مجله‌ی پژوهش‌های آماری ایران
پژوهشکده‌ی آمار
دوره‌ی ۱۷، شماره‌ی ۱، بهار و تابستان ۱۳۹۹، صص ۱۷۱–۱۹۰

J. Statist. Res. Iran **17** (2020): 171–190
DOI: 10.52547/jsri.17.1.171
DOR: 20.1001.1.25385771.2020.17.1.9.0

Non-Bayesian Estimation and Prediction under Weibull Interval Censored Data

Somayeh Ghafouri[†,*], Arezou Habibirad[‡], Fatemeh Yousefzadeh[⋆] and Reza Pakyari[◦]

[†] Arak University
[‡] Ferdowsi University of Mashhad
[⋆] University of Birjand
[◦] Qatar University

Received: 2022/09/03 Approved: 2022/19/12

Abstract. In this paper, a one-sample point predictor of the random variable X is studied. X is the occurrence of an event in any successive visits L_i and R_i :$i = 1, 2, \ldots, n$ (interval censoring). Our proposed method is based on finding the expected value of the conditional distribution of X given L_i and R_i $(i = 1, 2, \ldots, n)$. To make the desired prediction, our approach is on the basis of approximating the unknown Weibull parameters using the mid-point approximation and approximate maximum likelihood (AML). After obtaining the parameter estimation, the prediction of X can be made. Moreover, the 95% bootstrap confidence intervals of unknown parameters and the 95% bootstrap prediction bounds of X are presented. The performance of the proposed procedure based on the mean squared error (MSE) and the average width (AW) of the confidence interval is investigated by employing Monte Carlo simulation. A Real data set is also studied to illustrate the proposed procedure.

Keywords. Approximate maximum likelihood estimator, Bootstrap samples, interval censoring, mean squared prediction error, mid-point approximation, monte Carlo simulation, one-sample prediction.

[*] Corresponding author

MSC 2010: 60G25, 62F10, 62F40, 62N01, 65C05.

1 Introduction

Censoring is very common in reliability and survival analysis. If the lifetimes are censored before all failure times are observed, we will have the right censoring. However, sometimes, the desired event is only known to occur within an interval of time, say $[L, R]$. Namely, interval censored lifetime data arise when individuals in a study are monitored intermittently, not continuously, so that a lifetime observed (or a unit failed) lies between two successive times. This situation may occur in clinical trials when patients are visited only at a pre-scheduled visit say, L and R. If the event has not occurred at the time of visit L, but occurred by the following visit time R, then the event must belong to the interval $[L, R]$ (see Tomazella and Nadarajah (2015)). This type of censoring can be considered in lifetime study and it has attracted a lot of attention itself mainly in clinical biological studies where experimental units are human beings or animal subjects who are checked at discrete intervals (for example, quarterly, hourly or monthly check-ups). The desired events occur at some time between examinations (see Aggarwala (2001)). Also, according to Sun (2006), many areas including demographical, epidemiological, financial, medical, sociological, and engineering studies can utilize interval-censored failure time data (see Guure et al. (2013)). Thus, prediction of the event that occurred between the two observation periods has an important role in many fields such as medical sciences and reliability analysis. The focus of this paper is on interval censoring, which presumably is more important than right censoring. Sun (1997) has introduced a method based on a discrete logistic model (interval-censored data that arise from clinical trials with a discrete scale) for the regression analysis of interval-censored failure time data with a focus on the comparison of failure time distributions among different treatments. Lindsey and Ryan (1998) studied standard methods and software algorithms for analyzing interval-censored data (Aggarwala (2001)). As stated by Turnbull (1976), one could define an interval-censored observation as a union of several non-overlapping windows or intervals. Researchers who have discussed interval-censored data in the classical point of view are, Odell et al. (1992), Lindsey (1998), Scallan (1999), Lawless (2003) and Flygare and Buckwalter (2010) (see Guure et al. (2013)).

Pradhan and Kundu (2014) worked the Bayesian and Non-Bayesian analysis of interval-censored data for the unknown parameters, with Weibull distribution as the underlying lifetime distribution. For further investigation of inference under interval-censored lifetime, data see Gomez et al. (2004), Lawless and Babineau (2006), Dumbgen et al. (2006) and Singh and Totawattage (2013).

The problem of estimation for another type of interval-censored data, progressive type-I interval censoring has been considered for many other distributions, for more details, we refer to Xiang and Tse (2005), Lawless and Babineau (2006), Ng and Wang (2009), Chen and Lio (2010), Lio et al. (2011), Lin and Lio (2012), Pradhan and Kundu (2014) and Ahmadi and Yousefzadeh (2015). For more applications of interval-censored data on reliability, failure risk, repairable systems, etc. see Hashimoto et al. (2013), Garcia-Mora et al. (2015) and Peng et al. (2017).

Some prediction studies have also been conducted based on the maximum likelihood approach. For instance, Raqab (2004) based on a general multiple Type-II censored sample from a shifted exponential distribution, proposed some approximate predictors of missing failure times. Basak et al. (2006) studied the prediction of time to failure of units censored under a progressively censored sample from an absolutely continuous population. The best linear unbiased predictors (BLUPs), the ML predictors (MLPs) and the approximate MLPs of units under progressively censored sample for the Pareto distribution could be found in Raqab et al. (2010). Some other recent studies and applications of interval-censored data could be conducted in Zhang (2009), Shen (2014), Hyun et al. (2017), Wu and Cook (2020) and Yao et al. (2021).

In this paper, we focus on the prediction of an event (a failure time) say, X cannot be observed directly but can only be determined to lie in an interval $[L, R]$. Because the prediction of the value of X is important. Therefore, a simple proposed method such as the expectation of the conditional distribution of X given L and R can be used. To achieve this purpose, the estimation of unknown parameters is required. Then, using the estimated parameters, the prediction of the random variable X can be made. One of the advantages of this method is its simplicity. In addition to this prediction method for X, the possibility of obtaining other methods of prediction would be investigated by future researchs.

In summary, the estimation of unknown parameters (based on mid-point and AML methods) in the Weibull distribution can be obtained. Concerning

the estimated parameters, the point prediction of X using the expected value of the conditional distribution of X given L and R easily expressed. The rest of the paper is organized as follows.
At the beginning of Section 2, based on a sample of size n from intervals $[L_i, R_i]$ $(i = 1, 2, \ldots, n)$, the general likelihood equations derived to get estimations for unknown Weibull parameters. But, due to the simplicity of the extreme value distribution and its equivalence with the Weibel distribution, likelihood equations of extreme values presented. Subsection 2.1 discusses about mid-point approximation method for parameter estimation. Subsection 2.2, introduces approximate maximum likelihood estimation for unknown parameters. Finally, in subsection 2.3, using the estimated parameters and through the conditional distribution of X given L and R, the point predictor of a failed unit (say X) is introduced. In Section 3, a simulation study (and using the percentile bootstrap method for the prediction bounds of X) as well as an illustrative example have been analyzed for illustrative purposes.

2 ML Estimation of Parameters and Prediction for X

In this section, the likelihood equations for finding the estimation of unknown Weibull parameters are derived. In the following, the estimation of unknown Weibull parameters via mid-point and approximate ML are proposed. They will be used to get the average of the expected value of the conditional distribution of X given L_i and R_i, $i = 1, 2, \ldots, n$ as a predictor of X. One of the widely used distributions for analyzing lifetime data in reliability and survival analysis is the Weibull model. The corresponding probability density function (pdf) is

$$f(x; \alpha, \beta) \quad = \quad \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, \qquad x > 0,\ \alpha, \beta > 0. \tag{1}$$

Also, the likelihood function $L(\alpha, \beta; L_i, R_i; i = 1, 2, \ldots, n)$ under interval censoring is given by (see Gomez et al. (2004) and Guure et al. (2015))

$$L(\alpha, \beta; L_i, R_i, i = 1, 2, \ldots, n) = L = \prod_{i=1}^{n} \Big[F(R_i, \alpha, \beta) - F(L_i, \alpha, \beta) \Big]. \tag{2}$$

Therefore, from (1) and (2), the log likelihood function (LLF) will be of the

form

$$\ln L = \sum_{i=1}^{n} \ln \left(e^{-\alpha L_i^{\beta}} - e^{-\alpha R_i^{\beta}} \right).$$

Differentiating with respect to α and β, and then equating to zero, we get

$$\begin{cases} 1: \frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n} \frac{R_i^{\beta} \exp(-\alpha R_i^{\beta}) - L_i^{\beta} \exp(-\alpha L_i^{\beta})}{\exp(-\alpha L_i^{\beta}) - \exp(-\alpha R_i^{\beta})} = 0, \\ 2: \frac{\partial \ln L}{\partial \beta} = \alpha \sum_{i=1}^{n} \frac{R_i^{\beta} \exp(-\alpha R_i^{\beta}) \ln(R_i) - L_i^{\beta} \exp(-\alpha L_i^{\beta}) \ln(L_i)}{\exp(-\alpha L_i^{\beta}) - \exp(-\alpha R_i^{\beta})} = 0. \end{cases} \tag{3}$$

It is worthwhile to mention that instead of working with the Weibull model for X, it is often easier to work with the equivalent model for the log-lifetime $Y = \ln X$. Then, Y has an extreme value distribution with pdf f_1 and cumulative distribution function F_1, respectively, given by

$$\begin{aligned} f_1(y; \mu, \sigma) &= \frac{1}{\sigma} e^{(\frac{y-\mu}{\sigma}) - e^{(\frac{y-\mu}{\sigma})}}, \\ F_1(y; \mu, \sigma) &= 1 - e^{-e^{(\frac{y-\mu}{\sigma})}}, \quad -\infty < y < \infty, \ -\infty < \mu < \infty, \ \sigma > 0, \end{aligned}$$

where, $\sigma = \beta^{-1}$ and $\mu = \beta^{-1} \ln(\alpha^{-1})$.
Now, suppose $[L_i^*, R_i^*], i = 1, 2, \ldots, n$ denote the interval censored data from a distribution function F_1 with unknown parameters. Let Y_i represent the unknown time, that is, $L_i^* \leq Y_i \leq R_i^*$, where L_i^* is the last inspection time and R_i^* is the state of end time.

Then, from (2) under the extreme value distribution, the likelihood function (LF) is written as

$$L_1(\mu, \sigma; L_i^*, R_i^*, i = 1, 2, \ldots, n) = L_1 = \prod_{i=1}^{n} \left[e^{-e^{(\frac{L_i^* - \mu}{\sigma})}} - e^{-e^{(\frac{R_i^* - \mu}{\sigma})}} \right].$$

The LLF is reduced to

$$\ln L_1 = \sum_{i=1}^{n} \ln \left[e^{-e^{(\frac{L_i^* - \mu}{\sigma})}} - e^{-e^{(\frac{R_i^* - \mu}{\sigma})}} \right]. \tag{4}$$

Differentiating with respect to μ and σ, and then setting to zero, possible

likelihood equations may be expressed as

$$\begin{cases} \frac{\partial \ln L_1}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n} \frac{e^{(\frac{L_i^*-\mu}{\sigma})}-e^{(\frac{L_i^*-\mu}{\sigma})}-e^{(\frac{R_i^*-\mu}{\sigma})}-e^{(\frac{R_i^*-\mu}{\sigma})}}{e^{-e^{(\frac{L_i^*-\mu}{\sigma})}}-e^{-e^{(\frac{R_i^*-\mu}{\sigma})}}} = 0, \\ \\ \frac{\partial \ln L_1}{\partial \sigma} = \sum_{i=1}^{n} \frac{(\frac{L_i^*-\mu}{\sigma^2})e^{(\frac{L_i^*-\mu}{\sigma})}e^{(\frac{L_i^*-\mu}{\sigma})}-(\frac{R_i^*-\mu}{\sigma^2})e^{(\frac{R_i^*-\mu}{\sigma})}e^{(\frac{R_i^*-\mu}{\sigma})}-e^{(\frac{R_i^*-\mu}{\sigma})}}{e^{-e^{(\frac{L_i^*-\mu}{\sigma})}}-e^{-e^{(\frac{R_i^*-\mu}{\sigma})}}} = 0. \end{cases} \tag{5}$$

Unfortunately, these equations could not be expressed in explicit forms.

2.1 Mid-point approximation method

This subsection, provides the estimation of the unknown parameters based on the mid-point approximation method. Suppose that $m_i = \frac{L_i^*+R_i^*}{2}$, the LLF can be approximated by

$$\begin{aligned} \ln L_1(\mu,\sigma;m_i,i=1,2,\ldots,n) &\sim \sum_{i=1}^{n} \ln f_1(m_i,\mu,\sigma) = \sum_{i=1}^{n}\left[-\ln\sigma + \frac{m_i-\mu}{\sigma} - e^{\frac{m_i-\mu}{\sigma}}\right] \\ &= -n\ln(\sigma) + \sum_{i=1}^{n}\frac{m_i-\mu}{\sigma} - \sum_{i=1}^{n} e^{\frac{m_i-\mu}{\sigma}}. \end{aligned} \tag{6}$$

Differentiating with respect to μ and σ, and then equating to zero, possible likelihood equations can be written as

$$\begin{cases} \frac{\partial \ln L_1(\mu,\sigma;m_i,i=1,2,\ldots,n)}{\partial \mu} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} e^{(\frac{m_i-\mu}{\sigma})} = 0, \\ \\ \frac{\partial \ln L_1(\mu,\sigma;m_i,i=1,2,\ldots,n)}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{i=1}^{n}(\frac{m_i-\mu}{\sigma^2}) + \sum_{i=1}^{n}(\frac{m_i-\mu}{\sigma^2})e^{(\frac{m_i-\mu}{\sigma})} = 0. \end{cases} \tag{7}$$

Therefore, $\hat{\mu}$ can be obtained as $\hat{\mu} = \hat{\sigma}\ln(\frac{1}{n}\sum_{i=1}^{n} e^{\frac{m_i}{\hat{\sigma}}})$. By substituting $\hat{\mu}$ in the second equation of (7), $\hat{\sigma}$ is given by $\hat{\sigma} = \frac{1}{n}\left(\frac{\sum_{i=1}^{n} m_i e^{\frac{m_i}{\hat{\sigma}}}}{\frac{1}{n}\sum_{i=1}^{n} e^{\frac{m_i}{\hat{\sigma}}}} - \sum_{i=1}^{n} m_i\right)$.

2.2 Approximate Maximum Likelihood Method

By letting $m_i = \frac{L_i^* + R_i^*}{2}$, the LF reduces to

$$L_1(\mu, \sigma; m_i, i = 1, 2, \ldots, n) \quad \propto \quad \prod_{i=1}^{n} f_1(m_i; \mu, \sigma). \tag{8}$$

Also, by assumption of $g(x) = e^{x - e^x}$ and $z_i = \frac{m_i - \mu}{\sigma}$, (8) changes to

$$L_1(\mu, \sigma; m_i, i = 1, 2, \ldots, n) \quad \propto \quad \frac{1}{\sigma^n} \prod_{i=1}^{n} g(z_i; \mu, \sigma). \tag{9}$$

Then,

$$\ln L_1(\mu, \sigma; m_i, i = 1, 2, \ldots, n) = -n \ln \sigma + \sum_{i=1}^{n} \ln g(z_i), \tag{10}$$

and

$$\begin{cases} E1 : \frac{\partial \ln L_1(\mu, \sigma; m_i, i=1,2,\ldots,n)}{\partial \mu} = -\frac{1}{\sigma} \sum_{i=1}^{n} \frac{g'(z_i)}{g(z_i)} = 0, \\ \\ E2 : \frac{\partial \ln L_1(\mu, \sigma; m_i, i=1,2,\ldots,n)}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{i=1}^{n} \frac{g'(z_i) z_i}{\sigma g(z_i)} = 0. \end{cases} \tag{11}$$

Under the extreme value distribution, let G be the cumulative density function. Then, $G^{-1}(u) = \ln(-\ln(1-u))$, $0 < u < 1$. If m_i's ordered (therefore, z_i's are ordered), the function $\frac{g'(z_i)}{g(z_i)}$ will be approximated by expanding it in a Taylor series around the point $\mu_i = G^{-1}(u_{mi}) = \ln(-\ln(1 - u_{mi}))$, where $u_{mi} = \frac{u_i + u_{i-1}}{2}$; $u_i = \frac{i}{n}$ for $i = 1, 2, \ldots n$, (see Ahmadi and Yousefzadeh (2015)). Because, by law of large numbers, as n large enough, empirical distribution function $G_n(z_i)$ tends to $G(z_i)$. Then, $\mu_i = G_n^{-1}(u_{mi})$ tends to z_i and the expansion of $\frac{g'(z_i)}{g(z_i)}$ in a Taylor series around the point μ_i is rational. Note that in order to use $u_i = G_n(z_i) = \frac{i}{n}$, m_i's are arranged in ascending order. Now, by expanding the function $\frac{g'(z_i)}{g(z_i)}$ around μ_i and keeping only the first two terms, this function may be approximated by

$$\frac{g'(z_i)}{g(z_i)} \approx \frac{g'(\mu_i)}{g(\mu_i)} + (z_i - \mu_i)\big(\frac{g'(\mu_i)}{g(\mu_i)}\big)' = \alpha_i - \beta_i z_i, \tag{12}$$

where

$$\alpha_i = \frac{g'(\mu_i)}{g(\mu_i)} - \mu_i\left(\frac{g''(\mu_i)}{g(\mu_i)} - \big(\frac{g'(\mu_i)}{g(\mu_i)}\big)^2\right) = 1 + \ln q_{mi}[1 - \ln(-\ln q_{mi})], \quad q_{mi} = 1 - u_{mi},$$

$$\beta_i = \big(\frac{g'(\mu_i)}{g(\mu_i)}\big)^2 - \frac{g''(\mu_i)}{g(\mu_i)} = -\ln q_{mi}.$$

Using these linear approximations, the likelihood equations $E1$ and $E2$ in (11) can be written as

$$\begin{cases} E1 : -\frac{1}{\sigma}\sum_{i=1}^{n}(\alpha_i - \beta_i z_i) = 0, \\ \\ E2 : -\frac{n}{\sigma} - \sum_{i=1}^{n}\frac{z_i}{\sigma}(\alpha_i - \beta_i z_i) = 0. \end{cases} \tag{13}$$

Substituting $z_i = \frac{m_i - \mu}{\sigma}$ into E1 of the Equation (13) and after algebraic simplification, the AMLE of μ is obtained as $\check{\mu} = A_L - \check{\sigma} B_L$, where

$$A_L = \frac{\sum_{i=1}^{n} m_i \beta_i}{\sum_{i=1}^{n} \beta_i}, \quad B_L = \frac{\sum_{i=1}^{n} \alpha_i}{\sum_{i=1}^{n} \beta_i}. \tag{14}$$

Similarly, substituting $z_i = \frac{m_i - \mu}{\sigma}$ into E2 of the Equation (13) and simplifying, it can be transformed to the approximate likelihood equation form of σ as $n\sigma^2 + D_L\sigma - F_L = 0$, where

$$\begin{aligned} D_L &= \sum_{i=1}^{n} \alpha_i (m_i - A_L), \\ F_L &= \sum_{i=1}^{n} \beta_i (m_i - A_L)^2 > 0. \end{aligned} \tag{15}$$

Therefore, (15) yields the AMLE of σ to be

$$\check{\sigma} = \frac{-D_L + \sqrt{D_L^2 + 4nF_L}}{2n}, \tag{16}$$

which is the only positive root (see Bayat Mokhtari et al. (2011)).

2.3 How to Predict X

Let $[L_i^*, R_i^*]$ $i = 1, 2, \ldots, n$ be the interval censored survival data where L_i^* is the last observed time for the i^{th} individual before the event Y_i has occurred and R_i^* indicates the first time the event Y_i has been observed. According to Self and Grossman (1986), suppose that censoring occurs noninformative in the sense that for any y, l_i^*, r_i^* such that $l_i^* \leq y \leq r_i^*$, the conditional density of Y given L_i^* and R_i^*, $f_{[Y|L_i^*,R_i^*]}(y|l_i^*; r_i^*)$, satisfies

$$f_{[Y|L_i^*,R_i^*]}(y|l_i^*; r_i^*) = \frac{f_Y(y)}{P(Y \in [l_i^*, r_i^*])} = \frac{f_Y(y)}{F_Y(r_i^*) - F_Y(l_i^*)},\ i = 1, 2, \ldots, n. \tag{17}$$

Therefore, after substituting mid-point approximate estimators or AMLEs of unknown parameters, the expected value of the conditional distribution of Y given L_i^* and R_i^* can be introduced as a predictor of Y. Let the predictor of failed unit Y^* in the interval $[L_i^*, R_i^*]$ under the extreme value distribution be

$$\begin{aligned} \hat{Y_i^*} = E_{[Y^*|L_i^*,R_i^*]}(y^*|l_i^*; r_i^*) &= \frac{E(Y^*)}{e^{-e^{(\frac{l_i^*-\mu}{\sigma})}} - e^{-e^{(\frac{r_i^*-\mu}{\sigma})}}} \\ &= \frac{\int_{l_i^*}^{r_i^*} \frac{y}{\sigma} e^{(\frac{y^*-\mu}{\sigma}) - e^{(\frac{y^*-\mu}{\sigma})}}\, dy^*}{e^{-e^{(\frac{l_i^*-\mu}{\sigma})}} - e^{-e^{(\frac{r_i^*-\mu}{\sigma})}}},\ i = 1, 2, \ldots, n. \end{aligned} \tag{18}$$

Then, a predictor for Y can be introduced as

$$\bar{Y}^* = \frac{1}{n} \sum_{i=1}^{n} \hat{Y_i^*} \tag{19}$$

It is clear that by transformations $\bar{X} = \exp(\bar{Y}^*)$, $\hat{\beta} = 1/\hat{\sigma}$ and $\hat{\alpha} = \exp(-\hat{\mu}/\hat{\sigma})$, the corresponding predictor of X and the estimators (mid-point or AMLE) of α and β can be found, respectively. Also, using these estimators (mid-point or AMLE) of α and β, the predictor of X and the percentile bootstrap (p-boot) method, we can get the 95% confidence bounds of α and β as well

as 95% prediction bounds of X.

It is worthwhile to mention that if we denote the 95% confidence bounds of α, β and 95% prediction bounds of Y by $\big(l(\hat{\mu}), u(\hat{\mu})\big)$, $\big(l(\hat{\sigma}), u(\hat{\sigma})\big)$ and $\big(l(\bar{Y}^*), u(\bar{Y}^*)\big)$, respectively, then $\big((u(\hat{\mu}))^{-1}, (l(\hat{\mu}))^{-1}\big)$ and $\big(\exp\{l(\bar{Y}^*)\}, \exp\{u(\bar{Y}^*)\}\big)$ will be the 95% confidence bounds of β and prediction bounds of X based on the Weibull distribution, respectively. Also, if $\frac{l(\hat{\mu})}{u(\hat{\sigma})} < \frac{u(\hat{\mu})}{l(\hat{\sigma})}$, then the corresponding 95% confidence bounds of α will be $\Big(\exp\{-\frac{u(\hat{\mu})}{l(\hat{\sigma})}\}, \exp\{-\frac{l(\hat{\mu})}{u(\hat{\sigma})}\}\Big)$, otherwise the 95% confidence bounds of α can be obtained as $\Big(\exp\{-\frac{u(\hat{\mu})}{u(\hat{\sigma})}\}, \exp\{-\frac{l(\hat{\mu})}{l(\hat{\sigma})}\}\Big)$.

3 Numerical Studies

In this section, the performance of the proposed procedures is investigated by a simulation study and a real data set. All codes in this numerical study are conducted in **R** software (**R** i386 4.1.2) and they can be obtained from the authors upon request.

3.1 Simulation Results

This subsection is devoted to testing the performance of the obtained predictor of X, based on a sample of size n from intervals $[L_i, R_i]$ $(i = 1, 2, \ldots, n)$, the AMLEs of α and β, the prediction bounds of X as well as the 95% confidence bounds of α and β. The predictor of X and its mean square prediction errors (MSPEs), as well as $\hat{\alpha}$, $\hat{\beta}$ and their mean squared errors (MSEs) are calculated. In addition, the average widths (AWs) for the prediction bounds of X as well as 95% confidence bounds of α and β are computed.

Each data set contains $n = 25, 50, 100$ interval-censored observations. Let the true survival time follow an extreme value distribution.

It is worthwhile to mention that there are some generation methods of interval-censored data with similar consequences (see e.g. Kiani and Arasan (2012)). Here, the generation of interval-censored is according to Guure et al. (2013) and Guure et al. (2015). It is involved the following steps.

(1) Generate a single observation from the extreme value distribution say, y with initial values of $\mu = 1$ and $\sigma = 0.5, 1.5$ (arbitrary initial values of μ and σ).

(2) Generate a vector say V, for a set of clinic visits. Assume there are 10 clinic visits, for the extreme value distribution, take the first visit to be $V[1]$ and generate it from $U(0, b)$(b is an arbitrary natural number which is clinical visit step). The next visit of $V[2]$ is also generated from $U(V[1], V[1] + b)$. Subsequent generations are carried out with a similar approach.

(3) Generate a set of matrix called bounds for each of the data set. The lower and upper bounds can be defined as:

$$Bounds[i,1] = \begin{cases} 0 & if \quad y[i] < V[1], i = 1, 2, \ldots, n, \\ V[j] & if \quad V[j] < y[i] < V[j+1], \ where \ j = 1, 2, \ldots, 9, \\ V[10] & if \quad y[i] > V[10], \end{cases} \tag{20}$$

$$Bounds[i,2] = \begin{cases} V[1] & if \quad y[i] < V[1], i = 1, 2, \ldots, n, \\ V[j+1] & if \quad V[j] < y[i] < V[j+1], \ where \ j = 1, 2, \ldots, 9, \\ 1000 & if \quad y[i] > V[10]. \end{cases} \tag{21}$$

(4) An indicator is defined such that

$$Indicator[i] = \begin{cases} 0 & if \quad Bounds[i,2] = 1000, \\ 1 & otherwise. \end{cases} \tag{22}$$

(5) The predictor of Y is given according to Equation (19). Also, the mid-point estimators and AMLEs of μ and σ are computed.

(6) For $nboot = 200$ bootstrap iterations, find the two-sided 95% percentile bootstrap confidence bounds of Y and the 95% percentile bootstrap confidence bounds of μ, σ.

(7) Steps (1)-(6) are repeated $B = 3 \times 10^3$ times.

(8) Suitable mentioned transformations carried out to get the corresponding predictor of X, the estimators (mid-point as well as AMLE) of α and β, the 95% confidence bounds of α, β and 95% prediction bounds of X.

(9) The average point predictor of X, the mean square prediction errors (MSPEs) of $\hat{X}$ along with mid-point estimators and AMLEs of α and β, the MSEs for α and β and the AWs of all confidence intervals are reported. The results with $B = 3 \times 10^3$ Monte Carlo replications and 200 bootstrap samples are presented in Tables 1 and 2.

Remark: To apply AML method in both the Monte Carlo simulation study

Table 1. Simulation results for $b = 2$, $n = 25, 50, 100$ with $\alpha = 0.818, 0.406, 0.156$ and $\beta = 1, 1.428$ ($\mu = 0.2, 0.9, 1.3$ and $\sigma = 0.7, 1$), via AMLE method for parameters with $nboot = 200$ bootstrap iterations.

			Estimate (MSPE)	Estimate (MSE)		AW		
α	β	n	$\hat{X}$	$\hat{\alpha}$	$\hat{\beta}$	X	α	β
0.156	1.428	25	3.219(0.923)	0.055(0.010)	1.546(0.081)	4.197	0.088	0.809
		50	3.254(0.950)	0.056(0.010)	1.482(0.027)	4.386	0.061	0.534
		100	3.215(0.884)	0.057(0.009)	1.450(0.011)	4.361	0.043	0.366
0.406	1	25	2.615(1.072)	0.078(0.107)	1.538(0.414)	2.971	0.108	0.819
		50	2.583(1.076)	0.082(0.105)	1.438(0.232)	2.987	0.075	0.518
		100	2.608(1.074)	0.084(0.103)	1.393(0.172)	3.099	0.052	0.353
0.818	1	25	1.753(0.889)	0.066(0.567)	2.580(3.261)	1.359	0.134	1.575
		50	1.742(0.903)	0.073(0.555)	2.303(2.052)	1.323	0.095	0.876
		100	1.753(0.876)	0.078(0.549)	2.153(1.464)	1.396	0.067	0.552

and the following illustrative example, m_i's should be arranged in ascending order.

From Tables 1 and 2, it is interesting to see that the larger sample n, the smaller MSPE of predictor X and the shorter length of 95% prediction bounds for X. Also, in both mid-point and AML methods, as $\alpha = 0.156$ and $\beta = 1.428$ ($\sigma = 0.7$ and $\mu = 1.3$), it is clear that the smaller MSPE of predictor X and MSE of $\hat{\alpha}$ as well as $\hat{\beta}$, roughly. When $\alpha = 0.818$ and $\beta = 1$ ($\sigma = 1$ and $\mu = 0.2$), the shorter length of 95% prediction bounds for X under mid-point and AML methods, can be resulted. On the other hand, we can observe that the AML method, for almost all sample sizes and selected values of α and β, tends to give better results than the mid-point method for the MSPE of predictor X, the AWs of confidence intervals of α and β and the MSE of $\hat{\beta}$. However, this situation is reversed for the MSE of $\hat{\alpha}$ and the AW of prediction interval for X for almost selected values of α and β under all sample sizes. As expected from the simulation study, Figure 1 presents the mid-point method is the better method for estimating α. Figure 2 displays that for estimation of β the AML method and mid-point method are not much different, although the AML method is slightly better. Finally, Figure 3 shows the AML method often performs better than the mid-point method.

Table 2. Simulation results for $b = 2$, $n = 25, 50, 100$ with $\alpha = 0.818, 0.406, 0.156$ and $\beta = 1, 1.428$ ($\mu = 0.2, 0.9, 1.3$ and $\sigma = 0.7, 1$), via mid-point method for parameters with $nboot = 200$ bootstrap iterations.

			Estimate (MSPE)	Estimate (MSE)		AW		
α	β	n	$\hat{X}$	$\hat{\alpha}$	$\hat{\beta}$	X	α	β
0.156	1.428	25	3.198(0.921)	0.114(0.002)	1.562(0.083)	2.861	0.208	0.924
		50	3.164(0.869)	0.119(0.001)	1.497(0.030)	2.860	0.152	0.579
		100	3.203(0.900)	0.122(0.001)	1.464(0.012)	2.961	0.110	0.389
0.406	1	25	2.600(1.184)	0.162(0.061)	1.546(0.431)	2.029	0.243	1.161
		50	2.622(1.093)	0.176(0.054)	1.433(0.227)	2.076	0.182	0.671
		100	2.651(1.149)	0.179(0.051)	1.398(0.176)	2.159	0.131	0.444
0.818	1	25	1.763(0.911)	0.139(0.465)	2.572(3.261)	0.915	0.237	2.449
		50	1.758(0.885)	0.161(0.434)	2.240(1.874)	0.900	0.191	1.417
		100	1.753(0.901)	0.175(0.415)	2.075(1.289)	0.893	0.145	0.887

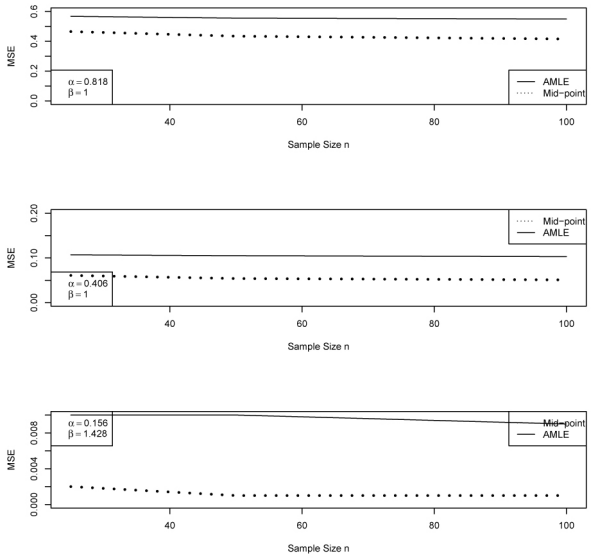

Figure 1. Plots of MSEs for α estimates versus sample size n under different values of α and β with AML and mid-point methods.

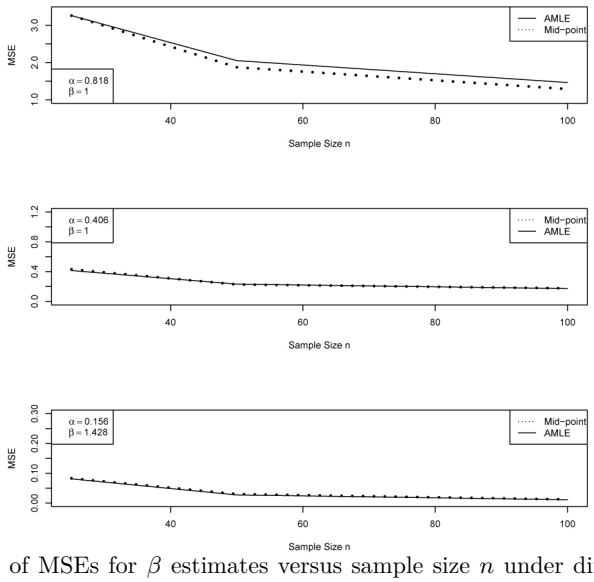

Figure 2. Plots of MSEs for β estimates versus sample size n under different values of α and β with AML and mid-point methods.

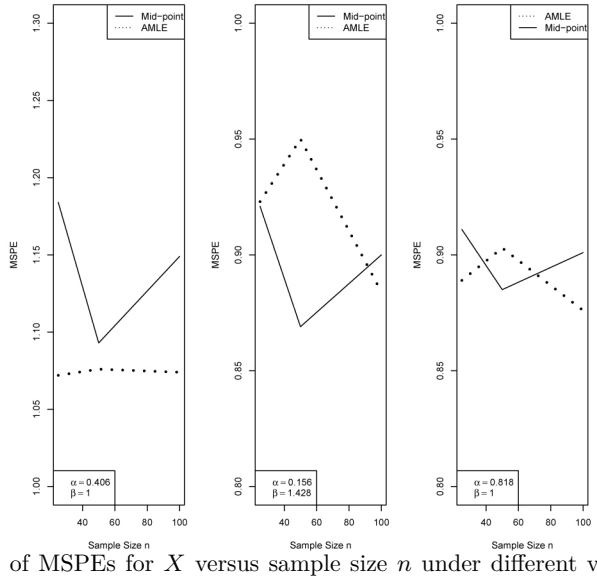

Figure 3. Plots of MSPEs for X versus sample size n under different values of α and β with AML and mid-point methods.

Table 3. Radiotherapy and chemotherapy data.

(8, 12]	(0, 5]	(30, 34]
(0, 22]	(5, 8]	(13, ∞]
(24, 31]	(12, 20]	(10, 17]
(17, 27]	(11, ∞]	(8, 21]
(17, 23]	(33, 40]	(4, 9]
(24, 30]	(31, ∞]	(11, ∞]
(16, 24]	(13, 39]	(14, 19]
(13, ∞]	(19, 32]	(4, 8]
(11, 13]	(34, ∞]	(34, ∞]
(16, 20]	(13, ∞]	(30, 36]
(18, 25]	(16, 24]	(18, 24]
(17, 26]	(35, ∞]	(16, 60]
(32, ∞]	(15, 22]	(35, 39]
(23, ∞]	(11, 17]	(21, ∞]
(44, 48]	(22, 32]	(11, 20]
(14, 17]	(10, 35]	(48, ∞]

3.2 Real Data Analysis

In this section, a data set from Guure et al. (2013) and Guure et al. (2015) for illustration and comparative purposes is analyzed. The data is a retrospective study taken from Lawless (2003). Initially, it was implemented to compare the cosmetic effects of radiotherapy versus radiotherapy and adjuvant chemotherapy on women with early breast cancer. Here, only a retrospective study of 48 radiation plus chemotherapy patients was considered. Patients were examined initially for every 46 months, but, when their recovery progressed, the interval between visits lengthened. The desired event was the time for the first appearance of moderate (severe) breast retraction. When the patients were examined only at some random times, the exact time, X_i, of breast retraction is only discovered to fall within the interval between visits.

Patients with no moderate (severe) breast retraction until the last visit were known as right-censored data. Then the end point of their intervals was presumed to be $R_i = \infty$ and L_i was presumed as the time from the beginning to the last visit. To achieve the objective of this paper, the data have been adapted to exclude right-censored observations. The data are presented in

Table 4. Numerical results for radiotherapy and chemotherapy data under mid-point and AML methods with $nboot = 200$ bootstrap iterations.

	Estimates			95%Confidence interval (Width)		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{X}$	α	β	X
AML	0.001	1.627	27.320	(0.000, 0.006)(0.006)	(1.310,2.147)(0.836)	(7.329, 64.992)(57.663)
Mid-point	0.004	1.654	18.162	(0.000, 0.021)(0.021)	(1.291, 2.214)(0.923)	(5.142, 44.453)(39.310)

Table 3.

First, the Kolmogorov-Smirnov (K-S) procedure was implemented to test whether the Weibull model with 0.001 (scale) and 2.029 (shape) parameters fitted to the data of Table 3 (using some R packages for interval-censored such as "interval", "survival", "fitdistrplus" and "Icens"). The corresponding Kolmogorov–Smirnov (K-S) statistic, i.e. D_n is 0.162355 and the K-S P-value obtained 0.997. For the Type-1 errors, 0.05 and 0.1, the critical values of K-S test table are shown that the Weibull distribution fits the data of Table 3. It can be seen $n = 35$ (due to discarding infinite upper bounds). Based on initial arbitrary value $b = 2$ and with 10 clinic visits (arbitrary), under $nboot = 200$ bootstrap iterations, the results are displayed in Table 4.

The above results confirm the simulation results which emphasize the AML method has better results for getting the AWs of confidence intervals of α and β and the mid-point method has the better performance in comparison to the AML method.

4 Conclusion

In this paper, the concentration was on the prediction of a failure time that cannot be observed directly but can only be determined to lie in an interval $[L, R]$. Based on a sample of size n from intervals $[L_i, R_i]$ $i = 1, 2, \ldots, n$, the AMLEs for unknown Weibull parameters were obtained. Then, we found that the prediction of failed unit X, using the conditional distribution of X given L and R is quite straightforward. A simulation study based on the Monte Carlo and the p-boot method is carried out. Moreover, when n increases, as we would expect, better results (the smaller MSPE of predictor X and length of 95% prediction bounds for X), especially under the AML approximation method can be concluded. It is even possible to use

other methods for estimation and prediction, such as the Bayesian method to compare with the proposed method for the estimation of parameters and prediction of the random variable X.

Acknowledgements

The authors thank the Associate Editor and anonymous referees for their comments and suggestions that helped to improve the earlier version of the paper.

References

Aggarwala, R. (2001). Progressive Interval Censoring: Some Mathematical Results with Applications to Inference, *Comm. Statist. Theory Methods*, **30**, 1921-1935.

Ahmadi, K. and Yousefzadeh, F. (2015). Estimation for the Parameters of Generalized Half-Normal Distribution based on Progressive Type-I Interval Censoring, *Comm. Statist. Simulation Comput.*, **44**, 2671-2695.

Basak, I., Basak, P. and Balakrishnan, N. (2006). On some Predictors of Times to Failure of Censored Items in Progressively Censored Samples, *Comput. Statist. Data Anal.*, **50**, 1313-1337.

Bayat Mokhtari, E., Habibi Rad, A. and Yousefzadeh, F. (2011). Inference for Weibull Distribution based on Progressively Type-II Hybrid Censored Data, *J. Statist. Plann. Inference*, **141**, 2824-2838.

Chen, D.G. and Lio, Y.L. (2010). Parameter Estimations for Generalized Exponential Distribution under Progressive Type-I Interval Censoring, *Comput. Statist. Data Anal.*, **54**, 1581-1591.

Dumbgen, L., Freitag-Wolf, S. and Jongbloed, G. (2006). Estimating a Unimodal Distribution from Interval-Censored Data, *J. Amer. Statist. Assoc.*, **101**, 1094-1106.

Flygare, M.E. and Buckwalter, J.A. (2010). Maximum Likelihood Estimation for the 2-parameter Weibull Distribution based on Interval Data, *IEEE Trans. Reliab.*, **34**, 57-60.

Garcia-Mora, B., Debon, A., Santamaria, C. and Carrion, A. (2015). Modelling the Failure Risk for Water Supply Networks with Interval-censored Data, *Reliability Engineering and System Safety*, **144**, 311-318.

Gomez, G., Calle, M.L. and Oller, R. (2004). Frequentist and Bayesian Approaches for Interval-censored Data, *Statist. Papers*, **45**, 139-173.

Guure, C.B., Ibrahim, N.A. and Adam, M.B. (2013). Bayesian Inference of the Weibull Model based on Interval Censored Survival Data, *Comput. Math. Methods Med.*, ID 849520, 10 pages.

Guure, C.B., Ibrahim, N.A., Dwomoha, D. and Bosompraha, S. (2015). Bayesian Statistical Inference of the Loglogistic Model with Interval Censored Lifetime Data, *J. Stat. Comput. Simul.*, **85**, 1567-1583.

Hashimoto, E.M., Ortega, E.M.M., Cancho, V.G. and Cordeiro, G.M. (2013). On Estimation and Diagnostics Analysis in Log-generalized Gamma Regression Model for Interval-censored Data, *Statistics; A Journal of Theoretical and Applied Statistics*, **47**, 379 -398.

Hyun, N., Cheung, L.C., Pan, Q., Schiffman, M. and Katki, H.A. (2017). Flexible Risk Prediction Models for Left or Interval-Censored Data from Electronic Health Records, *Annals of Applied Statistics*, **11**, 1063-1084.

Kiani, K. and Arasan, J. (2012). Simulation of Interval Censored Data in Medical and Biological Studies, *International Journal of Modern Physics: Conference Series*, International Conference Mathematical and Computational Biology 2011, **9**, 112-118.

Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, New York, John Wiley.

Lawless, J.F. and Babineau, D. (2006). Models for Interval Censoring and Simulation-based Inference for Lifetime Distributions, *Biometrika*, **93**, 671-686.

Lin, Y.J. and Lio, Y.L. (2012). Bayesian Inference under Progressive Type-I Interval Censoring, *J. Appl. Stat.*, **39**, 1811-1824.

Lindsey, J.K. (1998). A study of Interval Censoring in Parametric Regression Models, *Lifetime Data Anal.*, **4**, 329-354.

Lindsey, J.C. and Ryan, L.M. (1998). Tutorial in Biostatistics Methods for Interval Censored Data, *Stat. Med.*, **17**, 219-238.

Lio, Y.L., Chen, D.J. and Tsai, T.R. (2011). Parameter Estimations for Generalized Rayleigh Distribution under Progressively Type-I Interval Censored Data, *American Open Journal of Statistics*, **1**, 46-57.

Odell, P.M., Anderson, K.M. and D'Agostino, R.B. (1992). Maximum Likelihood Estimation for Interval Censored Data using a Weibull based Accelerated Failure Time Model, *Biometrics*, **48**, 951- 959.

Ng, H.K.T. and Wang, Z. (2009). Statistical Estimation for the Parameters of Weibull Distribution based on Progressively Type-I Interval Censored Sample, *J. Stat. Comput. Simul.*, **79**, 145-159.

Peng, Y., Wang, Y., Zi, Y.Y., Tsui, K.L. and Zhang, C. (2017). Dynamic Reliability Assess-

ment and Prediction for Repairable Systems with Interval-censored Data, *Reliability Engineering and System Safety*, **159**, 301-309.

Pradhan, B. and Kundu D. (2014). Analysis of Interval-Censored Data with Weibull Lifetime Distribution, *Sankhya: The Indian Journal of Statistics*, **76**B (Part-1), 120-139.

Raqab, M.Z. (2004). Approximate Maximum Likelihood Predictors of Future Failure Times of Shifted Exponential Distributions under Multiple Type II Censoring, *Stat. Methods Appl.*, **13**, 41-52.

Raqab, M.Z., Asgharzadeh , A. and Valiollahi, R. (2010). Prediction for Pareto Distribution based on Progressively Type-II Censored Samples, *Comput. Statist. Data Anal.*, **54**, 1732-1743.

Scallan, A.J. (1999). Regression Modelling of Interval Censored Failure Time Data using the Weibull Distribution, *J. Appl. Stat.*, **26**, 613-618.

Self, S.G. and Grossman, E.A. (1986). Linear Rank Tests for Interval Censored Data with Application to PCB Levels in Adipose Tissue of Transformer Repair Workers, *Biometrics*, **42**, 521-530.

Shen, P.S. (2014). Proportional Hazards Regression with Interval-Censored and Left-Truncated Data, *J. Stat. Comput. Simul.*, **84**, 264-272.

Singh, R.S. and Totawattage, D.P. (2013).The Statistical Analysis of Interval-Censored Failure Time Data with Applications, *Open Journal of Statistics*, **3**, 155-166.

Sun, J. (1997). Regression Analysis of Interval Censored Failure Time Data, *Stat. Med.*, **16**, 497-504.

Sun, J. (2006). *The Statistical Analysis of Interval Censored Failure Time Data*, New York, Springer.

Turnbull, B.W. (1976). The Empirical Distribution with Arbitrarily Grouped Censored and Truncated Data, *J. Roy. Statist. Soc. Ser. B*, **38**, 290-295.

Tomazella, V.L.D. and Nadarajah, S. (2015). Estimation of parameters in Laplace Distributions with Interval Censored Data, *Braz. J. Probab. Stat*, 2**9**, 677-694.

Wu, Y. and Cook, R.J. (2020). Assessing the Accuracy of Predictive Models with Interval-Censored Data, *Biostatistics*, doi:10.1093/biostatistics/kxaa011.

Xiang, L. and Tse, S.K. (2005). Maximum Likelihood Estimation in Survival Studies under Progressive Interval Censoring with Random Removals, *Journal of Biopharmaceutical Statistics*, **15**, 981-991.

Yao, W., Frydman, H. and Simonoff, J.S. (2021). An Ensemble Method for Interval-Censored Time-to-Event Data, *Biostatistics*, **22**, 198-213.

Zhang, Z. (2009). Linear Transformation Models for Interval-Censored Data: Prediction of Survival Probability and Model Checking , *Statistical Modelling*, **9**, 321-343.

Somayeh Ghafouri
Department of Mathematics,
Faculty of Sciences,
Arak University,
Arak, Iran.
email: *somayyeh.ghafoori@gmail.com*

Arezou Habibirad
Department of Statistics,
School of Mathematical Sciences,
Ferdowsi University of Mashhad,
Mashhad, Iran.
email: *ahabibi@um.ac.ir*

Fatemeh Yousefzadeh
Department of Statistics,
School of Mathematical Sciences,
University of Birjand,
Birjand, Iran.
email: *fyousefzadeh@birjand.ac.ir*

Reza Pakyar
Department of Mathematics, Statistics and Physics,
College of Arts and Sciences,
Qatar University,
Doha, Qatar.
email: *rpakyari@qu.edu.qa*