(t_0) > 1) = .9$, for some $t_0 < x_1$.

5. A Numerical Example

A type-2 censored sample with entries $n = 8$ and $r = 3$ was generated from an $\exp(10,1)$ distribution. The first three failure times were $x_1 = 2.02, x_2 = 7.68$ and $x_3 = 9.91$. We shall derive the relative profile and marginal likelihoods of $R(t)$ for $t = 4$ and $t = 8$ (i.e., for the case where $t > x_1$), plot these likelihoods and present a table displaying various likelihood intervals.

Note that the MLE $\hat{R} = \hat{R}(t)$ defined in (4.1) maximizes the relative profile likelihood in (3.10). The value of R which maximizes the marginal likelihood in (3.15) is called maximum marginal likelihood estimate (MMLE) and is denoted by $\hat{R}_M = \hat{R}_M(t)$. As already indicated earlier, the MMLE, as opposed to the MLE, cannot be solved analytically but only numerically. Based on the above data, the following table compares, for $t = 4,8$, the true value of $R(t)$ and the numerical values of the MLE and the MMLE.

For this specific censored sample and $t = 8$ both MLE and MMLE deviate significantly from the true value. This deviation seems to be mainly due to the relatively small censored sample size. Note however that for both cases of t, the numerical values of the MMLE are closer to the true values of $R(t)$ than those of the MLE.

Expressions of the relative profile and marginal likelihoods of $R = R(t)$ for $t = 4,8$, as extracted from (3.10) and (3.16) are given, respectively, by

$$RP(R(4)) = -14268R^{26.768}\ln^3 R\,, \tag{5.1}$$

$$RP(R(8)) = -517.9R^{8.8629}\ln^3 R, \tag{5.2}$$

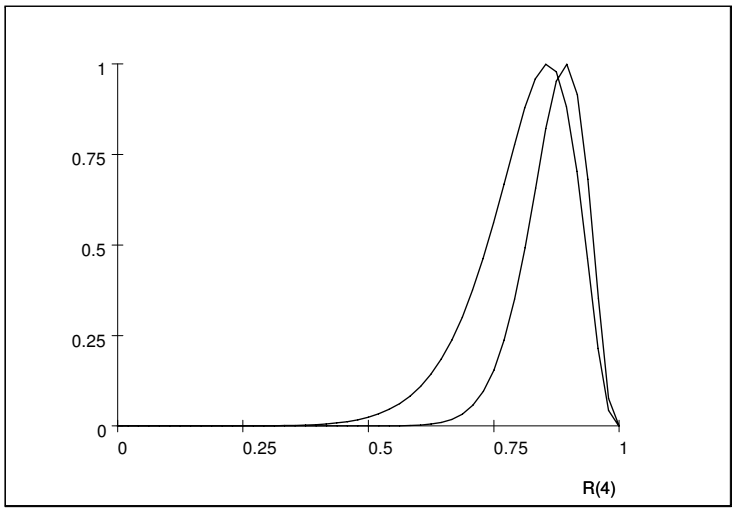

Figure 2. Relative profile (dash) and marginal (solid) likelihoods of $R(4)$

$$RM(R(4)) = 6.213R^8 - 6.213R^{26.768} + 116.6R^{26.768}\ln R - 1094.2R^{26.768}\ln^2 R \quad (5.3)$$

and

$$RM(R(8)) = 4471R^8 - 4471R^{8.8629} + 3857.7R^{8.8629}\ln R - 1664.4R^{8.8629}\ln^2 R. \quad (5.4)$$

Figures 2 and 3 plot, respectively, $RP(R(4))$ and $RM(R(4))$ versus $R(4)$ and $RP(R(8))$ and $RM(R(8))$ versus $R(8)$. It can be seen from Figure 3 that $RP(R(8))$ and $RM(R(8))$ almost coincide and are rather symmetrical around their maximizing values (MLE and MMLE, respectively). The plausibilities of the true value $R(8) = .4966$ (Table 2) under $RP(R(8))$ and $RM(R(8))$ are .359 and .389, respectively. The situation in Figure 2 is rather different; whereas $RP(R(4))$ is almost symmetrical, $RP(R(4))$ has much slower tailing off for smaller values of $R(4)$. Moreover, the plausibilities of the true value $R(4) = .7408$ under $RP(R(4))$ and $RM(R(4))$ are .125 and .518, respectively, implying that this true value of $R(4)$ is 4.1 times more likely under $RM(R(4))$ than under $RP(R(4))$. The latter result seems to be related to the fact that the marginal likelihood was shown to contain, at least by one criterion, all of the available information on $R(t)$ that can be extracted from the sample.

Characteristics similar to those demonstrated in Figures 2 and 3 can also be seen from Table 3 which displays 10% and 90% relative profile and marginal likelihood intervals.

For the sake of completeness, we also present 90% confidence intervals for $R(t)$, $t = 4, 8$, based on the above data and Grubbs' fiduicial approach (c.f., ((4.2)). These 90% fiduicial intervals are (.7451,1) for $R(4)$ and (.5984,1) for $R(8)$.

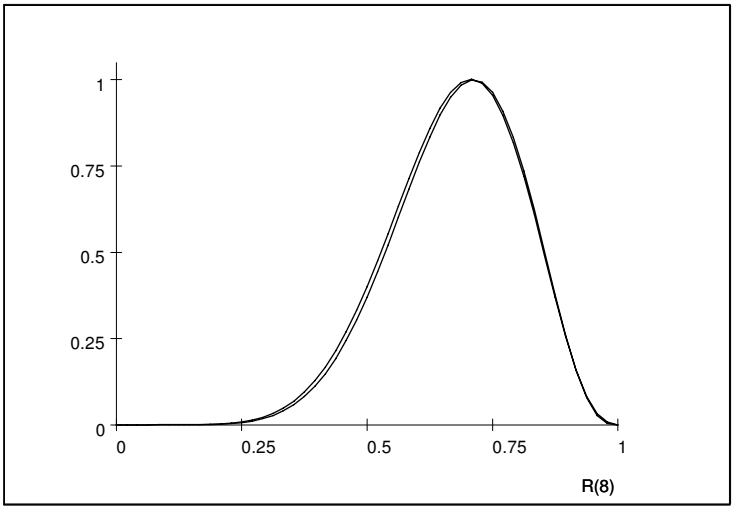

Figure 3. Relative profile (dash) and marginal (solid) likelihoods of $R(8)$

Table 3. 10% and 90% relative profile and marginal likelihood intervals

Interval type	$t = 4$	$t = 8$
10% profile likelihood interval	(.7310,.9766)	(.3882,.9310)
10% marginal likelihood interval	(.5971,.9698)	(.3785,.9321)
90% profile likelihood interval	(.8655,.9766)	(.6465,.7738)
90% marginal likelihood interval	(.8169,.8924)	(.6391,.7699)

6. Some Concluding Remarks

In this note we have been trying to invigorate the use of the likelihood principle by applying it to derive likelihood intervals for the reliability parameter of the location and scale exponential distributions. The resulting likelihood intervals provide at least a rough idea of reasonable and non-reasonable values of the parameter involved. However, a 10% likelihood interval is not comparable with a 90% level two-sided confidence interval. These two intervals have different meanings and interpretations. Whereas confidence intervals are based on hypothetically many repetitions of the same experiment, likelihood intervals are based on a particular experiment and parameter values are ranked by how likely they make an observed sample.

The question whether to use the likelihood-based approach for inference or the more commonly used frequency-based approach has no simple answer. Many of the commonly used criteria for evaluating various statistical procedures, such as variance, biasedness and coverage probabilities, may be justified only by repeated sampling. If repetitions are not made or planned, then to this end at least, the likelihood approach seems to be more appropriate. In his monograph on Statistical Evidence, Royal (1997) strongly supports the law of likelihood for likelihood inferential statements. While commenting on the strength of statistical evidence he states (p. 11): ”How strong is the evidence when the likelihood ratio

is 2?...Or 20? Many scientists (and journal editors) are comfortable interpreting a statistical significance level of 0.05 to mean that the observations are 'pretty strong evidence' against the null hypothesis, and a level of 0.01 to mean 'very strong evidence'. Are there reference values of likelihood ratios where corresponding interpretations are appropriate?" His monograph is devoted to providing a definitively affirmative response to the latter question. He states that (p.31): "The law of likelihood is intuitively reasonable, consistent with the rules of probability theory, and empirically meaningful. It is, however, incompatible with today's dominant statistical theory and methodology, which do not conform to the law's general implications, the irrelevance of the sample space and the likelihood principle, and which are articulated in terms of probabilities, which measure uncertainty, rather than likelihood ratios, which measure evidence".

Classical practitioners have refrained though from using likelihood-based methods not only because these methods stem from a different approach but possibly also because of the computational complexity involved. However, such complexity seems to be resolved with the present availability of computers and adequate mathematical software. Indeed, the computations in this paper have been conducted easily with a MAPLE package installed in a personal computer. In conclusion, we believe that although the likelihood principle-based approach for inference, cannot serve as a replacement for the traditional classical approach, it has its own merits and can be viewed as complementary to it.

It is however beyond the scope of this note to deeply discuss the various aspects of the likelihood approach for inference and the reader is referred to the references cited in this note. Additional references which advocate the use of likelihood-based methods are Basu (1977), Ghosh (1988). The monograph by Royall (1997) contains a rich list of further advocating references. Arguments against the use of the likelihood-based approach can be found in Berger and Wolpert (1988).

Acknowledgement

S.K. Bar-Lev was partially supported by NWO grant no. B61-493.

References

[1] Balasubramanian, K. and Balakrishnan, N. (1992). Estimation for one- and two- parameter exponential distributions under multiple type-II censoring. *Statistische Hefte*, **33**, 203-216.

[2] Bar-Lev, S.K. (2003). Likelihood-based inference for the shape parameter of a two parameter Weibull distribution. To appear in *Lifetime Data Analysis*.

[3] Barndorff-Nielsen, O. (1978). *Information and Exponential Families*, Wiley, Chichester

[4] Basu, D. (1977). On the elimination of nuisance parameters. *Journal of the American Statistical Association*, **72**, 355-366.

[5] Berger, J.O. and Wolpert, R.L. (1988). *The Likelihood Principle*, 2nd edition, IMS, Hayward.

[6] Engelhardt, M. and Bain, L.J. (1978). Tolerance limits and confidence limits on reliability for the two-parameter exponential distribution. *Technometrics*, **20**, 37-39.

[7] Fisher, R.A. (1934). Two new properties of mathematical likelihoods. *Proceedings of the Royal Society of London*, **A 144**, 285-307.

[8] Ghosh, J.K. (ed.) (1988). *Statistical Information and likelihood*, *Lecture Notes in Statistics*, **45**, Springer-Verlag, Heidelberg.

[9] Grubbs, F.E. (1971). Fiducial bounds on reliability for the two parameter negative exponential distribution. *Tecnometrics*, **13**, 873-876.

[10] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions* V.1, John Wiley, New York.

[11] Jorgensen, B. (1993). A review of conditional inference: is there a universal definition of nonformation?. *Bulletin of the International Statistics Institute*, **55**, 323-340.

[12] Kalbfleisch, J.D. and Sprott, D.A. (1973). Marginal and conditional likelihoods. *Sankhya A*, **35**, 311-328.

[13] Kalbfleisch, J.G. (1985). *Probability and statistical inference* V.2, Springer-Verlag, New York.

[14] Pace, L. and Salvan, A. (1997). *Principles of statistical inference*, World scientific, Singapore.

[15] Pierce, D.A. (1973). Fiducial, frequency, and Bayesian inference on the reliability for the two parameter exponential distribution. *Tecnometrics*, **15**, 249-253.

[16] Pugh, E.L. (1963). The best estimate of reliability in the exponential case. *Operations Research*, **11**, 57-61.

[17] Royall, R. (1997). *Statistical Evidence*, Chapman and Hall: London.

[18] Severini, T.A. (2000). *Likelihood Methods in Statistics*, Oxford University Press, New York.

[19] Sprott, D.A. (1975). Marginal and conditional sufficiency. *Biometrika*, **62**, 599-605.

[20] Sprott, D.A. and Kalbfleisch, J.D. (1969). Examples of likelihoods and comparisons with point estimates and large sample approximations. *Journal of the American Statistical Association*, **64**, 468-484.