Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-021-01510-3

ORIGINAL ARTICLE

Prediction of future failures in the log-logistic distribution based on hybrid censored data

Wassim R. Abou Ghaida[1] · Ayman Baklizi[1]

Received: 30 August 2021 / Revised: 3 November 2021 / Accepted: 8 November 2021

Abstract We consider the prediction of future observations from the log-logistic distribution. The data is assumed hybrid right censored with possible left censoring. Different point predictors were derived. Specifically, we obtained the best unbiased, the conditional median, and the maximum likelihood predictors. Prediction intervals were derived using suitable pivotal quantities and intervals based on the highest density. We conducted a simulation study to compare the point and interval predictors. It is found that the point predictor BUP and the prediction interval HDI have the best overall performance. An illustrative example based on real data is given.

Keywords Hybrid censoring · Log-logistic distribution · Point prediction · Prediction intervals

1 Introduction

In designing a life test to investigate the lifetime distribution of a certain product, information on future observations can help us in determining the cost of the testing process and whether actions are needed to redesign the test (Valiollahi et al. 2017). Manufacturers can predict the number and times of future failures of a product using the past record of failures. Such predictions are useful to quantify future warranty costs and insure that a sufficient number of spare parts is available.

Prediction helps in developing plans and reduce the probability of risks in future. In manufacturing companies, financial managers need prediction bounds on the costs of future warranty (Meeker and Escobar 1998). Prediction gives managers the ability to make informed decisions and develop data-driven strategies. Therefore, it is necessary to develop point and interval predictors for future failure times.

Several studies have been conducted in the literature to study point and interval predictors for various types of lifetime distributions under a variety of experimental conditions. For example, Asgharzadeh et al. (2015) considered prediction of future failures for the two parameter Weibull random variable under hybrid censoring. They obtained three classical point predictors, namely the MLP, the BUP, and the CMP in addition to a Bayesian point predictor. Moreover, Valiollahi et al. (2017) have studied prediction based on censored samples from the generalized exponential distribution. Similar work on prediction for future failures from the exponential distribution can be found in Lawless (1971) and Ebrahimi (1992). The Weibull case was considered by Kundu and Raqab (2012). The one and two sample Bayesian prediction problems with hybrid censored data were considered by Shafay et al. (2012). Classical and Bayesian prediction in the Bur III model were discussed by Singh et al. (2019). ChauPattnaik et al. (2021) discuss component based reliability prediction using Markov chains techniques. Prediction of remaining useful life in some distributions was investigated using Artificial neural networks (ANN) by Farsi and Hosseini (2019).

It appears that little attention was given to developing and comparing the performance of point and interval predictors for future failures in the loglogistic model when the

✉ Ayman Baklizi
a.baklizi@qu.edu.qa

Wassim R. Abou Ghaida
Wassim_abughaida@hotmail.com

1 Statistics Program, Department of Mathematics, Statistics and Physics, College of Arts and Science, Qatar University, 2713 Doha, Qatar

data are both left censored and right hybrid censored. In this paper, we developed various types of point predictors as well as prediction intervals. Using simulation techniques, the bias and mean squared error performance of the point predictors was investigated. The prediction intervals where compared based on their observed coverage probabilities and estimated expected interval lengths. The application of the methods studied in tis paper to real data was discussed using two examples.

The log-logistic distribution is an important statistical model with several applications in various fields. It is used in modeling population growth as first suggested by Verhulst (1838). It is known as the Fisk distribution in economic literature after Fisk (1961) where it was use for studying income inequality. In addition, it has several applications in reliability and life testing studies and sometimes used instead of the log-normal distribution as it is right skewed and has an increasing hazard rate function, it has heavier tail than the lognormal and its cumulative distribution functionand hazard function are available in simple closed forms, see Tahir et al (2014) and Akhtar and Khan (2014). The probability density, the cumulative distribution, and the reliability functions are given by (Lawless, 2003)

$$f(t) = \begin{cases} \dfrac{\alpha\beta(\alpha t)^{\beta-1}}{\left(1+(\alpha t)^{\beta}\right)^2} t \geq 0, \alpha >, \beta > 0, \\ 0 otherwise \end{cases} \quad (1)$$

$$F(t) = P(T \leq t) = \frac{(\alpha t)^{\beta}}{1+(\alpha t)^{\beta}}, \quad (2)$$

$$S(t) = 1 - F(t) = 1 - \frac{(\alpha t)^{\beta}}{1+(\alpha t)^{\beta}} = \frac{1}{1+(\alpha t)^{\beta}}, \quad (3)$$

The hybrid censoring scheme is as follows: Assume that n units are put on test, **t**he researcher terminates the experiment either when $m < n$ units fail or when a certain time τ for the experiment is reached. If X_m is the time of the m^{th} failure the life test will be stopped at time $T^* = \min(X_m, \tau)$. When n units are put on life test, and r out of n units have failure time less than a certain time t_0, the available information about the m units is that $0 \leq t_i \leq t_0$, $i = 1, \ldots, r$ where t_i is the failure time of the i^{th} unit. The failure times of the r units, $t_1, \ldots, t_r$ are left censored.

The remaining part of the paper is structured as follows. In Sect. 2, the likelihood equations are derived. In Sect. 3, predictive likelihood equations are obtained in addition to the point predictors. In Sect. 4, the prediction intervals are derived. In Sect. 5, we performed simulations to compare the performance of different predictors. An illustrative example based on real data is given in Sect. 6 . The conclusions and suggestions for further study are presented in the final section.

2 Likelihood construction and the maximum likelihood estimator

Now we will construct the likelihood function in the log-logistic case assuming a hybrid censored sample that include possibly left censored observations. Similar likelihood construction were considered by other authors in other situations including Ahmadi et al. (2012) for the proportional hazards model and Kundu and Mitra (2016) for the Weibull distribution. Assume that n units are placed on life test, assume that r units are left censored at timet_0. The contribution to the likelihood from a left censored observation is $L_i(\alpha, \beta) = \frac{(\alpha t_0)^{\beta}}{1+(\alpha t_0)^{\beta}}$ and hence the Likelihood based on the left censored observations is $\prod_{i=1}^{r} \frac{(\alpha t_0)^{\beta}}{1+(\alpha t_0)^{\beta}} = \frac{(\alpha t_0)^{r\beta}}{\left(1+(\alpha t_0)^{\beta}\right)^r}.$ Now we will consider hybrid censoring, we have two cases:

Case 1: There are m units with corresponding failure times: $t_{r+1}, t_{r+2}, t_{r+3}, \ldots, t_{r+m}$ where $t_i < \tau$ for all $i = r+1, \ldots, r+m.$

Case 2: We obtained s failures before time τ with failure times: $t_{r+1}, t_{r+2}, t_{r+3}, \ldots, t_{r+s}$ where $s < m$ and $t_{r+s} < \tau < t_{r+s+1}$.

The hybrid censoring scheme is given in Figure 1 in Kundu and Pradhan (2009). A description of time terminated and failure terminated censoring types is given in several sources including Kececioglu (1991).

Combining the two cases as follows: Let d be the number of observed failures and τ^* be the experiment termination time, we have

$$d = \begin{cases} m \; case \; 1 \\ s \; case \; 2 \end{cases}$$

$$\tau^* = \begin{cases} t_{r+m} \; case \; 1 \\ \tau \; case \; 2 \end{cases}$$

The d observations contribution to the likelihood is $\prod_{i=r+1}^{r+d} f(t_i) = \prod_{i=r+1}^{r+d} \frac{\alpha\beta(\alpha t_i)^{\beta-1}}{\left(1+(\alpha t_i)^{\beta}\right)^2}$. The $n-(r+d)$ censored data contribution to the likelihood is $(1-F(\tau^*))^{n-(r+d)}$.

The full likelihood function for both cases is

$$(\alpha,\beta)=\frac{(\alpha t_0)^{r\beta}}{\left(1+(\alpha t_0)^{\beta}\right)^{r}}\cdot\prod_{i=r+1}^{r+d}\frac{\alpha\beta(\alpha t_i)^{\beta-1}}{\left(1+\left(\alpha t_i\right)^{\beta}\right)^{2}}\cdot\left\{\left(\frac{1}{1+(\alpha\tau^*)^{\beta}}\right)^{n-(r+d)}\right\}$$
$$=\frac{(\alpha t_0)^{r\beta}}{\left(1+(\alpha t_0)^{\beta}\right)^{r}}\cdot\prod_{i=r+1}^{r+d}\frac{\alpha\beta(\alpha t_i)^{\beta-1}}{\left(1+\left(\alpha t_i\right)^{\beta}\right)^{2}}\cdot\left\{\left(1+(\alpha\tau^*)^{\beta}\right)^{(r+d)-n}\right\} \tag{4}$$

The Log-Likelihood function is

$$\ln L(\alpha,\beta)=r\beta\ln(\alpha t_0)-r\ln(1+\left(\alpha t_0\right)^{\beta})+\sum_{i=r+1}^{i=r+d}\ln(\alpha\beta(\alpha t_i)^{\beta-1})-2\sum_{i=r+1}^{i=r+d}\ln((1+\left(\alpha t_i\right)^{\beta})-(n-(r+d))\ln(1+\left(\alpha\tau^*\right)^{\beta}), \tag{5}$$

The log-likelihood function derivatives are given by

$$\frac{\partial L}{\partial\alpha}=\frac{r\beta}{\alpha}-\frac{r\beta(\alpha t_0)^{\beta}}{\alpha\left(1+(\alpha t_0)^{\beta}\right)}+\frac{d\beta}{\alpha}-2\sum_{i=r+1}^{i=r+d}\frac{\beta(\alpha t_i)^{\beta}}{\alpha\left(1+(\alpha t_i)^{\beta}\right)}-\frac{(n-(r+d))\beta(\alpha\tau^*)^{\beta}}{\alpha\left(1+(\alpha\tau^*)^{\beta}\right)}. \tag{6}$$

$$\frac{\partial L}{\partial\beta}=r\ln(\alpha t_0)-\frac{r(\alpha t_0)^{\beta}\ln(\alpha t_0)}{\left(1+(\alpha t_0)^{\beta}\right)}+\sum_{i=r+1}^{i=r+d}\frac{1+\beta\ln(\alpha t_i)}{\beta}-2\sum_{i=r+1}^{i=r+d}\frac{(\alpha t_i)^{\beta}\ln(\alpha t_i)}{\left(1+(\alpha t_i)^{\beta}\right)}-\frac{(n-(r+d))(\alpha\tau^*)^{\beta}\ln(\alpha\tau^*)}{\left(1+(\alpha\tau^*)^{\beta}\right)} \tag{7}$$

It is clear that the maximum likelihood estimator can't be obtained explicitly. Therefore, numerical techniques can be used like the Newton Raphson technique or the Expectation–Maximization algorithm of Dempster et al. (1977) to find the MLEs.

3 Point prediction of future observations

We consider prediction of $Y=T_{p+(r+d)}$ where $p=1,\ldots,n-(r+d)$ based on $T=(T_{r+1},\ldots,T_{r+d})$. Due to the Markov property for censored order statistics (Aggarwala and Balakrishnan 1998), $Y|T=t$ has the same distribution as the p^{th} order statistic of a sample of size $n-(r+d)$ from the distribution $G(y)=\frac{F(y)-F(\tau^*)}{1-F(\tau^*)}$ for all $y>\tau^*$. The density function is $g(y)=\frac{d}{dy}G(y)=\frac{f(y)}{1-F(\tau^*)}$. Therefore, the conditional density of $Y=T_{p+(r+d)}$ given $T=t=(t_{r+1},\ldots,t_{r+d})$ for all $y>\tau^*$ is given by (Valiollahi et al. 2017):

$$f(y|t)=\frac{(n-(r+d))!}{(p-1)!(n-(r+d)-p)!}(G(y))^{p-1}(1-G(y))^{n-(r+d)-p}g(y)$$

Replace $G(y) with \frac{F(y)-F(\tau^*)}{1-F(\tau^*)} and g(y) with \frac{f(y)}{1-F(\tau^*)}$, we get:

$$(y|t)=p\binom{n-(r+d)}{p}\left[(\alpha y)^{\beta}-(\alpha\tau^*)^{\beta}\right]^{p-1}\left(1+(\alpha y)^{\beta}\right)^{-n+(r+d)-1}\cdot\left(1+(\alpha\tau^*)^{\beta}\right)^{n-(r+d)-p+1}\cdot\alpha\beta(\alpha y)^{\beta-1}, \tag{8}$$

for all $y>\tau^*$, $p=1,\ldots,n-(r+d)$.

3.1 The maximum likelihood predictor

Referring to Asgharzadeh et al. (2015), the predictive likelihood function (PLF) of $Y and (\alpha,\beta)$ is given by

$$L(y,\alpha,\beta|t)=f(y|t,\alpha,\beta)\cdot g(t|\alpha,\beta)$$

where $f(y|t,\alpha,\beta)$ is the conditional density of Y and $g(t|\alpha,\beta)$ is the likelihood function of the log-logistic distribution. Hence,

$$L(y,\alpha,\beta)=p\binom{n-(r+d)}{p}\left[(\alpha y)^{\beta}-(\alpha\tau^*)^{\beta}\right]^{p-1}\left(1+(\alpha y)^{\beta}\right)^{-n+(r+d)-1}\left(1+(\alpha\tau^*)^{\beta}\right)^{n-(r+d)-p+1}\alpha\beta(\alpha y)^{\beta-1}\frac{(\alpha t_0)^{r\beta}}{\left(1+(\alpha t_0)^{\beta}\right)^{r}}\prod_{i=r+1}^{r+d}\frac{\alpha\beta(\alpha t_i)^{\beta-1}}{\left(1+\left(\alpha t_i\right)^{\beta}\right)^{2}}\left\{\left(1+(\alpha\tau^*)^{\beta}\right)^{(r+d)-n}\right\} \tag{9}$$

$$\ln L(y,\alpha,\beta)=(p-1)\ln\left[(\alpha y)^{\beta}-(\alpha\tau^*)^{\beta}\right]+((r+d)-n-1)\ln(1+(\alpha y)^{\beta})+(1-p).\ln\left(1+(\alpha\tau^*)^{\beta}\right)+(d+1)\ln(\alpha\beta)+(\beta-1)\ln(\alpha y)+r\beta\ln(\alpha t_0)-r\ln\left(1+(\alpha t_0)^{\beta}\right)+(\beta-1)\sum_{i=r+1}^{i=r+d}\ln(\alpha t_i)-2\sum_{i=r+1}^{i=r+d}\ln(1+(\alpha t_i)^{\beta}) \tag{10}$$

It follows that

The maximum likelihood predictor (MLP) is the value ($\tilde{Y}_{MLP}$) in the maximizer of this function. Theerfore we obtain the derivatives of the log predictive likelihood as follows:

$$\frac{\partial \ln L(y,\alpha,\beta)}{\partial y} = \frac{1}{y}\left[\frac{(p-1)\beta(\alpha y)^{\beta}}{(\alpha y)^{\beta}-(\alpha\tau^{*})^{\beta}}+\frac{(r+d-n-1)\beta(\alpha y)^{\beta}}{1+(\alpha y)^{\beta}}+\beta-1\right] \tag{11}$$

$$\frac{\partial \ln L(y,\alpha,\beta)}{\partial \alpha} = \frac{\beta}{\alpha}\left((p-1)+\frac{(r+d-n-1)(\alpha y)^{\beta}}{1+(\alpha y)^{\beta}}+\frac{(1-p)(\alpha\tau^{*})^{\beta}}{1+(\alpha\tau^{*})^{\beta}}+(d+r+1)-\frac{r(\alpha t_0)^{\beta}}{1+(\alpha t_0)^{\beta}}-2\sum_{i=r+1}^{i=r+d}\frac{(\alpha t_i)^{\beta}}{1+(\alpha t_i)^{\beta}}\right) \tag{12}$$

$$\begin{aligned}\frac{\partial \ln L(y,\alpha,\beta)}{\partial \beta} &= (p-1)\frac{(\alpha y)^{\beta}\ln(\alpha y)-(\alpha\tau^{*})^{\beta}\ln(\alpha\tau^{*})}{(\alpha y)^{\beta}-(\alpha\tau^{*})^{\beta}} \\ &+(r+d-n-1)\frac{(\alpha y)^{\beta}\ln(\alpha y)}{1+(\alpha y)^{\beta}} \\ &+(1-p)\frac{(\alpha\tau^{*})^{\beta}\ln(\alpha\tau^{*})}{1+(\alpha\tau^{*})^{\beta}}+\frac{(d+1)}{\beta} \\ &+\ln(\alpha y)+r\ln(\alpha t_0)-\frac{r(\alpha t_0)^{\beta}\ln(\alpha t_0)}{1+(\alpha t_0)^{\beta}} \\ &+\sum_{i=r+1}^{i=r+d}\ln(\alpha t_i)-2\sum_{i=r+1}^{i=r+d}\frac{(\alpha t_i)^{\beta}\ln(\alpha t_i)}{1+(\alpha t_i)^{\beta}}\end{aligned} \tag{13}$$

Solving the predictive likelihood equations will give the maximum likelihood predictor of Y ($\tilde{Y}_{MLP}$)

3.2 The best unbiased predictor

A predictor could be obtained using the conditional distribution of $Y = T_{p+(r+d)}$ given $T = (T_{r+1},\ldots,T_{r+d})$, see Valiollahi et al. (2019). This predictor Y_{Cond} is the best unbiased predictor (BUP) of Y. It is given by: $Y_{Cond} = E(Y|T) = \int_{\tau^*}^{\infty} y f(y|t,\alpha,\beta)dy$ where τ^* is the experiment termination time. It follows that

$$\begin{aligned}Y_{Cond} &= \int_{\tau^*}^{\infty} y\cdot p\binom{n-(r+d)}{p}\cdot\left[(\alpha y)^{\beta}-(\alpha\tau^{*})^{\beta}\right]^{p-1} \\ &\cdot\left(1+(\alpha y)^{\beta}\right)^{-n+(r+d)-1}\cdot\left(1+(\alpha\tau^{*})^{\beta}\right)^{n-(r+d)-p+1} \\ &\cdot\alpha\beta(\alpha y)^{\beta-1}\cdot dy = p\binom{n-(r+d)}{p} \\ &\cdot\left(1+(\alpha\tau^{*})^{\beta}\right)^{n-(r+d)-p+1}\cdot I_1,\end{aligned}$$

where $I_1 = \int_{\tau^*}^{\infty} y\cdot \left[(\alpha y)^{\beta}-(\alpha\tau^{*})^{\beta}\right]^{p-1}\cdot \left(1+(\alpha y)^{\beta}\right)^{-n+(r+d)-1}\cdot\alpha\beta(\alpha y)^{\beta-1}\cdot dy$,

Using the binomial theorem we obtain

$$\left[(\alpha y)^{\beta}-(\alpha\tau^{*})^{\beta}\right]^{p-1} = \sum_{k=0}^{p-1}\binom{p-1}{k}(-1)^{p-1-k}\cdot\left((\alpha y)^{\beta}\right)^{k}\cdot\left((\alpha\tau^{*})^{\beta}\right)^{p-1-k}, I_1 = \sum_{k=0}^{P-1}\binom{p-1}{k}(-1)^{p-1-k}\cdot\left((\alpha\tau^{*})^{\beta}\right)^{p-1-k}\cdot\beta\int_{\tau^*}^{\infty}\left((\alpha y)^{\beta}\right)^{k+1}\cdot\left(1+(\alpha y)^{\beta}\right)^{-n+(r+d)-1}dy,$$

Let $t = \frac{1}{1+(\alpha y)^{\beta}}$ then $(\alpha y)^{\beta} = \frac{1-t}{t}$ and $1+(\alpha y)^{\beta} = \frac{1}{t}$ and $dy = \frac{-dt}{\alpha\beta(\alpha y)^{\beta-1}t^2}$ as $y = \tau^*, t = \frac{1}{1+(\alpha\tau^{*})^{\beta}}$, as $y\to\infty, t\to 0$ and the integral $I_2 = \int_{\tau^*}^{\infty}\left((\alpha y)^{\beta}\right)^{k+1}\cdot\left(1+(\alpha y)^{\beta}\right)^{-n+(r+d)-1}dy$ will become

$$\begin{aligned}I_2 &= \frac{1}{\alpha\beta}\cdot\int_0^{\frac{1}{1+(\alpha\tau^{*})^{\beta}}}(t)^{n-(r+d)-k-1-\frac{1}{\beta}}\cdot(1-t)^{k+\frac{1}{\beta}}dt, Y_{Cond} \\ &= p\binom{n-(r+d)}{p}\cdot\left(1+(\alpha\tau^{*})^{\beta}\right)^{n-(r+d)-p+1}\cdot\sum_{k=0}^{p-1}\binom{p-1}{k} \\ &(-1)^{p-1-k}\cdot\left((\alpha\tau^{*})^{\beta}\right)^{p-1-k}\cdot\frac{1}{\alpha}\cdot\int_0^{\frac{1}{1+(\alpha\tau^{*})^{\beta}}}(t)^{n-(r+d)-k-1-\frac{1}{\beta}}\cdot(1-t)^{k+\frac{1}{\beta}}dt,\end{aligned} \tag{14}$$

We can replace α and β by their MLEs to obtain Y_{Cond}. Note that, I_2 can be expressed as $\frac{1}{\alpha\beta}\bullet B(\frac{1}{1+(\alpha\tau^{*})^{\beta}}; n-(r+d)-k-\frac{1}{\beta}, k+\frac{1}{\beta}+1)$ where $B(z;a,b)$ is the incomplete beta function: $B(z;a,b) = \int_0^z (t)^{a-1} \cdot (1-t)^{b-1} dt$ for $0 \le z < 1$.

3.3 The conditional median predictor

Another predictor is the conditional median predictor Y_{med}. It is defined as Y_{med} satisfying

$$P(Y \le Y_{med} | T = (T_{r+1}, \ldots, T_{r+d})) = P(Y \ge Y_{med} | T = (T_{r+1}, \ldots, T_{r+d})),$$

(Valiollahi et al., 2017). Consider the distribution of $B(y) = \frac{F(y) - F(\tau^*)}{1 - F(\tau^*)}$ where F is the cumulative distribution function of the log-logistic distribution.

$$B(y) = \left(\frac{(\alpha y)^\beta}{1 + (\alpha y)^\beta} - \frac{(\alpha \tau^*)^\beta}{1 + (\alpha \tau^*)^\beta} \right) \cdot \frac{1}{1 - F(\tau^*)} = \left(\frac{(\alpha y)^\beta}{1 + (\alpha y)^\beta} - \frac{(\alpha \tau^*)^\beta}{1 + (\alpha \tau^*)^\beta} \right) \cdot \left(1 + (\alpha \tau^*)^\beta \right) = \frac{(\alpha y)^\beta - (\alpha \tau^*)^\beta}{1 + (\alpha y)^\beta}$$

Following Asgharzadeh et. A. (2015), $B(y)$ can be considered as having a Beta distribution $B(p, n-r-d-p+1)$. Consider:

$$P(Y \le Y_{med} | T = (T_{r+1}, \ldots, T_{r+d})) = P\left(\frac{(\alpha Y)^\beta - (\alpha \tau^*)^\beta}{1 + (\alpha Y)^\beta} \le \frac{(\alpha Y_{med})^\beta - (\alpha \tau^*)^\beta}{1 + (\alpha Y_{med})^\beta} | T = (T_{r+1}, \ldots, T_{r+d}) \right),$$

Equivalently, $P\left(B \le \frac{(\alpha Y_{med})^\beta - (\alpha \tau^*)^\beta}{1 + (\alpha Y_{med})^\beta} \right) = 0.5$ and hence $\frac{(\alpha Y_{med})^\beta - (\alpha \tau^*)^\beta}{1 + (\alpha Y_{med})^\beta} = \text{Med}(B)$. Therefore,

$Y_{med} = \frac{1}{\alpha} \left[\frac{\text{Med}(B) + (\alpha \tau^*)^\beta}{1 - \text{Med}(B)} \right]^{\frac{1}{\beta}}$. (15).

Where the unknown parameters are substituted by their MLEs.

4 Prediction intervals

We consider deriving prediction intervals for $Y = T_{p+(r+d)}$ based on the data $T = (T_{r+1}, \ldots, T_{r+d})$. Let $Z = B(y) = \frac{(\alpha y)^\beta - (\alpha \tau^*)^\beta}{1 + (\alpha y)^\beta}$, it is clear that Z has a Beta distribution $B(p, n-r-d-p+1)$. It could be used as a pivot to obtain $(1-\gamma)100\%$ prediction interval of Y (Asgharzadeh et al. 2015). If B_γ is the $100^{\gamma th}$ percentile of $B(p, n-r-d-p+1)$ then $(1-\gamma)100\%$ PI of Y is (a_1, b_1) where.

$a_1 = \frac{1}{\alpha} \left[\frac{B_{\frac{\gamma}{2}} + (\alpha \tau^*)^\beta}{1 - B_{\frac{\gamma}{2}}} \right]^{\frac{1}{\beta}}$ and $a_2 = \frac{1}{\alpha} \left[\frac{B_{1-\frac{\gamma}{2}} + (\alpha \tau^*)^\beta}{1 - B_{1-\frac{\gamma}{2}}} \right]^{\frac{1}{\beta}}$. (16).

The parameters α and β are to be replaced by their MLEs. Another Prediction interval is based on the highest density interval, see Valiollahi et al. (2017). It is the interval consisting of all points with density function higher than that of other values. The HDI for unimodal distributions produces the shortest such interval. Since $B(y)$ is distributed as $B(p, n-r-d-p+1)$, the $(1-\gamma)100\%$ HDI for prediction of Y is (a_2, b_2)

where $a_2 = \frac{1}{\alpha} \left[\frac{w_1 + (\alpha \tau^*)^\beta}{1 - w_1} \right]^{\frac{1}{\beta}}$ and $b_2 = \frac{1}{\alpha} \cdot \left[\frac{w_2 + (\alpha \tau^*)^\beta}{1 - w_2} \right]^{\frac{1}{\beta}}$, (17).where w_1, w_2 are defined as: $\int_{w_1}^{w_2} g(z) dz = 1 - \gamma$ and $g(w_1) = g(w_2)$. Solving $g(w_1) = g(w_2)$, we get

$$\left(\frac{w_1}{w_2} \right)^{p-1} = \left(\frac{1 - w_2}{1 - w_1} \right)^{n-r-d-1}, \quad (18)$$

It is clear that if $p = 1$, we get $(\frac{1-w_2}{1-w_1})^{n-r-d-1} = 1$, hence $w_1 = w_2$ and no prediction interval can be obtained. For other values of p, we obtain the solutions w_1 and w_2 of the above equation and then substitute MLEs of the parameters α and β.

5 Simulation study

We will design a simulation study to investigate the performance of the predictors. Hybrid censored samples of size n from log-logistic distribution will be generated. The parameters (α, β) from which the samples are generated are taken as $(3, 2)$, the sample size n is taken as 20,30,50 and 80.

Following Piegorsch (1987), we used 2000 replications, based on each sample we obtain the point and interval predictors for the future values of $Y = T_{p+(r+d)}$,$p = 1, 2, \ldots, n-(r+d)$. We reported the bias and mean square prediction error (MSPE) for the point predictors MLP, BUP and CMP. The Bias and MSPE of each predictor value are calculated as follows: If $\tilde{y}_i$ is the value of the predictor of $Y = T_{p+(r+d)}$ obtained from the i^{th} iteration,$i = 1, \ldots, 2000$, then Bias $= \frac{1}{N} (\sum_{i=1}^N (\tilde{y}_i - Y))$ and MSPE $= \frac{1}{N} (\sum_{i=1}^N ((\tilde{y}_i - Y)^2)$.

We also investigated the performance of the prediction intervals developed in this paper. We calculated their empirical coverage probabilities and simulated expected lengths. The results for the point and interval predictors are given in Tables 1, 2, 3.

Table 1 Performance of point and interval predictors for $p = 1$

n	t_0	t_1	m	Point predictors				Interval prediction		
					MLP	BUP	CMP		Pivotal	HDI
30	1/9	1/3	12	Bias	− 0.0356428	− 0.0001214	− 0.0067327	Cov.Prob	0.9170	–
				MSPE	0.0023130	0.0000000	0.0000453	Length	0.0794	–
30	0	1/3	12	Bias	0.0209951	− 0.0010612	− 0.0064391	Cov.Prob	0.9065	–
				MSPE	0.0023301	0.0000011	0.0000415	Length	0.0649	–
50	1/9	1/3	20	Bias	− 0.0491196	0.0003674	− 0.0036448	Cov.Prob	0.9285	–
				MSPE	0.0030274	0.0000001	0.0000133	Length	0.0470	–
50	0	1/3	20	Bias	− 0.0180144	− 0.0005035	− 0.0038210	Cov.Prob	0.9200	–
				MSPE	0.0016416	0.0000003	0.0000146	Length	0.0394	–
80	1/9	1/3	32	Bias	− 0.0279471	− 0.0000879	− 0.0025565	Cov.Prob	0.9425	–
				MSPE	0.0011871	0.0000000	0.0000065	Length	0.0297	–
80	0	1/3	32	Bias	− 0.0021231	− 0.0002719	− 0.0023719	Cov.Prob	0.9380	–
				MSPE	0.0009112	0.0000001	0.0000056	Length	0.0251	–
30	1/9	0.509	18	Bias	− 0.0988914	− 0.0021778	− 0.0141224	Cov.Prob	0.9380	–
				MSPE	0.0126730	0.0000047	0.0001994	Length	0.1428	–
30	0	0.509	18	Bias	− 0.0961920	− 0.0007020	− 0.0094334	Cov.Prob	0.9345	–
				MSPE	0.0136949	0.0000005	0.0000890	Length	0.1030	–
50	1/9	0.509	30	Bias	− 0.0430364	− 0.0005838	− 0.0076966	Cov.Prob	0.9260	–
				MSPE	0.0037569	0.0000003	0.0000592	Length	0.0830	–
50	0	0.509	30	Bias	− 0.1058517	0.0001221	− 0.0051086	Cov.Prob	0.9385	–
				MSPE	0.0140694	0.0000000	0.0000261	Length	0.0616	–
80	1/9	0.509	48	Bias	− 0.0072073	− 0.0008286	− 0.0052486	Cov.Prob	0.9445	–
				MSPE	0.0012743	0.0000007	0.0000275	Length	0.0520	–
80	0	0.509	48	Bias	0.0077276	− 0.0000515	− 0.0033197	Cov.Prob	0.9440	–
				MSPE	0.0021962	0.0000001	0.0000110	Length	0.0390	–
20	0	1/3	8	Bias	− 0.0073401	− 0.0030650	− 0.0107117	Cov.Prob	0.893	–
				MSPE	0.0026192	0.0000094	0.0001147	Length	0.1039	–
20	0	0.509	12	Bias	0.0370314	− 0.0003054	− 0.0132941	Cov.Prob	0.9025	–
				MSPE	0.0070952	0.0000001	0.0001767	Length	0.0937	–

6 A real data example

We will consider the real data set used by Schmee and Nelson (1977) and later discussed by Lawless (2003). It represents the number of miles to failure of 96 locomotive controls. The experiment was stopped after 135,000 miles and the 37 failure times were recorded (in thousands):

22.5, 37.5, 46.0, 48.5, 51.5, 53.0, 54.5, 57.5, 66.5, 68.0, 69.5, 76.5, 77.0, 78.5, 80.0,

81.5, 82.0, 83.0, 84.0, 91.5, 93.5, 102.5, 107.0, 108.5, 112.5, 113.5, 116.0, 117.0,

118.5, 119.0, 120.0, 122.5, 123.0, 127.5, 131.0, 132.5, 134.0.

Hybrid censored samples will be obtained from this data set. We will predict T_p for $p = 1, 2, 3, 4$ and 5. We notice that the (MLEs) obtained are $\widehat{\alpha} = 0.00623$ and $\widehat{\beta} = 2.5941$. The following data were generated according to the scheme where $t_0 = 0$, $t_1 = 135$ and $m_1 = 25$:

Data values: 22.5, 37.5, 46.0, 48.5, 51.5, 53.0, 54.5, 57.5, 66.5, 68.0, 69.5, 76.5, 77.0, 78.5, 80.0, 81.5, 82.0, 83.0, 84.0, 91.5, 93.5, 102.5, 107.0, 108.5, 112.5

Using the formulas (10), (14), (15), (16) and (17) derived in this paper, we obtained the following prediction results.

We note that all Point Predictors produce accurate results. Moreover, both Prediction Intervals contain the

Table 2 Performance of point and interval predictors for $p = 2$

n	t_0	t_1	m	Point predictors				Interval prediction		
					MLP	BUP	CMP		Pivotal	HDI
30	1/9	1/3	12	Bias	− 0.0798906	− 0.0003420	− 0.0078126	Cov.Prob	0.9040	0.8970
				MSPE	0.0077366	0.0000001	0.0000610	Length	0.1228	0.1080
30	0	1/3	12	Bias	− 0.1098029	− 0.0041781	− 0.0097948	Cov.Prob	0.8820	0.8600
				MSPE	0.0142840	0.0000175	0.0000959	Length	0.0961	0.0852
50	1/9	1/3	20	Bias	− 0.0255834	0.0004877	− 0.0038237	Cov.Prob	0.9115	0.9090
				MSPE	0.0014081	0.0000002	0.0000146	Length	0.0697	0.0616
50	0	1/3	20	Bias	− 0.0105547	− 0.0014895	− 0.0050050	Cov.Prob	0.9145	0.9085
				MSPE	0.0015578	0.0000022	0.0000250	Length	0.0582	0.0516
80	1/9	1/3	32	Bias	− 0.0081806	0.0001233	− 0.0024999	Cov.Prob	0.9345	0.9355
				MSPE	0.0004983	0.0000000	0.0000062	Length	0.0437	0.0387
80	0	1/3	32	Bias	0.0481446	− 0.0007660	− 0.0029591	Cov.Prob	0.9220	0.9140
				MSPE	0.0032718	0.0000006	0.0000088	Length	0.0370	0.0328
30	1/9	0.509	18	Bias	− 0.1722274	− 0.0010769	− 0.0159727	Cov.Prob	0.9210	0.9125
				MSPE	0.0336022	0.0000012	0.0002551	Length	0.2291	0.1994
30	0	0.509	18	Bias	− 0.0629317	− 0.0018666	− 0.0118709	Cov.Prob	0.9035	0.8940
				MSPE	0.0094738	0.0000035	0.0001409	Length	0.1602	0.1404
50	1/9	0.509	30	Bias	0.0727522	− 0.0019253	− 0.0100715	Cov.Prob	0.9370	0.9315
				MSPE	0.0074687	0.0000037	0.0001014	Length	0.1290	0.1131
50	0	0.509	30	Bias	0.0462383	− 0.0008137	− 0.0065900	Cov.Prob	0.9255	0.9185
				MSPE	0.0054594	0.0000007	0.0000434	Length	0.0933	0.0822
80	1/9	0.509	48	Bias	0.0729730	− 0.0011922	− 0.0060945	Cov.Prob	0.9300	0.9345
				MSPE	0.0067363	0.0000014	0.0000371	Length	0.0791	0.0696
80	0	0.509	48	Bias	− 0.0040232	− 0.0003281	− 0.0038653	Cov.Prob	0.9385	0.9380
				MSPE	0.0022064	0.0000001	0.0000149	Length	0.0577	0.0510
20	0	1/3	8	Bias	0.0247661	− 0.0043042	− 0.0130384	Cov.Prob	0.8445	0.8265
				MSPE	0.0034740	0.0001852	0.0001700	Length	0.1447	0.1277
20	0	0.509	12	Bias	− 0.0055823	− 0.0057162	− 0.0217901	Cov.Prob	0.9105	0.8940
				MSPE	0.0072028	0.0000327	0.0000474	Length	0.2511	0.2185

true "future" failure. The HDI intervals are shorter than the corresponding pivotal intervals (Table 4).

7 Conclusions and suggestions for further research

In this paper, we studied prediction of future values of the log-logistic distribution under hybrid censored data with possible left censoring. We obtain the predictive likelihood function and use it to obtain the MLP. The conditional density of the future failure was used to obtain the BUP, and CMP. Prediction intervals based on pivotal quantities were constructed. Moreover, we obtained the HDI interval. We investigated and compared the performance of the point predictors in terms of their Biases and MSPE. The two prediction intervals were compared in terms of coverage probability and average width.

Based on the results of the simulation study, it appears that the BUP has the best performance in terms of Bias and MSPE, followed by the CMP. As for prediction intervals, it appears that the interval based on pivotal quantity when $p = 1$ has an observed coverage probability that is close to the nominal level for all combinations of n, t_0, t_1, and m. $n \geq 30$. As the value of p gets larger, the observed coverage probability of both intervals gets close to the nominal level for large values of n. The expected length of the HDI interval is smaller than that of the interval based on pivotal quantity for $p > 1$.

Table 3 Performance of point and interval predictors for $p = 3$

n	t_0	t_1	m	Point predictors				Interval prediction		
					MLP	BUP	CMP		Pivotal	HDI
30	1/9	1/3	12	Bias	− 0.0002865	0.0009938	− 0.0073408	Cov.Prob	0.8860	0.8860
				MSPE	0.0018606	0.0000010	0.0000539	Length	0.1590	0.1458
30	0	1/3	12	Bias	0.0402289	− 0.0034564	− 0.0094984	Cov.Prob	0.8550	0.8440
				MSPE	0.0044099	0.0000119	0.0000902	Length	0.1230	0.1135
50	1/9	1/3	20	Bias	0.0218879	− 0.0001042	− 0.0046385	Cov.Prob	0.9100	0.9050
				MSPE	0.0013513	0.0000000	0.0000215	Length	0.0892	0.0822
50	0	1/3	20	Bias	0.0449275	− 0.0017215	− 0.0053526	Cov.Prob	0.8890	0.8820
				MSPE	0.0037440	0.0000030	0.0000287	Length	0.0735	0.0680
80	1/9	1/3	32	Bias	− 0.0304175	0.0007796	− 0.0019608	Cov.Prob	0.9290	0.9240
				MSPE	0.0014383	0.0000006	0.0000038	Length	0.0550	0.0507
80	0	1/3	32	Bias	− 0.0116306	− 0.0003288	− 0.0025965	Cov.Prob	0.9220	0.9125
				MSPE	0.0011558	0.0000001	0.0000067	Length	0.0459	0.0424
30	1/9	0.509	18	Bias	− 0.1929948	− 0.0002218	− 0.0189897	Cov.Prob	0.8985	0.9040
				MSPE	0.0426268	0.0000000	0.0003606	Length	0.3256	0.2989
30	0	0.509	18	Bias	− 0.0178158	− 0.0037171	− 0.0153040	Cov.Prob	0.9010	0.8940
				MSPE	0.0067576	0.0000138	0.0002342	Length	0.2153	0.1970
50	1/9	0.509	30	Bias	0.0327140	− 0.0006912	− 0.0098018	Cov.Prob	0.9150	0.9075
				MSPE	0.0038957	0.0000005	0.0000961	Length	0.1711	0.1564
50	0	0.509	30	Bias	0.0235009	− 0.0025488	− 0.0087723	Cov.Prob	0.9200	0.9110
				MSPE	0.0041981	0.0000065	0.0000770	Length	0.1199	0.1101
80	1/9	0.509	48	Bias	− 0.0471465	− 0.0000941	− 0.0053663	Cov.Prob	0.9345	0.9405
				MSPE	0.0038174	0.0000000	0.0000288	Length	0.1007	0.0924
80	0	0.509	48	Bias	0.0519966	− 0.0011418	− 0.0048357	Cov.Prob	0.9350	0.9335
				MSPE	0.0050858	0.0000013	0.0000234	Length	0.0735	0.0676
20	0	1/3	8	Bias	0.0110992	− 0.0083584	− 0.0180271	Cov.Prob	0.8135	0.7960
				MSPE	0.0044274	0.0000699	0.0003249	Length	0.1904	0.1751
20	0	0.509	12	Bias	0.0567831	− 0.0043021	− 0.0249301	Cov.Prob	0.8715	0.8550
				MSPE	0.0126722	0.0000185	0.0006215	Length	0.3603	0.3306

Table 4 The values of point predictors and the 95% prediction intervals

	Exact Value	MLP	BUP	CMP	PIVOTAL	HDI
$p = 1$	113.5	113.5	114.7	114.1	(112.5,120.6)	–
$p = 2$	116.0	115.6	116.9	116.2	(113.1,125.6)	(112.6,123.7)
$p = 3$	117.0	116.7	119.2	118.5	(114.0,127.6)	(113.0,127.6)
$p = 4$	118.0	118.2	121.4	120.7	(115.2,133.0)	(114.3,131.3)
$p = 5$	119.0	118.8	123.6	123.0	(116.5,136.4)	(115.6,134.7)

Under the effect of left censoring, the expected length of both intervals get slightly larger while the coverage probabilities get closer to the nominal level. Left censoring does not appear to have clear fixed on the performance of point predictors.

As m and t_1 increase, the bias and MSPE ofthe point predictors are almost unchanged. However, it is slightly increasing for the CMP. On the other hand, the lengths of the pivotal and HDI intervals gets slightly larger and the coverage probability of both intervals gets closer to the nominal.

In conclusion, we recommend that the best point predictor is the BUP as it has the least bias and MSPE, and the best interval prediction method is the HDI method for $p > 1$.

The work in this paper can be extended in several ways. For example, it may be of interest to consider prediction under other types of censoring that occur frequently in life testing experiments. Another possibility is to consider Bayesian prediction techniques. Moreover, it is of interest to consider prediction and the performance of the prediction techniques developed in this paper under step-stress life testing models.

Acknowledgments The authors would like to thank the two referees for their suggestions and thoughtful comments that resulted in a much improved version of the paper.

Funding Open Access funding provided by the Qatar National Library.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Aggarwala R, Balakrishnan N (1998) Some properties of progressive censored order statistics from arbitrary and uniform distributions with applications to inference and simulation. J Statist Planning Inference 70(1):35–49

Ahmadi J, Doostparast A, Parsian A (2012) Estimation with left-truncated and right censored data: A comparison study. Statist Probab Lett 82:1391–1400

Akhtar MT, Khan AA (2014) Log-logistic Distribution as a Reliability Model: A Bayesian Analysis. Am J Math Statist 4(3):162–170

Asgharzadeh A, Valiollahi R, Kundu D (2015) Prediction for future failures in Weibull distribution under hybrid censoring. J Stat Comput Simul 85(4):824–838

ChauPattnaik S, Ray M, Nayak M (2021) Component based reliability prediction. Int J Syst Assur Eng Manag 12(3):391–406

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc B 39(1):1–38

Ebrahimi N (1992) Prediction intervals for future failures in the exponential distribution under hybrid censoring. IEEE Trans Reliab 41(1):127–132

Farsi M, Hosseini S (2019) Statistical distributions comparison for remaining useful life prediction of components via ANN. Int J Syst Assur Eng Manag 10(3):429–436

Fisk PR (1961) The graduation of income distribution. Econometrica 29:171–185

Kececioglu, D. (2002). *Reliability Engineering Handbook*: v. 1. DEStech Publications, Inc

Kundu D, Mitra D (2016) Bayesian inference of Weibull distribution based on left truncated and right censored data. Comput Stat Data Anal 99:38–50

Kundu D, Pradhan B (2009) Estimating the Parameters of the Generalized Exponential Distribution in Presence of Hybrid Censoring. Commun Statist Theory Methods 38(12):2030–2041

Kundu D, Raqab MZ (2012) Bayesian inference and prediction for a Type-II censored Weibull distribution. J Statist Planning Inference 142:41–47

Lawless JF (1971) A prediction problem concerning samples from the exponential distribution, with application in life testing. Technometrics 13(4):725–730

Lawless JF (2003) Statistical Models and Methods for Lifetime Data, 2nd edn. John Wiley & Sons Inc, Hoboken, New Jersey

Piegorsch W (1987) Performance of Likelihood-Based Interval Estimates for Two-Parameter Exponential Samples Subject to Type I Censoring. Technometrics 29(1):41–49

Schmee, J. and Nelson, W. (1977). Estimates and approximate confidence limits for (log) normal life distributions from singly censored samples by maximum likelihood. General Electric C. R. & D. TIS Report 76CRD250. Schenectady, New York.

Shafay AR, Balakrishnan N (2012) One- and two-sample Bayesian prediction intervals Based on Type-I hybrid censored data. Commun Stat Simul Comput 41:65–88

Singh S, Belaghi R, Mehri N (2019) Estimation and prediction using classical and Bayesian approaches for Burr III model under progressive type-I hybrid censoring. Int J Syst Assur Eng Manag 10(4):746–764

Tahir MH, Mansoor M, Zubair M, Hamedani GG (2014) McDonald Log-Logistic distribution. J Statist Theory Appl 13:65–82

Valiollahi R, Asgharzadeh A, Kundu D (2017) Prediction of future failures for generalized exponential distribution under Type-I or Type-II hybrid censoring. Braz J Probability Statist 31(1):41–61

Valiollahi R, Asgharzadeh A, Raqab M (2019) Prediction of Future Failures Times based on Type-I Hybrid Censored Samples of Random Sample Sizes. Commun Statist Simulation Comput 48(1):109–125

Verhulst PF (1838) Notice sur la loi que la population suit dans son accroissement. Corresp Math Phys 10:113–121