Check for updates

A Weighted Likelihood Approach to Problems in Survival Data

Adhidev Biswas and Ayanendranath Basu
Indian Statistical Institute, Kolkata, India
Suman Majumder
North Carolina State University, Raleigh, USA
Pratim Guha Niyogi
Michigan State University, East Lansing, USA

Abstract

This work is motivated by the need to perform the appropriate "robust" analysis on right-censored survival data. As in other domains of application, modelling and analysis of data generated by medical and biological studies are often unstable due to the presence of outliers and model misspecification. Use of robust techniques is helpful in this respect, and has often been the default in such situations. However, a large contaminating set of observations can often mean that the group is generated systematically by a model which is different from the one to which the majority of the data are attributed, rather than being stray outliers. The method of weighted likelihood estimating equations might provide a solution to this problem, where the different roots obtained can indicate the presence of distinct parametric clusters, rather than providing a single robust fit which ignores the observations incompatible with the major fitted component. Efron's (J. Am. Stat. Assoc. **83**, 402, 414–425, 1988) head-and-neck cancer data provide an ideal scenario for the application of such a method. A recently developed variant of the weighted likelihood method provides a nice illustration of the presence of different clusters in Efron's data, and highlights the benefits of the weighted likelihood method in relation to classical robust techniques.

AMS (2000) *subject classification.* Primary 62F35; Secondary 62N02.
Keywords and phrases. Censored survival data, Kaplan-Meier estimator, Root selection, Weibull distribution, Weighted likelihood estimation

1 Introduction

Analyzing survival data is a problem of immense practical interest. Survival data naturally occur in diverse medical research areas such as epidemiology, cancer research, biomedical experiments, etc., as well as in various

fields such as industrial life testing, criminology, engineering, sociology, marketing and insurance. Motivated by the need to provide appropriate analysis of right-censored survival data involving a large proportion of outliers, the present paper develops a robust method of analysis for such right-censored survival data. In particular, we have attempted to provide a neat analysis of Efron's head-and-neck cancer data set, Arm A (Efron, 1988) which exhibits several right-censored observations as well as an appreciable proportion of large outlying observations.

The application of robust statistics comes into play in such situations where one or more standard assumptions of parametric inference fail to hold. Robust procedures protect the analysis from degenerating into useless numbers under the presence of a group of observations which are incompatible with the rest of the data under the assumed model. Often robust procedures provide a nice fit to the majority of the observations in the sample while completely sacrificing a smaller but discordant minority. The application of robust techniques is increasingly becoming more popular as the need for stability under model misspecification and data contamination is becoming more and more real in the age of big data.

A robust method, however, generally tends to provide a single meaningful fit to the given data on the basis of the assumed model. Observations which stand out subsequent to the model fit may be termed as outliers. Yet it is possible that the group of outliers that are discarded by the robust model fit may themselves form a separate homogeneous cluster which may be appropriately described by a different model and is, sometimes, also of interest to us. Thus, apart from simply looking at a single robust fit to the entire data set, our aim, at times, is to uncover the different segments of the data set in question which clearly appears to be generated by different models. The primary aim of this paper is to demonstrate that our method can be useful when there are different groups in the data.

Efron's head-and-neck cancer (Arm A) data set (Efron, 1988) is an obvious and prominent example of the type of situations we want to handle. It has a small but non-negligible proportion of moderately extreme outliers as well as a small proportion of censored observations. In addition, the largest observation in this case is an actual failure (and not a censored value). Thus we have a well defined and complete empirical estimate of the theoretical survival curve through the Kaplan-Meier product-limit estimator (Kaplan and Meier, 1958).

The rest of the paper is organized as follows. In Section 2 we describe the Arm A data set for the head-and-neck cancer study considered by Efron (1988). Section 3 is devoted to the description of the concept of the weighted

likelihood estimation procedure developed in Biswas et al. (2015) and Majumder et al. (2016). Necessary adjustments via the Kaplan-Meier product-limit estimator have been done in Section 4 in order to accommodate censored observations in the analysis. Section 5 contains necessary theoretical results for the weighted likelihood estimator under the right-censored survival set up. The statistical analysis of Efron's Arm A data set has been presented in Section 6. Some simulation results have been provided for exponential and Weibull distributions in Section 7. A tuning parameter selection technique is described in Section 8. Finally, Section 9 has some concluding remarks. All the numerical computations have been performed using the R programming language.

It is also pertinent to mention here that what we have demonstrated in this paper is an application of the robust and efficient weighted likelihood methodology that is general and not specific to this particular variant (Biswas et al., 2015; Majumder et al., 2016) of the weighted likelihood method alone. This inference procedure would work equally well with the weighted likelihood approach of Markatou et al. (1997) and Markatou et al. (1998). In addition, our method will work perfectly well as the robust estimator of choice when there is a single and prominent majority component in the data with the other observations being stray outliers, and a single robust fit is desired. We primarily emphasize the property of identifying discordant model clusters, rather than the robustness issue per se, because we want to highlight its novelty with relation to the former.

2 Efron's Head-and-Neck Cancer Data

Efron (1988) considered two batches of head-and-neck cancer patients for the experiment conducted by the Northern California Oncology Group. The two batches, denoted Arm A and Arm B, differ in the medication administered to them. The first batch of patients was treated only with radiation therapy whereas the second batch received both radiation therapy and chemotherapy. Both batches have some large outlying observations as well as a few censored observations. We wish to study such data, particularly that in Arm A, in the present exposition. The survival times (in days) recorded for the patients in Arm A are provided in Table 1 ("+" indicates a censored observation).

The data in Arm B can also be studied by the method presented here. However, in case of Arm B, the last observation is censored (and is not an actual failure), so we do not have a complete estimate of the survival curve through the Kaplan-Meier approach, and some artificial method of completing it has to be adopted. We plan to tackle such cases in a sequel paper.

Table 1: Efron's Head-and-Neck Cancer Data (Arm A, in Days)

7	108	149	218	405	1116+
34	112	154	225	417	1146
42	129	157	241	420	1226+
63	133	160	248	440	1349+
64	133	160	273	523	1412+
74+	139	165	277	523+	1417
83	140	173	279+	583	
84	140	176	297	594	
94	146	185+	319+	1101	

A quick look at the data, as well as the kernel density plots of Figs. 1 and 2 appear to indicate there are two distinct clusters in the data, one constituted by the first forty four observations, the other by the last seven. We will make use of this fact later in this paper.

3   Weighted Likelihood Estimation

The likelihood is normally defined as the product of the density values at the data points under the assumed model. All the observations in the sample get uniform treatment in the construction of the likelihood. Maximizing the
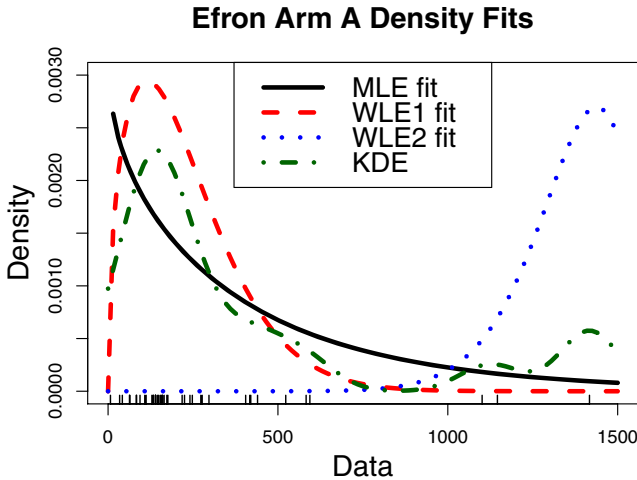

Figure 1: Fitted Weibull densities including those corresponding to the MLE, WLE1 = $(1.4651, 251.4444)$ and WLE2 = $(10.5922, 1451.7040)$) for Efron's data (Arm A), together with a kernel density estimate (KDE). The forty two actual failure times are indicated in the base of the plot
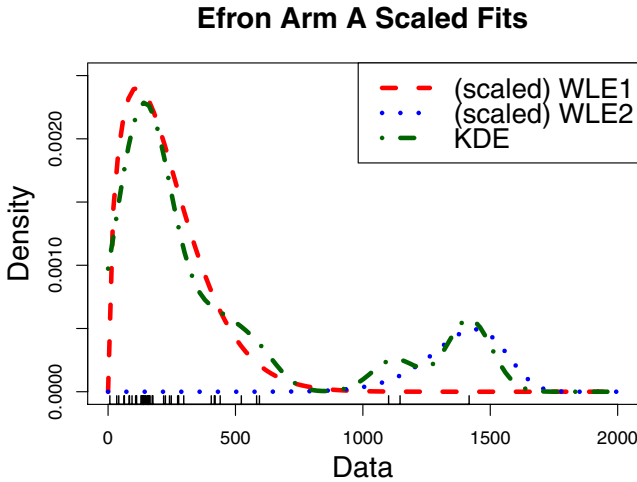
Efron Arm A Scaled Fits

Figure 2: Fitted Weibull densities (WLE1 $= (1.4651, 251.4444)$, WLE2 $= (10.5922, 1451.7040)$) scaled by the proportion of weights attributed to each for Efron's Arm A data, together with a kernel density estimate

likelihood with respect to the parameter, therefore, leads to an estimator in which contributions from all the sample points get the same, uniform weight. Similarly, in the maximum likelihood score equation, the score function is the ordinary sum of the scores of each observation. In case observations incompatible with respect to the model are present in the sample, there may be a considerable amount of influence of these points on the estimation process which may often lead to inaccurate inference. Consequently, the usefulness of the maximum likelihood estimator (MLE) is often questionable in such cases.

One of the objectives of robust statistical procedures in these situations is to develop an estimator on which the impact of the outliers should not be as high as their impact on the MLE. At the same time, the asymptotic efficiency of the estimator under the true model should not be unduly low in comparison with the MLE. Weighting the likelihood so that the observations incompatible with the model get lower weights and do not significantly affect the estimation process is one such strategy.

A variety of weighted likelihood estimation methods have been developed in the literature. In the weighted likelihood strategy proposed by Biswas et al. (2015), we obtain an estimator which is simultaneously robust and asymptotically fully efficient. The idea and form of this method closely resembles that of Markatou et al. (1998), which is another approach which could have

produced similar results in this context. It has been shown in Majumder et al. (2016) that the Biswas et al. estimator enjoys many important theoretical properties (as also does the Markatou et al., 1998, method), e.g., full asymptotic efficiency at the model and location-scale equivariance. Instead of the density function, the sample ($F_n(\cdot)$ or $S_n(\cdot)$) and model ($F_\theta(\cdot)$ or $S_\theta(\cdot)$) cumulative distribution functions (CDF) and survival functions have been employed in Biswas et al. (2015). The idea is to construct a residual function $\tau_{n,\theta}(\cdot)$, similar to the Pearson residual (Lindsay, 1994), which is defined by

$$
\tau_{n,\theta}(x) = \begin{cases} \dfrac{F_n(x)}{F_\theta(x)} - 1, & \text{if } F_\theta(x) \leqslant p, \\ 0, & \text{if } p < F_\theta(x) < 1 - p, \\ \dfrac{S_n(x)}{S_\theta(x)} - 1, & \text{if } F_\theta(x) \geqslant 1 - p, \end{cases} \tag{3.1}
$$

where $p \in (0, 0.5]$ is a tuning parameter determining the proportion of observations to be downweighted. So, essentially we are keeping the contribution of the sample points coming from the central $100 \times (1 - 2p)\%$ area of the model distribution intact. Observations representing the tails of the distribution are possible candidates for downweighting. But we will downweight them only if they appear to be incompatible with the model in relation to the rest of the data. Based on the residual in Eq. (3.1), the observations are weighted by a suitable function $H(\cdot)$. The function $H(\cdot)$ is chosen to be a nonnegative unimodal function, with the maximum value 1 when the residual $\tau_{n,\theta}(x)$ equals zero; the tails decrease smoothly on both sides to downweight the effect of discordant observations. In particular, the function $H(\cdot)$ drops to zero as its argument runs away to (positive) infinity. Detailed construction of the weight function $H(\cdot)$ along with some examples has been provided in Majumder et al. (2016). One such example of H-function obtained as suitable ratios of the gamma density $g_{\alpha,\beta}$ is given by

$$
H_{\text{Gamma}}(x) = \frac{g_{\alpha,\beta}(x+1)}{g_{\alpha,\beta}(1)} = e^{-(\beta-1)x}(x+1)^{\beta-1}, \tag{3.2}
$$

where $\beta > 1$ is the shape parameter and $\alpha = 1/(\beta-1)$ is the scale parameter necessary for the function $H_{\text{Gamma}}(\cdot)$ to have the right properties. Another choice for H may be on the basis of the Weibull density $h_{\alpha,\beta}$, with shape parameter β and scale parameter $\alpha = (1 - 1/\beta)^{-1/\beta}$, $\beta > 1$, which leads to the weight function

$$
H_{\text{Weibull}}(x) = \frac{h_{\alpha,\beta}(x+1)}{h_{\alpha,\beta}(1)} = (x+1)^{\beta-1} \exp\left[-(\beta-1)\left\{\frac{(x+1)^\beta - 1}{\beta}\right\}\right]. \tag{3.3}
$$

Based on the above formulation, the proposed weight function is then defined by

$$w_\theta(x) = H(\tau_{n,\theta}(x)). \tag{3.4}$$

We now introduce the basic mathematical set up for the subsequent weighted likelihood analyses. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed (i.i.d.) observations from an unknown distribution, which we model by the family $\mathcal{F}_\Theta = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^d\}$. The aim is to estimate the unknown parameter θ that best fits the data at hand. The maximum likelihood estimator of θ is obtained by solving the estimating equation

$$\sum_{i=1}^n u_\theta(X_i) = 0, \tag{3.5}$$

where $u_\theta(x) = \frac{\partial}{\partial \theta} \log f_\theta(x)$ is the likelihood score function. In weighted likelihood estimation, the estimator is the solution to

$$\sum_{i=1}^n w_\theta(X_i) u_\theta(X_i) = 0, \tag{3.6}$$

where the weight $w_\theta(\cdot)$ is as in Eq. (3.4). This strategy downweights outlying observations, but does so adaptively, so that it does not necessarily downweight a proportion of the observations. This provides an intuitive explanation of why the estimators are simultaneously robust and asymptotically fully efficient.

4 Weighted Likelihood Estimation for Right-Censored Data

In the earlier section, we described the weighted likelihood strategy developed for i.i.d. data. In this section we are going to describe the appropriate extension under which it may be used in the survival analysis situation to accommodate the censored observations. The precise mathematical set up is as follows.

Let $T_1, T_2, \ldots, T_n$ represent the lifetimes of a group of individuals (or the failure times of a certain brand of products) obeying the law $F_\theta(\cdot)$ (with density $f_\theta(\cdot)$ and survival function $S_\theta(\cdot)$). Letting T denote the lifetime variable, F_θ and S_θ are defined as

$$F_\theta(t) = \mathbb{P}_\theta(T \leqslant t) \ \text{ and } \ S_\theta(t) = \mathbb{P}_\theta(T \geqslant t).$$

It may not always be possible to monitor all the items till their actual failure times; as a result, some censoring is often introduced in the sampling scheme.

Here we assume that there is an independent censoring mechanism and the censoring time for each item is randomly governed by the distribution $F_C(\cdot)$ (with density $f_C(\cdot)$ and survival function $S_C(\cdot)$). Let $C_1, C_2, \ldots, C_n$ be the corresponding censoring times for the items under consideration. In practice, we have information about whether a particular item is censored and the corresponding time. In other words, what we know about the data set are the variables

$$X_i = \min\{T_i, C_i\} \text{ and } \Delta_i = \mathbf{1}_{\{T_i \leqslant C_i\}}; \quad i \in \{1, 2, \ldots, n\}.$$

The latter variable is the indicator of the event $\{T_i \leqslant C_i\}$. For a particular realization, we shall consider the lower case versions, i.e., the pair (x_i, δ_i) will denote the ith observation.

Next we propose the weighted likelihood strategy by defining the residual function. In the survival analysis scenario, we consider the Kaplan-Meier estimate of the survival function, defined by

$$\hat{S}_{KM}(x) = \begin{cases} \prod\limits_{i:X_{(i)}\leqslant x} \left(\dfrac{n-i}{n-i+1}\right)^{\delta_{[i]}}, & \text{if } x \leqslant X_{(n)}, \\ 0, & \text{if } \delta_{[n]} = 1 \text{ for } x > X_{(n)}, \\ \text{undefined}, & \text{if } \delta_{[n]} = 0 \text{ for } x > X_{(n)}, \end{cases}$$

where $X_{(i)}$, $i \in \{1, 2, \ldots, n\}$, are the order statistics of $(X_1, X_2, \ldots, X_n)$ and $\delta_{[i]}$ is the ith concomitant corresponding to $X_{(i)}$, $i \in \{1, 2, \ldots, n\}$ (i.e., $\delta_{[i]}$ is the indicator corresponding to X_j, where X_j is the i-th order statistic). In case the largest value is a censored observation (rather than being an actual failure), the survival curve remains incomplete, and artificial methods are required to complete them. We will not bring that situation within the ambit of the present discussion. However, in case of the Arm A data set of Efron (1988), the final observation is indeed an uncensored observation. So the product-limit estimator does not need any modification in this case.

Notice that in this definition $\hat{S}_{KM}(x)$ is the estimated probability $\mathbb{P}(T > x)$. For example, if the first observation is a failure at x, then $\hat{S}_{KM}(x) = (n-1)/n$. On the other hand, $\hat{F}_{KM}(x) = 1 - \hat{S}_{KM}(x)$ is an estimate of the probability $\mathbb{P}(T \leqslant x)$. In order to get the estimated probabilities corresponding to $\mathbb{P}(T \geqslant x)$ we actually use $\hat{S}^*_{KM}(x)$ in our subsequent analysis, obtained by adding the Kaplan-Meier mass, equal to the jump of the original

Kaplan-Meier curve at that point, to each failure point x. This leads us to the definition of the modified residual function as

$$\tau_{KM,\theta}(x) = \begin{cases} \dfrac{\hat{F}_{KM}(x)}{F_\theta(x)} - 1, & \text{if } F_\theta(x) \leqslant p, \\ 0, & \text{if } p < F_\theta(x) < 1 - p, \\ \dfrac{\hat{S}^*_{KM}(x)}{S_\theta(x)} - 1, & \text{if } F_\theta(x) \geqslant 1 - p. \end{cases} \qquad (4.1)$$

Having set up the above concepts, we now proceed to define the estimating equation in the right-censored, independent censoring scenario. The likelihood function is given by

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_\theta^{\delta_i}(x_i) S_\theta^{1-\delta_i}(x_i).$$

We further consider the log-likelihood function in order to define the score equation and this leads to the expression

$$\begin{aligned} \ell(\theta) \equiv \ell(\theta|\boldsymbol{x}, \boldsymbol{\delta}) &= \sum_{i=1}^{n} \{\delta_i \log f_\theta(x_i) + (1 - \delta_i) \log S_\theta(x_i)\} \\ &\quad + \sum_{i=1}^{n} \{\delta_i \log S_C(x_i) + (1 - \delta_i) \log f_C(x_i)\}. \end{aligned} \qquad (4.2)$$

We obtain the maximum likelihood estimator of θ by maximizing this log-likelihood with respect to $\theta \in \Theta$ or, equivalently, solving the estimating equation given by

$$\nabla\ell(\theta) = \sum_{i=1}^{n} \{\delta_i u_{1,\theta}(x_i) + (1 - \delta_i) u_{2,\theta}(x_i)\} = \sum_{i=1}^{n} U_\theta(x_i, \delta_i) = 0, \qquad (4.3)$$

where ∇ represents derivative with respect to θ, $u_{1,\theta}(x) = \nabla \log f_\theta(x)$, $u_{2,\theta}(x) = \nabla \log S_\theta(x)$ and

$$U_\theta(x, \delta) = \begin{cases} u_{1,\theta}(x), & \text{if } \delta = 1, \\ u_{2,\theta}(x), & \text{if } \delta = 0. \end{cases}$$

Hence, the proposed weighted likelihood estimating equation turns out to be

$$\sum_{i=1}^{n} w_\theta(X_i) U_\theta(X_i, \delta_i) = 0. \qquad (4.4)$$

The weighted likelihood estimator (WLE) of θ is a suitable root of the estimating Eq. (4.4).

5 Theoretical Properties of the Proposed Weighted Likelihood Estimator

In this paper, we have primarily focused on the methodology and application of the weighted likelihood method in the right-censoring scenario. The theoretical properties, viz., the influence function and the asymptotic distribution of the corresponding weighted likelihood estimators have been dealt with elsewhere by Biswas et al. (2018). Here, we briefly mention the results for the sake of completion and understanding. We first begin with the following important lemma.

Lemma 1. *The proposed weighted likelihood estimator is Fisher-consistent.*

Using the above lemma, we write down the expression of the influence function under the right-censored set up as below.

Theorem 1 (Influence Function of the Proposed Estimator)**.** *The influence function of the proposed weighted likelihood estimator is given by*

$$IF(y; T, G) \equiv \theta'_\epsilon\big|_{\epsilon=0} = D^{-1}N, \tag{5.1}$$

where

$$
\begin{aligned}
D &= \int_{\mathcal{X}_1} H'(\tau(x)) \frac{G(x)}{F_{\theta_g}(x)} \frac{\nabla F_{\theta_g}(x)}{F_{\theta_g}(x)} u_{1,\theta_g}(x) S_C(x) dG(x) \\
&+ \int_{\mathcal{X}_2} H'(\tau(x)) \frac{\bar{G}(x)}{S_{\theta_g}(x)} \frac{\nabla S_{\theta_g}(x)}{S_{\theta_g}(x)} u_{2,\theta_g}(x) \bar{G}(x) dF_C(x) \\
&- \int_{\mathcal{X}} H(\tau(x)) \nabla u_{1,\theta_g}(x) S_C(x) dG(x) - \int_{\mathcal{X}} H(\tau(x)) \nabla u_{2,\theta_g}(x) \bar{G}(x) dF_C(x),
\end{aligned}
\tag{5.2}
$$

and

$$
\begin{aligned}
N &= \int_{\mathcal{X}_1} H'(\tau(x)) \frac{\mathbf{1}_{\{x \geqslant y\}} - G(x)}{F_{\theta_g}(x)} u_{1,\theta_g}(x) S_C(x) dG(x) \\
&+ \int_{\mathcal{X}_2} H'(\tau(x)) \frac{\mathbf{1}_{\{x < y\}} - \bar{G}(x)}{S_{\theta_g}(x)} u_{1,\theta_g}(x) S_C(x) dG(x) \\
&+ H(\tau(y)) u_{1,\theta_g}(y) S_C(y) \\
&+ \int_{\mathcal{X}_1} H'(\tau(x)) \frac{\mathbf{1}_{\{x \geqslant y\}} - G(x)}{F_{\theta_g}(x)} u_{2,\theta_g}(x) \bar{G}(x) dF_C(x) \\
&+ \int_{\mathcal{X}_2} H'(\tau(x)) \frac{\mathbf{1}_{\{x < y\}} - \bar{G}(x)}{S_{\theta_g}(x)} u_{2,\theta_g}(x) \bar{G}(x) dF_C(x) \\
&+ \int_{\mathcal{X}} H(\tau(x)) u_{2,\theta_g}(x) \mathbf{1}_{\{x < y\}} dF_C(x),
\end{aligned}
\tag{5.3}
$$

where $\mathcal{X}_1 = \{x : F_\theta(x) \leqslant 1/2\}$, $\mathcal{X}_2 = \{x : F_\theta(x) > 1/2\}$, $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$, $\bar{G}(x) = \mathbb{P}(T > x)$ under the true distribution G, and θ_g is the solution of the weighted likelihood estimating Eq. (4.4) *under the true distribution G.*

If the true distribution G belongs to the model, i.e., $G = F_{\theta_0}$ for some $\theta_0 \in \Theta$, then the influence function turns out to have the simpler form

$$IF(y; T, F_{\theta_0}) = -\left[\int_{\mathcal{X}} \nabla u_{1,\theta_0}(x) S_C(x) dF_{\theta_0}(x) + \int_{\mathcal{X}} \nabla u_{2,\theta_0}(x) S_{\theta_0}(x) dF_C(x)\right]^{-1}$$
$$\times \left[u_{1,\theta_0}(y) S_C(y) + \int_{\mathcal{X}} u_{2,\theta_0}(x) \mathbf{1}_{\{x<y\}} dF_C(x)\right], \qquad (5.4)$$

which is nothing but the influence function of the maximum likelihood estimator.

To prove the asymptotic normality of the proposed weighted likelihood estimator under the right-censored scenario, we need the following assumptions. For simplicity the results are presented with a scalar parameter in mind; however they may be extended to the vector parameter case with appropriate generalizations in the notation.

(A1) The weight function $H(\tau)$ is non-negative, bounded and twice continuously differentiable such that $H(0) = 1$ and $H'(0) = 0$.

(A2) The functions $H'(\tau)(1 + \tau)$ and $H''(\tau)(1 + \tau)^2$ are bounded. Further, $H''(\tau)$ is continuous in τ.

(A3) For every $\theta_0 \in \Theta$, there is a neighbourhood $N(\theta_0)$ such that for every $\theta \in N(\theta_0)$, the quantities $\left|\tilde{U}_\theta(x)\nabla U_\theta(x, \delta)\right|$, $\left|\tilde{U}_\theta^2(x)U_\theta(x, \delta)\right|$, $\left|\nabla\tilde{U}_\theta(x)U_\theta(x,\delta)\right|$ and $|\nabla_2 U_\theta(x,\delta)|$ are bounded by $M_1(x), M_2(x), M_3(x)$ and $M_4(x)$ respectively, with $\mathbb{E}_{\theta_0}[M_i(X)] < \infty$ for all $i \in \{1, 2, 3, 4\}$.

(A4) $\mathbb{E}_{\theta_0}[\tilde{U}_\theta^2(X)U_\theta^2(X, \Delta)] < \infty$ and $\mathbb{E}_{\theta_0}[\nabla U_{\theta_0}^2(X, \Delta)] < \infty$.

(A5) The analogue of Fisher information in the survival context, $J(\theta) = \mathbb{E}_\theta[U_\theta^2(X, \Delta)]$ is non-zero and finite for all $\theta \in \Theta$.

(A6) $S_C(x) \geqslant S_{\theta_0}^\epsilon(x)$, for some $\epsilon > 0$.

Here, the function $\tilde{U}_\theta$ is defined as

$$\tilde{U}_\theta(x) = \begin{cases} \frac{\nabla F_\theta(x)}{F_\theta(x)}, & \text{if } F_\theta(x) \leqslant \frac{1}{2}, \\ \frac{\nabla S_\theta(x)}{S_\theta(x)}, & \text{if } F_\theta(x) > \frac{1}{2}. \end{cases}$$

We now present the main result in the following theorem.

Theorem 2. *Let the true distribution G of the failure time data belong to the model, with θ_0 being the true parameter value. Let $\hat{\theta}_{n,WLE}$ be the proposed weighted likelihood estimator. Define*

$$A_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} H(\tau_{KM,\theta}(X_i)) U_\theta(X_i, \Delta_i),$$

$$B_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \nabla \left(H(\tau_{KM,\theta}(X_i)) U_\theta(X_i, \Delta_i) \right), \text{ and}$$

$$C_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \nabla_2 \left(H(\tau_{KM,\theta}(X_i)) U_\theta(X_i, \Delta_i) \right),$$

where ∇_2 represents the second derivative with respect to θ. Then, under Assumptions (A1)–(A6) listed above, the following results hold.

(a) *$\sqrt{n} \left| A_n(\theta_0) - \frac{1}{n} \sum_{i=1}^{n} U_{\theta_0}(X_i, \Delta_i) \right| = o_P(1)$.*

(b) *$\left| B_n(\theta_0) - \frac{1}{n} \sum_{i=1}^{n} \nabla U_{\theta_0}(X_i, \Delta_i) \right| = o_P(1)$.*

(c) *$C_n(\theta') = O_P(1)$, where θ' lies on the line segment joining θ_0 and $\hat{\theta}_{n,WLE}$.*

Based on the above theorem, we present two very important corollaries in the following.

Corollary 1. *There exists a sequence $\{\hat{\theta}_{n,WLE}\}_{n\in\mathbb{N}}$ of roots of the weighted likelihood estimating* Eq. (4.4) *such that*

$$\hat{\theta}_{n,WLE} \xrightarrow{P} \theta_0.$$

Corollary 2. *$\sqrt{n}(\hat{\theta}_{n,WLE} - \theta_0) \xrightarrow{\mathcal{D}} N(0, J^{-1}(\theta_0))$.*

6 Analysis of Efron's Data

This section is devoted to the statistical analyses of the Arm A data set considered by Efron (1988). While monitoring 51 of the patients from Arm A, it was found that 9 (about 18%) of the values were censored. In order to analyze these data, we will apply the proposed weighted likelihood method under the Weibull distribution model.

The Weibull density, with shape parameter b and scale parameter a, is given by

$$f_{b,a}(x) = \left(\frac{b}{a} \right) \left(\frac{x}{a} \right)^{b-1} \exp\left[-\left(\frac{x}{a} \right)^b \right], \quad x > 0, \ a, b > 0. \tag{6.1}$$

We have also considered the Weibull model (and some other models) for a brief simulation study in Section 7. The Eqs. (7.1) and (7.2) in Section 7.2 have to be iteratively solved to obtain the WLEs of a and b. In the present section, the weight function in Eq. (3.4) is chosen based on the gamma density function as defined by H_{Gamma} in Eq. (3.2), with the shape parameter β of the gamma density being the tuning constant. Throughout the present analysis and in the subsequent simulation studies as well, the tuning parameter p of the weighting scheme is fixed at 0.5. The resulting estimates of a and b (rounded up to 4th decimal places), for various choice of the initial values, are presented in Table 2.

The MLE of the parameter vector (b, a) for the fifty one observations of the Arm A data under the Weibull model is $(0.9301, 426.9934)$. A closer look at the WLEs in Arm A for varying tuning parameters reveals that there is an increasing pattern in the values of the shape parameter as the tuning constant β increases. This is true for all the initial values, viz., the MLE, the pair $(1.44, 257.33)$ and the pair $(9.30, 1427.04)$. The last two pairs are the MLEs of the first forty four observations and the last seven observations, respectively. Also, in general the roots are fairly close to the initial values in each case.

The final fitted weights, corresponding to all the observations for each of the three initial values are provided in Table 3 corresponding to the tuning parameter $\beta = 1.05$. The column 'status' indicates the censoring status, with '0' indicating a censored observation. It may be observed that when the MLEs represent the set of initial values all the final fitted weights are close to 1 and the root is an MLE-like root. With the second set of initial values, the final fitted weights for the last seven observations are all equal to zero (rounded up to the fourth decimal place) although the first forty four observations get weights close to 1. The situation reverses for the third set of initial values, where all the first forty four observations are now sacrificed.

In Fig. 1, we plot the densities based on the MLE and the WLEs $(1.4651, 251.4444)$ and $(10.5922, 1451.7040)$. We also report a kernel density estimate for the Kaplan-Meier estimator (Wand and Jones, 1994) in the same plot which uses the Kaplan-Meier masses of the forty two fully observed failure times in its construction. For the kernel density estimate, we use the bandwidth

$$h = 0.9 \times \min\{\text{sd}(\text{Data}), \text{IQR}(\text{Data})/1.34\} \times n^{-0.2},$$

where $\text{sd}(\cdot)$ and $\text{IQR}(\cdot)$ refer to the standard deviation and interquartile range of the data respectively and n is the sample size.

Table 2: WLEs with different initial estimates for arm A data

Tuning Constant (β)	Shape (b)	Scale (a)
Initial estimate:	($b = 0.93, a = 426.99$)	
1.01	0.9307	426.4139
1.02	0.9314	425.8148
1.03	0.9321	425.1946
1.04	0.9327	424.5519
1.05	0.9335	423.8850
1.06	0.9342	423.1919
1.07	0.9350	422.4705
1.08	0.9358	421.7183
1.09	0.9366	420.9324
1.10	0.9375	420.1097
Initial estimate:	($b = 1.44, a = 257.33$)	
1.01	1.4461	256.4093
1.02	1.4507	255.2861
1.03	1.4550	254.1302
1.04	1.4597	252.8597
1.05	1.4651	251.4444
1.06	1.4712	249.8387
1.07	1.4784	247.9677
1.08	1.4875	245.6899
1.09	1.4999	242.6716
1.10	1.5232	237.4208
Initial estimate:	($b = 9.30, a = 1427.04$)	
1.010	9.4070	1430.5239
1.015	9.4694	1432.3907
1.020	9.5405	1434.3601
1.025	9.6228	1436.4542
1.030	9.7202	1438.7053
1.035	9.8390	1441.1647
1.040	9.9913	1443.9242
1.045	10.2053	1447.1866
1.050	10.5922	1451.7040

While Fig. 1 is useful in demonstrating that the two components WLE1 = $(1.4651, 251.4444)$ and WLE2 = $(10.5922, 1451.7040)$ represent two different segments of the data, these densities have to be appropriately scaled to highlight the degree of their match with the different patterns in the data.

Table 3: Final fitted weights corresponding to the three different roots for the Arm A observations

Observations	Status	MLE-like Weights	WLE1 Weights	WLE2 Weights
7	1	0.9998	0.9316	0.0000
34	1	0.9866	0.9982	0.0000
42	1	0.9922	0.9993	0.0000
63	1	0.9906	0.9956	0.0000
64	1	0.9952	0.9985	0.0000
74	0	0.9927	0.9957	0.0000
83	1	0.9945	0.9962	0.0000
84	1	0.9971	0.9983	0.0000
94	1	0.9977	0.9981	0.0000
108	1	0.9978	0.9973	0.0000
112	1	0.9987	0.9982	0.0000
129	1	0.9985	0.9971	0.0000
133	1	0.9997	0.9988	0.0000
133	1	0.9997	0.9988	0.0000
139	1	0.9999	0.9990	0.0000
140	1	0.9999	0.9998	0.0000
140	1	0.9999	0.9998	0.0000
146	1	0.9998	0.9999	0.0000
149	1	0.9995	1.0000	0.0000
154	1	0.9993	1.0000	0.0000
157	1	0.9989	1.0000	0.0000
160	1	0.9978	0.9998	0.0000
160	1	0.9978	0.9998	0.0000
165	1	0.9975	0.9998	0.0000
173	1	0.9974	0.9998	0.0000
176	1	0.9968	0.9997	0.0000
185	0	0.9975	0.9999	0.0000
218	1	0.9987	0.9996	0.0000
225	1	0.9984	0.9996	0.0000
241	1	0.9986	0.9993	0.0000
248	1	0.9983	0.9994	0.0000
273	1	0.9987	0.9982	0.0000
277	1	0.9984	0.9986	0.0000
279	0	0.9985	0.9991	0.0000
297	1	0.9984	0.9976	0.0000
319	0	0.9983	0.9959	0.0000

Table 3: (continued)

Observations	Status	MLE-like Weights	WLE1 Weights	WLE2 Weights
405	1	0.9998	0.9671	0.0000
417	1	0.9996	0.9655	0.0000
420	1	0.9991	0.9696	0.0000
440	1	0.9988	0.9641	0.0000
523	1	0.9996	0.8934	0.0000
523	0	0.9996	0.8934	0.0000
583	1	0.9998	0.8071	0.0000
594	1	0.9991	0.8128	0.0000
1101	1	0.9823	0.0000	0.5377
1116	0	0.9882	0.0000	0.5933
1146	1	0.9854	0.0000	0.6794
1226	0	0.9881	0.0000	0.8630
1349	0	0.9743	0.0000	0.9749
1412	0	0.9646	0.0000	0.9708
1417	1	0.9638	0.0000	0.9718

To facilitate this we consider the scaled densities corresponding to WLE1 and WLE2 in Fig. 2. To do this we multiply these two densities with the sum of the Kaplan-Meier masses for the observations getting non-zero weights in the corresponding fit. Thus we multiply the WLE1 density with the sum of the masses of the thirty nine failure times among the first forty four observations (the other five observations are censored); similarly the WLE2 density is scaled with the sum of the masses of the three failure times among the last seven observations. For future reference, we refer to these two scaling constants as γ and $1 - \gamma$. In Fig. 2 we plot the kernel density estimate of the full data, as well as the two scaled densities. It is clearly seen that the scaled densities provide nice and close fits to the kernel density estimate in the two segments of the data. The peaks of both the components are very closely matched.

Since the fitted curves (or the scaled curves) corresponding to WLE1 and WLE2 are associated with Weibull densities, they are technically supported over the entire range $[0, \infty)$. For all practical purposes, however, the densities f_1 and f_2 corresponding to WLE1 and WLE2, respectively, are disjoint, as is observed in Figs. 1 and 2. In Fig. 3 we plot the mixture $f_{\text{mix}} = \gamma f_1 + (1 - \gamma) f_2$ of the two Weibull densities (with γ being the sum of the masses of the first component as defined earlier) and the two component Weibull mixture fit for the full data as reported in Basu et al. (2006). It
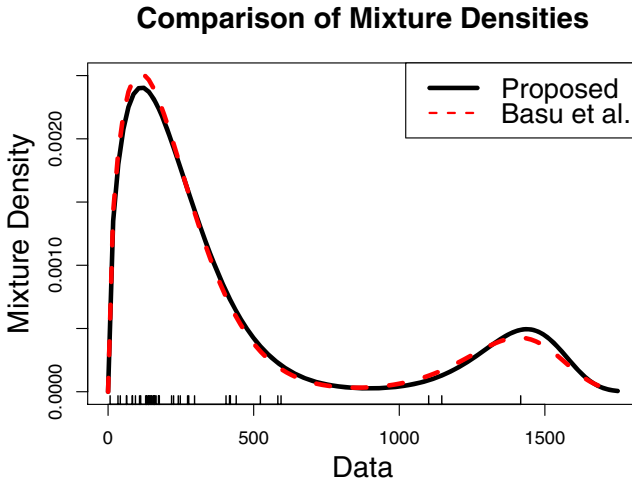
Comparison of Mixture Densities

Figure 3: Mixture of the two densities corresponding to WLE1 = $(1.4651, 251.4444)$ and WLE2 = $(10.5922, 1451.7040)$ for Efron's Arm A data, indicated as "Proposed" in the legend together with an estimated mixture density as reported in Basu et al. (2006)

is obvious that the two curves are extremely close to each other. Thus the mixture of the fits obtained by the two non MLE-like roots is entirely successful in replicating the fit of the mixture without getting into the (quite substantial) complications of Weibull mixture modelling in this case.

7   Simulation Results

To study the performance of the estimators in a little more detail, we take up a simulation study. In the following we compare the empirical efficiency, i.e., the ratio of the mean squared error (MSE) of the MLE to the MSE of the WLE for different levels of contamination and censoring. Various values of the tuning parameters are also explored in this investigation.

We consider the following two distributions prevalent in modelling survival data.

(i) The exponential distribution with rate λ, whose density is given by

$$f_\lambda(x) = \lambda e^{-\lambda x}, \quad \lambda > 0, x > 0.$$

(ii) The Weibull distribution as described in Eq. (6.1).

We want to study the performance of the method when the data are both contaminated and censored. We take up the two distributions separately.

In order to evaluate the empirical MSE for both the estimators (MLE and WLE), we draw 5000 samples with different levels (10% and 20%) of censoring and different proportions (0%, 10%, 15% and 20%) of contamination. It is to be noted that we have employed two different weighting functions for the definition of $H(\cdot)$ as discussed in Section 3 in the following analyses, viz., one based on the gamma density given by Eq. (3.2), and the other on the basis of the Weibull density as illustrated in Eq. (3.3). All the estimators corresponding to every sample are calculated for different values of the tuning constants.

7.1. *Exponential Distribution* For ease of description, let us denote the exponential distribution with rate parameter λ by $Exp(\lambda)$. The weighted likelihood estimating equation in this case boils down to

$$\sum_{i=1}^{n} w_\lambda(x_i) \left\{ \frac{\delta_i}{\lambda} - x_i \right\} = 0,$$

where $w_\lambda(x_i)$ is as described in Eq. (4.4) with $\lambda = \theta$, and this leads to the solution

$$\lambda = \sum_{i=1}^{n} w_\lambda(x_i)\delta_i \Big/ \sum_{i=1}^{n} w_\lambda(x_i)x_i.$$

First we simulate from the $Exp(5)$ distribution, with, say, ϵ proportion of contamination from $Exp(1.5)$. We also introduce censoring at the appropriate proportion in our system. Let the failure time distribution T be $Exp(\lambda)$; if the censoring variable C, independently of the failure time distribution, has distribution $Exp(\kappa)$, and if the censoring proportion is ϱ, then, an evaluation of the expression

$$\mathbb{P}[T > C] = \varrho$$

gives,

$$\kappa = \frac{5\varrho}{1 - \varrho}.$$

So, in case of 10 % censoring, we choose the rate parameter of the censoring distribution to be 5/9, and to be 5/4 for 20% censoring.

Tables 4 and 5 represent the empirical efficiencies of the WLE with respect to the MLE. As is evident from the tables and our theoretical intuitions, the efficiency is less than 1 at the 0% contamination, i.e., under pure data. The efficiency increases as the contamination increases, although the pattern with respect to increase in the tuning constant is not quite uniform. Under contamination, the efficiency decreases as we increase the censoring proportion. For each contamination level and weight function combination, the highest empirical efficiency is highlighted.

Table 4: Empirical efficiencies of WLE in the exponential model under 10% censoring and Exponential(1.5) contamination (rounded to 4th decimal places)

Sample size	Weight function	Tuning parameter (β)	Contamination proportion (ϵ)			
			0%	10%	15%	20%
50	Gamma	1.05	0.9077	1.1941	1.2660	1.3040
		1.07	0.8738	1.1949	1.2891	1.3488
		1.09	0.8382	1.1814	1.3032	1.3728
	Weibull	1.05	0.8974	1.1680	1.2435	1.3166
		1.07	0.8600	1.1615	1.2641	1.3607
		1.09	0.8061	1.1371	1.2590	1.3797

7.2. Weibull Distribution For streamlining the following discussion, the Weibull density is denoted by $W(b, a)$, where b is the shape parameter and a is the scale parameter. Here also, as before, we obtain the estimating equation for applying an algorithm similar to the *Iteratively Reweighted Estimating Equation* (IREE) algorithm (Basu and Lindsay, 2004). The equations turn out to be

$$a = \left(\sum_{i=1}^{n} w_{b,a}(x_i) x_i^b \Big/ \sum_{i=1}^{n} w_{b,a}(x_i)\delta_i \right)^{1/b} \tag{7.1}$$

$$1/b = \sum_{i=1}^{n} w_{b,a}(x_i) x_i^b \log x_i \Big/ \sum_{i=1}^{n} w_{b,a}(x_i) x_i^b$$
$$- \sum_{i=1}^{n} w_{b,a}(x_i)\delta_i \log x_i \Big/ \sum_{i=1}^{n} w_{b,a}(x_i)\delta_i. \tag{7.2}$$

Table 5: Empirical efficiencies of WLE in the exponential model under 20% censoring and Exponential(1.5) contamination (rounded to 4th decimal places)

Sample size	Weight function	Tuning parameter (β)	Contamination Proportion (ϵ)			
			0%	10%	15%	20%
50	Gamma	1.05	0.9433	1.0648	1.1044	1.1230
		1.07	0.9202	1.0687	1.1174	1.1470
		1.09	0.8986	1.0689	1.1275	1.1632
	Weibull	1.05	0.9403	1.0463	1.1025	1.1222
		1.07	0.9120	1.0406	1.1140	1.1400
		1.09	0.8800	1.0269	1.1169	1.1528

We simulate pure data from $W(5, 2)$ distribution, in which we impose ϵ proportion of contamination from $Exp(1.5)$ distribution. We choose $Exp(\lambda(\varrho))$ for censoring at proportion ϱ. As in the case of exponential distribution, we solve for $\lambda(\varrho)$ and, in respect of the original $W(5, 2)$ data generating distribution, it turns out to be the solution of

$$\int_0^\infty \exp[-\lambda(\varrho)t]f_{5,2}(t)dt = 1 - \varrho,$$

which we solve numerically in order to obtain

$$\lambda(0.1) = 0.0576 \text{ and } \lambda(0.2) = 0.1226.$$

Tables 6 and 7 display the empirical efficiencies. The increasing pattern of efficiency with increasing contamination remains the same. The other observations also are, in general, similar. The pure data efficiencies of the shape parameter appear to be lower than the scale parameter.

7.3. *Half Cauchy Contamination* The simulation study presented in the previous subsections involves the exponential distribution as the contaminating distribution in both exponential and Weibull models. As advised by a referee, we have further investigated the performance of the weighted likelihood estimators of the parameters in these two models when the contaminating

Table 6: Empirical efficiencies of WLE in Weibull distribution under 10% censoring and Exponential(1.5) contamination (rounded to 4th decimal places)

Parameter	Weight function	Tuning parameter (β)	Contamination Proportion (ϵ)			
			0%	10%	15%	20%
Scale, $\hat{a}$	Gamma	1.02	0.9901	2.7412	2.9967	2.5686
		1.04	0.9673	3.1635	4.1441	3.9781
		1.06	0.9417	3.2746	4.7763	5.2933
	Weibull	1.02	0.9781	2.7924	3.1949	2.5547
		1.04	0.9440	3.1803	4.5717	4.1709
		1.06	0.9068	3.2659	5.2205	5.6464
Shape, $\hat{b}$	Gamma	1.02	0.8645	3.5462	3.5754	2.8468
		1.04	0.7404	3.4737	4.2011	3.6922
		1.06	0.5884	2.8384	3.7441	3.5647
	Weibull	1.02	0.8569	3.6834	3.7681	2.8458
		1.04	0.7079	3.5309	4.4880	4.0968
		1.06	0.5999	2.6944	3.2690	3.6214

Table 7: Empirical efficiencies of WLE in the Weibull model under 20% censoring and Exponential(1.5) contamination (rounded to 4th decimal places)

Parameter	Weight function	Tuning parameter (β)	Contamination Proportion (ϵ)			
			0%	10%	15%	20%
Scale, $\hat{a}$	Gamma	1.02	0.9916	2.4845	2.6651	2.2426
		1.04	0.9693	2.8743	3.6104	3.3720
		1.06	0.9398	2.9984	4.2144	4.4185
	Weibull	1.02	0.9843	2.5537	2.7848	2.2792
		1.04	0.9516	2.9008	3.9751	3.5978
		1.06	0.9107	3.0230	4.6290	4.7146
Shape, $\hat{b}$	Gamma	1.02	0.8927	3.4033	3.2426	2.5637
		1.04	0.7596	3.4173	3.7808	3.4275
		1.06	0.6078	2.7205	3.3235	3.2253
	Weibull	1.02	0.9470	3.4298	3.3442	2.6088
		1.04	0.7865	3.3025	3.7743	3.6628
		1.06	0.6231	2.4417	2.9703	3.1255

distribution is the one-sided Cauchy distribution with parameters (μ, σ) with the probability density function given by

$$f_{\mathrm{HC}}(x) = \begin{cases} \frac{2}{\pi} \times \frac{\sigma}{\sigma^2 + (x-\mu)^2}, & \text{if } x \geqslant 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (7.3)$$

We undergo the same simulation study under this modified contamination. To make the distribution comparable with the exponential contamination, we choose $\mu = 0$ and $\sigma = 1/\pi$. The results are summarized in Tables 8, 9, 10 and 11. The patterns are, on the whole, similar to the exponential contamination patterns.

8   Selection of the Tuning Parameter

In the analysis of the Efron data set in Section 6 and the simulation studies in Section 7 we employ both the Gamma and Weibull type weights with a variety of tuning parameters for the evaluation of the weighted likelihood estimates. While it is our experience that the behaviour of the estimator only differs marginally over the choice of the weight function type (Gamma, Weibull or other) with comparable tuning parameter, the tuning parameter itself has a much more pronounced impact on the estimate. In general, the finite sample efficiency of the estimator appears to decrease with increase in the value of the tuning parameter β. On the other hand, larger values of β

Table 8: Empirical efficiencies of WLE in the exponential model under 10% censoring and one-sided Cauchy contamination (rounded to 4th decimal places)

Sample size	Weight function	Tuning parameter (β)	Contamination Proportion (ϵ)			
			0%	10%	15%	20%
50	Gamma	1.05	0.9077	1.4444	1.6966	1.8214
		1.07	0.8738	1.4152	1.6999	1.8566
		1.09	0.8382	1.3629	1.6789	1.8753
	Weibull	1.05	0.8974	1.4078	1.6880	1.7987
		1.07	0.8600	1.3669	1.6673	1.8127
		1.09	0.8061	1.2739	1.6270	1.7956

appear to give greater stability under data contamination. It is, therefore, extremely important to choose the tuning parameter judiciously. One would want to choose a larger tuning parameter for stability in estimation only if it is necessary.

In connection with choosing the optimal tuning parameter for the minimum divergence estimator based on the density power divergence (DPD; Basu et al., 1998), an algorithm was presented by Warwick and Jones (2005) which minimizes an empirical measure of mean squared error (MSE); also see Basak et al. (2019). In calculating this empirical measure of the MSE for any given real data set, the Warwick-Jones algorithm exploits the theoretical form of the asymptotic variance of the estimator, the empirical distribution, and a robust pilot estimator for the computation of the bias. Then the

Table 9: Empirical efficiencies of WLE in the exponential model under 20% censoring and one-sided Cauchy contamination (rounded to 4th decimal places)

Sample size	Weight function	Tuning parameter (β)	Contamination Proportion (ϵ)			
			0%	10%	15%	20%
50	Gamma	1.05	0.9433	1.0970	1.1796	1.2415
		1.07	0.9202	1.0785	1.1887	1.2632
		1.09	0.8986	1.0664	1.1903	1.2700
	Weibull	1.05	0.9403	1.0632	1.1837	1.2482
		1.07	0.9120	1.0403	1.1869	1.2657
		1.09	0.8800	1.0176	1.1850	1.2611

Table 10: Empirical efficiencies of WLE in Weibull distribution under 10% censoring and one-sided Cauchy contamination (rounded to 4th decimal places)

Parameter	Weight function	Tuning parameter (β)	Contamination Proportion (ϵ)			
			0%	10%	15%	20%
Scale, $\hat{a}$	Gamma	1.02	0.9901	2.4646	2.2750	1.9754
		1.04	0.9673	2.6619	1.8223	1.7754
		1.06	0.9417	2.6730	2.4601	1.8126
	Weibull	1.02	0.9781	2.8714	1.8715	1.8639
		1.04	0.9440	3.1737	2.1991	2.2983
		1.06	0.9068	3.3561	2.2354	2.7687
Shape, $\hat{b}$	Gamma	1.02	0.8645	2.2395	2.4204	2.2422
		1.04	0.7404	2.2117	2.6783	2.8180
		1.06	0.5884	2.0142	2.4713	2.7552
	Weibull	1.02	0.8569	2.3023	2.5497	2.2093
		1.04	0.7079	2.2303	2.7087	2.8254
		1.06	0.5999	1.9355	2.4134	2.6988

empirical mean squared error is the estimated, summed, asymptotic variance plus the estimated squared bias computed against an initial robust pilot estimator.

Table 11: Empirical efficiencies of WLE in the Weibull model under 20% censoring and one-sided Cauchy contamination (rounded to 4th decimal places)

Parameter	Weight Function	Tuning Parameter (β)	Contamination Proportion (ϵ)			
			0%	10%	15%	20%
Scale, $\hat{a}$	Gamma	1.02	0.9916	2.2236	2.5992	1.9965
		1.04	0.9693	2.3809	3.2115	2.6055
		1.06	0.9398	2.4264	3.2515	2.8774
	Weibull	1.02	0.9843	2.0491	2.2575	1.9254
		1.04	0.9516	1.9800	3.2729	2.6881
		1.06	0.9107	1.9964	3.5663	3.1690
Shape, $\hat{b}$	Gamma	1.02	0.8927	2.3704	2.3613	2.1060
		1.04	0.7596	2.3507	2.6031	2.6705
		1.06	0.6078	2.0727	2.3915	2.6221
	Weibull	1.02	0.9470	2.2939	2.5130	2.0903
		1.04	0.7865	2.2096	2.7467	2.6910
		1.06	0.6231	1.9252	2.4213	2.4860

In this paper, we propose to extend this idea to make it suitable for our situation. Notice that unlike the case of the DPD, where the asymptotic variance was a function of the tuning parameter α, our estimators are all first-order efficient, and hence there is no asymptotic discrimination between the estimators in terms of the variance. The finite sample variance varies quite a bit, however, and we try to make use of this fact in the development of our algorithm, which proceeds along the following steps for any of our weight functions which depends on a single parameter β, when a real data set of size n is available.

1. We choose a fine sequence of values of the tuning parameter β over the appropriate range.

2. We select B bootstrap samples of the same size n from the given data set.

3. We compute the weighted likelihood estimators of the parameters for each bootstrap sample and each value of β in our sequence.

4. For each value of β, the variance of the estimator is calculated as the observed variance of the parameter estimates over the bootstrap samples. This is following the idea of Léger and Romano (1990) to consider the bootstrap estimate of the variance.

5. We compute the empirical mean squared error at any given value of β as the sum of the estimated bootstrap variances at that β and the squared bias estimated as the squared deviation of the estimator for that β against a suitable pilot estimator.

6. The optimal value of the tuning parameter is then chosen as the value of β which minimizes the above criterion.

In case of the Efron data (Arm A), we illustrate our tuning parameter selection technique with Gamma weights. Here we have used a fine sequence of β values between 1.00 and 1.05. We restrict the upper limit of the tuning parameter at 1.05 because for larger values of β the pure data efficiency of the WLEs appears to be unacceptably low, particularly for the shape parameter estimate. Note that the sample size (51) for Efron data (Arm A) is approximately equal to the sample size (50) used in our simulations as reported in Tables 4 to 11. The optimal value of β through the above exercise is found to be 1.05 for the Efron (Arm A) data, indicating a strongly robust estimator is needed in this case. Interestingly, if we were to exclude

the seven large observations of this data set (i.e., the observations belonging to the second cluster), the optimal value of the tuning parameter turns out to be 1.00 which generates the MLE of this truncated group. In both cases, the pilot estimator was the WLE at $\beta = 1.05$ with Gamma weights. Thus, at least in this example, the algorithm suggested by us appears to be successful in detecting where a robust estimator is necessary and where it is not. A further refinement of the above procedure is possible which selects the tuning parameter adaptively by letting the optimal solution at any stage be the pilot for the next iteration; see Basak et al. (2019). We hope to pursue such a procedure in the future.

9   Concluding Remarks

This article has been devoted to analyzing right-censored survival data by adapting a weighted likelihood approach. Studies of various theoretical properties of this newly devised estimator have been provided in Biswas et al. (2018). While obtaining the numerical findings, it has been noticed that the convergence of the IREE type algorithm (Basu and Lindsay, 2004) heavily depends upon the choice of the starting values, and, to a lesser extent, on the value of the tuning parameter. A suitable robust estimator should be employed for the quick convergence of the IREE type algorithm.

The scope of this method has been illustrated, apart from simulation, through Efron's (Efron, 1988) head-and-neck cancer data. In Efron's data (Arm A), the two clusters in the data were easily separable, and, consequently, the roots representing these clusters could be easily identified. In general, we need some algorithm to uncover all the roots of the system. For this we recommend the bootstrap root search algorithm as described in Markatou et al. (1998).

The problem of early failure in reliability theory is a common one and seems an area worth exploring for the purpose of applying robust statistical methods. In this situation, failure of some items may occur at a very primary period of the experiment. These observations, we suspect, are exceptional when considering the appropriate statistical model. We wish to take up the problem of analyzing data with early failures in a subsequent work.

Another possible approach of study, which could be useful for the situation considered in this paper, was suggested by a referee. This would involve performing an initial cluster analysis, followed by an application of the weighted likelihood method of estimation for each of the identified clusters. This would, however, require a robust clustering technique appropriate for right-censored data. We hope to take it up in a sequel paper.

References

BASAK, S, BASU, A and JONES, MC (2019). On the "optimal" density power divergence tuning parameter. Tech. Rep. ISRU/2019/1, Interdisciplinary Statistical Research Unit, Indian Statistical Institute. https://www.isical.ac.in/isru/tr19a.pdf.

BASU, A. and LINDSAY, B.G. (2004). The iteratively reweighted estimating equation in minimum distance problems. *Computational Statistics & Data Analysis* **45**, 2, 105–124.

BASU, A., HARRIS, I.R., HJORT, N.L. and JONES, M.C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85**, 3, 549–559.

BASU, S., BASU, A. and JONES, M.C. (2006). Robust and efficient parametric estimation for censored survival data. *Ann. Inst. Stat. Math.* **58**, 2, 341–355.

BISWAS, A., ROY, T., MAJUMDER, S. and BASU, A. (2015). A new weighted likelihood approach. *Stat* **4**, 1, 97–107.

BISWAS, A., GUHA NIYOGI, P., MAJUMDER, S., BHANDARI, S.K., GHOSH, A. and BASU, A. (2018). Theoretical properties of a new weighted likelihood estimator for right censored data. Tech. Rep. ISRU/2018/2, Interdisciplinary Statistical Research Unit, Indian Statistical Institute. https://www.isical.ac.in/isru/tr18a.pdf.

EFRON, B. (1988). Logistic regression, survival analysis, and the kaplan-Meier curve. *J. Am. Stat. Assoc.* **83**, 402, 414–425.

KAPLAN, E.L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 282, 457–481.

LÉGER, C. and ROMANO, J.P. (1990). Bootstrap choice of tuning parameters. *Ann. Inst. Stat. Math.* **42**, 4, 709–735.

LINDSAY, B.G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Stat.* **22**, 2, 1081–1114.

MAJUMDER, S, BISWAS, A, ROY, T, BHANDARI, S and BASU, A (2016). Statistical inference based on a new weighted likelihood approach. arXiv:161007949.

MARKATOU, M., BASU, A. and LINDSAY, B.G. (1997). Weighted likelihood estimating equations: the discrete case with applications to logistic regression. *Journal of Statistical Planning and Inference* **57**, 2, 215–232.

MARKATOU, M., BASU, A. and LINDSAY, B.G. (1998). Weighted likelihood equations with bootstrap root search. *J. Am. Stat. Assoc.* **93**, 442, 740–750.

WAND, M.P. and JONES, M.C. (1994). Kernel Smoothing. CRC Press.

WARWICK, J. and JONES, M.C. (2005). Choosing a robustness tuning parameter. *J. Stat. Comput. Simul.* **75**, 7, 581–588.

ADHIDEV BISWAS
AND AYANENDRANATH BASU
INTERDISCIPLINARY STATISTICAL RESEARCH
UNIT, INDIAN STATISTICAL INSTITUTE,
203 B.T. ROAD, KOLKATA, PIN - 700108,
INDIA
E-mail: ayanbasu@isical.ac.in

SUMAN MAJUMDER
DEPARTMENT OF STATISTICS,
NORTH CAROLINA STATE UNIVERSITY,
2311 STINSON DRIVE, RALEIGH,
NC 27695, USA

PRATIM GUHA NIYOGI
DEPARTMENT OF STATISTICS
AND PROBABILITY, MICHIGAN STATE
UNIVERSITY, 619 RED CEDAR ROAD,
EAST LANSING, MI 48824, USA