Robust Weighted Likelihood Estimation of Exponential Parameters

Ejaz S. Ahmed, Andrei I. Volodin, and Abdulkadir. A. Hussein

***Abstract*—The problem of estimating the parameter of an exponential distribution when a proportion of the observations are outliers is quite important to reliability applications. The method of weighted likelihood is applied to this problem, and a robust estimator of the exponential parameter is proposed. Interestingly, the proposed estimator is an α-trimmed mean type estimator. The large-sample robustness properties of the new estimator are examined. Further, a Monte Carlo simulation study is conducted showing that the proposed estimator is, under a wide range of contaminated exponential models, more efficient than the usual maximum likelihood estimator in the sense of having a smaller risk, a measure combining bias & variability. An application of the method to a data set on the failure times of throttles is presented.**

***Index Terms*—Exponential distribution, maximum likelihood estimation, robust estimation, weighted likelihood method.**

ACRONYMS[1]

- a.s. almost surely
- WLE weighted likelihood estimator
- MLE maximum likelihood estimator

NOTATION

- $X^{(n)} = (X_1, \ldots, X_n)$ random sample size n
- θ parameter of interest
- Θ parametric space
- $\{F(x;\theta);\ \theta \in \Theta\}$ a parametric family of cumulative distribution functions
- $\{f(x;\theta);\ \theta \in \Theta\}$ a parametric family of probability density functions
- $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ the usual maximum likelihood estimator of parameter θ
- θ_n^* the proposed estimator of the parameter θ
- $\ln(\cdot)$ the natural logarithm function
- $t_i = t_i(X^{(n)})$ special weights that we use for the estimation method
- ε an obstructing parameter
- $\mathcal{G} = \{G_\varepsilon(x) = (1-\varepsilon)F(x;\theta)+\varepsilon F(x;\theta_1),\ 0 \le \varepsilon \le 1\}$ the set of obstructing distributions where $\theta_1 \in \Theta$, and $\theta_1 \neq \theta$
- Δ, λ, α fixed positive constants
- $\bar{X}$ sample mean
- $\bar{Y}$ α-trimmed sample mean
- $R(\cdot)$ quadratic risk of a statistics
- $O(a)$, $o(a)$ Convergence as fast as a, convergence faster than a

Manuscript received August 28, 2003. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Associate Editor: M. Xie.

S. E. Ahmed and A. A. Hussein are with the Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario, N9B 3P4 Canada.

A. I. Volodin is with the Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, S4S 4G2 Canada.

Digital Object Identifier 10.1109/TR.2005.853276

[1]The singular and plural of an acronym are always spelled the same.

I. INTRODUCTION

THE EXPONENTIAL distribution is the most commonly used model in reliability & life-testing analysis [1], [2]. We propose a method of weighted likelihood for robust estimation of the exponential distribution parameter. We establish that the suggested weighted likelihood method provides α-trimmed mean type estimators. Further, we investigate the robustness properties of the weighted likelihood estimator (WLE) in comparison with the usual maximum likelihood estimator (MLE). This can be viewed as an application of the classical likelihood that was introduced by [3] as *the relevance weighted likelihood* to problems of robust estimation of parameters. The weighted likelihood method was introduced as a generalization of the local likelihood method. For further discussion of the local likelihood method in the context of nonparametric regression, see [4]–[6]. In contrast to the local likelihood, the weighted likelihood method can be global, as demonstrated by one of the applications in [7] where the James-Stein estimator is found to be a maximum weighted likelihood estimator with weights estimated from the data.

The theory of weighted likelihood enables a bias-precision trade of to be made without relying on a Bayesian approach. The latter permits the bias-variance trade of to be made in a conceptually straightforward manner. However, the reliance on empirical Bayes methods softens the demands for realistic prior modeling in complex problems. The weighted likelihood theory offers a simple alternative to the empirical Bayesian approach for many complex problems. At the same time, it links within a single formal framework a diverse collection of statistical domains, such as weighted least squares, nonparametric regression, meta-analysis, and shrinkage estimation. The weighted likelihood principle comes with an underlying general theory that extends Wald's theory for maximum likelihood estimators, as shown in [8].

We propose applying the method of weighted likelihood to the problem of the robust estimation of the parameter of an exponential distribution by assigning zero weights to observations with small likelihood. Interestingly, the weighted likelihood method yields α-trimmed mean type estimators of the parameter of interest. The statistical asymptotic properties of the WLE is developed, and a simulation study is conducted to appraise the behavior of the proposed estimators for moderate sample sizes in comparison with the usual maximum likelihood estimator.

An analogous idea appears in [9]. However, these authors use a different type of likelihood equation in which they do not weigh the likelihood of each observation, but rather the logarithmic derivatives. Hence, the proposed estimation strategy is a completely different approach.

II. The Proposed Weighted Likelihood Estimator

Let in general $\{F(x;\theta);\ \theta \in \Theta\}$ be a parametric family of cumulative distribution functions with corresponding probability density functions $f(x;\theta)$.

Estimators of θ obtained by maximizing the weighted log-likelihood function

$$L\left(\theta|X^{(n)}\right) = \sum_{i=l}^{n} t_i \ln f(X_i;\,\theta)$$

where t_i depends on the sample $t_i = t_i(X^{(n)})$; are called Weighted Likelihood Estimators (WLE) of θ. If all the weights are equal to unity, then the resulting estimator is the usual maximum likelihood estimator.

In this paper, we suggest a special choice of weights t_i; connected with the theory of robust estimation, and based on the maximum likelihood method with rejection of spurious observations. Let $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ be the usual maximum likelihood estimator of the parameter θ. The weight t_i that corresponds to the observation X_i is assumed to be 1, if its estimated likelihood is suffciently large, and 0 elsewhere; that is

$$t_i = \begin{cases} 1 & \text{if } f(X_i;\,\hat{\theta}_n) > C \\ 0 & \text{elsewhere.} \end{cases}$$

Consequently, we delete all *improbable* observations from the sample (the choice of C is discussed in the next paragraph). Not surprisingly, it can be seen that in the case of a uni-modal probability density function f we reject only extreme order statistics. However, this may not be the case for multi-modal probability density functions.

The proposed estimator θ_n^* of the parameter θ is defined as the solution of the equation

$$\sum_{k=1}^{m} \frac{\partial \ln f(X_{i_k};\,\theta)}{\partial\theta} = 0$$

where $X_{i1}, \ldots, X_{im}$ are the remaining observations in the sample after the rejection procedure.

In the case of the exponential distribution, these estimators are obtained by using in the above equations the appropriate distribution, and density functions given by, respectively

$$F(x;\,\theta) = 1 - \exp\left\{\frac{-x}{\theta}\right\}, \quad f(x;\,\theta) = \theta^{-1}\exp\left\{\frac{-x}{\theta}\right\}$$

where $x \geq 0, \theta > 0$.

Regarding the issue of the choice of C, we suggest $C = a/\hat{\theta}_n$; where a can be chosen as in the selection of the critical constant in the criterion of elimination of outliers. Therefore, in the exponential case, a is chosen from the condition of a small probability of rejection of an observation assuming that the observation has come from the above exponential cumulative distribution function $F(x;\theta)$. That is, a is defined by the given small probability α in the equation

$$P\left(\max_{1\leq i\leq n} X_i > -\hat{\theta}_n \ln a\right) \approx 1 - \prod_{i=1}^{n} P(X_i \leq -\theta \ln a) = 1-(1-a)^n = \alpha.$$

From this equation, we obtain $a = 1-(1-\alpha)^{1/n} \approx \alpha/n$. This approximation of a is reasonable because in our calculations we will replace $\hat{\theta}_n$ by the limit in probability (even almost surely) of this estimator, i.e., by θ: So, we choose the value of C as

$$C = \frac{\alpha}{n\hat{\theta}_n}.$$

Hence, we reject an observation from the sample if $X_k > -\overline{X}\ln(C\overline{X}) = \overline{X}\ln(n/\alpha)$. The probability α is user-dependent as in [9], and can be considered as a measure of the level of desired robustness.

Example: Here we consider a data set from [1] (also used by [10]) on the failure times of throttles. The data reports the failure times of 25 units where time is measured in Kilometers driven prior to failure. According to exponential model fit, the estimated rate of failure for this data is $1/\hat{\theta}_{25} = .0002$. We added 5% contamination from a different exponential with rate .000 05. That is equivalent to augmenting the data by one observation. The estimate of θ after contamination is $\bar{X} \approx 7327$. At $\alpha = .05$ level of robustness, the rejection criteria is $X_i > 50\,634.85$. Thus, only the one extra observation will be Removed, and now $\bar{X} \approx 4909$.

Furthermore, to illustrate that the proposed procedure & the α-trimmed estimators do not necessarily coincide, we notice that a .01-trimming procedure would not reject the contaminant observation, whereas our robust procedure with $\alpha = 0.01$ would reject it.

In the following section, we study the robustness properties of the $\hat{\theta}_n$ & θ_n^* by using contaminated (obstructing) distributions.

III. Robustness of the Proposed Estimator

In general, define a set of *obstructing distributions* as

$$\mathcal{G} = \{G_\varepsilon(x) = (1-\varepsilon)F(x;\,\theta) + \varepsilon F(x;\,\theta_1),\ 0 \leq \varepsilon \leq 1\} \tag{1}$$

where F denotes the cumulative distribution function with density function f, $\theta_1 \in \Theta$, and $\theta_1 \neq \theta$: As it is commonly done in the classical robustness studies (c.f., [11]), under the assumption that the sample is taken from the distribution G_ε with fixed ε one would find the limits in probability of the estimates $\hat{\theta}_n$ & θ_n^* for $n \to 1$, and compare their biases. The values of θ for which $|\theta_n^* - \theta|$ is less than $|\hat{\theta}_n - \theta|$, indicate the regions in the parameter space where the estimate θ_n^* is to be favored over the estimate $\hat{\theta}_n$. Along the same lines, we shall also compare the quadratic risks of these estimates. The quadratic risk of an estimator is the expected (average) squared distance between the estimator, and the parameter being estimated. In other words, the quadratic risk is an omnibus measure of the performance of an estimator which takes into consideration the bias & the precision (variance) of the estimator.

Let $\theta > 0$; $\Delta > 0$, and $0 < \alpha < 1$ be any positive Numbers, and put $\theta_1 = \theta(1+\Delta)$. Although in practice α depends on the sample size, n, here for the sake of brevity we assume that it is a constant.

In the exponential distribution case, and in virtu of (1), we assume that the sample $(X_1, \cdots, X_n)$ is taken from the distribution

$$G(x) = G_\varepsilon(x) = 1 - \exp\left\{\frac{-x}{\theta}\right\} -\varepsilon\left(\exp\left\{\frac{-x}{\theta(1+\Delta)}\right\} - \exp\left\{\frac{-x}{\theta}\right\}\right).$$

Under this contaminated model, note that

$$\hat{\theta}_n = \overline{X} = n^{-1}\sum_{i=1}^{n} X_i$$

and by the strong law of large numbers

$$\hat{\theta}_n \to \theta_\epsilon = E(\overline{X}) = (1-\varepsilon)\theta + \varepsilon\theta(1+\Delta) = \theta(1+\varepsilon\Delta) \text{ a.s.}$$

On the other hand, the estimate θ_n^* converges in probability (even almost surely) to some value $\overline{\theta}_\varepsilon$ which can be calculated as the limit of the expected values truncated at the point $A = -\theta_\varepsilon \ln(C\theta_\varepsilon) = \theta_\varepsilon \ln(n/\alpha)$ of the distribution $\overline{G}(x) = G(x)/G(A)$, $0 < x \leq A$. That is

$$\overline{\theta}_\varepsilon = \lim_{n\to\infty} \frac{1}{G(A)} \int_0^A x\left[\frac{1-\varepsilon}{\theta}\exp\left\{-\frac{x}{\theta}\right\} + \frac{\varepsilon}{\theta(1+\Delta)} \times \exp\left\{-\frac{x}{\theta(1+\Delta)}\right\}\right] dx. \quad (2)$$

Indeed, these integrals can be evaluated in closed form. However the solution is cumbersome, thus rendering the comparison of the relative biases $(\overline{\theta}_\varepsilon - \theta)/\theta$ & $(\theta_\varepsilon - \theta)/\theta = \varepsilon\Delta$ quite difficult. In fact

$$G(A) = 1 - e^{\frac{-A}{\theta}} - \varepsilon\left(e^{\frac{-A}{\theta(1+\Delta)}} - e^{\frac{-A}{\theta}}\right)$$

and the integral in the numerator of (2) equals

$$\theta_\varepsilon - (1-\varepsilon)(A+\theta)e^{\frac{-A}{\theta}} - \varepsilon\Big(A + \theta(1+\Delta)\Big)e^{-\frac{A}{\theta(1+\Delta)}}.$$

Therefore, an exact calculation of the gain in bias, and reduction in risk of the proposed estimator as compared to the usual MLE, may not be possible. Here we shall give only approximate analysis concerning these quantities under convenient assumptions, such as

$$\varepsilon = \frac{\lambda}{n}$$

for some $\lambda > 0$, or equivalently, $\epsilon = O(1/n)$. This requirement is not restrictive at all, as can be seen from the simulation results in Section IV. In fact, in practice, ϵ is a constant, and $n\epsilon$ is bounded for all reasonable sample sizes.

Recall that we reject observations satisfying $X_k > \overline{X}\ln(n/\alpha)$, and note that

$$\overline{X}\ln\left(\frac{n}{\alpha}\right) \sim \theta(1+\varepsilon\Delta)\ln\left(\frac{n}{\alpha}\right) = A, \text{ almost surely.}$$

The probability of rejecting an observation in the obstructed model is asymptotically equal to

$$\begin{aligned} P(X_1 > A) &= (1-\varepsilon)e^{-\frac{A}{\theta}} + \varepsilon e^{-\frac{A}{\theta(1+\Delta)}} \\ &= (1-\varepsilon)\left(\frac{\alpha}{n}\right)^{1+\varepsilon\Delta} + \varepsilon\left(\frac{\alpha}{n}\right)^{\frac{1+\varepsilon\Delta}{1+\Delta}} \\ &= \frac{\alpha}{n} + O\left(n^{-\frac{2+\Delta}{1+\Delta}}\right). \end{aligned}$$

Hence, the asymptotic distribution of θ_n^* equals the distribution of the α-trimmed sample mean

$$\overline{Y} = \frac{1}{n(1-\alpha)}\sum_{k=1}^{n(1-\alpha)} Y_k$$

of a random sample of size $n(1-\alpha)$ from the distribution concentrated on the interval (0, A). The probability density of this distribution is positive only on this interval, and has the form

$$f_A(x;\theta) = \frac{1}{G_\alpha}\left[\frac{1-\varepsilon}{\theta}e^{-\frac{x}{\theta}} + \frac{\varepsilon}{\theta(1+\Delta)}e^{-\frac{x}{\theta(1+\Delta)}}\right]$$

where $G_\alpha = 1 - \alpha/n + o(1/n)$.

Denote by $E_\alpha(X)$, and $Var_\alpha(X)$ the mathematical expectation, and variance, respectively, of a random variable X relative to the distribution with probability density function $f_A(x;\theta)$. With the preliminaries accounted for, the first result can now be presented.

Theorem 1: The maximum likelihood estimator $\hat{\theta}_n$ under the obstructing model has the relative bias

$$\varepsilon\Delta = \frac{\lambda\Delta}{n}$$

and the estimator θ_n^* has the relative bias

$$\varepsilon\Delta - \frac{\alpha}{n}\ln\frac{n}{\alpha} + O\left(\frac{\varepsilon\ln n}{n}\right) = \frac{\lambda\Delta}{n} - \frac{\alpha}{n}\ln\frac{n}{\alpha} + O\left(\frac{\varepsilon\ln n}{n}\right).$$

The above relation reveals that the relative bias of the proposed estimator θ_n^* in comparison with relative bias of the MLE $\hat{\theta}_n$ decreases in the value by the order $(\alpha/n)\ln(n/\alpha)$. Therefore, it is apparent that the value of α must be chosen as a quantity of order $O(1/\ln n)$; otherwise the values of θ will be under-estimated (negative bias). A possibility is to choose α from the relation $\lambda\Delta \approx \alpha\ln(n/\alpha)$, if the Δ & ϵ can be appropriately guessed. On the other hand, the difference between the relative bias of $\hat{\theta}_n$, and that of θ_n^* is positive, meaning there is always some gain in terms of bias reduction, as long as $\epsilon\alpha$ is not too small compared to $\alpha\ln(n/\alpha)$.

Next, we provide approximate expressions for the quadratic risks of the proposed estimator.

Theorem 2: The maximum likelihood estimator $\hat{\theta}_n$ has the quadratic risk

$$R(\hat{\theta}_n) = \frac{\theta^2}{n}\left[1 + 2\varepsilon\Delta(1+\Delta)\right] \quad (3)$$

and the estimator θ_n^* has the asymptotic representation of the quadratic risk

$$R(\theta_n^*) \sim \frac{\theta^2}{n(1-\alpha)}\left[1 + 2\varepsilon\Delta(1+\Delta) - \frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right]. \quad (4)$$

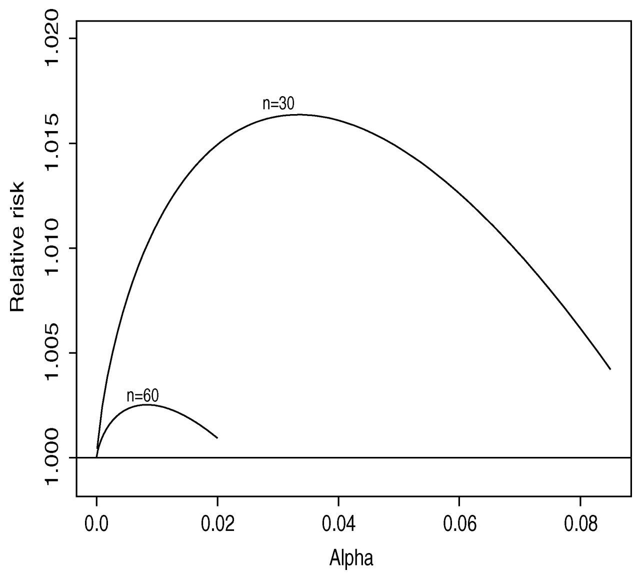

Fig. 1. The approximate relative risk $\mathcal{E}_n = R(\hat{\theta}_n)/R(\theta_n^*)$ as function of α. The line through $y = 1$ is the benchmark.

The relative efficiency of the estimator θ_n^* in comparison with maximum likelihood estimator $\hat{\theta}_n$ has the asymptotic representation

$$\varepsilon_n = \frac{R(\hat{\theta}_n)}{R(\theta_n^*)} \sim (1-\alpha)\left(1 + \frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right).$$

In passing, we note that more precise results can be obtained based on the joint distribution of the extreme terms of the sample, and the sample mean.

By examining the formula for $\mathcal{E}_n$, one notes that most of the risk reduction will be achieved for small samples, and for small α as compared to the size of the sample. The following Fig. 1 shows such a pathology for $n = 30$, 60, and for small α values. From the graph, apparently, the gain in risk will decrease as α increases, or as n increases, for fixed α. However, the simulations reported in the next section show a richer set of the parameters Δ, ϵ, α over which the proposed estimator has lesser risk than the MLE.

IV. Monte-Carlo Simulations

For the purposes of simulation, we fix $\theta > 0$, $0 < \varepsilon < 1$, and $\Delta > 0$. Recall that, if a random variable U has the uniform distribution on the interval [0, 1], then $X = -\theta\ln(U)$ will have the exponential distribution with parameter θ. We need this result for generating random numbers from the exponential distribution.

We generate random numbers from the obstructed distribution

$$G(x) = G_\varepsilon(x) = 1 - e^{\frac{-x}{\theta}} - \varepsilon\left(e^{\frac{-x}{\theta(1+\Delta)}} - \exp\left\{\frac{-x}{\theta}\right\}\right).$$

This procedure is organized in the following way. The main idea is that, with probability $1-\varepsilon$, we need to generate a random number from the exponential distribution with parameter $\theta_1 = \theta$; and with probability ε, generate a random number from the exponential distribution with parameter $\theta_1 = \theta(1+\Delta)$.

TABLE I
Simulated & Asymptotic Bias for Estimators $\hat{\theta}$ & θ^*, for $n = 200$ & $\Delta = 1.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	.0194	.0100	.0282	.0300	.0422	.0500	.0670	.0700	.0821	.0900
$\alpha = .01$	.0188	.0095	.0266	.0295	.0391	.0495	.0656	.0695	.0795	.0895
$\alpha = .03$	.0076	.0087	.0183	.0287	.0507	.0487	.0609	.0687	.0839	.0887
$\alpha = .05$	.0059	.0079	.0289	.0279	.0479	.0409	.0567	.0679	.0774	.0879
$\alpha = .07$	.0128	.0072	.0256	.0272	.0445	.0472	.0667	.0672	.0825	.0872
$\alpha = .09$	.0054	.0065	.0341	.0265	.0442	.0465	.0630	.0665	.0809	.0865

TABLE II
Simulated & Asymptotic Risk for Estimators $\hat{\theta}$ & θ^*, for $n = 200$ & $\Delta = 1.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	.0057	.0052	.0060	.0056	.0077	.0060	.0114	.0064	.0129	.0068
$\alpha = .01$	.0057	.0052	.0058	.0056	.0072	.0060	.0113	.0064	.0125	.0068
$\alpha = .03$	.0062	.0053	.0054	.0057	.0083	.0061	.0160	.0070	.0056	.0058
$\alpha = .05$	.0056	.0054	.0078	.0058	.0074	.0062	.0106	.0066	.0135	.0071
$\alpha = .07$	.0057	.0055	.0059	.0059	.0078	.0063	.0113	.0068	.0135	.0072
$\alpha = .09$	.0052	.0056	.0066	.0060	.0067	.0064	..0117	.0069	.0138	.0073

TABLE III
Differences Between Biases $\text{Bias}(\hat{\theta}) - \text{Bias}(\theta^*)$ & Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 30$ & $\Delta = 1.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0000	.0000	.0034	-.0008	.0000	.0000	.0060	-.0001	.0000	.0000
$\alpha = .03$	.0000	.0000	.0035	.0001	.0023	-.0014	.0051	.0051	.0249	.0035
$\alpha = .05$	.0076	.0040	.0022	-.0009	.0109	.0016	.0084	.0002	.0208	.0006
$\alpha = .07$	.0059	-.0006	.0089	.0018	.0301	.0022	.0126	.0005	.0184	.0095
$\alpha = .09$	.0037	.0008	.0034	.0008	.0079	.0009	.0387	.0007	.0124	-.0003

TABLE IV
Differences Between Biases $\text{Bias}(\hat{\theta}) - \text{Bias}(\theta^*)$ & Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 30$ & $\Delta = 3.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0007	-.003	.0282	.0240	.0601	.0213	.0588	.0319	.0598	.0458
$\alpha = .03$	.0084	.0108	.0572	.0191	.0870	.0413	.0391	.0208	.0822	.0490
$\alpha = .05$	.0321	.0084	.0134	-.0002	.0516	.0167	.0729	.0381	.0984	.0719
$\alpha = .07$	.0217	.0049	.0589	.0248	.0940	.0560	.0991	.0364	.1176	.0760
$\alpha = .09$	.0279	-.0021	.0709	.0219	.0731	.0364	.0802	.0480	.1136	.0653

To achieve this, we generate a random number u from the uniform distribution on the interval [0, 1], and compare u with ε. If $u > \varepsilon$, then take $\theta_1 = \theta$. Otherwise $\theta_1 = \theta(1+\Delta)$. Next, generate a random number x from the exponential distribution with the parameter θ_1. The random number x has the cumulative distribution function $G(x)$. All the following simulation results are based on 5000 replicates.

Tables I and II show for $n = 200$ the extent to which the simulated & the asymptotic biases, and risks of the proposed estimator & the MLE agree. The asymptotic biases & risks are calculated using the relations in Theorems 1 & 2. Note that the risk of the estimator $\hat{\theta}$ does not depend on α. From the tables, we see that the asymptotic results are accurate to the second decimal digit except for small α & ϵ in the case of bias, and for large α & ϵ in the case of risk.

To examine the performance of the proposed estimator in terms of bias & risk, we reported in Tables III–XI the differences between the biases $Bias(\hat{\theta}) - Bias(\theta^*)$, and the risks $R(\hat{\theta}) - R(\theta^*)$ of the two estimators for different values of n, Δ,

TABLE V
DIFFERENCES BETWEEN BIASES $\mathrm{Bias}(\hat{\theta}) - \mathrm{Bias}(\theta^*)$ & RISKS $R(\hat{\theta}) - R(\theta^*)$ FOR $n = 30$ & $\Delta = 5.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0215	.0087	.0477	.0231	.0838	.0396	.1496	.1344	.1770	.1197
$\alpha = .03$	.0334	.0161	.0966	.0641	.1717	.1334	.2675	.3158	.1661	.1403
$\alpha = .05$	.0258	.0089	.0898	.0432	.1254	.0718	.1686	.1293	.2009	.1569
$\alpha = .07$	.0272	.0042	01369	.0899	.1719	.1257	.1913	.1679	.2497	.2106
$\alpha = .09$	.0292	.0083	.0963	.0745	.1768	.0832	.2070	.1603	.2761	.2070

TABLE VI
DIFFERENCES BETWEEN BIASES $\mathrm{Bias}(\hat{\theta}) - \mathrm{Bias}(\theta^*)$ & RISKS $R(\hat{\theta}) - R(\theta^*)$ FOR $n = 100$ & $\Delta = 1.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0024	.0000	.0047	.0003	.0013	.0003	.0039	.0010	.0033	0001
$\alpha = .03$	.0025	.0001	.0083	-.0001	.0032	.0001	.0083	.0007	.0049	.0001
$\alpha = .05$	.0049	.0002	.0056	-.0001	.0098	.0000	.0143	.0020	.0174	.0024
$\alpha = .07$	.0085	.0003	.0065	.0001	.0134	.0024	.0165	.0026	.0187	.0035
$\alpha = .09$	.0068	.0000	.0065	.0005	.0110	.0009	.0130	.0018	.0129	.0022

TABLE VII
DIFFERENCES BETWEEN BIASES $\mathrm{Bias}(\hat{\theta}) - \mathrm{Bias}(\theta^*)$ & RISKS $R(\hat{\theta}) - R(\theta^*)$ FOR $n = 100$ & $\Delta = 3.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0148	.0033	.0341	.0094	.0521	.0190	.0722	.0348	.0551	.0308
$\alpha = .03$	.0272	.0032	.0419	.0112	.0593	.0196	.0693	.0327	.0764	.0468
$\alpha = .05$	.0177	.0014	.0560	.0064	.0592	.0230	.0654	.0307	.0930	.0505
$\alpha = .07$	.0122	.0000	.0560	.0131	.0592	.0231	.0654	.0326	.0930	.0503
$\alpha = .09$	.0117	-.0004	.0475	.0111	.0664	.0190	.0817	.0324	.1043	.0483

TABLE VIII
DIFFERENCES BETWEEN BIASES $\mathrm{Bias}(\hat{\theta}) - \mathrm{Bias}(\theta^*)$ & RISKS $R(\hat{\theta}) - R(\theta^*)$ FOR $n = 100$ & $\Delta = 5.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0229	.0028	.0880	.0373	.0865	.0404	.1565	.1089	.1337	.1153
$\alpha = .03$	.0237	.0056	.0939	.0374	.1559	.1105	.1646	.1193	.2179	.1929
$\alpha = .05$	.0404	.0065	.1052	.0405	.1584	.0796	.1760	.1146	.2240	.1928
$\alpha = .07$	.0291	.0078	.0888	.0297	.1503	.0752	.1786	.1074	.2354	.2013
$\alpha = .09$	.0391	.0033	.1085	.0379	.1388	.0572	.1999	.1267	.2285	.1808

TABLE IX
DIFFERENCES BETWEEN BIASES $\mathrm{Bias}(\hat{\theta}) - \mathrm{Bias}(\theta^*)$ & RISKS $R(\hat{\theta}) - R(\theta^*)$ FOR $n = 200$ & $\Delta = 1.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0000	.0000	.0013	.0000	.0031	.0004	.0012	.0001	.0033	.0004
$\alpha = .03$	.0016	.0000	.0048	.0003	.0048	.0011	.0075	.0015	.0070	.0009
$\alpha = .05$	.0015	.0001	.0041	.0002	.0105	.0018	.0081	.0012	.0100	.0020
$\alpha = .07$	.0030	-.0002	.0065	.0003	.0076	.0005	.0139	.0016	.0151	.0027
$\alpha = .09$	.0031	-.0001	.0082	.0005	.0111	.0002	.0112	.0011	.0123	.0018

ε, and α. These tables show the advantage of the estimator θ^* in the majority of the cases considered. The pathology pointed out in the previous section is clearly seen in these tables. For instance, the gain in terms of risk reduction in favor of θ_n^* decreases as α gets larger while fixing all the remaining parameters. Similar behavior can be observed for increasing n while keeping all other parameters fixed. Also, as intuitively expected, the gain in terms of bias & risk in favor of the proposed estimator increases as the parameter of the contaminating observations gets farther away from that of the main stream observations (i.e., as Δ becomes larger).

TABLE X
DIFFERENCES BETWEEN BIASES $\mathrm{Bias}(\hat{\theta}) - \mathrm{Bias}(\theta^*)$ & RISKS $R(\hat{\theta}) - R(\theta^*)$ FOR $n = 200$ & $\Delta = 3.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0113	.0014	.0249	.0048	.0438	.0137	.0481	.0204	.0617	.0315
$\alpha = .03$	.0107	.0010	.0351	.0069	.0419	.0119	.0641	.0235	.0692	.0369
$\alpha = .05$	.0117	.0012	.0462	.0094	.0510	.0144	.0664	.0238	.0785	.0387
$\alpha = .07$	.0162	.0018	.0412	.0092	.0575	.0185	.0738	.0265	.0852	.0474
$\alpha = .09$	.0166	.0018	.0378	.0067	.0540	.0129	.0852	.0340	.0885	.0421

TABLE XI
DIFFERENCES BETWEEN BIASES $\mathrm{Bias}(\hat{\theta}) - \mathrm{Bias}(\theta^*)$ & RISKS $R(\hat{\theta}) - R(\theta^*)$ FOR $n = 200$ & $\Delta = 5.0$

	$\varepsilon = .01$		$\varepsilon = .03$		$\varepsilon = .05$		$\varepsilon = .07$		$\varepsilon = .09$	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
$\alpha = .01$	.0233	.0028	.0696	.0212	.1014	.0498	.1255	.0787	.1535	.1299
$\alpha = .03$	.0304	.0038	.0743	.0218	.1200	.0564	.1536	.0927	.1765	.1637
$\alpha = .05$	.0358	.0058	.0903	.0218	.1363	.0641	.1677	.1017	.2034	.1596
$\alpha = .07$	.0420	.0058	.0931	.0257	.1376	.0617	.1912	.1017	.2045	.1584
$\alpha = .09$	.0326	.0042	.0871	.0216	.1465	.0588	.1684	.0916	.2159	.1677

V. CONCLUSION

The estimation of the parameter of an exponential distribution under contamination (i.e., in the presence of outliers) is quite relevant to reliability, and various robust estimation strategies have been suggested in the recent past. However, there is no one best estimation strategy for all situations. In this paper we suggested a robust weighted likelihood estimator (WLE) for this problem. We examined the risk (a measure combining bias & variability) of the estimator as compared to the usual maximum likelihood estimator (MLE), and illustrated through approximate formulae & simulations that WLE performs better than the MLE under various contamination scenario.

The methodology of the paper, and the estimator proposed, can be extended to the case of censored reliability data. Suppose that the observed data is subject to independent Type I right-censoring. Therefore the observed data comes in the form of pairs (X_i, δ_i), where $X_i = \min(\mathrm{T_i}, \mathrm{C_i})$, T_i is the failure time, C_i is the censoring time, and $\delta_i = 1$ if $X_i = T_i$ or zero otherwise. Here, the MLE of θ is $\hat{\theta}_{nc} = n\bar{X}/d$, where $d = \sum_i \delta_i$, $\bar{X} = \sum_i X_i/n$ are, respectively, the number of failed items, and average of the observed times.

Defining $C = a/\hat{\theta}_{nc}^{\delta_i}$, and following the arguments in Section II, the criteria for rejection of extreme observations would be $X_i > \hat{\theta}_{nc} ln(\alpha/n)$. Although the proof may be more involving, it is very plausible that the optimality & robustness properties noticed for the estimator of this paper will also be true for its counterpart for the censored failure data. Other types of censoring can be handled in a similar way.

A further interesting generalization of the suggested methodology can be made to the case of the two-parameter Weibull distribution. See [2] for detailed treatment of the Weibull distribution & related issues. The criteria for a Weibull (β, δ) would be to reject any observation such that $\log x^{(\delta-1)} - (x/\beta)^\delta > (1/C)\beta^\delta/\delta$. One may expand this expression, and isolate x or neglect either the term $\log x^{(\delta-1)}$ or $(x/\beta)^\delta$ depending on the values of $\delta - 1$, and then proceed as in the paper to obtain the robustness level, and finally replace δ & β by their MLE. The robust estimators so obtained should then be compared to other estimators of the Weibull parameters such as moment, least squares, probability weighted moment, and maximum likelihood estimators (see for example [12]–[14]. Models that are modifications of the Weibull distribution, and their censored versions, are also fertile areas to which the methods of the paper can be extended [15]–[17].

APPENDIX A
PROOF OF THEOREM 1

For the asymptotic bias of the estimator θ_n^*, we have

$$\begin{aligned} E\theta_n^* &\sim E_\alpha \overline{Y} \\ &= \frac{\theta}{G_\alpha}\left[1+\varepsilon\Delta-(1-\varepsilon)\left(\frac{A}{\theta}+1\right)\exp\left\{-\frac{A}{\theta}\right\}\right. \\ &\qquad \left. -\varepsilon(1+\Delta)\left(\frac{A}{\theta(1+\Delta)}+1\right)\exp\left\{-\frac{A}{\theta(1+\Delta)}\right\}\right] \\ &\sim \frac{\theta}{G_\alpha}\left[1+\varepsilon\Delta-(1-\varepsilon)\frac{\alpha}{n}\left((1+\varepsilon\Delta)\ln\frac{n}{\alpha}+1\right)\right. \\ &\qquad \left. -\varepsilon(1+\Delta)\left(\frac{\alpha}{n}\right)^{\frac{1}{(1+\Delta)}}\left(\frac{1+\varepsilon\Delta}{1+\Delta}\ln\frac{n}{\alpha}+1\right)\right] \\ &= \theta\left(1+\frac{\lambda\Delta}{n}-\frac{\alpha}{n}\ln\frac{n}{\alpha}\right) \end{aligned}$$

with the remainder term $O(n^{-(2+\Delta)/(1+\Delta)}\ln n)$

The MLE $\hat{\theta}_n$ has mathematical expectation $\theta(1+\varepsilon\Delta)$, and the weighted likelihood estimator θ_n^* has mathematical expectation

$$E\theta_n^* = \theta\left(1+\varepsilon\Delta-\frac{\alpha}{n}\ln\frac{n}{\alpha}+O\left(\varepsilon\frac{\ln n}{n}\right)\right).$$

Hence, the gain in bias has the order $(\alpha/n)\ln(n/\alpha)$.

APPENDIX B
PROOF OF THEOREM 2

The proof is based on the following two elementary integrals

$$\int x\mathrm{e}^{ax}dx = \mathrm{e}^{ax}\left(\frac{x}{a}-\frac{1}{a^2}\right)+c$$

and

$$\int x^2\mathrm{e}^{ax}dx = \mathrm{e}^{ax}\left(\frac{x^2}{a}-\frac{2x}{a^2}+\frac{2}{a^3}\right)+c.$$

We mainly use the case $a=-1$.

The quadratic risk of $\hat{\theta}_n$ is

$$\begin{aligned} R(\hat{\theta}_n) &= \frac{E(X_1-\theta)^2}{n} \\ &= \frac{E\Big(X_1-\theta(1+\varepsilon\Delta)\Big)^2+\Big(\theta-\theta(1+\varepsilon\Delta)\Big)^2}{n} \\ &= \frac{\theta^2}{n}\left(Var\left(\frac{X}{\theta}\right)+\varepsilon^2\Delta^2\right). \end{aligned}$$

Direct computations show that the second moment of the normalized random variable X/θ with distribution G equals $2(1+\varepsilon\Delta(2+\Delta))$. Hence

$$Var\left(\frac{X}{\theta}\right) = 1+2\varepsilon\Delta(1+\Delta)-\varepsilon^2\Delta^2.$$

Finally,

$$R(\hat{\theta}_n) = \frac{\theta^2}{n}\left[1+2\varepsilon\Delta(1+\Delta)\right]$$

which proves (3).

We have

$$\begin{aligned} R(\theta_n^*) &\sim E_\alpha\left[\frac{1}{n(1+\alpha)}\sum_{k=1}^{n(1-\alpha)}(Y_k-\theta)\right]^2 \\ &= \frac{1}{n(1-\alpha)}Var_\alpha(Y)+(\mu-\theta)^2 \end{aligned}$$

where

$$\mu = E_\alpha Y = \theta\left(1+\varepsilon\Delta-\frac{\alpha}{n}\ln\frac{n}{\alpha}\right).$$

Further, $Var_\alpha(Y)=E_\alpha(Y^2)-\mu^2$, so

$$R(\theta_n^*) \sim \frac{1}{n(1-\alpha)}\Big[E_\alpha(Y^2)-\mu^2\Big]+(\mu-\theta)^2.$$

With some algebra, one can show that

$$\begin{aligned} E_\alpha(Y^2) &= \frac{\theta^2}{G_\alpha}\Big\{2\big(1+\varepsilon\Delta(2+\Delta)\big)\Big\} \\ &\quad -\frac{\theta^2}{G_\alpha}\int_{\frac{A}{\theta}}^{\infty}x^2\left[(1-\varepsilon)e^{-x}+\frac{\varepsilon}{(1+\Delta)}e^{-\frac{x}{(1+\Delta)}}\right]dx \\ &\sim \theta^2\left[2\big(1+\varepsilon\Delta(2+\Delta)\big)-\frac{\alpha}{n}\ln^2\frac{n}{\alpha}-2\frac{\alpha}{n}\ln\frac{n}{\alpha}\right] \end{aligned}$$

with a remainder term of the order $O(n^{-(2+\Delta)/(1+\Delta)}\ln^2 n)$. Hence we have established that

$$R(\theta_n^*) \sim \frac{\theta^2}{n(1-\alpha)}\left[1+2\varepsilon\Delta(1+\Delta)-\frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right]$$

which completes the proof of (4).

The asymptotic relative efficiency of the estimator θ_n^* to the estimator $\hat{\theta}_n$ is

$$\begin{aligned} ARE_n &= \frac{R(\hat{\theta}_n)}{R(\theta_n^*)} \\ &\sim (1-\alpha)\frac{1+2\varepsilon\Delta(1+\Delta)}{1+2\varepsilon\Delta(1+\Delta)-\frac{\alpha}{n}\ln^2\frac{n}{\alpha}} \\ &\sim (1-\alpha)\Big[1+2\varepsilon\Delta(1+\Delta)\Big] \\ &\qquad \times\left[1-2\varepsilon\Delta(1+\Delta)+\frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right] \\ &\sim (1-\alpha)\left(1+\frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right). \end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank the associate editor and anonymous referees for their constructive comments.

REFERENCES

[1] W. R. Blischke and D. N. P. Murthy, *Reliability: Modeling, Prediction and Optimization*: Wiley, 2000.
[2] D. N. P. Murthy, M. Xie, and R. Jiang, *Weibull Models*: Wiley, 2004.
[3] F. Hu, "Symptotic properties of relevance weighted likelihood estimations," *Canad. J. Statist.*, vol. 25, pp. 45–60, 1997.
[4] R. Tibshirani and T. Hastie, "Local likelihood estimation," *J. Amer. Statist. Assoc.*, vol. 82, pp. 559–567, 1987.
[5] J. G. Staniswalis, "The kernel estimate of a regression function in likelihood-based models," *J. Amer. Statist. Assoc.*, vol. 84, pp. 276–283, 1989.
[6] J. Fan, N. E. Heckman, and W. P. Wand, "Local polynomial kernel regression for generalized linear models and quasilikelihood functions," *J. Amer. Statist. Assoc.*, vol. 90, pp. 826–838, 1995.
[7] F. Hu and J. V. Zidek, *The Relevance Weighted Likelihood With Applications*, S. E. Ahmed and N. Reid, Eds: Springer Verlag, 2001, vol. 25, Empirical Bayes and Likelihood Inference.
[8] ——, "The weighted likelihood," *Canad. J. Statist.*, vol. 30, pp. 347–372, 2002.
[9] D. J. Dupuis and S. Morgenthaler, "Robust weighted likelihood estimators with an application to bivariate extreme value problems," *Canad. J. Statist.*, vol. 30, pp. 17–36, 2002.
[10] R. Jiang and D. N. P. Murthy, "Graphical representation of two mixed Weibull distributions," *IEEE Trans. Reliab.*, vol. 44, pp. 477–488, 1995.
[11] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[12] J. H. Gove, "Moment and maximum likelihood estimators for Weibull distributions under length- and area-biased sampling," *Environmental and Ecological Statistics*, vol. 10, no. 4, pp. 455–467, 2003.
[13] S. Mahdi and F. Ashkar, "Exploring generalized probability weighted moments, generalized moments and maximum likelihood estimating methods in two-parameter Weibull model," *Journal of Hydrology*, vol. 285, pp. 62–75, 2004.
[14] K. R. Skinner, J. B. Keats, and W. J. Zimmer, "A comparison of three estimators of the Weibull parameters," *Qual. Reliab. Eng. Int.*, vol. 17, no. 4, pp. 249–256, 2001.
[15] C. D. Lai, M. Xie, and D. N. P. Murthy, "A modified weibull distribution," *IEEE Trans. Reliab.*, vol. 52, no. 1, pp. 33–37, 2003.
[16] M. Xie, Y. Tang, and T. N. Goh, "A modified weibull extension with bathtub-shaped failure rate function," *Reliab. Eng. Syst. Saf.*, vol. 76, no. 3, pp. 279–285, 2002.
[17] U. Balasooriya and C. K. Low, "Competing causes of failure and reliability tests for weibull lifetimes under type i progressive censoring," *IEEE Trans. Reliab.*, vol. 53, no. 1, pp. 29–36, 2004.

Ejaz S. Ahmed is Professor and Head of the Department of Mathematics and Statistics and Director of the Statistical Consulting, Research and Services Centre at the University of Windsor. Before joining the University of Windsor, he was Professor and Head of the Department of Mathematics and Statistics at University of Regina, and had a faculty position at the University of Western Ontario. His area of expertise includes statistical inference, multivariate analysis and asymptotic theory. He has published numerous articles in scientific journals, both collaborative and methodological. Further, he co-edited two volumes and organized many sessions and workshops. He made several invited scholarly presentations in various countries and supervised graduate students both at masters' and Ph.D. level. Dr. Ahmed serves on the editorial board of many statistical journals.

Andrei I. Volodin (B.S. in Mathematics, Kazan University, Russia, 1983; Kandidant Nouk in Mathematics and Physics (Russian analog of Ph.D.)(1999) Vilnius University (Lithuania); Ph.D. in Statistics, University of Regina, Canada, 2002) is an assistant professor in Statistics at the University of Regina, Regina, Saskatchewan, Canada. His research topics are: Bootstrap estimation; Shrinkage estimation; Limit theorems; Distributions on abstract spaces.

Abdulkadir A. Hussein (B.S. in Applied Mathematics, University of Trieste, Italy, 1996; M.Sc. (1999) and Ph.D. (2003) in Statistics, University of Alberta, Edmonton, Canada) is an assistant professor in Statistics at the University of Windsor in Windsor, Ontario, Canada. His research topics are: Techniques for the analyzes of censored data and their applications in medical and industrial problems. Change-point and Sequential analyzes; Health outcomes and policy research.