Robust Fitting of a Weibull Model with Optional Censoring

Jingjing Yang\textsuperscript{a,*} and David W. Scott\textsuperscript{a}
\textsuperscript{a}Rice University, Department of Statistics, 6100 Main St. MS-138, Houston, TX 77005, U.S.A.

Abstract

The Weibull family is widely used to model failure data, or lifetime data, although the classical two-parameter Weibull distribution is limited to positive data and monotone failure rate. The parameters of the Weibull model are commonly obtained by maximum likelihood estimation; however, it is well-known that this estimator is not robust when dealing with contaminated data. A new robust procedure is introduced to fit a Weibull model by using $L_2$ distance, i.e. integrated square distance, of the Weibull probability density function. The Weibull model is augmented with a weight parameter to robustly deal with contaminated data. Results comparing a maximum likelihood estimator with an $L_2$ estimator are given in this article, based on both simulated and real data sets. It is shown that this new $L_2$ parametric estimation method is more robust and does a better job than maximum likelihood in the newly proposed Weibull model when data are contaminated. The same preference for $L_2$ distance criterion and the new Weibull model also happens for right-censored data with contamination.

Keywords
Weibull distribution; $L_2$ distance; Robust estimator; Maximum likelihood; Right-censored data; Contamination

1. Introduction

Failure data, or survival data, are usually collected during research about the safe limit of using some component or system, survival time of some medical therapy, and product lifetime in business. Researchers are often interested in finding a threshold to ensure safety and to model failure probabilities. The standard two-parameter Weibull distribution with cumulative density function (CDF): $F(t; \alpha, \beta) = 1 - \exp\left(-\frac{t}{\alpha}\right)^\beta, \ t \geq 0; \ \alpha, \beta > 0$, is one of the commonly used families for modeling such data, even though it is limited to positive data values ($t > 0$) and a monotone hazard function: $h(t) = \frac{dF(t)}{dt} = \frac{\beta t^{\beta-1}}{\alpha^\beta}$ (Weibull, 1951). The Weibull distribution is one of the stable distributions of extreme value theory and a member of the exponential family, so it has many nice properties. Therefore, it is an important practical problem to robustly and efficiently estimate the parameters of this distribution.

\*Corresponding author. Tel: +1 713 348 6032 (Office), +1 713 348 5476 (Fax). jy13@rice.edu (Jingjing Yang), scottdw@rice.edu (David W. Scott).

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
A number of extensions to the original Weibull distribution have appeared in the literature. For example, Smith & Naylor (1987) compared the maximum likelihood estimator (MLE) and Bayesian estimator for the three-parameter Weibull distribution with CDF \( F(t; \alpha, \beta, \theta) = 1 - \exp \left\{ -\left[ t - \theta \right]/\alpha \right\}^{\beta}, t \geq \theta; \alpha, \beta > 0, \theta \in \mathbb{R} \), based on two samples of fibre strength data. Due to the high correlation among the three parameters \((\alpha, \beta, \theta)\), we have reason to question the accuracy of this model. In addition, the third parameter \(\theta\) in this extended Weibull distribution only shifts data origin and allows negative data in the model. Moreover, it does not play a meaningful role in the statistical modeling, because it neither provides useful nice properties nor helps to explain data characteristics. The issue of negative data can be solved by transforming the data to ensure nonnegativity before applying the model. Gurvich et al. (1997) proposed a different three-parameter Weibull distribution \( F(t; \alpha, \beta, \kappa) = 1 - \exp\{(t/\alpha)^\beta \exp(t/\kappa), t \geq 0; \alpha, \beta, \kappa > 0 \) in order to model the random strength of a brittle material; this parameterization can account for a wide range of relations regarding the strength-size effect.

In this article, we propose a new augmented Weibull model that retains the same scale and shape parameters, but has an extra weight parameter that can detect the percentage of contamination. This idea was first proposed by Scott (2004) for fitting the Gaussian mixture model. We investigate parameter estimation in the classical two-parameter and the new Weibull model by minimizing the \(L_2\) distance (Scott, 2001) between the PDFs. Comparisons of \(MLE\) and \(L_2\) estimator (\(L_2E\)) with several data sets are presented in this paper. The results from case studies on both simulated and real data indicate that the \(L_2E\) is indeed more robust and accurate with contaminated data. This nice behavior of \(L_2E\) in the new Weibull model is preserved as well when modeling right-censored data.

Here, we briefly review several other extensions of the Weibull distribution proposed since the 1970’s, some of which have proven very useful in modeling of complex lifetime data, especially survival data from biomedical studies. When the hazard rate function of the data is not monotone, some generalized Weibull distributions may be able to capture the main shape of the hazard function. More details about types of Weibull probability density functions and corresponding hazard functions are described by Lai et al. (2011), who also give suggestions about which Weibull distribution should be used to model data with a particular shaped hazard rate. For example, the exponentiated Weibull family (Mudholkar & Srivastava, 1993) is capable of modeling constant, decreasing, increasing, unimodal and bathtub shapes of failure rate, through different ranges of the parameter values. An application to the head-and-neck cancer censored data (Efron, 1988) was conducted by Mudholkar et al. (1996). In addition, Wahed & Jeong (2009) proposed another generalization of the Weibull distribution, the Beta-Weibull distribution, incorporating the exponentiated Weibull family as a special case.

Nikulin & Haghighi (2006) presented the power-generalized Weibull family and applied it to model the head-and-neck cancer censored data (Efron, 1988), resulting in an unimodal shaped hazard rate function and agreeing with the earlier analyses of Efron (1988) and Mudholkar et al. (1996). In this article, we reanalyze the head-and-neck cancer data set by our \(L_2\) method, and the results are promising. Furthermore, our method can be extended to any of the generalized three-parameter Weibull distributions discussed above, and will behave more robustly because of the properties of the \(L_2\) distance criterion.

This paper is organized as follows. First, the detailed Weibull models and \(L_2\) criterions are stated in Section 2, along with some preliminary examples. Section 3 briefly describes some nice properties of \(L_2E\). In Section 4, we compare \(MLE\) and \(L_2E\) in some case studies with uncensored data. Section 5 extends the model to deal with right-censored data. In Section 6 we present results about comparisons of Weibull modeling with \(MLE\) and \(L_2E\) on simulated data.
2. Weibull modeling with the $L_2$ distance criterion

2.1. MLE for the classical two-parameter Weibull model

Consider the classical or standard two-parameter Weibull model for a group of uncensored observations $t = \{t_1, \ldots, t_n\}$ with probability density function (PDF):

$$f(t; \alpha, \beta) = \frac{dF(t)}{dt} = \beta \alpha \beta^{-1} \exp \left( -\left( \frac{t}{\alpha} \right)^\beta \right), \quad t \geq 0; \quad \alpha, \beta > 0. \quad (1)$$

The log likelihood function for this data set is:

$$l(\alpha, \beta; t) = \sum_{i=1}^{n} \left[ \log(\beta) - \beta \log(\alpha) + (\beta - 1) \log(t) - \left( \frac{t_i}{\alpha} \right)^\beta \right]. \quad (2)$$

The MLE parameters $\hat{\alpha}, \hat{\beta}$ of this model can be easily found by numerical maximization of (2).

2.2. $L_2E$ for the classical two-parameter Weibull model

Let $(\alpha_0, \beta_0)$ and $(\alpha, \beta)$ denote the unknown true and possible parameter values, respectively. It is intuitive that we would like the PDF with estimated parameters $f(t; \hat{\alpha}, \hat{\beta})$ to be the closest one to the PDF with unknown true parameter values $f(t; \alpha_0, \beta_0)$ across the whole domain. The $L_2$ distance between those two PDFs, i.e. $\int_0^\infty \left[ f(t; \alpha, \beta) - f(t; \alpha_0, \beta_0) \right]^2 dt$, is considered as our criterion. Obviously, we do not know the true parameter values, but we can find the $L_2E$ parameters $\hat{\alpha}, \hat{\beta}$ by minimizing an estimate of this $L_2$ distance over all possible values of $(\alpha, \beta)$. The estimate of this $L_2$ distance can be written as:

$$\text{Est} \left\{ \int_0^\infty \left[ f(t; \alpha, \beta) - f(t; \alpha_0, \beta_0) \right]^2 dt \right\}$$

The second term in (4), we can estimate $E[f(T; \alpha, \beta)]$ by the sample average of $f(t; \alpha_0, \beta_0)$ for the given random sample:

$$\text{Est} \left\{ \int_0^\infty f(t; \alpha, \beta)^2 dt \right\} - 2 \text{Est} \left\{ E[f(T; \alpha, \beta)] \right\} \quad (3)$$

regarding the second term in (4), we can estimate $E[f(T; \alpha, \beta)]$ by the sample average of $f(t; \alpha_0, \beta_0)$ for the given random sample:

$$\text{Est} \left\{ E[f(T; \alpha, \beta)] \right\} = -\frac{1}{n} \sum_{i=0}^{n} f(t; \alpha, \beta) = -\frac{1}{n} \sum_{i=0}^{n} \beta \beta^{-1} \alpha^{-\beta} \exp \left[ -(t_i/\alpha)^\beta \right]. \quad (5)$$

For the classical two-parameter Weibull model, the first integration term in (3) has an analytical integrand expression under the condition $\beta > \frac{1}{2}$.
If $0 < \beta \leq \frac{1}{2}$, then $f(t; \alpha, \beta)$ is not square integrable, and $L_2E$ cannot be applied. However, this case is rarely seen with real data as the density has a dramatically monotone decreasing shape.

After plugging (6) and (5) into the minimization objective function (4), the $L_2E$ parameters $(\bar{\alpha}, \bar{\beta})$ can be found numerically. For the case studies in this paper, we used the function `nlminb` in the R language (R Core Team, 2013) for this purpose.

### 2.3. Mixture Weibull model for contaminated data

Real lifetime data often contain contaminated data for a variety of reasons. Therefore, we suppose that any group of samples is a mix of “good” and “bad” observations. Specifically, we assume that each observation is “good” with probability $(1 - w)$; good observations follow a standard Weibull PDF, $f(t; \alpha, \beta)$, while “bad” ones come from some other PDF, $f^*(t)$, whose parametric form may or may not be known. Equivalently, we could say that all observations are independent, identically distributed (i.i.d.) with the common mixture distribution:

$$f_{mix}(t; \alpha, \beta, w) = wf(t; \alpha, \beta) + (1 - w)f^*(t), \quad t \geq 0; \alpha, \beta > 0, w \in (0, 1).$$

(7)

If we have some knowledge about the contamination that can lead us to assume a proper model for $f^*(t; \theta)$, then we can fit the mixture model by the methods illustrated in Sections 2.1 and 2.2: maximizing the likelihood $\prod_{i=1}^{n} f_{mix}(t_i; \alpha, \beta, w, \theta)$, or minimizing the estimate of $L_2$ distance $\int_0^\infty \left[ f_{mix}(t; \alpha, \beta, w, \theta) - f_{mix}(t; \alpha_0, \beta_0, w_0, \theta_0) \right]^2 dt$.

To motivate our approach, we display a few examples from our simulation study in Section 4.1 comparing MLE and $L_2E$ estimates in three cases: data with no contamination, contamination that can be modeled explicitly, or contamination whose distribution is unknown. In Figure 1, we display results from all three cases. We generated two groups of data: an uncontaminated data set $T_1$ and a contaminated data set $T_2$. $T_1$ consisted of 900 data points from Weibull distribution $f(t; 4, 2)$. $T_2$ consisted of $T_1$ and 100 points of contamination data from a normal distribution $N(11, 1/2)$.

We fitted the classical two-parameter Weibull model to both datasets $T_1$ and $T_2$; see Figures 1(a) and 1(c), respectively. Clearly, both MLE and $L_2E$ provide very good estimates at the model, as in Figure 1(a). However, Figure 1(c) shows that $L_2E$ gives a more robust fit than MLE with contaminated data. Intuitively, a minimum distance criterion will try to reproduce the larger component, ignoring the smaller contamination cluster; Donoho & Liu (1989) showed this is a general property of all minimum distance criteria. On the other hand, if the correct model for the contamination is known, then Figure 1(b) illustrates that both MLE and $L_2E$ accurately fit the contaminated and uncontaminated data. We discuss Figure 1(d) in the next section.

### 2.4. New augmented Weibull model for contaminated data

In most cases we would not have enough knowledge about the contamination to make a parametric assumption about the density $f^*(t)$. We propose to fit the main (Weibull) part of the mixture model by minimizing the estimate of $L_2$ distance between $wf(t; \alpha, \beta)$ and the true
mixture density \( f_{\text{mix}}(t; \alpha_0, \beta_0, w_0) \), where \((\alpha_0, \beta_0, w_0)\) and \( f^*(t) \) are unknown. However, we make no attempt to model \( f^*(t) \). We have already seen in Figure 1(c) that \( L_2E \) with \( w = 1 \) provides a useful fit.

The \( L_2E \) parameters \((\tilde{\alpha}, \tilde{\beta}, \tilde{w})\) can be easily found implementing similar calculations as in Section 2.2. The new \( L_2E \) criterion with weight \( w \) can be expressed as follows:

\[
\int_0^\infty \left\{ w f(t; \alpha, \beta) - f_{\text{mix}}(t; \alpha_0, \beta_0, w_0) \right\}^2 dt
\]

\[
= w^2 \int_0^\infty f(t; \alpha, \beta)^2 dt - 2w \text{Est} \left\{ E \left[ f(T_{\text{mix}}; \alpha, \beta) \right] \right\} + \int_0^\infty f_{\text{mix}}(t; \alpha_0, \beta_0, w_0)^2 dt, \tag{8}
\]

where the first term is integrable and has an analytical integrand expression for \( \beta > \frac{1}{2} \); \( T_{\text{mix}} \) is a random variable with mixture density \( f_{\text{mix}}(t; \alpha_0, \beta_0, w_0) \), hence the expectation of the second term can be estimated by taking the sample average over the entire dataset; finally the third term is irrelevant to the minimization.

By adding the weight parameter \( w \), this new augmented Weibull model can detect contamination and estimate the proportion of contamination in the data set simultaneously by at least \((1 - \tilde{w})\). Depending on the degree of overlap between the Weibull density and the unspecified contamination density \( f^*(t) \), \( \tilde{w} \) will be biased upwards. We derive two benefits of this new model: (i) the shape parameter will not be inflated due to the requirement that the density integrates to one as in the two-parameter model; (ii) it also avoids the need to explicitly model the contaminated data information \( f^*(t) \), so the \( L_2E \) in this new augmented model should be more robust than other parametric estimators of the standard model. An example R code is given in Appendix A.

Of course, we can also attempt to fit the main (Weibull) component of the mixture model by maximizing \( \prod_{i=1}^n w f(t_i; \alpha, \beta) \), but this will result in the same estimates as the classical Weibull model since the \( MLE \) must always be 1.

In order to show the robustness of fitting the new Weibull model by \( L_2E \), we fitted the new model on the contaminated data set \( T_2 \). Figure 1(d) shows that fitting the new augmented Weibull model with \( L_2E \) provides the most robust fit for the contaminated data. Notice the fitted \( \tilde{w} = 0.914 \) is slightly biased upward from the true value of 0.900. However, the parameters \((\alpha, \beta)\) are clearly better estimated by allowing \( \tilde{w} < 1 \). The theoretical target values are derived in Appendix B, together with other asymptotic properties.

### 3. Properties of \( L_2E \)

Wojciechowski & Scott (1999) explained that both \( MLE \) and \( L_2E \) are special cases of the minimum density power divergence estimator (Basu et al., 1998), a particular family of M-estimators (Huber & Ronchetti, 2009). Basu et al. (1998) showed the minimum density power divergence estimator is asymptotically normal and consistent at the correct model. This family is indexed by a parameter \( \gamma \), which controls a trade-off between efficiency and robustness. \( MLE \) corresponds to \( \gamma = 0 \), and is most efficient if the model is correct and no contamination exists in the data. \( L_2E \) is the case when \( \gamma = 1 \), which is more robust than \( MLE \) but less efficient. Basu et al. (1998) also generally observed that values of \( \gamma < 1/4 \) provided sufficient robustness. Wojciechowski & Scott (1999) noted that computational considerations recommend the choice \( \gamma = 1 \) since the relevant integral \( \int f(t)^{1+\gamma} dt \) is more likely to have a closed form expression if the density has many parameters, for example, a mixture of normals, when \( \gamma = 1 \).
The robustness property of $L_2E$ can be understood through examining the influence functions (Huber & Ronchetti, 2009) for both criteria of maximum likelihood and minimum $L_2$ distance. Here we have two parameters and the influence functions have similar shapes as the partial derivatives of the objective function (or its logarithm) with respect to $\alpha$ and $\beta$, respectively. From Figure 2, it is easily observed that the influence functions for MLE are unbounded, while those for $L_2E$ are bounded. For the $L_2E$ criterion, the curves also redescend, which is a property sometimes specified for M-estimators, but is only implicit in the $L_2E$ formulation as a consequence of the Weibull parametric assumption. Therefore, outliers in the data, especially those from the region with large absolute maximum likelihood influence function values, would affect the MLE much more than $L_2E$. Consequently, $L_2E$ should be more robust when the data set is contaminated. In fact, Donoho & Liu (1989) have shown that robustness is a general property of any distance criterion.

There are many other robust estimators such as Huber’s P15, Andrew’s AMT, and Tukey’s Biweight, which are versions of M-estimators as well (Stigler, 1977). Usually, several steps of scaling iterations are required to find the values of these estimators. Another nice property of $L_2E$ is that the the robust scaling issues for the influence function are automatic and do not require iteration, because the analytical expression of the estimate of the $L_2$ distance criterion obviates the need for specification of the shape and scale of an influence function at each iteration (Scott, 2001).

4. Case studies with uncensored data

4.1. Simulation study with good and contaminated Weibull data

We repeated the classical and new augmented Weibull model fitting on simulated Weibull data in order to confirm the robustness of $L_2E$ and contamination detecting property of our new model. This simulation will also explore the efficiency of the MLE and $L_2E$.

The uncontaminated data $T_1$ consisted of 900 data points from $f(t;\alpha=4,\beta=2)$. The contaminated data $T_2$ here consisted of 1000 data points, of which $m_1 \sim \text{Binomial}(1000,0.9)$ points are generated from $f(t;\alpha=4,\beta=2)$ and $(1000 - m_1)$ data points are generated from a normal distribution $N(\mu=11, \sigma=1/2)$. We repeatedly generated $T_1$ & $T_2$, fitting both of the classical Weibull model and the new augmented Weibull model to $T_1$ and $T_2$ for 1000 times. Each optimization problem was solved numerically by the R function nlminb with the same corresponding starting value ($\alpha=4,\beta=2$) or ($\alpha=4,\beta=2, w=0.9$).

The mean together with standard deviations of those estimates are shown in Table 1. Although the standard deviations of $L_2Es$ are slightly larger than the standard deviations of MLEs, $L_2E$ in the new model detected most of the contamination (9.1% vs 10%) and gave more accurate estimates ($\hat{\alpha}=4.037, \hat{\beta}=1.973$) for the contaminated data; these agree with the values given in the final paragraph of Appendix B. The accuracy can also be examined by the root mean square error (RMSE) of each vector of estimates. For example, the RMSE for MLE ($T_2$) is 0.638, while the RMSE for $L_2Ew$ ($T_2$) is only 0.027.

4.2. Weibull model fitting on the Fibre strength data

Our first real case study involves fitting the Weibull model to two samples of fibre strength data (list in Appendix C), which were examined by Smith & Naylor (1987). The samples are experimental data of the strength of glass fibres of two lengths, 1.5 cm, and 15.0 cm, from the National Physical Laboratory in England.
First, we fitted the classical two-parameter Weibull model with both MLE and $L_2E$. Histograms for each of the two data sets and the estimated Weibull PDFs (MLE short dashed line, $L_2E$ solid line) are shown in Figure 3 (a) & (b). It is obvious that both estimates of $\alpha$ are approximately equal to 1.6 for sample 1 and 1.2 for sample 2, but the estimates of $\beta$ are different for each data set, especially for sample 1. Based on a preliminary inspection of the data, Smith & Naylor (1987) stated there are possible outliers in the lower end of the sample. Thus we can explain why the $L_2E$ PDFs from both data sets are slightly right shifted and higher peaked.

Second, we fitted the new Weibull model to both fibre sample data sets, since the results from previous model fitting implied the data sets are contaminated. As we can see from Figure 3 (a) & (b), the $L_2Ew$ (long dashed line) detects data contamination ($\hat{\alpha}_1<1$) and has different estimated values for $(\alpha, \beta)$, which we believe are closer to the true $(\alpha_0, \beta_0)$.

Finally, we note that since the $L_2E$ criterion is not convex, there may be more than one local minima. Using random starting values allows us to explore these alternative possible solutions. For each model fitted, we generated a number of random starting values, and the corresponding estimates were calculated numerically. Since no extreme estimates happened in our study, we took the mean of those estimates as our final estimate, whose values are shown in the legend of Figure 3 (a) & (b).

5. Extension of $L_2E$ for right-censored data

Methods described in Section 2 are conducted when the lifetime data are uncensored. However, in most biomedical studies, censoring occurs quite often. Thus we would like to extend the $L_2E$ for censored data. In this paper, we only consider right-censored data, the most common type. Suppose $t = \{t_1, \ldots, t_n\}$ is a lifetime data set, and $c = \{c_1, \ldots, c_n\}$, where $c_i$ denotes whether the $i^{th}$ corresponding observation $t_i$ is right-censored ($c_i = 1$) or not ($c_i = 0$).

The common maximum likelihood method for right-censored data is to maximize

$$
\prod_{i=1}^{n} \left[ f(t_i; \alpha, \beta) \right]^{1-c_i} \left[ 1 - F(t_i; \alpha, \beta) \right]^{c_i}. \tag{9}
$$

This criterion uses the product of likelihood function and survival probabilities as the objective likelihood to be maximized. Clearly, the censored data points are not as informative as the uncensored, but are useful in calibrating the size of the right tail of the PDF. What we present here is an extension of the $L_2E$ idea to the Weibull models for right-censored data, including censored by one fixed value (type I) or censored randomly (type II). The type II censoring means all data have one fixed censoring value but start at random times. The $L_2E$ will not only minimize the $L_2$ distance between the estimated Weibull PDF and the truth, but also minimize the $L_2$ distance between the estimated survival probability and the true one for each data point. This means we want the estimated Weibull PDF to be the one closest to the true PDF, while ensuring the survival probabilities corresponding to the censored data are also as close as possible to the truth.

5.1. $L_2$ criterion for type I censored data

Let $(\alpha_0, \beta_0)$ and $(\alpha, \beta)$ denote the unknown true and any feasible values of the parameters respectively. Let $t_c$ denote the fixed censoring value, i.e. all censored data values are equal to $t_c$ for Type I censoring. Then the unknown true survival probability for censored data is $p_0 = 1 - F(t_c; \alpha_0, \beta_0)$. The censoring indicators $(c_1, \ldots, c_n)$ can be thought as a sample i.i.d. with a Bernoulli distribution and true success (survival) probability $p_0$. The $L_2E$ criteria of
the Bernoulli part is to minimize an estimate of the $L_2$ distance between the probability mass function (PMF) of Bernoulli distribution with success probability $p = 1 - R(t_c; \alpha, \beta)$ and the truth. Since the PMF for Bernoulli distribution with success probability $p$ is $B(c; p) = cp + (1-c)(1-p)$, so an estimate of the $L_2$ distance can be expressed as:

$$\text{Est} \left( \sum_{c=0}^{1} \left[ B(c; p) - B(c; p_0) \right]^2 \right) = p^2 + (1-p)^2 - 2\text{Est} \left[ E \left[ B(C; p) \right] \right] + \sum_{c=0}^{1} B(c; p_0)^2 ,$$  

(10)

where $C$ denotes a Bernoulli random variable with success probability $p_0$ and 

$$\text{Est} \left[ E \left[ B(C; p) \right] \right] = \frac{1}{n} \sum_{i=0}^{n} B(c_i; p) .$$  

As in the continuous distribution case, the final term, here $\sum_{c=0}^{1} B(c; p_0)^2$, is free of the parameters and will not affect the minimization problem.

With censoring the $L_2E$ criterion for the Weibull PDF has an additional factor, an identity function $I(T \leq t_c)$, which represents the censoring mechanism. The modified $L_2E$ criterion of the Weibull part becomes:

$$\text{Est} \left[ \int_0^\infty \left[ f(t; \alpha, \beta) - f(t; \alpha_0, \beta_0) \right]^2 I(t \leq t_c) dt \right] = \int_0^\infty \left[ f(t; \alpha, \beta)^2 dt - 2\text{Est} \left[ E \left[ f(T; \alpha, \beta) I(T \leq t) \right] \right] + \int_0^\infty f(t; \alpha, \beta)^2 dt \right] .$$

It is equivalent to minimize

$$\int_0^\infty f(t; \alpha, \beta)^2 dt - 2\frac{1}{n} \sum_{i=1}^{n} (1 - c_i) f(t_i; \alpha, \beta) .$$  

(11)

We use the summation of $L_2E$ criteria (10) and (11) as our $L_2$ criterion to model type I censored data:

$$\int_0^t f(t; \alpha, \beta)^2 dt - 2\frac{1}{n} \sum_{i=1}^{n} (1 - c_i) f(t_i; \alpha, \beta) + p^2 + (1-p)^2 - 2\frac{1}{n} \sum_{c=0}^{1} B(c; p_0) , \quad p = 1 - F(t_c; \alpha, \beta).$$  

(12)

When the type I censored data are contaminated at some unknown percentage, we can introduce the $w$ trick as mentioned in Section 2.4 to the Weibull portion of the criterion (11); compare to equation (8). More sophisticated model can be achieved by introducing $w$ properly into the Bernoulli portion of the criterion, but we defer this to later research.

5.2. $L_2$ criterion for type II censored data

If the data are type II censored, those censored values are not necessarily the same. However, the random starting times $s = \{ s_1, \ldots, s_n \}$ are collected and the fixed censoring value $t_c$ is given. It is assumed that all starting times are less than $t_c$. Then the $i^{th}$ observed data $t_i$ is equal to $(t_c - s_i)$ if censored with $c_i = 1$, otherwise $t_i < t_c - s_i$ and $c_i = 0$.

Thus for each data point $t_i$, the corresponding censoring value is $(t_c - s_i)$ and the survival probability is $p_i = 1 - F(t_c - s_i; \alpha, \beta)$ with given $(\alpha, \beta)$. The $L_2E$ criteria for a single data point $t_i$ is
With all \( n \) observations considered, the \( L_2E \) criterion for type II censored data becomes the average of the \( n \) sample-size-one criteria given in (13), namely,

\[
\int_0^{t_{c}^{-1}} f(t;\alpha,\beta)^2 dt - 2 (1 - c_i) f(t;\alpha,\beta) + p_i^2 + (1 - p_i)^2 - 2B(c_i; p_i). \tag{13}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \left[ \int_0^{t_{c}^{-1}} f(t;\alpha,\beta)^2 dt - 2 (1 - c_i) f(t;\alpha,\beta) + p_i^2 + (1 - p_i)^2 - 2B(c_i; p_i) \right], p_i=1 - F(t_c - s_i; \alpha, \beta). \tag{14}
\]

Note that the criterion (14) will collapse to (12) when the starting times are all 0, i.e. the data are basically type I censored. Similarly, for contaminated type II censored data, we can also introduce \( w \) to (14).

6. Case studies with right censored data

6.1. Simulation studies with right censored data

In this set of simulations, we want to explore the efficiency and robustness of MLE and \( L_2E \) for censored data of both types. Since outliers will be mixed with censored data if contamination happens on the right side of the data, we also want to simulate data with contamination occurring on the left side. A Weibull distribution with scale \( \alpha = 4 \) and shape \( \beta = 10 \) is utilized as the true model.

We generated a group of uncontaminated data \( T = \{ t_1, \ldots, t_{900} \} \) consisting of 900 lifetimes from a standard Weibull distribution with scale \( \alpha = 4 \) and shape \( \beta = 10 \). Let \( T_3 \) denote the type I censored data from \( T \) with 15\% censored at one fixed value, namely, the 0.15\% sample upper quantile of \( T \). For randomly censored data (denoted by \( T_4 \)), we generated 900 waiting times \( s = \{ s_1, \ldots, s_{900} \} \) from the uniform \( U(0,0.45) \), and used the true 0.05\% upper quantile of \( f(t;\alpha = 4, \beta = 10) \) as censoring time: \( t_c = 4.464 \). The uniform distribution \( U(0,0.45) \) is chosen to control the percentage of random censoring. Then \( t_i \) was set to be \((t_c - s_i)\) if \( t_i > t_c - s_i \) and \( c_i = 1 \), otherwise \( t_i \) did not change and \( c_i = 0 \). Now we obtained randomly censored (type II) lifetime data \( T_4 \).

We also generated data \( T_c \) with 10\% contamination, which consisted of \( m_1 \sim \text{Binomial}(1000,0.9) \) lifetimes from \( f(t;\alpha = 4, \beta = 10) \), and \( (1000 - m_1) \) lifetimes that were the absolute values of samples from a normal distribution \( N(1,1/2) \). Then censored data \( T_{3c} \) (type I) and \( T_{4c} \) (type II) with contamination were obtained from \( T_c \) similarly as in the uncontaminated case. We fitted the classical Weibull model (by both methods of MLE and \( L_2E \)), and the new Weibull model to datasets \( T_3, T_4, T_{3c}, T_{4c} \). Each optimization problem was solved numerically by the \texttt{R} function \texttt{nlminb} with the same corresponding starting value \((\alpha = 4, \beta = 10)\) or \((\alpha = 4, \beta = 10, w = 0.9)\).

The above generation process and model fitting of \((T_3, T_4, T_{3c}, T_{4c})\) were repeated for 1000 times. As a result, the percentage of random censoring for \( T_4 \) has mean 17.9\% and standard deviation 0.013, and the percentage of random censoring for \( T_{4c} \) has mean 16.2\% and standard deviation 0.012. Means and standard deviations of those MLEs, \( L_2E \)s and \( L_2E \)s are shown in Table 2 and 3.

Histograms of typical generated data sets and model fitting results with the mean estimates are given in Figures 4, 5, 6 and 7. The left frame in each figure displays a histogram of censored data, the true PDF (dotted line), the PDFs with the MLE estimates (short dashed
line), the $L_2E$ estimates (solid line) and the $L_2Ew$ estimates (long dashed line), while the right frame displays the corresponding log Weibull hazard functions.

For the uncontaminated cases, all types of lines are overlapped in the plots because all estimates are close to the truth; while for the cases with 10% contamination, the long dashed line and the dotted line are almost overlapped in the Figures 5 and 6, which shows that the $L_2Ew$ of the new augmented Weibull fitting is the best.

One can easily observe that MLE is more efficient than $L_2E$ if the data are uncontaminated, but $L_2E$ can robustly give more accurate estimates for contaminated data sets, in which cases the estimates by MLE are way off the truth. For example, the RMSE for $MLE(T3c)$ is 4.120, while the RMSE for $L_2Ew(T3c)$ is 0.130. The standard deviations of MLE and $L_2E$ are approximately the same for each type of data set, although most of the standard deviations of $L_2E$ are slightly larger.

6.2. The $L_2$ method on the head-and-neck data

Our second real case study is conducted on a group of randomly right-censored cancer survival times. The survival times $t$ (in days) for 51 patients of Arm A in the head-and-neck cancer trial were first analyzed by Efron (1988), and re-examined by Mudholkar et al. (1996) and Nikulin & Haghighi (2006) using MLE with two different generalized Weibull models.

We scaled the data by 30.438 for converting the daily data to monthly as suggested by Efron (1988). These are type II censored data, which have random starting times and were censored by the same time. Efron (1988) did not state the exact starting times $s$ and censoring time $t_c$, thus we made a reasonable conjecture about $s$ and $t_c$. We took the ceiling of the maximum data value 47 as $t_c$, then generated $s_i$ uniformly from $U(0, t_c - t_i)$ if the $i^{th}$ observation is uncensored ($c_i = 0$), otherwise $s_i = t_c - t_i$.

With $t$, $s$, $c$, and $t_c$, we fitted the classical Weibull model by the $L_2E$ criterion (14) and maximum likelihood (9), and the new Weibull model to the head-and-neck cancer data. The results are shown in Figure 8. The left frame in Figure 8 displays a histogram of the monthly head-and-neck cancer data, together with the fitted PDFs using MLE estimates (short dashed line), $L_2E$ estimates (solid line) and $L_2Ew$ estimates (long dashed line), while the right frame displays the corresponding log Weibull hazard functions.

Previous analyses of this data set concluded that both the PDF and hazard function should be unimodal. Due to the limitation of monotone hazard function in the classical two-parameter Weibull model, neither MLE nor $L_2Ew$ might be satisfactory. However, the PDF shape with $L_2Ew$ estimate is unimodal and much closer to the one given by Nikulin & Haghighi (2006), while the PDF shape with MLE estimates is clearly very different. This study showed that $L_2Ew$ is more proper to model the hand-and-neck cancer trial data set, comparing with MLE.

7. Discussion

$L_2E$ performs very well in all of the real cases and simulation studies, especially when the data are contaminated. The proposed new augmented Weibull model has a nice property of detecting the percentage of contamination. The third parameter in this model is $w$, which allows $L_2E$ to avoid some inflation in the variance of the fitted Weibull in certain situations. A more robust and accurate $L_2$ estimate will result from this new model. We advocate using both MLE and $L_2E$ approaches simultaneously. When the results are close, it is safe to use MLE, when they are not, careful exploration of the data should be undertaken. Then if one
concludes the data are contaminated at some level, the newly proposed Weibull model with an extra parameter w will be a good choice.

Although we only investigated the classical two-parameter Weibull distribution in this paper, the idea of $L_2E$ and the w trick can be easily adapted into any of the generalized three-parameter Weibull distributions proposed in the literature. The only problem in the further adaption will be that most of these generalized distributions are not square integrable in closed form, so numerical integration will be required for each evaluation of the criterion function during the optimization. In any case, we can weigh the complexity of the adapted $L_2E$ method and the robustness case by case.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by NIH grant 2PO1CA082710 and NSF grant DMS-09-07491. We especially wish to thank Dr. Dennis Cox for his wonderful suggestions and support via the NIH P01 grant. We also appreciate the referees’ valuable suggestions that improved the motivation of $L_2E$.

This research was also supported by a research contract to the University of Pennsylvania and the University of California from the Intelligence Advanced Research Projects Activity (IARPA) via the Department of Interior National Business Center contract number D11PC20061. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions expressed herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

Appendix A: Sample L2E function

The following example function fits the Weibull parameters ($a,b,w$) using the $L_2E$ criterion and the R function `nlminb`. The inputs are the data vector, x, and the vector, x0, of initial guesses for the three parameters.

```r
weibull.l2e = function( x, x0 ) {
  if(length(x0)!=3) stop("x0 = (a,b,w) not of length 3")
  mycrit = function(x,X=X) {a=x[1]; b=x[2]; w=x[3]; c=2-1/b
    w^2 * b/a/2^c*gamma(c) - 2*w*mean(dweibull(X,b,a)) }
  ans = nlminb(x0,mycrit,X=x,lower=c(.01,.51,.01),
    upper=c(Inf,Inf,1))
  a = ans$spar[1]; b = ans$spar[2]; w = ans$spar[3]
  return(c(a,b,w))
}
```

Appendix B: Asymptotic normality of L2E

We give a brief outline of the general asymptotic normality of the $L_2E$ parameters, assuming any regularity and smoothness conditions required. The data are sampled from the density $g(x)$, and the model is denoted by $f_\theta(x)$; the parameter w may or may not be included in $\theta \in \mathbb{R}^\theta$. The $L_2E$ criterion (equation 8) divided by 2 may be written as

$$
C(\theta) - \frac{1}{n} \sum_{i=1}^{n} f_\theta(x_i) = C(\theta) - \frac{1}{n} \sum_{i=1}^{n} \left[ f_\theta(x_i) - M(\theta) \right] - M(\theta), \quad (B.1)
$$

\textit{Comput Stat Data Anal. Author manuscript; available in PMC 2014 November 01.}
where \( C(\theta) = \frac{1}{2} \int f_\theta(x)^2 \, dx \) and \( M(\theta) = \int f_\theta(x)g(x) \, dx \). Let \( \hat{\theta}_n \) denote a minimizer of (B.1) close to \( \theta_0 \), which is a (perhaps local) minimum of \( C(\theta) - M(\theta) \). Thus \( \theta_0 \) is a root of \( \nabla_\theta C(\theta) - \nabla_\theta M(\theta) = 0 \).

For simplicity, we show the asymptotic normality of \( \hat{\theta}_n \) when \( p = 1 \). From equation (B.1), the minimizer \( \hat{\theta}_n \) is a solution to the equation

\[
\sqrt{n} \left( C'(\hat{\theta}_n) - M'(\overline{\theta}_n) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ f'_\theta(x_i) - M'(\overline{\theta}_n) \right] \approx Z \sqrt{\text{Var}_{\theta_0}(X)}, \quad (B.2)
\]

where \( Z \sim N(0,1) \). Assuming \( \hat{\theta}_n \) converges to a particular root \( \theta_0 \), a Taylor’s Series of the left hand side gives

\[
C'(\hat{\theta}_n) - M'(\overline{\theta}_n) = C'(\theta_0) - M'(\theta_0) + \left( C''(\theta_0) - M''(\theta_0) \right)(\hat{\theta}_n - \theta_0) + \cdots;
\]

but \( C'(\theta_0) = M'(<\theta_0) \). Hence, \( \sqrt{n}(\hat{\theta}_n - \theta_0) \) is asymptotically normal with variance \( \text{Var}(X)/\left( C''(\theta_0) - M''(\theta_0) \right)^2 \).

For the two examples with data \( T2 \) in Section 2.3, the minimizers \( \theta_0 \) may be found numerically to be (a\(_0\) = 4.202, b\(_0\) = 1.860) or in the 3-parameter case, (a\(_0\) = 4.038, b\(_0\) = 1.967, w\(_0\) = 0.910). The upward bias in the weight parameter increases as the separation between the Weibull data and contamination decreases, and vice versa. For example, if the mean of the contamination is shifted to \( \mu = 12 \), then (a\(_0\) = 4.011, b\(_0\) = 1.989, w\(_0\) = 0.903). When \( \mu > 13.5 \), the parameters almost exactly match truth.

**Appendix c: Real data sets**

### Table C.4

Experimental data for the strength of glass fibre of length 1.5 cm (Smith & Naylor, 1987)

<table>
<thead>
<tr>
<th>Sample 1 — Strengths of 1.5 cm fibre</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.55 0.74 0.77 0.81 0.84 0.93 1.04 1.11 1.13 1.24 1.25 1.27 1.28 1.29 1.30 1.36</td>
</tr>
<tr>
<td>1.39 1.42 1.48 1.48 1.49 1.49 1.50 1.50 1.51 1.52 1.53 1.54 1.55 1.55 1.58 1.59</td>
</tr>
<tr>
<td>1.60 1.61 1.61 1.61 1.62 1.62 1.63 1.64 1.66 1.66 1.67 1.68 1.68 1.69</td>
</tr>
<tr>
<td>1.70 1.70 1.73 1.76 1.76 1.77 1.78 1.81 1.82 1.84 1.84 1.84 1.89 2.00 2.01 2.24</td>
</tr>
</tbody>
</table>

### Table C.5

Experimental data for the strength of glass fibre of length 15 cm (Smith & Naylor, 1987)

<table>
<thead>
<tr>
<th>Sample 2 — Strengths of 15 cm fibre</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.37 0.40 0.70 0.75 0.80 0.81 0.83 0.86 0.92 0.92 0.94 0.95 0.98 1.03 1.06 1.06</td>
</tr>
<tr>
<td>1.08 1.09 1.10 1.10 1.13 1.14 1.15 1.17 1.20 1.20 1.21 1.22 1.25 1.28 1.28 1.29</td>
</tr>
<tr>
<td>1.29 1.30 1.35 1.35 1.37 1.37 1.38 1.40 1.40 1.42 1.43 1.51 1.53 1.61</td>
</tr>
</tbody>
</table>
Table C.6

Survival times (days) for patients at Arm A of the head-and-neck cancer trial, where the sign * indicates the observations lost to follow up (Efron, 1988)

<table>
<thead>
<tr>
<th>Arm A:</th>
</tr>
</thead>
</table>

References


Figure 1.
(a) & (c) Classical two-parameter Weibull model fits to data $T_1$ and $T_2$, respectively, with parameters $(\alpha, \beta)$; (b) mixture Weibull-Normal model fits to $T_2$ with parameters $(\alpha, \beta, w, \mu, \sigma)$; (a) & (d) new augmented Weibull model fits (with area $w$ under the curve) to $T_1$ and $T_2$, respectively, with parameters $(\alpha, \beta, w)$. 
Figure 2.
Shape of influence functions for MLE ((a) & (b)) and $L_2E$ ((c) & (d)) with $\alpha = 4$, $\beta = 2$. 
Figure 3.
Classical two-parameter \((a, \beta)\) Weibull model fits and the new \((a, \beta, w)\) Weibull model fits on both fibre samples, with area \(w\) under the curve.
Figure 4.
Weibull model fits on simulated type I censored data with no contamination.
Figure 5.
Weibull model fits on simulated type II censored data with no contamination.
Figure 6.
Weibull model fits (with area \( w \) under the curve) on simulated type I censored data with 10% contamination.
Figure 7.
Weibull model fits (with area $w$ under the curve) on simulated type II censored data with 10% contamination.
Figure 8.
Compare Weibull model fitting by MLE, \( L_2E \) and \( L2Ew \) on the head-and-neck cancer data (type II censored).
Table 1

Mean of the 1000 estimates of $MLE$, $L_2E$ and $L_2Ew$ for $(\alpha, \beta, w)$, with the standard deviation in the parentheses

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$w$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MLE(T1)$</td>
<td>4.001 (0.071)</td>
<td>2.005 (0.054)</td>
<td>-</td>
<td>0.004</td>
</tr>
<tr>
<td>$L_2E(T1)$</td>
<td>4.000 (0.084)</td>
<td>2.007 (0.069)</td>
<td>-</td>
<td>0.005</td>
</tr>
<tr>
<td>$L_2Ew(T1)$</td>
<td>3.997 (0.086)</td>
<td>2.011 (0.072)</td>
<td>0.997 (0.005)</td>
<td>0.007</td>
</tr>
<tr>
<td>$MLE(T2)$</td>
<td>4.807 (0.101)</td>
<td>1.595 (0.031)</td>
<td>-</td>
<td>0.638</td>
</tr>
<tr>
<td>$L_2E(T2)$</td>
<td>4.202 (0.102)</td>
<td>1.864 (0.067)</td>
<td>-</td>
<td>0.172</td>
</tr>
<tr>
<td>$L_2Ew(T2)$</td>
<td>4.037 (0.101)</td>
<td>1.973 (0.079)</td>
<td>0.909 (0.012)</td>
<td>0.027</td>
</tr>
</tbody>
</table>
Table 2
Means of the 1000 estimates of $MLEs$, $L_2Es$ and $L_2Ews$ for uncontaminated data, with the corresponding standard deviation in the parentheses

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$w$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MLE(T3)$</td>
<td>4.000 (0.015)</td>
<td>10.021 (0.313)</td>
<td>-</td>
<td>0.015</td>
</tr>
<tr>
<td>$L_2E(T3)$</td>
<td>4.000 (0.017)</td>
<td>10.033 (0.375)</td>
<td>-</td>
<td>0.023</td>
</tr>
<tr>
<td>$L_2Ew(T3)$</td>
<td>3.999 (0.017)</td>
<td>10.066 (0.397)</td>
<td>0.996 (0.006)</td>
<td>0.038</td>
</tr>
<tr>
<td>$MLE(T4)$</td>
<td>4.000 (0.015)</td>
<td>10.026 (0.319)</td>
<td>-</td>
<td>0.018</td>
</tr>
<tr>
<td>$L_2E(T4)$</td>
<td>4.000 (0.017)</td>
<td>10.042 (0.389)</td>
<td>-</td>
<td>0.030</td>
</tr>
<tr>
<td>$L_2Ew(T4)$</td>
<td>4.000 (0.017)</td>
<td>10.070 (0.410)</td>
<td>0.996 (0.005)</td>
<td>0.040</td>
</tr>
</tbody>
</table>
### Table 3

Means of the 1000 estimates of $\text{MLEs}$, $L_2\text{Es}$ and $L_2\text{Ews}$ for contaminated data, with the corresponding standard deviation in the parentheses

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\omega$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{MLE(}\ T_3c\ )$</td>
<td>3.897 (0.018)</td>
<td>4.175 (0.283)</td>
<td>-</td>
<td>4.120</td>
</tr>
<tr>
<td>$L_2\text{E(}\ T_3c\ )$</td>
<td>3.994 (0.016)</td>
<td>9.333 (0.368)</td>
<td>-</td>
<td>0.472</td>
</tr>
<tr>
<td>$L_2\text{Ew(}\ T_3c\ )$</td>
<td>3.985 (0.016)</td>
<td>10.224 (0.454)</td>
<td>0.882 (0.014)</td>
<td>0.130</td>
</tr>
<tr>
<td>$\text{MLE(}\ T_4c\ )$</td>
<td>3.903 (0.025)</td>
<td>4.158 (0.283)</td>
<td>-</td>
<td>4.131</td>
</tr>
<tr>
<td>$L_2\text{E(}\ T_4c\ )$</td>
<td>3.993 (0.016)</td>
<td>9.301 (0.362)</td>
<td>-</td>
<td>0.494</td>
</tr>
<tr>
<td>$L_2\text{Ew(}\ T_4c\ )$</td>
<td>3.999 (0.017)</td>
<td>9.995 (0.441)</td>
<td>0.901 (0.012)</td>
<td>0.003</td>
</tr>
</tbody>
</table>