AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(10): 28508–28525.
DOI: 10.3934/math.20241383
Received: 24 July 2024
Revised: 28 September 2024
Accepted: 30 September 2024
Published: 08 October 2024

Research article

Predictive analysis of doubly Type-II censored models

Young Eun Jeon[1], **Yongku Kim**[2,3,*] **and Jung-In Seo**[1,*]

[1] Department of Data Science, Andong National University, Andong, 36729, Korea
[2] Department of Statistics, Kyungpook National University, Daegu, 41566, Korea
[3] KNU G-LAMP Research Center, Institute of Basic Sciences, Kyungpook National University, Daegu, 41566, Korea

* **Correspondence:** Email: kim.1252@knu.ac.kr, leehoo1928@gmail.com.

Abstract: The application of a doubly Type-II censoring scheme, where observations are censored at both the left and right ends, is often used in various fields including social science, psychology, and economics. However, the observed sample size under this censoring scheme may not be large enough to apply a likelihood-based approach due to the occurrence of censoring at both ends. To effectively respond to this difficulty, we propose a pivotal-based approach within a doubly Type-II censoring framework, focusing on two key aspects: Estimation for parameters of interest and prediction for missing or censored samples. The proposed approach offers two prominent advantages, compared to the likelihood-based approach. First, this approach leads to exact confidence intervals for unknown parameters. Second, it addresses prediction problems in a closed-form manner, ensuring computational efficiency. Moreover, novel algorithms using a pseudorandom sequence, which are introduced to implement the proposed approach, have remarkable scalability. The superiority and applicability of the proposed approach are substantiated in Monte Carlo simulations and real-world case analysis through a comparison with the likelihood-based approach.

Keywords: doubly Type-II censoring scheme; generalized confidence interval; likelihood-based inference; pivotal quantity; prediction
Mathematics Subject Classification: 62F10, 62N01, 62N02

1. Introduction

In many practical studies, extreme sample values are often excluded when the observations are not well known or the presence of outliers is suspected. In addition, carelessness or lack of experimental preparation at the beginning of an experiment can lead to the censoring of the smallest few values in the observed sample, and limited time or cost issues can lead to experiment termination without observing the largest values. When both conditions occur simultaneously, to be more accurate, a scenario where the smallest and largest values in an ordered sample are censored at both ends is called a doubly Type-II (DT-II) censoring scheme. Under this censoring scheme, the observed sample size is not often sufficient to apply valid statistical inference, requiring a very challenging task. This challenge has made the DT-II censoring scheme less studied than other censoring schemes, and most of the studies related to it have focused on likelihood-based and Bayesian inference. For example, Fernández [1] introduced a Bayesian inference for a Rayleigh distribution with a scale parameter. Raqab and Madi [2] discussed the prediction problem for the total amount of the remaining testing time from the same distribution and resolved it using Gibbs sampling. Khan et al. [3] provided a prediction for a single future response in a two-parameter Rayleigh distribution, using a Bayesian approach. Kim and Song [4] discussed parameter estimation for a generalized exponential distribution using the likelihood-based and Bayesian methods. Kotb and Raqab [5] derived maximum likelihood estimators (MLEs) as well as Bayes estimators for a modified Weibull distribution. Panahi [6] provided likelihood-based and Bayesian estimation methods for a Burr Type XII distribution. Sindhu and Hussain [7] dealt with the estimation problem of a power function II distribution, using non-informative and informative priors in a Bayesian approach. Besides these methods, some studies have employed a pivotal quantity to obtain a joint confidence region for unknown parameters under the DT-II censoring scheme. Wu [8] obtained a joint confidence region in an exponential distribution with two parameters. Wu [9] discussed interval inference for a Pareto distribution with two parameters.

However, the likelihood-based approach has the drawback of being able to cause substantial bias and invalid inference results for a small sample size. In particular, this approach for interval inference yields approximate, rather than exact, confidence intervals (CIs) for unknown parameters, and these CIs may not meet a nominal level in small sample scenarios. On the other hand, the Bayesian approach has stress related to the elicitation of prior information for each parameter and requires very demanding calculations such as the expected Fisher information for objective Bayesian inference if there is not enough prior information.

To provide solutions to these challenging problems in a frequentist approach, this study proposes a pivotal-based approach under the DT-II censoring scheme, yielding valid statistical inference results even in a situation where the available sample size falls short of being sufficiently large. Notably, the proposed approach leads to not only exact CIs for unknown parameters but also closed-form results in prediction problems for missing or censored samples, unlike the likelihood-based approach. Furthermore, this study introduces novel algorithms using a pseudorandom sequence to implement the proposed approach, which is attractive due to its excellent scalability. The superiority and validity of the proposed approach are demonstrated in the simulation study and real-world case analysis, showing that its performance does not lag behind and rather performs better compared with the most popular likelihood-based approach for small and medium sample sizes.

The study is outlined as follows: Section 2 briefly describes a distribution family used for inference and the DT-II censoring scheme. Section 3 provides an estimation method for unknown parameters and a prediction method for missing or censored samples based on a likelihood function within the DT-II censoring framework. Section 4 proposes the parameter estimation and prediction methods using pivotal quantities within the same framework. Section 5 demonstrates the superiority and validity of the proposed method by applying the theoretical results to simulated and real-life datasets. Finally, Section 6 concludes the study with a concise recapitulation of the results.

2. Model description

Let $\Theta = (\alpha, \beta)$ be an unknown parameter vector. To implement our argument, it is assumed that a continuous random variable X has the following family of distributions with Θ:

$$F(x; \Theta) = 1 - [G(x; \alpha)]^{h(\beta)}, \quad x > 0,\ \alpha > 0,\ \beta > 0, \tag{2.1}$$

where $G(x; \alpha)$ denotes a function of X involving α, and $h(\beta)$ denotes a function of β. The distribution (2.1) is a cumulative distribution function (CDF) of X, and its members are given in Appendix (Table 5).

Suppose that $X_1, \ldots, X_n$ is a random sample with CDF (2.1), and $X_{(1)} \le \cdots \le X_{(n)}$ is an order statistic from this sample. In experiments concerning life-testing or reliability, negligence or poor preparation at the start of the experiment may result in censoring the first few observations, while the limitation of the experiment duration may result in censoring the last few observations. Specifically, when the r smallest and the s largest observations are simultaneously censored in a situation where n identical components are placed in the test, it is called a DT-II censoring scheme. In addition, the remaining $\boldsymbol{X} = (X_{(r+1)}, \ldots, X_{(n-s)})$ can be regarded as a DT-II censored sample.

3. Likelihood-based inference

This section provides the most widely used likelihood-based approach for comparison with the pivotal-based approach to be proposed in Section 4.

3.1. Estimation

The likelihood and its logarithm functions for the DT-II censored sample $\boldsymbol{X}$ from the probability distribution with CDF (2.1) are formulated as

$$L(\Theta) \propto \left(1 - [G(x_{(r+1)}; \alpha)]^{h(\beta)}\right)^r [G(x_{(n-s)}; \alpha)]^{sh(\beta)} [h(\beta)]^{n-s-r} \prod_{i=r+1}^{n-s} [G(x_{(i)}; \alpha)]^{h(\beta)-1} [-g(x_{(i)}; \alpha)], \tag{3.1}$$

and

$$\begin{aligned} \ell(\Theta) \propto\ & r \log\left(1 - [G(x_{(r+1)}; \alpha)]^{h(\beta)}\right) + sh(\beta) \log G(x_{(n-s)}; \alpha) + (n - s - r) \log h(\beta) \\ & + (h(\beta) - 1) \sum_{i=r+1}^{n-s} \log G(x_{(i)}; \alpha) + \sum_{i=r+1}^{n-s} \log [-g(x_{(i)}; \alpha)], \end{aligned} \tag{3.2}$$

respectively, where $\ell(\cdot) = \log L(\cdot)$, and $g(x_{(i)}; \alpha) = \partial G(x_{(i)}; \alpha)/\partial x_{(i)}$. Then, the MLEs $\hat{\alpha}$ and $\hat{\beta}$ are obtained by optimizing the log-likelihood function (3.2) for α and β, respectively. In addition, by the asymptotic normality of the MLEs, the approximate $100(1 - \gamma)\%$ CIs are formulated as

$$\hat{\alpha} \pm z_{\gamma/2} \sqrt{\widehat{\text{Var}}(\hat{\alpha})},$$

and

$$\hat{\beta} \pm z_{\gamma/2} \sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)},$$

for $0 < \gamma < 1$, where $z_{\gamma/2}$ is the $(\gamma/2)$th upper percentile of a standard normal distribution, and $\widehat{\mathrm{Var}}(\cdot)$ is the diagonal element of the asymptotic variance-covariance matrix of the MLEs

$$\widehat{\Sigma} = \begin{pmatrix} \widehat{\mathrm{Var}}\,(\hat{\alpha}) & \widehat{\mathrm{Cov}}\left(\hat{\alpha}, \hat{\beta}\right) \\ \widehat{\mathrm{Cov}}\left(\hat{\alpha}, \hat{\beta}\right) & \widehat{\mathrm{Var}}\left(\hat{\beta}\right) \end{pmatrix},$$

which is obtained by inverting the observed Fisher information matrix with the MLEs

$$I\left(\hat{\alpha}, \hat{\beta}\right) = \left.\begin{pmatrix} -\frac{\partial^2 \ell(\Theta)}{\partial \alpha^2} & -\frac{\partial^2 \ell(\Theta)}{\partial \alpha \partial \beta} \\ -\frac{\partial^2 \ell(\Theta)}{\partial \alpha \partial \beta} & -\frac{\partial^2 \ell(\Theta)}{\partial \beta^2} \end{pmatrix}\right|_{\left(\alpha=\hat{\alpha},\, \beta=\hat{\beta}\right)}.$$

3.2. Prediction

In the presence of the DT-II censored sample, the predictive likelihood function requires two conditional probability density functions (PDFs) because censoring occurs at both ends. One is the conditional PDF of $X_{(l)}$ given $X_{(r+1)} = x_{(r+1)}$ for $l = 1, \ldots, r$, and the other is the conditional PDF of $X_{(l)}$ given $X_{(n-s)} = x_{(n-s)}$ for $l = n - s + 1, \ldots, n$. The former is formulated as

$$\begin{aligned} f\left(x_{(l)} \mid X_{(r+1)} = x_{(r+1)}\right) = & \frac{r!}{(l-1)!(r-l)!} \frac{h(\beta)\left[G(x_{(l)};\alpha)\right]^{h(\beta)-1}\left[-g(x_{(l)};\alpha)\right]}{1-\left[G(x_{(r+1)};\alpha)\right]^{h(\beta)}} \\ & \times \left\{\frac{\left[G(x_{(l)};\alpha)\right]^{h(\beta)} - \left[G(x_{(r+1)};\alpha)\right]^{h(\beta)}}{1-\left[G(x_{(r+1)};\alpha)\right]^{h(\beta)}}\right\}^{r-l} \left\{\frac{1-\left[G(x_{(l)};\alpha)\right]^{h(\beta)}}{1-\left[G(x_{(r+1)};\alpha)\right]^{h(\beta)}}\right\}^{l-1}. \end{aligned} \tag{3.3}$$

The latter is formulated as

$$\begin{aligned} f\left(x_{(l)} \mid X_{(n-s)} = x_{(n-s)}\right) = & \frac{s!}{(l-n+s-1)!(n-l)!} \frac{h(\beta)\left[G(x_{(l)};\alpha)\right]^{h(\beta)-1}\left[-g(x_{(l)};\alpha)\right]}{\left[G(x_{(n-s)};\alpha)\right]^{h(\beta)}} \\ & \times \left\{\frac{\left[G(x_{(n-s)};\alpha)\right]^{h(\beta)} - \left[G(x_{(l)};\alpha)\right]^{h(\beta)}}{\left[G(x_{(n-s)};\alpha)\right]^{h(\beta)}}\right\}^{l-n+s-1} \left[\frac{G(x_{(l)};\alpha)}{G(x_{(n-s)};\alpha)}\right]^{(n-l)h(\beta)}. \end{aligned} \tag{3.4}$$

Then, under the DT-II censoring scheme, prediction is made from the following predictive likelihood function:

$$L(x_{(l)}, \Theta) \propto \begin{cases} [h(\beta)]^{n-s-r+1}\left[G(x_{(l)};\alpha)\right]^{h(\beta)-1}\left[-g(x_{(l)};\alpha)\right]\left[G(x_{(n-s)};\alpha)\right]^{sh(\beta)} \prod_{i=r+1}^{n-s}\left[G(x_{(i)};\alpha)\right]^{h(\beta)-1}\left[-g(x_{(i)};\alpha)\right] \\ \times\left\{1-\left[G(x_{(l)};\alpha)\right]^{h(\beta)}\right\}^{l-1}\left\{\left[G(x_{(l)};\alpha)\right]^{h(\beta)}-\left[G(x_{(r+1)};\alpha)\right]^{h(\beta)}\right\}^{r-l}, \qquad l = 1, \ldots, r, \\ [h(\beta)]^{n-s-r+1}\left[G(x_{(l)};\alpha)\right]^{(n-l+1)h(\beta)-1}\left[-g(x_{(l)};\alpha)\right] \prod_{i=r+1}^{n-s}\left[G(x_{(i)};\alpha)\right]^{h(\beta)-1}\left[-g(x_{(i)};\alpha)\right] \\ \times\left\{1-\left[G(x_{(r+1)};\alpha)\right]^{h(\beta)}\right\}^{r}\left\{\left[G(x_{(n-s)};\alpha)\right]^{h(\beta)}-\left[G(x_{(l)};\alpha)\right]^{h(\beta)}\right\}^{l-n+s-1}, \quad l = n-s+1, \ldots, n, \end{cases} \tag{3.5}$$

which is derived by multiplying the likelihood function (3.1) with the conditional PDFs (3.3) and (3.4). The point prediction for $X_{(l)}$, denoted by $\hat{X}_{(l)}$, and the predictive MLEs for α and β are obtained by optimizing the natural logarithm of the predictive likelihood function (3.5) for $X_{(l)}$, α, and β, respectively. Here, our focus is on the prediction of $X_{(l)}$, so only the results related to it are reported in Section 5.

4. Pivotal-based inference

The goal of this section is to propose an estimation method for α and β and a prediction method for missing or censored samples, using pivotal quantities within the DT-II censoring framework. The foremost advantage of the proposed approach is its provision of exact CIs for α and β, unlike the likelihood-based approach in Section 3. Moreover, the proposed approach excels in resolving prediction problems for missing or censored samples by yielding closed-form results.

Before introducing the essential pivotal quantities used for inference in the subsequent subsection, we define the following notations for some distributions to simplify their expressions.

- χ^2_v : Chi-squared distribution with v degrees of freedom.
- $\mathcal{B}e(a, b)$: Beta distribution with parameters (a, b).

4.1. Pivotal quantity

Let

$$\begin{aligned} Y_{(i)} &= -\log\left[1 - F(x_{(i)}; \Theta)\right] \\ &= -h(\beta)\log G(x_{(i)}; \alpha),\ \ i = r+1, \ldots, n-s. \end{aligned}$$

Since a DT-II censored sample $Y_{(r+1)} \leq \cdots \leq Y_{(n-s)}$ has a standard exponential distribution with a mean $E(Y_{(i)}) = \sum_{j=1}^{i}(n-j+1)^{-1}$, it induces the normalized spacings

$$\begin{aligned} \mathcal{S}_i &= (n-i+1)\left(Y_{(i)} - Y_{(i-1)}\right) \\ &= (n-i+1)h(\beta)\log\left(\frac{G(x_{(i-1)}; \alpha)}{G(x_{(i)}; \alpha)}\right),\ \ i = r+2, \ldots, n-s, \end{aligned}$$

which are standard exponential random variables, all independent and identically distributed (iid). Then, the following pivotal quantity can be led by the independence of the spacings:

$$\begin{aligned} T_j(\Theta) &= 2\sum_{i=r+2}^{r+1+j}\mathcal{S}_i \\ &= -2(n-r)Y_{(r+1)} + 2(n-r-j)Y_{(r+j+1)} + 2\sum_{i=r+1}^{r+j}Y_{(i)} \\ &= 2h(\beta)M_{1,j}(\boldsymbol{x}; \alpha),\ \ j = 1, \ldots, n-s-r-1, \end{aligned}$$

which are independent random variables from χ^2_{2j}, where

$$M_{1,j}(\boldsymbol{x}; \alpha) = (n-r)\log G(x_{(r+1)}; \alpha) - (n-r-j)\log G(x_{(r+j+1)}; \alpha) - \sum_{i=r+1}^{r+j}\log G(x_{(i)}; \alpha).$$

The subsequent lemma provides some pivotal quantities derived from the pivotal quantity $T_j(\Theta)$, which are instrumental in inference for α and β.

Lemma 4.1. *Let $X_{(r+1)} \leq \cdots \leq X_{(n-s)}$ be a DT-II censored sample from the probability distribution with CDF* (2.1). *Then,*

$$(i)\ W_1(\Theta) = 2h(\beta)M_2(\boldsymbol{x};\alpha),$$

and

$$(ii)\ W_2(\alpha) = -2\sum_{j=1}^{n-s-r-2}\log\left(\frac{M_{1,j}(\boldsymbol{x};\alpha)}{M_2(\boldsymbol{x};\alpha)}\right),$$

which have $\chi^2_{2(n-s-r-1)}$ and $\chi^2_{2(n-s-r-2)}$, respectively, where

$$M_2(\boldsymbol{x};\alpha) = (n-r)\log G(x_{(r+1)};\alpha) - (s+1)\log G(x_{(n-s)};\alpha) - \sum_{i=r+1}^{n-s-1}\log G(x_{(i)};\alpha).$$

Proof. $W_1(\Theta)$ is clear from $W_1(\Theta) = T_{n-s-r-1}(\Theta)$. In addition, since $\left(T_j(\Theta)/T_{j+1}(\Theta)\right)^j$ $(j = 1, \ldots, n-s-r-2)$ are standard uniform random variables, (ii) is derived as

$$\begin{aligned} W_2(\alpha) &= -2\sum_{j=1}^{n-s-r-2} j\log\left(\frac{T_j(\Theta)}{T_{j+1}(\Theta)}\right) \\ &= -2\sum_{j=1}^{n-s-r-2}\log\left(\frac{T_j(\Theta)}{T_{n-s-r-1}(\Theta)}\right). \end{aligned}$$

This completes the proof. □

It is worth noting that the pivotal quantity $W_2(\alpha)$ is a function of α only, not dependent on β. This quantity plays a very important role in inference not only for α itself but also for the parameters of interest or functions of interest, even in a situation where α is a nuisance parameter.

The process of deriving the essential pivotal quantities used for inference under the DT-II censoring scheme is summarized in Figure 1, where Exp(1) denotes a standard exponential distribution.

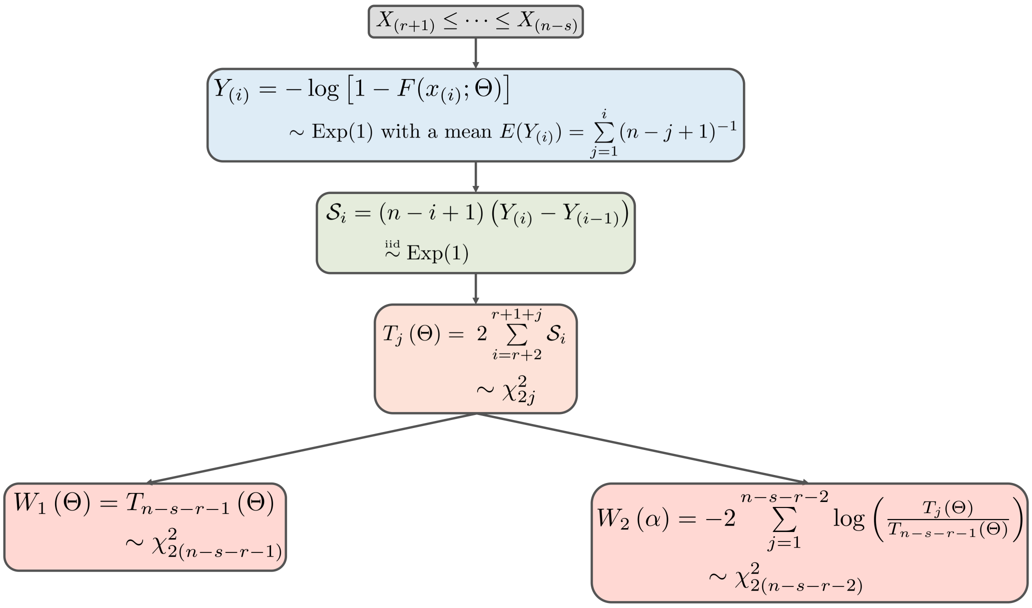


Figure 1. Flowchart of the essential pivotal quantities used for inference under the DT-II censoring scheme.

4.2. Estimation

Using the pivotal quantities in Lemma 4.1, exact CIs for α and β are provided now. For α, since the pivotal quantity $W_2(\alpha)$ in Lemma 4.1 has $\chi^2_{2(n-s-r-2)}$, it leads to

$$1-\gamma = P\left[\chi^2_{1-\gamma/2,2(n-s-r-2)} < W_2(\alpha) < \chi^2_{\gamma/2,2(n-s-r-2)}\right],$$

for $0 < \gamma < 1$, where $\chi^2_{\gamma/2,v}$ is the $(\gamma/2)$th upper percentile of χ^2_v. Then, an exact $100(1-\gamma)\%$ CI of α is formulated as

$$\left[W_2^{-1}\left(\chi^2_{1-\gamma/2,2(n-s-r-2)}\right), W_2^{-1}\left(\chi^2_{\gamma/2,2(n-s-r-2)}\right)\right],$$

or

$$\left[W_2^{-1}\left(\chi^2_{\gamma/2,2(n-s-r-2)}\right), W_2^{-1}\left(\chi^2_{1-\gamma/2,2(n-s-r-2)}\right)\right],$$

when the pivotal quantity $W_2(\alpha)$ in Lemma 4.1 is the increasing or decreasing functions of α, respectively, where $W_2^{-1}(k)$ is a solution of α for the equation $W_2(\alpha) = k$. In a similar manner, using the pivotal quantity $W_1(\Theta)$ in Lemma 4.1, an exact $100(1-\gamma)\%$ CI of β for known α can be formulated as

$$\left[h^{-1}\left(\frac{\chi^2_{1-\gamma/2,2(n-s-r-1)}}{2M_2(\boldsymbol{x};\alpha)}\right), h^{-1}\left(\frac{\chi^2_{\gamma/2,2(n-s-r-1)}}{2M_2(\boldsymbol{x};\alpha)}\right)\right],$$

or

$$\left[h^{-1}\left(\frac{\chi^2_{\gamma/2,2(n-s-r-1)}}{2M_2(\boldsymbol{x};\alpha)}\right), h^{-1}\left(\frac{\chi^2_{1-\gamma/2,2(n-s-r-1)}}{2M_2(\boldsymbol{x};\alpha)}\right)\right],$$

for $0 < \gamma < 1$ when $h(\beta)$ is the increasing or decreasing functions of β, respectively, where $h^{-1}(t)$ is the solution of β for the equation $h(\beta) = t$. However, if the nuisance parameter α is unknown, this CI cannot be used. To resolve this issue, we propose a method of constructing a generalized CI for β by defining a generalized pivotal quantity (GPQ) and introducing a new algorithm based on it.

Let α^* be a unique solution of α for the equation $W_2(\alpha) = \omega$, where ω is a realization from $\chi^2_{2(n-s-r-2)}$. Then, from the pivotal quantity $W_1(\Theta)$ in Lemma 4.1, we define a GPQ for β as

$$\psi_1 \left(\beta \mid \alpha^*\right) = h^{-1}\left(\frac{W_1(\alpha^*, \beta)}{2M_2(\boldsymbol{x}; \alpha^*)}\right),$$

with the observed values $\boldsymbol{x}$, according to the argument of Weerahandi [10, 11]. To obtain its percentile required to construct a generalized CI for β, we propose utilizing a pseudorandom sequence from χ^2_2 based on the principles used to derive the pivotal quantities $W_1(\Theta)$ and $W_2(\alpha)$ in Lemma 4.1. This approach is detailed in Algorithm 1.

Algorithm 1.

Step 1. Generate $\zeta_1, \zeta_2, \ldots, \zeta_n$ from χ^2_2.
Step 2. Compute $\upsilon_j = \sum_{i=r+2}^{r+1+j} \zeta_i$ for $j = 1, \ldots, n-s-r-1$ and $\omega = -2 \sum_{j=1}^{n-s-r-2} \log\left(\frac{\upsilon_j}{\upsilon_{n-s-r-1}}\right)$.
Step 3. Obtain α^* by solving the equation $W_2(\alpha) = \omega$ for α.
Step 4. Compute $\psi_1 \left(\beta \mid \alpha^*\right) = h^{-1}\left(\frac{\upsilon_{n-s-r-1}}{2M_2(\boldsymbol{x}; \alpha^*)}\right)$.
Step 5. Repeat $N(\geq 10,000)$ times.

Algorithm 1 generates a sequence $\boldsymbol{\psi}_1 = \{\psi_{1,1}\left(\beta \mid \alpha^*\right), \ldots, \psi_{1,N}\left(\beta \mid \alpha^*\right)\}$. Then, a generalized $100(1-\gamma)\%$ CI of β is formulated as

$$\left(\psi_{1,([(\gamma/2)N])}\left(\beta \mid \alpha^*\right), \psi_{1,([(1-\gamma/2)N])}\left(\beta \mid \alpha^*\right)\right),$$

for $0 < \gamma < 1$, where $\psi_{1,([(\gamma/2)N])}\left(\beta \mid \alpha^*\right)$ is the $[(\gamma/2)N]$th smallest of the sequence $\boldsymbol{\psi}_1$, and $[\cdot]$ denotes the greatest integer function.

Furthermore, the pivotal quantities $W_1(\Theta)$ and $W_2(\alpha)$ in Lemma 4.1 lead to a valid and statistically good estimation equation for α and β. When the parameter of interest is α, the pivotal quantity $W_2(\alpha)$ plays a very important role in estimating α since it is independent of β, unlike the pivotal quantity $W_1(\Theta)$. To derive an estimator of α, the following lemma is introduced.

Lemma 4.2. *Let $V(\alpha) = W_2(\alpha) / [2(n-s-r-3)]$. Then, $V(\alpha)$ converges to one in probability as $n-s-r \to \infty$.*

Proof. Since $V(\alpha)$ has a gamma distribution with parameters $(n-s-r-2, n-s-r-3)$, it follows that

$$\begin{aligned} P\left(|V(\alpha) - 1| > \epsilon\right) &\leq \frac{4\mathrm{Var}\left(V(\alpha)\right)}{\epsilon^2} \\ &\to 0, \end{aligned}$$

for any $\epsilon > 0$ as $n-s-r \to \infty$. This completes the proof. □

From Lemma 4.2, the estimation equation for α is derived as $W_2(\alpha) = 2(n - s - r - 3)$. In addition, the subsequent theorem establishes the uniqueness of the solution to the equation.

Theorem 4.3. *The equation $W_2(\alpha) = 2(n - s - r - 3)$ has a unique solution $\tilde{\alpha}$ for α.*

This requires different proofs depending on the form of the probability distribution with CDF (2.1), and the proofs are given in Appendix.

If the parameter of interest is β for known α, then the following estimator of $h(\beta)$ can be used:

$$\tilde{h}_\beta(\alpha) = \frac{n - s - r - 2}{M_2(\boldsymbol{x}; \alpha)}, \tag{4.1}$$

which has an inverse gamma distribution with parameters $(n - s - r - 1, (n - s - r - 2)h(\beta))$ from the pivotal quantity $W_1(\Theta)$ in Lemma 4.1. It is noted that the estimator (4.1) has unbiasedness. From it, an estimation equation for β is derived and the consequent estimator of β is formulated as

$$\tilde{\beta}(\alpha) = h^{-1}\left(\frac{n - s - r - 2}{M_2(\boldsymbol{x}; \alpha)}\right). \tag{4.2}$$

For unknown α in the estimator (4.2), it can be substituted with the estimator $\tilde{\alpha}$. The validity of the estimators $\tilde{\alpha}$ and $\tilde{\beta}(\tilde{\alpha})$ is examined by Monte Carlo simulations, and the excellence is demonstrated through a comparison with the MLEs $\hat{\alpha}$ and $\hat{\beta}$ in Section 5.

4.3. Prediction

To predict missing or censored samples using the pivotal quantity, the following Lemmas are first introduced.

Lemma 4.4. *Let* $U_L = \dfrac{1 - [G(x_{(l)}; \alpha)]^{h(\beta)}}{1 - [G(x_{(r+1)}; \alpha)]^{h(\beta)}}$ *for $l = 1, \ldots, r$ in the conditional PDF* (3.3). *Then, U_L has* $\mathcal{B}e(l, r - l + 1)$.

Lemma 4.5. *Let* $U_R = \left[\frac{G(x_{(l)};\alpha)}{G(x_{(n-s)};\alpha)}\right]^{h(\beta)}$ *for $l = n - s + 1, \ldots, n$ in the conditional PDF* (3.4). *Then, U_R has* $\mathcal{B}e(n - l + 1, l - n + s)$.

Lemmas 4.4 and 4.5 are easily proved by the variable transformation in the conditional PDFs (3.3) and (3.4), respectively. From these Lemmas, the prediction for $X_{(l)}$ is formulated as

$$X_{(l)} = \begin{cases} G^{-1}\left(\left(1 - U_L\left\{1 - [G(x_{(r+1)}; \alpha)]^{h(\beta)}\right\}\right)^{1/h(\beta)}\right), & l = 1, \ldots, r, \\ G^{-1}\left(U_R^{1/h(\beta)} G(x_{(n-s)}; \alpha)\right), & l = n - s + 1, \ldots, n, \end{cases} \tag{4.3}$$

where $G^{-1}(\cdot)$ is an inverse function of $G(x; \alpha)$. Note that the proposed approach induces a closed-form prediction result for $X_{(l)}$, unlike the likelihood-based approach in Section 3.

To obtain the predicted value for $X_{(l)}$ in (4.3), the realizations of U_L and U_R are required. These can be obtained from the fact that if $\mathfrak{X}_1$ and $\mathfrak{X}_2$ are independent random variables from χ^2_{2a} and χ^2_{2b}, respectively, then $\mathfrak{X}_1/(\mathfrak{X}_1 + \mathfrak{X}_2)$ has $\mathcal{B}e(a, b)$. In this study, as in the case of Algorithm 1, a pseudorandom sequence is used to generate these realizations, and the detailed procedure is provided in Algorithm 2.

Moreover, α and β in (4.3) can be substituted with $\tilde{\alpha}$ and $\tilde{\beta}(\tilde{\alpha})$, respectively, then the consequent point prediction is denoted by $\widetilde{X}_{(l)}$.

Additionally, a predictive interval (PI) for $X_{(l)}$ can be constructed based on its GPQ $\psi_2(X_{(l)})$ which is defined by substituting α and β with α^* and $\psi_1(\beta \mid \alpha^*)$ in (4.3), respectively, unlike the likelihood-based approach in Section 3. The percentiles of $\psi_2(X_{(l)})$, required to construct the PI for $X_{(l)}$, are obtained along the lines of Algorithm 1, with the realizations of U_L and U_R generated from two independent χ^2 random variables as in the case of $\widetilde{X}_{(l)}$. The implementation process is detailed step-by-step within Algorithm 2.

Algorithm 2.

Step 1. Generate $\zeta_1, \zeta_2, \ldots, \zeta_n$ from χ^2_2.
Step 2. Compute $\upsilon_j = \sum_{i=r+2}^{r+1+j} \zeta_i$ for $j = 1, \ldots, n-s-r-1$ and $\omega = -2 \sum_{j=1}^{n-s-r-2} \log\left(\frac{\upsilon_j}{\upsilon_{n-s-r-1}}\right)$.
Step 3. Obtain α^* by solving the equation $W_2(\alpha) = \omega$ for α.
Step 4. Compute $\psi_1(\beta \mid \alpha^*) = h^{-1}\left(\dfrac{\upsilon_{n-s-r-1}}{2M_2(\boldsymbol{x}; \alpha^*)}\right)$.
Step 5. Substitute α and β with α^* and $\psi_1(\beta \mid \alpha^*)$ in (4.3), respectively.
Step 6. Compute $u_L = \dfrac{\sum_{i=1}^{l} \zeta_i}{\sum_{i=1}^{l} \zeta_i + \sum_{i=l+1}^{r+1} \zeta_i}$ or $u_R = \dfrac{\sum_{i=l}^{n} \zeta_i}{\sum_{i=l}^{n} \zeta_i + \sum_{i=n-s}^{l-1} \zeta_i}$.
Step 7. Substitute U_L and U_R with u_L and u_R in (4.3), respectively, to compute $\psi_2(X_{(l)})$.
Step 8. Repeat $N(\geq 10,000)$ times.

Algorithm 2 generates a sequence $\boldsymbol{\psi}_2 = \{\psi_{2,1}(X_{(l)}), \ldots, \psi_{2,N}(X_{(l)})\}$. Then, a generalized $100(1-\gamma)\%$ PI of $X_{(l)}$ is formulated as

$$\left(\psi_{2,([(\gamma/2)N])}(X_{(l)}), \psi_{2,([(1-\gamma/2)N])}(X_{(l)})\right),$$

for $0 < \gamma < 1$, where $\psi_{2,([(\gamma/2)N])}(X_{(l)})$ is the $[(\gamma/2)N]$th smallest of $\boldsymbol{\psi}_2$. It is worth noting that Algorithm 2 includes some steps of Algorithm 1. So, it can be viewed as an extended version of Algorithm 1. This means that if the function of interest contains α and β, then its percentile and the consequent generalized PI can be obtained without difficulty, as in the case of prediction using Algorithm 2.

5. Application

This section substantiates the superiority and validity of the proposed method by making a comparison with the likelihood-based method in Monte Carlo simulations and real-world case analysis.

5.1. Simulation study

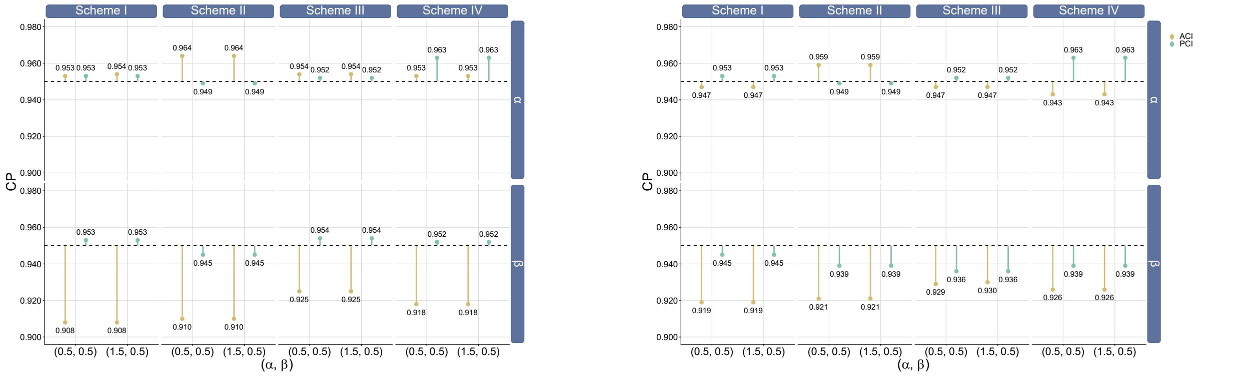
In the probability distribution with CDF (2.1), Weibull and bathtub-shaped distributions are prominent statistical probability distributions employed in life data analysis. In addition, most distributions in this family have an unknown parameter α as a shape parameter which describes the shape of the distributions. This parameter may be an important parameter of interest because it allows us to model various characteristics of lifetime distributions. For example, the hazard function of a Weibull distribution increases over time for $\alpha > 1$, and vice versa for $\alpha < 1$. A bathtub-shaped distribution has an increasing hazard function for $\alpha \geq 1$, while it has a bathtub shape for $\alpha < 1$. Based on those characteristics, results for the Weibull and bathtub-shaped distributions with $\alpha = (0.5, 1.5)$ and $\beta = 0.5$ are only reported here. Additionally, to highlight the superiority of the proposed method for small and medium sample sizes, the following censoring schemes are used:

$$\begin{aligned} \text{Scheme} \quad \text{I}&: n = 20,\ r = 2,\ s = 2,\\ \text{II}&: n = 20,\ r = 3,\ s = 3,\\ \text{III}&: n = 30,\ r = 3,\ s = 3,\\ \text{IV}&: n = 30,\ r = 4,\ s = 4. \end{aligned}$$

Monte Carlo simulations are conducted by generating 1,000 DT-II censored samples for each censoring scheme, and the results are presented through various statistical measures. First, the coverage probabilities (CPs) of 95% CIs provided in Sections 3 and 4 are reported in Figure 2. Here, the ACI and PCI denote the approximate and proposed CIs, respectively. In the case of the PCI for β, it is computed based on $N = 10,000$ in Algorithm 1. In addition, the mean squared errors (MSEs) and biases of the estimators are reported in Figures 3 and 4.

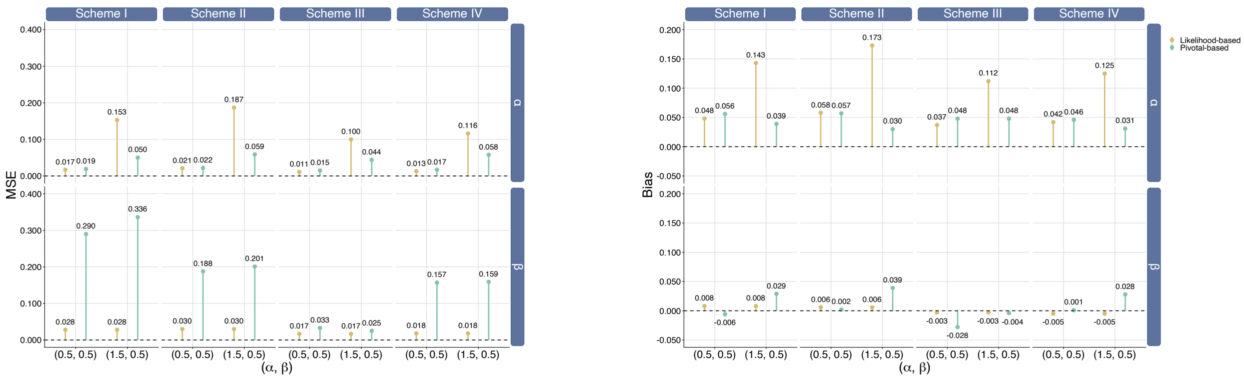
Through Figures 2–4, the following results are led: The CPs of the PCI are generally found to be more aligned with the desired nominal level of 0.95 than those of the ACI. Notably, for β, the PCI has a CP much closer to 0.95 than the ACI, indicating that the generalized CI obtained from the proposed Algorithm 1 yields superior results for small and medium sample sizes. In terms of the MSE and bias, $\tilde{\alpha}$ shows superior MSE and bias results to the MLE counterpart when $\alpha = 1.5$. For a Weibull distribution, the biases of $\tilde{\beta}(\tilde{\alpha})$ are generally more efficient than those of the MLE counterpart when $\alpha = 0.5$, despite performing poorly in terms of MSE. On the other hand, for a bathtub-shaped distribution, $\tilde{\beta}(\tilde{\alpha})$ exhibits generally superior performance in terms of bias when $\alpha = 0.5$ and 1.5, compared to the MLE counterpart.

Furthermore, to assess the validity of the pivotal-based prediction approach proposed, the prediction errors for censored samples are computed. Specifically, the prediction error is computed as $X_{(l)} - \hat{X}_{(l)}$ and $X_{(l)} - \widetilde{X}_{(l)}$ for likelihood-based and pivotal-based predictions, respectively. The resulting prediction errors for the simulated 1,000 DT-II censored samples are visualized in Figures 5–8 as box plots, which indicates that the medians of prediction errors obtained from the pivotal-based method are consistently closer to zero than those of prediction errors obtained from the likelihood-based method in all settings. This finding substantiates the superior predictive performance of the proposed method for small and medium sample sizes, revealing its competitiveness compared to the likelihood-based method.



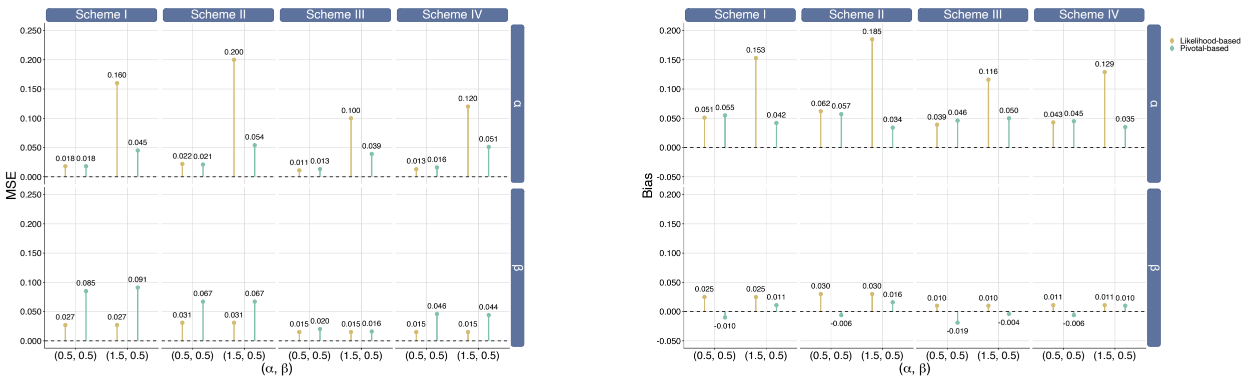
(a) **(b)**

Figure 2. CPs of 95% CIs for α and β in (a) Weibull and (b) bathtub-shaped distributions.



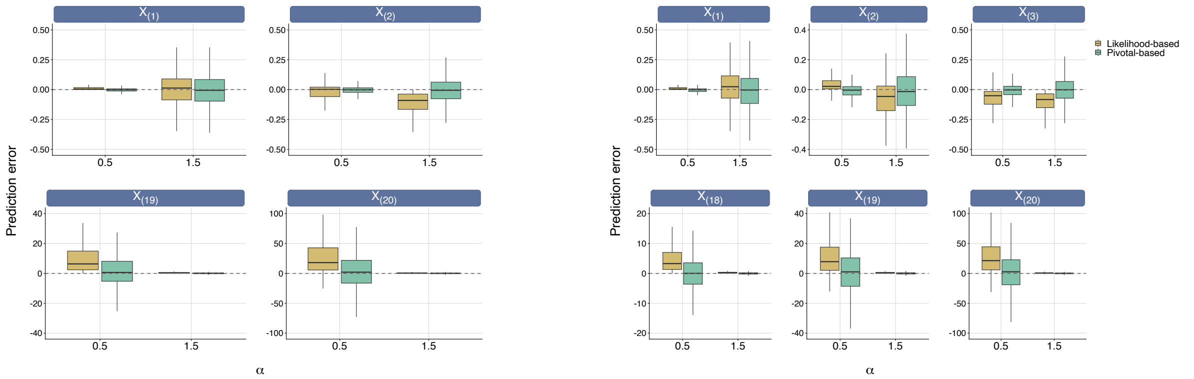
(a) **(b)**

Figure 3. (a) MSEs and (b) biases of estimators for α and β in a Weibull distribution.



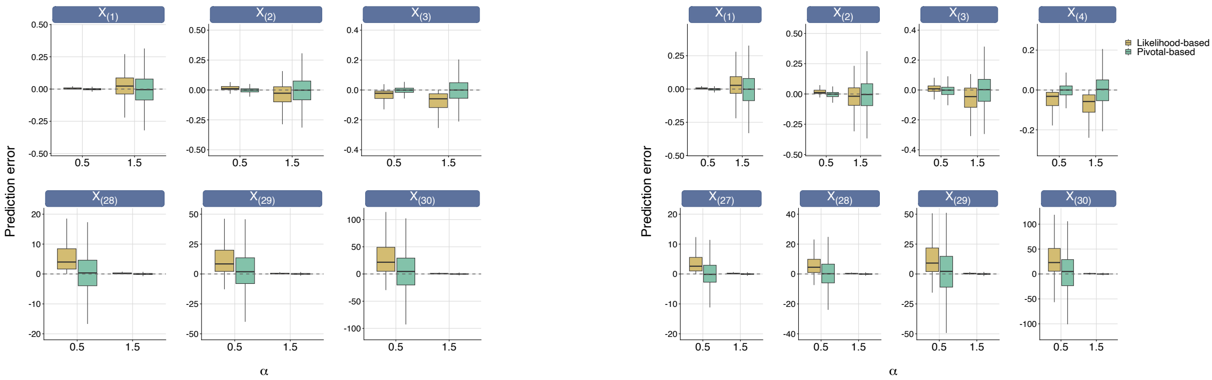
(a) **(b)**

Figure 4. (a) MSEs and (b) biases of estimators for α and β in a bathtub-shaped distribution.



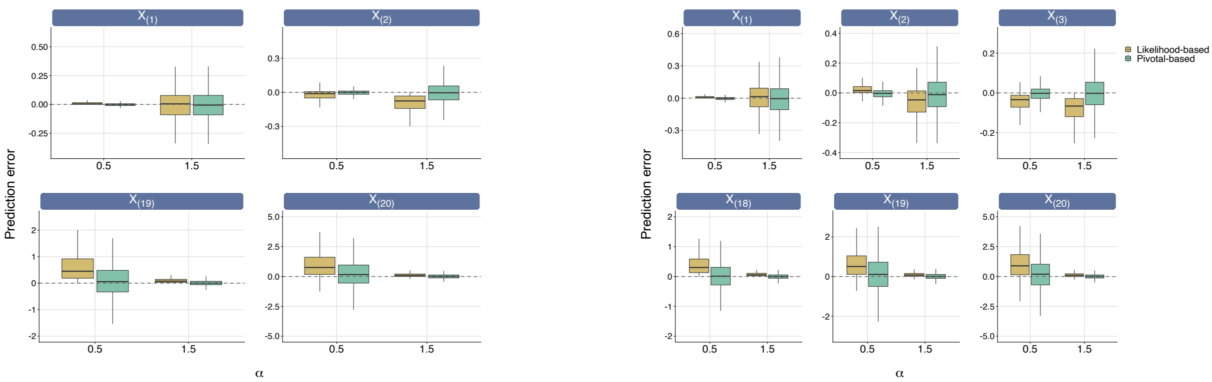
(a) **(b)**

Figure 5. Box plots of prediction errors for a Weibull distribution under (a) Scheme I and (b) Scheme II.



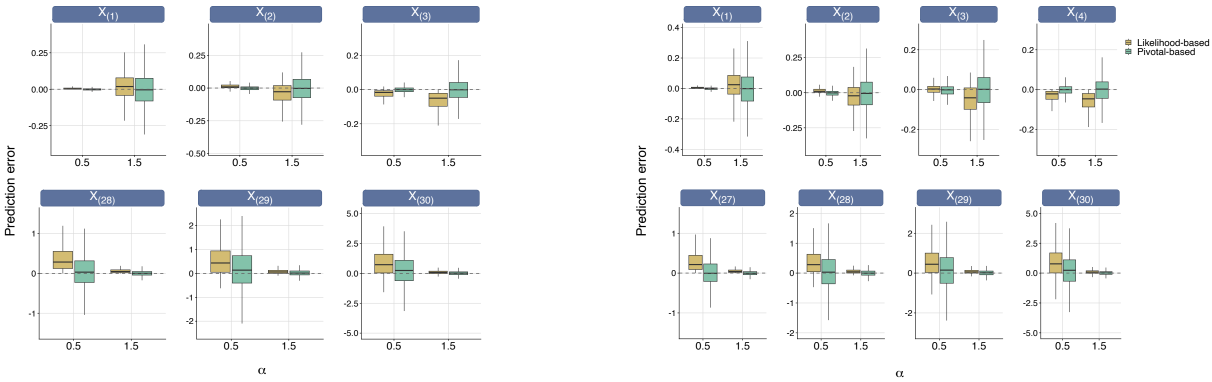
(a) **(b)**

Figure 6. Box plots of prediction errors for a Weibull distribution under (a) Scheme III and (b) Scheme IV.



(a) **(b)**

Figure 7. Box plots of prediction errors for a bathtub-shaped distribution under (a) Scheme I and (b) Scheme II.



(a) (b)

Figure 8. Box plots of prediction errors for a bathtub-shaped distribution under (a) Scheme III and (b) Scheme IV.

5.2. *Real data analysis*

To illustrate the practical application of the proposed pivotal-based approach, an analysis is conducted on a fraction of COVID-19 mortality rates provided by Alsuhabi et al. [12]. The COVID-19 mortality rates are obtained by calculating the ratio of the daily number of deaths to daily new cases in the United States of America (USA) from June 18 to July 7, 2020, and reported in Table 1. For analytical purposes, the data are multiplied by 100, from which a DT-II censored sample (Table 2) is generated through the setting of $r = 2$ and $s = 2$. Tables 3 and 4 provide the analysis results. Here, the PCIs for β are computed based on $N = 20,000$ in Algorithm 1. From Tables 3 and 4, it can be seen that the length of the PCI for α and β is longer compared with the ACI counterpart. However, it may be more reasonable than the ACI because the proposed method yields satisfactory performance in terms of CP according to the simulation results in Section 5.1.

Table 1. COVID-19 mortality rates in the USA from June 18 to July 7, 2020.

0.0259	0.0333	0.0318	0.0188	0.0172	0.0112	0.0155	0.0229	0.0184	0.0621
0.0146	0.0114	0.0216	0.0103	0.0129	0.0134	0.0117	0.0143	0.0032	0.0054

Table 2. DT-II censored sample generated from COVID-19 mortality rates.

1.03	1.12	1.14	1.17	1.29	1.34	1.43	1.46
1.55	1.72	1.84	1.88	2.16	2.29	2.59	3.18

Table 3. 95% CIs and estimates of α and β for a Weibull distribution.

95% CI				Estimates			
ACI for α	PCI for α	ACI for β	PCI for β	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\beta}$	$\tilde{\beta}(\tilde{\alpha})$
$(1.344, 2.997)$	$(0.000, 2.773)$	$(0.036, 0.393)$	$(0.103, 5.543)$	2.171	2.004	0.215	0.264

Table 4. 95% CIs and estimates of α and β for a bathtub-shaped distribution.

95% CI				Estimates			
ACI for α	PCI for α	ACI for β	PCI for β	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\beta}$	$\tilde{\beta}(\tilde{\alpha})$
(0.663, 1.135)	(0.000, 0.994)	(0.049, 0.281)	(0.120, 4.255)	0.899	0.787	0.165	0.266

Additionally, we evaluate the effectiveness of the proposed prediction approach using the COVID-19 mortality rates. The generalized 95% PIs for $X_{(l)}$ ($l = 1, 2, 19, 20$) are computed based on $N = 20,000$ in Algorithm 2, and the results are plotted in Figure 9. It shows that the observations for the censored samples and the predicted values obtained from the proposed method lie within a 95% interval well. In addition, the proposed method has superior predictive performance compared with the likelihood-based method because its prediction results are closer to the observations for the censored samples.

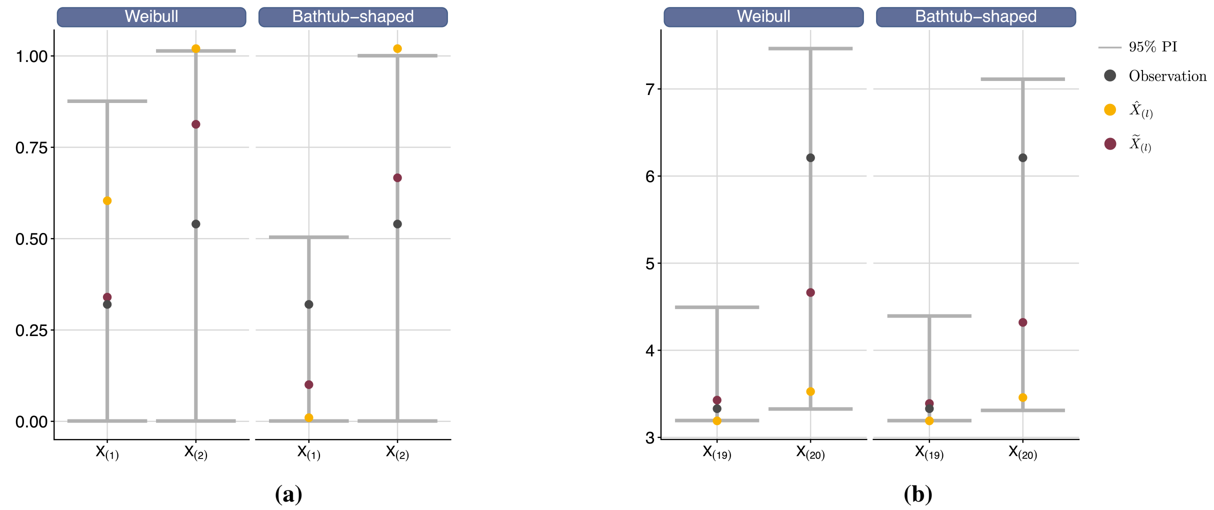


Figure 9. Generalized 95% PIs and predicted values for (a) left and (b) right censored samples.

6. Conclusions

The observed sample size under the DT-II censoring scheme is likely to be very small, which can yield highly biased results when using the likelihood-based method. Our proposed pivotal-based estimation and prediction methods overcome these obstacles by providing exact interval estimation and prediction results that are valid in small sample scenarios. The proposed prediction method is particularly valuable in that it yields closed-form results, unlike the likelihood-based method. In addition, novel algorithms utilizing a pseudorandom sequence, introduced to implement the proposed approach, offer outstanding scalability.

The superiority and effectiveness of the proposed method were demonstrated through a comparison with the likelihood-based method in Monte Carlo simulations and real-world case analysis. According to simulation results, the proposed intervals achieve closer CP to the considered nominal level than the ACIs for small and medium sample sizes. In addition, the proposed prediction method outperforms the likelihood-based method, showing that the medians of prediction errors for the proposed method are close to zero. These results reveal that the proposed method has superior competitiveness compared to the likelihood-based method when focusing on interval estimation and prediction. Based on these findings, we strongly recommend the use of the proposed method to estimate intervals for unknown parameters and predict censored samples, particularly in situations where the available sample size is not sufficiently large. Furthermore, the applicability of the proposed method was confirmed by analyzing COVID-19 mortality rates in the USA.

The proposed method is applicable to inference not only for unknown parameters but also for functions of interest involving these parameters, as demonstrated through Algorithm 2 in this study. For example, the proposed method can be extended to inference for more complex functions involving unknown parameters, such as the reliability function, entropy, and hazard rate under various censoring schemes. This scalability is highly valuable and presents promising opportunities for future research.

Appendix

Proof of Theorem 4.3.

The term $M_2(\boldsymbol{x};\alpha)/M_{1,j}(\boldsymbol{x};\alpha)$ in $W_2(\alpha)$ can be written as

$$\frac{M_2(\boldsymbol{x};\alpha)}{M_{1,j}(\boldsymbol{x};\alpha)} = 1 + \frac{\sum_{i=r+2+j}^{n-s} Q_{(i),(r+j+1)} + sQ_{(n-s),(r+j+1)} - (n-r-j-1)}{\sum_{i=r+1}^{r+1+j} Q_{(i),(r+j+1)} - (n-r)Q_{(r+1),(r+j+1)} + (n-r-j-1)},$$

where

$$Q_{(i),(j)} = \frac{\log G(x_{(i)};\alpha)}{\log G(x_{(j)};\alpha)}.$$

This proof can then be completed by simply showing that $Q_{(i),(j)}$ is a strictly increasing or decreasing function of α for any $\alpha > 0$. The quantity $Q_{(i),(j)}$ for each distribution is given in Table 5. For the Weibull distribution, it is clear that $Q_{(i),(j)}$ is a strictly increasing function of α. For the bathtub-shaped and Gompertz distributions, see Lemma 2 in Chen [13] and Example 3.1 in Wang et al. [14], respectively.

Table 5. Members of distribution family (2.1).

Distribution	$F(x;\Theta)$	$G(x;\alpha)$	$h(\beta)$	Parameter α	β	$Q_{(i),(j)}$
Weibull	$1-e^{-\beta x^{\alpha}},\ x>0$	$e^{-x^{\alpha}}$	β	Shape	Scale	$\left(\frac{x_i}{x_j}\right)^{\alpha}$
Bathtub-shaped	$1-e^{\beta\left(1-e^{x^{\alpha}}\right)},\ x>0$	$e^{1-e^{x^{\alpha}}}$	β	Shape		$\frac{1-e^{x_i^{\alpha}}}{1-e^{x_j^{\alpha}}}$
Lomax	$1-(1+\alpha x)^{-\beta},\ x>0$	$(1+\alpha x)^{-1}$	β	Scale	Shape	$\frac{\log(1+\alpha x_i)}{\log\left(1+\alpha x_j\right)}$
Gompertz	$1-e^{-(\beta/\alpha)(e^{\alpha x}-1)},\ x>0$	$e^{-(e^{\alpha x}-1)/\alpha}$	β	Shape	Scale	$\frac{e^{\alpha x_i}-1}{e^{\alpha x_j}-1}$
Burr-XII	$1-(1+x^{\alpha})^{-\beta},\ x>0$	$(1+x^{\alpha})^{-1}$	β	Shape	Shape	$\frac{\log\left(1+x_i^{\alpha}\right)}{\log\left(1+x_j^{\alpha}\right)}$

Author contributions

Young Eun Jeon: Formal analysis, methodology, software, writing-original draft, writing-review and editing; Yongku Kim: Conceptualization, methodology, supervision, writing-review and editing; Jung-In Seo: Conceptualization, formal analysis, methodology, software, supervision, writing-original draft, writing-review and editing. All authors have read and approved the final version of the manuscript for publication.

Acknowledgments

This research was supported by Global-Learning & Academic research institution for Master's·PhD students, and Postdocs (G-LAMP) Program of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. RS-2023-00301914).

Conflict of interest

All authors declare no conflict of interest in this paper.

References

1. A. J. Fernández, Bayesian inference from type II doubly censored Rayleigh data, *Stat. Probabil. Lett.*, **48** (2000), 393–399. https://doi.org/10.1016/S0167-7152(00)00021-3
2. M. Z. Raqab, M. T. Madi, Bayesian prediction of the total time on test using doubly censored Rayleigh data, *J. Stat. Comput. Sim.*, **72** (2002), 781–789. https://doi.org/10.1080/00949650214670
3. H. M. R. Khan, S. B. Provost, A. Singh, Predictive inference from a two-parameter Rayleigh life model given a doubly censored sample, *Commun. Stat.-Theor. M.*, **39** (2010), 1237–1246. https://doi.org/10.1080/03610920902871453

4. C. Kim, S. Song, Bayesian estimation of the parameters of the generalized exponential distribution from doubly censored samples, *Stat. Pap.*, **51** (2010), 583–597. https://doi.org/10.1007/s00362-008-0142-3
5. M. S. Kotb, M. Z. Raqab, Inference and prediction for modified Weibull distribution based on doubly censored samples, *Math. Comput. Simulat.*, **132** (2017), 195–207. https://doi.org/10.1016/j.matcom.2016.07.014
6. H. Panahi, Estimation for the parameters of the Burr Type XII distribution under doubly censored sample with application to microfluidics data, *Int. J. Syst. Assur. Eng.*, **10** (2019), 510–518. https://doi.org/10.1007/s13198-018-0735-8
7. T. N. Sindhu, Z. Hussain, Objective Bayesian analysis for the power function II distribution under doubly type II censored data, *Thail. Statist.*, **20** (2022), 710–731.
8. S. F. Wu, Interval estimation for the two-parameter exponential distribution based on the doubly type II censored sample, *Qual. Quant.*, **41** (2007), 489–496. https://doi.org/10.1007/s11135-006-9008-8
9. S. F. Wu, Interval estimation for a Pareto distribution based on a doubly type II censored sample, *Comput. Stat. Data An.*, **52** (2008), 3779–3788. https://doi.org/10.1016/j.csda.2007.12.015
10. S. Weerahandi, Generalized confidence intervals, *J. Am. Stat. Assoc.*, **88** (1993), 899–905. https://doi.org/10.2307/2290779
11. S. Weerahandi, *Generalized inference in repeated measures: Exact methods in MANOVA and mixed models*, New York: John Wiley & Sons, 2004.
12. H. Alsuhabi, I. Alkhairy, E. M. Almetwally, H. M. Almongy, A. M. Gemeay, E. H. Hafez, et al., A superior extension for the Lomax distribution with application to Covid-19 infections real data, *Alex. Eng. J.*, **61** (2022), 11077–11090. https://doi.org/10.1016/j.aej.2022.03.067
13. Z. Chen, A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function, *Stat. Probabil. Lett.*, **49** (2000), 155–161. https://doi.org/10.1016/S0167-7152(00)00044-4
14. B. X. Wang, K. Yu, M. C. Jones, Inference under progressively type II right-censored sampling for certain lifetime distributions, *Technometrics*, **52** (2010), 453–460. https://doi.org/10.1198/TECH.2010.08210