Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120452

Predicting Failure times for some Unobserved Events with Application to Real-Life Data

Mahmoud Mansour and Mohamed Aboshady[*]

Department of Basic Science, Faculty of Engineering, The British University in Egypt, El Sherook City, Cairo, Egypt

Received: 30 Aug. 2022, Revised: 10 Mar. 2023, Accepted: 12 Mar. 2023
Published online: 1 Apr. 2023

Abstract: This study aims to predict failure times for some units in some lifetime experiments. In some practical situations, the experimenter may not be able to register the failure times of all units during the experiment. Recently, this situation can be described by a new type of censored data called multiply-hybrid censored data. In this paper, the linear failure rate distribution is well-fitted to some real-life data and hence some statistical inference approaches are applied to estimate the distribution parameters. A two-sample prediction approach applied to extrapolate a new sample simulates the observed data for predicting the failure times for the unobserved units.

Keywords: Lifetime experiments, multiply hybrid censored scheme, linear failure rate distribution, statistical inference, two sample prediction

1 Introduction

Observations on all the sample components are typically not accessible in statistical inference because they are lost, not recorded, or because of cost and time constraints. Statistical inference is a method that uses data from a randomly collected sample to estimate the population parameters. Then, the information is referred to as a censored sample. According to the method used to end the experiment, censoring schemes are divided into two categories: Type-I censoring schemes and Type-II censoring schemes. If the experiment is ended after a predetermined amount of time, such as T, the test will be guaranteed to last but the efficiency will be ignored. On the other hand, the experiment is stopped after testing a predetermined number of components, let's say r, if the efficiency level is more important than the experiment's length. This system is referred as the Type-II censoring scheme, in which a particular efficiency is maintained and the test's duration is determined at random. Hybrid censoring, which combines these two censoring methods, is usually applied because it provides a more adaptable system and is more manageable. Throughout the hybrid censoring methods, a predetermined quantity of the components being tested, r, and a predetermined time for the experiment's conclusion, T, are both given. Assuming we are testing n items, let $X_{i,n}$ be the i^{th} ordered failed component. If the test ends at a time $T_1 = \min\{X_{r,n}, T\}$, the scheme is referred to as a Type-I hybrid censored scheme, and if it ends at a time $T_2 = \max\{X_{r,n}, T\}$, the scheme is referred to as Type-II hybrid censored scheme.Numerous academic works have investigated the estimate of unknown parameters for various distribution functions using the hybrid Type-I as well as the hybrid Type-II censored schemes, respectively, see [3,5,8,12,15,21].

It was necessary to establish a new scheme in order to improve the effectiveness of these schemes and take these losses into account because of the number of failures, let's say R_i, that could happen between any two consecutive observations without knowing the precise failure times for all of these units. The multiple Type-II hybrid censoring system is the name of the novel technique, which was initially presented by Lee and Lee [24]. In their discussion of the multiply Type-II hybrid censoring method, Jeon and Kang covered the estimation of parameters for two distinct distributions, see [17,18].

Due to its memory-less and constant failure rate characteristics, the exponential distribution—a specific instance of the Weibull, Gamma, Gompertz, exponential-geometric, and linear failure rate distributions—is regarded as one of the primary life distributions. In analyzing the characteristics of any lifetime phenomenon, it is crucial. Recent decades have seen the introduction of some modifications to exponential distributions; see [7,28,31].

[*] Corresponding author e-mail: mohamed.aboshady@bue.edu.eg

The simplest form of the linear failure rate (LFR) distribution, which is a generalization of the exponential distribution, is probably where it first appears in the reliability analysis literature. This distribution, which is a particular example of the quadratic failure rate model, contains two parameters, α, and β, and it is a member of the rising failure rate class. In contrast to the gamma as well as Weibull distributions, the LFR model fails with a positive failure rate at $t = 0$ whereas the gamma and Weibull distribution fail with a failure rate of zero. Due to its frequent application to data on human survival, the LFR distribution became well-known, and many researchers have looked into its characteristics, see [9]. In this work, we found that the LFR distribution fitted well with real-life survival data.

The random variable X has an LFR distribution with probability density function (PDF) and cumulative distribution function (CDF) defined as

$$f(x) = (\alpha+\beta x)e^{-\left(\alpha x+\frac{\beta}{2}x^2\right)}, \quad x>0,\ \alpha,\beta>0. \tag{1}$$

and

$$F(x) = 1-e^{-\left(\alpha x+\frac{\beta}{2}x^2\right)}, \quad x>0,\alpha,\beta>0. \tag{2}$$

This research will provide the classical and Bayesian estimation methods for the unknown parameters of the multiple Type-II hybrid censoring LFR distribution. Both the non-informative priors and the informative priors will be taken into account for the Bayesian methods. Additionally, a simulation study is conducted to evaluate how well the previous distributions were chosen and how this decision affected the accuracy of the outcomes. Additionally, a two-sample prediction study is used for the two models for clinical and industrial data that are proposed. For the parameters under discussion, Section 2 presents maximum likelihood estimates (MLEs) and estimated confidence intervals (ACIs). The MCMC method and Bayesian estimates are covered in Section 3. Good estimators can be applied in the Bayesian prediction study that is provided in Section 4. Section 5 explores two actual data sets as examples. The simulation research to assess the efficacy of the derived estimators is presented in Section 6. Finally, there are conclusions at the end of this paper.

2 Maximum-likelihood estimation

The log-likelihood functions are thought to be the foundation for determining the parameters' estimators given the data. Maximum likelihood estimators have numerous advantages, including the ability to satisfy invariant properties. They are also asymptotically unbiased, asymptotically normally distributed, and have asymptotically minimum variance; for more information on likelihood theory, see [4,31]. As illustrated in Figure 1, one can observe one of two types of censored data under the multiply Type-II hybrid censoring scheme:

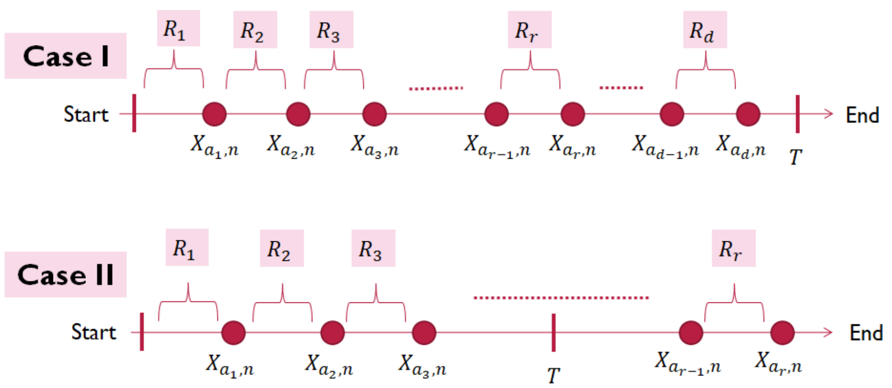

Fig. 1: Multiply Type-II hybrid censoring scheme.

The likelihood function is

Case I: $L \propto \prod_{i=1}^{d} f(x_{a_i:n}) \prod_{i=1}^{d-1} \left[F(x_{a_{i+1}:n}) - F(x_{a_i:n})\right]^{R_{i+1}} \left[F(x_{a_1:n})\right]^{R_1} \left[1-F(x_{a_{d}:n})\right]^{n-d}$,

Case II: $L \propto \prod_{i=1}^{r} f(x_{a_i:n}) \prod_{i=1}^{r-1} \left[F(x_{a_{i+1}:n}) - F(x_{a_i:n})\right]^{R_{i+1}} \left[F(x_{a_1:n})\right]^{R_1} \left[1-F(x_{a_{r}:n})\right]^{n-r}$, Joining cases I and II, we can rewrite the likelihood function as follows:

$$L \propto \prod_{i=1}^{m} f(x_{a_i:n}) \prod_{i=1}^{m-1} \left[F(x_{a_{i+1}:n}) - F(x_{a_i:n})\right]^{R_{i+1}} \left[F(x_{a_1:n})\right]^{R_1} \left[1 - F(x_{a_{m:n}})\right]^{n-m}, \tag{3}$$

where m is the number of failure items until the termination point occurred, $R_i = a_i - a_{i-1} - 1$, $a_0 = 0$.

The log-likelihood function is

$$\ln L \propto \sum_{i=1}^{m} \ln f(x_{a_i:n}) + \sum_{i=1}^{m-1} R_{i+1} \ln\left[F(x_{a_{i+1}:n}) - F(x_{a_i:n})\right] + R_1 \ln\left[F(x_{a_1:n})\right] + (n-m)\ln\left[1 - F(x_{a_{m:n}})\right] \tag{4}$$

Substituting by the PDF in (1) and the CDF in (2) in the log-likelihood function (4), we will get:

$$\ln L \propto \sum_{i=1}^{m} \left[\ln(\alpha + \beta x_{a_i:n}) - \left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)\right] + \sum_{i=1}^{m-1} R_{i+1} \ln\left[e^{-\left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)} - e^{-\left(\alpha x_{a_{i+1}:n} + \frac{\beta}{2} x_{a_{i+1}:n}^2\right)}\right]$$
$$+ R_1 \ln\left[1 - e^{-\left(\alpha x_{a_1:n} + \frac{\beta}{2} x_{a_1:n}^2\right)}\right] + (m-n)\left(\alpha x_{a_{m:n}} + \frac{\beta}{2} x_{a_{m:n}}^2\right), \tag{5}$$

and thus we have the likelihood equations for α and β respectively, as

$$\sum_{i=1}^{m} \frac{1 - x_{a_i:n}}{\left[(\alpha + \beta x_{a_i:n}) - \left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)\right]} + \sum_{i=1}^{m-1} \frac{R_{i+1}\left[x_{a_{i+1}:n} e^{-\left(\alpha x_{a_{i+1}:n} + \frac{\beta}{2} x_{a_{i+1}:n}^2\right)} - x_{a_i:n} e^{-\left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)}\right]}{\left[e^{-\left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)} - e^{-\left(\alpha x_{a_{i+1}:n} + \frac{\beta}{2} x_{a_{i+1}:n}^2\right)}\right]}$$
$$+ \frac{R_1 x_{a_1:n} e^{-\left(\alpha x_{a_1:n} + \frac{\beta}{2} x_{a_1:n}^2\right)}}{\left[1 - e^{-\left(\alpha x_{a_1:n} + \frac{\beta}{2} x_{a_1:n}^2\right)}\right]} + (m-n) x_{a_m:n} = 0, \tag{6}$$

and

$$\sum_{i=1}^{m} \frac{x_{a_i:n}\left(1 - \frac{1}{2} x_{a_i:n}\right)}{\left[(\alpha + \beta x_{a_i:n}) - \left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)\right]} + \sum_{i=1}^{m-1} \frac{R_{i+1}\left[\frac{1}{2} x_{a_{i+1}:n}^2 e^{-\left(\alpha x_{a_{i+1}:n} + \frac{\beta}{2} x_{a_{i+1}:n}^2\right)} - \frac{1}{2} x_{a_i:n}^2 e^{-\left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)}\right]}{\left[e^{-\left(\alpha x_{a_i:n} + \frac{\beta}{2} x_{a_i:n}^2\right)} - e^{-\left(\alpha x_{a_{i+1}:n} + \frac{\beta}{2} x_{a_{i+1}:n}^2\right)}\right]}$$
$$+ \frac{\frac{1}{2} R_1 x_{a_1:n}^2 e^{-\left(\alpha x_{a_1:n} + \frac{\beta}{2} x_{a_1:n}^2\right)}}{\left[1 - e^{-\left(\alpha x_{a_1:n} + \frac{\beta}{2} x_{a_1:n}^2\right)}\right]} + \frac{1}{2}(m-n) x_{a_m::n}^2 = 0 \tag{7}$$

Because solving the two nonlinear equations (6) and (7) in the two unknown parameters α and β simultaneously is too difficult, it is preferable to solve them numerically using the Newton Raphson method to obtain an approximate solution. See [14] for more information on the steps of the Newton Raphson algorithm. Finally, the Maximum Likelihood Estimators (MLEs) for the parameters α and β will be denoted as $\hat{\alpha}$ and $\hat{\beta}$.

2.1 Approximate confidence intervals

The entries of the inverse matrix of the Fisher information matrix $I_{ij} = E\{-[\partial^2 \ell(\Phi)/\partial \phi_i \partial \phi_j]\}$ gives the asymptotic variances and covariances of the MLEs, $\hat{\alpha}$ and $\hat{\beta}$, where $i, j = 1, 2$ and $\Phi = (\phi_1, \phi_2) = (\alpha, \beta)$. Unfortunately, obtaining the exact closed forms of the previous expectations is difficult. As a result, the observed Fisher information matrix $\hat{I}_{ij} = \left\{-\left[\partial^2 \ell(\Phi)/\partial \phi_i\, \partial \phi_j\right]\right\}_{\Phi=\hat{\Phi}}$, obtained by dropping the expectation operator E, will be used to construct confidence intervals for the parameters. The observed Fisher information matrix entries are simple second partial derivatives of the log-likelihood function. As a result,

$$\hat{I}(\alpha,\beta)=\begin{pmatrix}-\frac{\partial^2\ell}{\partial\alpha^2} & -\frac{\partial^2\ell}{\partial\alpha\partial\beta}\\ -\frac{\partial^2\ell}{\partial\beta\partial\alpha} & -\frac{\partial^2\ell}{\partial\beta^2}\end{pmatrix}_{(\alpha,\beta)=(\hat{\alpha},\hat{\beta})} \tag{8}$$

As a result, for the MLEs, the approximate (or observed) asymptotic variance-covariance matrix $[\hat{V}]$ is obtained by inverting the observed information matrix $\hat{I}(\alpha,\beta)$. Alternatively, equivalent

$$[\hat{V}]=\hat{I}^{-1}(\alpha,\beta)=\begin{pmatrix}\widehat{Var}(\hat{\alpha}) & cov(\hat{\alpha},\hat{\beta})\\ cov(\hat{\alpha},\hat{\beta}) & \widehat{Var}(\hat{\beta})\end{pmatrix} \tag{9}$$

Under certain regularity conditions, $(\hat{\alpha},\hat{\beta})$ is well known to be approximately distributed as multivariate normal with mean (α,β) and covariance matrix $I^{-1}(\alpha,\beta)$, see [23]. The $100(1-\gamma)\%$ two-sided confidence intervals for α and β can then be calculated as follows, where $Z_{\frac{\gamma}{2}}$ is the standard normal distribution's percentile with right-tail probability $\frac{\gamma}{2}$.

$$\widehat{\alpha}\pm Z_{\frac{\gamma}{2}}\sqrt{\widehat{Var}(\hat{\alpha})}\text{ and }\widehat{\beta}\pm Z_{\frac{\gamma}{2}}\sqrt{\widehat{Var}(\widehat{\beta})}, \tag{10}$$

3 Bayes estimation

The Bayesian estimates of the two unknown parameters α and β are obtained in this section using the squared error loss function. Assume the two parameters α and β are independent and follow the jeffrey prior distributions.

$$\begin{aligned}\pi_1(\alpha)&=\alpha^{-1} \qquad &, \alpha>0,\\ \pi_2(\beta)&=\beta^{-1} \qquad &, \beta>0.\end{aligned} \tag{11}$$

Most statistical inference researchers use Bayesian estimation for unknown parameters because it reduces the posterior expected value for loss functions, see [1,2,10,25,27]. The posterior distribution of the parameters α and β, denoted by $\pi^*(\alpha,\beta\mid\text{data})$, can be obtained using Bayes' theorem and up to proportionality by combining the likelihood function (3) with the prior (11) and will be written as

$$\pi^*(\alpha,\beta\mid\text{data})=\frac{\pi_1(\alpha)\,\pi_2(\beta)\,L(\alpha,\beta\mid\text{data})}{\int_0^\infty\int_0^\infty\pi_1(\alpha)\,\pi_2(\beta)\,L(\alpha,\beta\mid\text{data})\,d\alpha d\beta}. \tag{12}$$

A square error loss (SEL) function, which is commonly used, is a symmetric loss function because it assigns equal losses to over and under estimation. The square error loss function will be defined if $\hat{\phi}$ is the estimator for the parameter ϕ.

$$L(\phi,\hat{\phi})=(\hat{\phi}-\phi)^2, \tag{13}$$

As a result, for any function in α and β, say $g(\alpha,\beta)$, we can obtain the Bayes estimate using the SEL function as

$$\hat{g}_{BS}(\alpha,\beta\mid\text{data})=E_{\alpha,\beta|\text{data}}[g(\alpha,\beta)], \tag{14}$$

where

$$E_{\alpha,\beta|\underline{x}}[g(\alpha,\beta)]=\frac{\int_0^\infty\int_0^\infty g(\alpha,\beta)\,\pi_1(\alpha)\,\pi_2(\beta)\,L(\alpha,\beta\mid\text{data})d\alpha d\beta}{\int_0^\infty\int_0^\infty\pi_1(\alpha)\,\pi_2(\beta)\,L(\alpha,\beta\mid\text{data})\,d\alpha d\beta}. \tag{15}$$

Concerning the difficulty of solving the multiple integrals in Equation (15) analytically, the MCMC approximation method is proposed for generating samples from the joint posterior density function in Equation (12) and then using them to compute the Bayes estimate of α and β and construct the associated credible intervals. The joint posterior can be written as follows from Equation (12) up to proportionality.

$$\pi^*(\alpha,\beta \mid \text{data}) \propto \alpha^{-1}\beta^{-1}\prod_{i=1}^{m}(\alpha+\beta x_{a_i:n})e^{-\left(\alpha x_{a_i:n}+\frac{\beta}{2}x^2_{a_i:n}\right)} \times \prod_{i=1}^{m-1}\left[e^{-\left(\alpha x_{a_i:n}+\frac{\beta}{2}x^2_{a_i:n}\right)} - e^{-\left(\alpha x_{a_{i+1}:n}+\frac{\beta}{2}x^2_{a_{i+1}:n}\right)}\right]^{R_{i+1}} \times \left[1-e^{-\left(\alpha x_{a_1:n}+\frac{\beta}{2}x^2_{a_1:n}\right)}\right]^{R_1}\left[e^{(m-n)\left(\alpha x_{a_m:n}+\frac{\beta}{2}x^2_{a_m:n}\right)}\right] \quad (16)$$

Up to proportionality, the full conditionals for α and β can be written as

$$\pi_1^*(\alpha \mid \beta,\text{data}) \propto \alpha^{-1}\prod_{i=1}^{m}(\alpha+\beta x_{a_i:n})e^{-\left(\alpha x_{a_i:n}+\frac{\beta}{2}x^2_{a_i:n}\right)}\prod_{i=1}^{m-1}\left[e^{-\left(\alpha x_{a_i:n}+\frac{\beta}{2}x^2_{a_i:n}\right)} - e^{-\left(\alpha x_{a_{i+1}:n}+\frac{\beta}{2}x^2_{a_{i+1}:n}\right)}\right]^{R_{i+1}} \times \left[1-e^{-\left(\alpha x_{a_1:n}+\frac{\beta}{2}x^2_{a_1:n}\right)}\right]^{R_1}\left[e^{(m-n)\alpha x_{a_m:n}}\right] \quad (17)$$

and

$$\pi_2^*(\beta \mid \alpha,\text{data}) \propto \beta^{-1}\prod_{i=1}^{m}(\alpha+\beta x_{a_i:n})e^{-\left(\alpha x_{a_i:n}+\frac{\beta}{2}x^2_{a_i:n}\right)}\prod_{i=1}^{m-1}\left[e^{-\left(\alpha x_{a_i:n}+\frac{\beta}{2}x^2_{a_i:n}\right)} - e^{-\left(\alpha x_{a_{i+1}:n}+\frac{\beta}{2}x^2_{a_{i+1}:n}\right)}\right]^{R_{i+1}} \times \left[1-e^{-\left(\alpha x_{a_1:n}+\frac{\beta}{2}x^2_{a_1:n}\right)}\right]^{R_1}\left[e^{\frac{\beta(m-n)}{2}x^2_{a_m:n}}\right] \quad (18)$$

It should be noted that the conditional posteriors of α and β in Equations (17) and (18) are not in standard forms, making Gibbs sampling difficult. As a result, the Metropolis-Hastings (M-H) sampler is used in MCMC algorithm implementations. [26] presents a hybrid algorithm with M-H steps for updating α and β given the conditional distributions in Equations (17) and (18).

Because the initial value selections may affect convergence, we will discard M simulated points during the burn-in period. For sufficiently large N, the selected samples, $\alpha^{(j)}$ and $\beta^{(j)}$, $j = M+1,...,N$, will then form approximate posterior samples from which Bayesian inferences will be developed.

In reference to the SEL function in equation (15), the approximate Bayes estimates of $\varphi = \alpha,\beta$ will be written as

$$\hat{\varphi}_{BS} = \frac{1}{N-M}\sum_{j=M+1}^{N}\varphi^{(j)}, \quad (19)$$

4 Two sample prediction

In this section, we will be working on deriving the interval prediction of the future order statistics from a random sample following a LFR distribution based on multiply Type-II hybrid censoring scheme as it may be interested to get the failure times for some observations from a future sample.

Assume that the order statistics of a future random sample of size m are $Y_{1:m} \le Y_{2:m} \le ... \le Y_{m:m}$. Given a probability density function $f(x)$ and a continuous distribution's cumulative distribution function $F(x)$. The marginal density function of the s^{th} order statistic is then given for a random sample of size m from this distribution.

$$g_{Y_{s:m}}(y_s|\theta) = \frac{m!}{(s-1)!\,(m-s)!}[F(y_s)]^{s-1}[1-F(y_s)]^{m-s}f(y_s) = \sum_{q=0}^{m-s}\frac{(-1)^q\binom{m-s}{q}m!}{(s-1)!\,(m-s)!}[F(y_s)]^{s+q-1}f(y_s), \quad (20)$$

where $y_s \ge 0$ and $\theta = (\alpha,\beta)$, see [11]. Substituting by equations (1) and (2) in (20), the marginal density function of $Y_{s:m}$ becomes

$$g_{Y_{s:m}}(y_s \mid \alpha,\beta) = (\alpha+\beta y_s)e^{-(\alpha y_s+\frac{\beta}{2}y_s^2)}\sum_{q=0}^{m-s}\frac{(-1)^q\binom{m-s}{q}m!}{(s-1)!\,(m-s)!}\left[1-e^{-\left(\alpha y_s+\frac{\beta}{2}y_s^2\right)}\right]^{s+q-1} \quad (21)$$

Multiplying the marginal density function in equation (21) and the joint posterior in equation (16), then integrating this product over the set $\{(\alpha,\beta);\ 0<\alpha<\infty,\ 0<\beta<\infty\}$ we can get the predictive posterior density of future observations under Type-II multiply hybrid censoring scheme as follows:

$$g^*(y_s \mid \mathbf{x}) = \int_0^\infty \int_0^\infty g_{Y_{s:m}}(y_s \mid \mathbf{x})\pi^*(\alpha,\beta \mid \mathbf{x})d\alpha d\beta. \tag{22}$$

One can note that equation (22) is difficult to be solved analytically. Therefore, we will use MCMC samples through Gibbs algorithm to obtain an estimator for the predictive posterior $g^*(y_s \mid \mathbf{x})$. Assume that $\{(\alpha_i,\beta_i), i=1,2,\ldots,N\}$ be the MCMC samples derived from $\pi^*(\alpha,\beta \mid \mathbf{x})$, then we can get the consistent estimate of $g^*(y_s \mid \mathbf{x})$ as follows

$$\begin{aligned} g^*(y_s \mid \underline{x}) = & \frac{1}{N-M}\sum_{i=M+1}^{N}(\alpha^{(i)}+\beta^{(i)}y_s)e^{-(\alpha^{(i)}y_s+\frac{\beta^{(i)}}{2}y_s^2)} \\ & \times \sum_{q=0}^{m-s}\frac{(-1)^q\binom{m-s}{q}m!}{(s-1)!\,(m-s)!}[1-e^{-(\alpha^{(i)}y_s+\frac{\beta^{(i)}}{2}y_s^2)}]^{s+q-1}. \end{aligned} \tag{23}$$

In addition to their applications in Bayesian parameter estimation, loss functions also play a vital role in Bayesian prediction. Using a loss function, we will find a point predictor which minimizes risk among all other predictors, known as the Bayesian point predictor. To forecast a future observation, we will use an SEL function as well as the Bayesian point predictors of $Y_s, 1\le s\le N$, denoted as $\hat{Y}_{SELP}$ under the SEL function.

$$\hat{Y}_{SELP} = \int_0^\infty y_s g^*(y_s|\underline{x})dy_s = \frac{1}{N-M}\sum_{i=M+1}^{N}\int_0^\infty y_s\ g(y_s|\alpha^{(i)},\beta^{(i)},\underline{x})dy_s. \tag{24}$$

A prediction interval is an interval that relies on information from a sample provided to generate leads for a future sample out of a fixed population with a predetermined probability. Considering the density function $g(y_s \mid \alpha,\beta,\mathbf{x})$, the distribution function is given as

$$G(y_s \mid \alpha,\beta,\underline{x}) = \int_0^{y_s}\sum_{q=0}^{m-s}\frac{(-1)^q\binom{m-s}{q}m!}{(s-1)!\,(m-s)!}[F(y_s)]^{s+q-1}f(y_s) = \sum_{q=0}^{m-s}\frac{(-1)^q\binom{m-s}{q}m!}{(s-1)!\,(m-s)!}\int_0^{y_s}[F(y_s)]^{s+q-1}f(y_s). \tag{25}$$

Let $G^*_{Y_{s:m}}$ denotes the predictive distribution estimator for y_s, then a simulation consistent estimator of $G^*_{Y_{s:m}}(y_s \mid \mathbf{x})$ could be written as follows

$$G^*(y_s \mid \alpha,\beta,\underline{x}) = \frac{1}{N-M}\sum_{i=M+1}^{N}G(y_s|\alpha^{(i)},\beta^{(i)},\underline{x}). \tag{26}$$

Therefore, the $100(1-\gamma)\%$ confidence interval for the Bayesian predictive interval can be derived after solving the following two equations simultaneously:

$$G^*_{Y_{s:m}}(L_{y_{s:m}}|\underline{x}) = 1-\frac{\gamma}{2} \text{ and } G^*_{Y_{s:m}}(U_{y_{s:m}}|\underline{x}) = \frac{\gamma}{2}, \tag{27}$$

where $L_{y_{s:m}}$ and $U_{y_{s:m}}$ indicate the lower and upper bounds, respectively. It is evident that is not possible to compute the two equations in (27) analytically. Then, the MCMC method is suggested for constructing the Bayesian prediction intervals.

5 Application to real-life data

In this section, two real data sets will be discussed to explain the general methods in the previous sections. The first refer to failure times to some industrial data and the second is an application on some clinical data.

5.1 Example on Industrial data

First, we consider the classical real data-set in [20] on the times, in operating days, between successive failures of air conditioning equipment in an aircraft. This data is recorded as shown in Table 1. Comparing the empirical distribution for the failure data and the CDF for the LFR distribution as shown in Figure 2, it was found that the Kolmogorov Smirnov (K-S) distance between them is 0.102272 with a P-value equal to 0.891983. Therefore, the LFR distribution is fitting well to the mentioned data.

Table 1: Times between failures of air conditioning equipment in an aircraft

0.417	0.583	0.833	0.958	1.000	1.042	1.083	1.208	1.833	1.833
2.042	2.333	2.458	2.500	2.542	2.583	2.917	3.167	3.292	3.500
3.750	4.208	4.917	5.417	6.500	7.750	8.667	8.667	12.917	

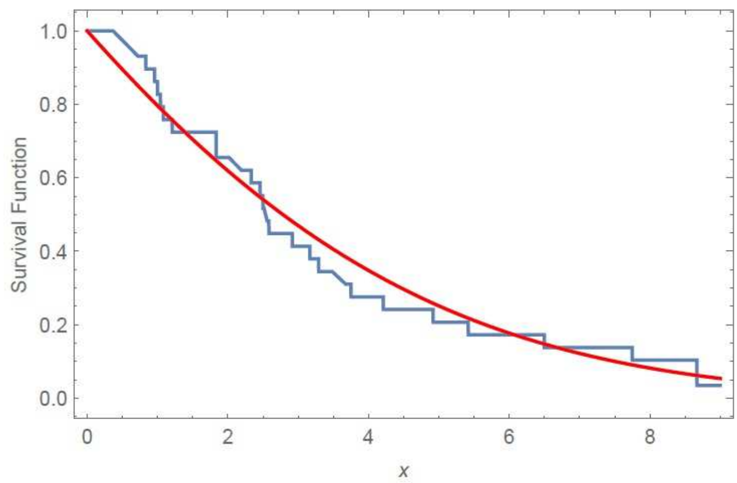

Fig. 2: Empirical and fitted survival functions of data-set in Table 1.

Then a multiply Type-II hybrid censored sample of effective size $m = 23$ was randomly selected from the 29 failure observations in Table 1 with the multiply Type-II hybrid censored scheme $R = (0,1,0,0,1,0,0,0,1,0,0,0,2,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0)$ as shown in Table 2.

Table 2: Multiply Type-II hybrid failures data

0.417	0.833	0.958	1.042	1.083	1.208	1.833	2.042	2.333	2.542
2.583	3.167	3.292	3.500	3.750	4.208	4.917	5.417	6.500	7.750
8.667	8.667	12.917							

The MLEs for the parameters, α and β, based on Multiply Type-II hybrid censored data in addition to the Bayes estimates relative to SEL function for the same parameters are displayed in Table 3. Also the 95% approximate confidence intervals and credible intervals for the parameters α and β are calculated and the results are also displayed in Table 3.

Table 3: The point estimates and 95% CIs for α and β

	Point estimate		95% CIs	
	MLE	SEL	MLE	MCMC
α	0.215785	0.214591	[0.0317858,0.399783]	[0.212914,0.216269]
β	0.0255161	0.0246476	[-0.0424651,0.0934974]	[0.0244673,0.0248278]

From Table 3, we see that the values of estimates are close together which indicates the good performance of the estimators.

Using the MCMC method, a two-sample Bayesian prediction is then computed and the corresponding summary statistics are presented in Table 4. It is noted that as s increases, the standard error as well as the confidence intervals widths increase which means that predictive densities get more and more larger as the order statistics become larger.

Table 4: The values of point predictions and 95% PIs for y_s.

		95% PIs					95% PIs		
s	SEL	Lower	Upper	Length	s	SEL	Lower	Upper	Length
1	0.251806	0.00655	0.90774	0.90118	10	3.04697	1.5711	4.8969	3.3258
2	0.51109	-0.05572	1.37661	1.43233	11	3.46596	1.86211	5.45045	3.58833
3	0.778955	0.1682	1.8007	1.6325	12	3.92753	2.18426	6.06427	3.88001
4	1.01567	0.3034	2.21117	1.90777	13	4.44512	2.5447	6.76104	4.21634
5	1.34586	0.46293	2.62059	2.15766	14	5.03974	2.95444	7.57723	4.62279
6	1.64831	0.6436	3.03674	2.39314	15	5.74726	3.43147	8.57767	5.14621
7	1.96635	0.8442	3.46591	2.62171	16	6.63705	4.00814	9.89545	5.8873
8	2.30283	1.06478	3.91421	2.84943	17	7.87495	4.75343	11.876	7.12256
9	2.66143	1.3064	4.38845	3.08205	18	10.746	5.87414	15.9654	10.0913

5.2 Example on Clinical data

The second data-set to be considered here is the survival times (in years) after diagnosis of 43 patients with a certain kind of leukemia as shown in Table 5, see [19]. The Kolmogorov Smirnov (K-S) distance between the empirical distribution for the failure data and the CDF for the LFR distribution, as shown in Figure 3, is 0.0816726 with a P-value of 0.914159. Therefore, the LFR distribution is a reasonable choice for this data.

Table 5: Survival times after diagnosis of 43 patients with a kind of leukemia

0.019	0.129	0.159	0.203	0.485	0.636	0.748	0.781	0.869	1.175
1.206	1.219	1.219	1.282	1.356	1.362	1.458	1.564	1.586	1.592
1.781	1.923	1.959	2.134	2.413	2.466	2.548	2.652	2.951	3.038
3.600	3.655	3.745	4.203	4.690	4.888	5.143	5.167	5.603	5.633
6.192	6.655	6.874							

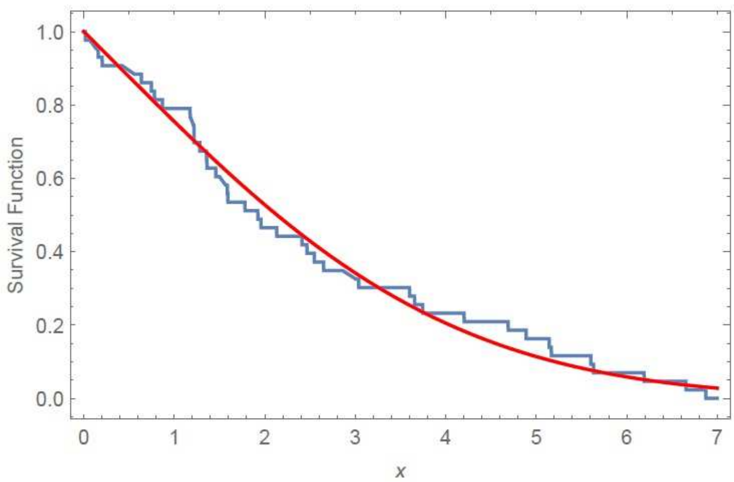

Fig. 3: Empirical and fitted survival functions of data-set in Table 5.

The 43 failure observations in Table 5 are then randomly selected for a multiply Type-II hybrid censored sample of effective size $m = 33$ using the multiply Type-II hybrid censored scheme

$R = (0,0,3,0,2,1,1,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$ as shown in Table 6.

Table 6: Multiply Type-II hybrid failures data

0.019	0.129	0.636	0.748	1.175	1.219	1.282	1.362	1.564	1.586
1.781	1.923	1.959	2.134	2.413	2.466	2.548	2.652	2.951	3.038
3.600	3.655	3.745	4.203	4.690	4.888	5.143	5.167	5.603	5.633
6.192	6.655	6.874							

Table 7 shows the MLEs for the parameters α and β based on Multiply Type-II hybrid censored data, as well as the Bayes estimates relative to the SEL function for the same parameters. The approximate confidence intervals and credible intervals for the parameters α and β are also calculated and displayed in Table 7.

Table 7: The point estimates and 95% CIs for α and β

	Point estimate		95% CIs	
	MLE	SEL	MLE	MCMC
α	0.24474	0.265739	[0.0363186,0.453161]	[0.248641,0.28227]
β	0.075861	0079844	[-0.0617529,0.213475]	[0.0766457,0.0841982]

The values of the estimates in Table 7 are close together, indicating that the estimators performed well.

A two-sample Bayesian prediction is then computed using the MCMC method, and its summary statistics are shown in Table 8. It is worth noting that as s increases, so do the standard error and confidence interval widths, implying that predictive densities grow larger and larger as the order statistics increase.

Table 8: The values of point predictions and 95% PIs for y_s.

		95% PIs					95% PIs		
s	SEL	Lower	Upper	Length	s	SEL	Lower	Upper	Length
1	0.182964	0.0049	0.64442	0.639522	12	2.3435	1.37295	3.4791	2.10615
2	0.364917	-0.04168	0.95305	0.994733	13	2.58715	1.56126	3.77906	2.2178
3	0.546774	0.1234	1.22096	1.09755	14	2.84894	1.76337	3.56278	1.79941
4	0.729419	0.21969	1.47136	1.25167	15	3.13419	1.98226	4.79029	2.80803
5	0.913746	0.33041	1.71323	1.38282	16	3.45066	2.22227	3.64974	1.42747
6	1.10068	0.45237	1.9516	1.49923	17	3.8104	2.48999	7.17473	4.68474
7	1.29122	0.58385	2.18994	1.60609	18	4.23389	2.79605	7.36996	4.57391
8	1.48646	0.72403	2.43111	1.70708	19	4.76065	3.15952	6.67321	3.51369
9	1.68766	0.87265	2.67773	1.80509	20	5.48471	3.62082	7.7991	4.17828
10	1.8963	1.02988	2.9325	1.90262	21	6.75234	4.30081	5.90783	1.60702
11	2.11417	1.19631	3.19843	2.00212					

6 Simulation Study

In this section, a simulation is performed to compare the estimators of the parameters on the basis of the samples generated from the LFR distribution. According to the Mean Square Errors (MSEs), average confidence interval lengths (ACILs) and the coverage probabilities (CPs), the performance of the competitive estimators have been compared.

Accordingly, a random sample of size n has been generated from the LFR distribution for fixed values of its parameters α and β. To study the effect of the hybrid censoring scheme on the performance of the estimators, different combinations of the censoring parameters (R,T) have been considered and simulation results are then summarized in Table 9 and Table 10.

Table 9. MSE of MLEs and Bayesian estimates for the parameter α, β with $\alpha_\circ = 2, \beta_\circ = 5$.

				α		β	
n	T	R_i	a_r	MLE	SEL	MLE	SEL
30	3	$R_2 = 2, R_5 = 1$	10	2.3855 (0.4031)	2.3855 (0.4031)	5.3006 (0.8326)	5.3007 (0.8323)
		$R_i = 0, i \neq 2,5$	18	2.3491 (0.3897)	2.3491 (0.3898)	5.3329 (0.803)	5.3329 (0.8032)
			25	2.3789 (0.4183)	2.379 (0.4183)	5.2453 (0.7485)	5.2453 (0.7484)
	5	$R_2 = 2, R_5 = 1$	10	2.3743 (0.4287)	2.3743 (0.4287)	5.2584 (0.7683)	5.2583 (0.7685)
		$R_i = 0, i \neq 2,5$	18	2.3768 (0.4211)	2.3768 (0.4211)	5.2689 (0.7412)	5.269 (0.7412)
			25	2.3799 (0.3873)	2.3798 (0.3873)	5.3006 (0.8011)	5.3005 (0.8011)
40	3	$R_4 = 2, R_9 = 1$	18	2.2745 (0.2968)	2.2745 (0.2968)	5.2792 (0.7653)	5.2792 (0.7656)
		$R_i = 0, i \neq 2,9$	30	2.2849 (0.2944)	2.2848 (0.2944)	5.2671 (0.7794)	5.2673 (0.7795)
			35	2.2443 (0.2731)	2.2443 (0.2731)	5.329 (0.8159)	5.329 (0.816)
	5	$R_4 = 2, R_9 = 1$	18	2.2781 (0.2949)	2.2781 (0.2949)	5.3002 (0.8198)	5.3001 (0.8198)
		$R_i = 0, i \neq 2,9$	30	2.255 (0.2914)	2.255 (0.2915)	5.3123 (0.7931)	5.3124 (0.7931)
			35	2.2821 (0.3118)	2.2821 (0.3118)	5.2793 (0.7458)	5.2793 (0.7458)

Table 10. ACL and CP of 95% CIs for the parameters α and β

				α		β	
				MLE	MCMC	MLE	MCMC
n	T	R_i	a_r				
30	3	$R_2 = 2, R_5 = 1$	10	3.397 (0.9456)	0.0015 (0.9358)	13.9187 (0.9312)	0.0061 (0.9294)
		$R_i = 0, i \neq 2,5$	18	3.3796 (0.9433)	0.0014 (0.9329)	13.8409 (0.971)	0.0058 (0.9606)
			25	3.3822 (0.9738)	0.0015 (0.9666)	13.8007 (0.9538)	0.0059 (0.9483)
	5	$R_2 = 2, R_5 = 1$	10	3.3913 (0.9329)	0.0015 (0.9316)	13.8519 (0.9489)	0.006 (0.9508)
		$R_i = 0, i \neq 2,5$	18	3.3932 (0.9377)	0.0014 (0.9713)	13.8733 (0.9568)	0.006 (0.9436)
			25	3.3956 (0.9400)	0.0015 (0.9406)	13.9008 (0.9576)	0.0059 (0.9695)
40	3	$R_4 = 2, R_9 = 1$	18	2.9366 (0.9303)	0.0013 (0.9571)	11.8872 (0.9504)	0.005 (0.9483)
		$R_i = 0, i \neq 2,9$	30	2.9514 (0.9273)	0.0013 (0.9513)	11.9389 (0.9533)	0.0051 (0.9269)
			35	2.929 (0.9694)	0.0012 (0.9742)	11.8674 (0.9462)	0.005 (0.9666)
	5	$R_4 = 2, R_9 = 1$	18	2.9452 (0.9688)	0.0013 (0.9746)	11.9416 (0.9648)	0.0051 (0.9324)
		$R_i = 0, i \neq 2,9$	30	2.9269 (0.9690)	0.0013 (0.9637)	11.8574 (0.9538)	0.005 (0.9512)
			35	2.9454 (0.9732)	0.0013 (0.9461)	11.9353 (0.9467)	0.0051 (0.9373)

From the reported values, we observed the following:

– As the sample size increases, the MSEs for most of the estimators decreases, for the same values of α and β.

–For different values of n, T and a_r, the MSEs of Bayes estimators and MLEs are nearly the same.
–The Bayes estimators obtained are nearly the same to the MLEs. Also it is noticed that providing prior information gives more accurate results for the Bayesian estimates of the parameters.
–The credible intervals width is less than that of the asymptotic confidence intervals in all the cases. However, it is noted that the confidence intervals width decreases as the sample size increases.

7 Conclusion

The Linear failure rate distribution is well-fitted to some real-life data which leads to obtaining good estimators for its parameters. Multiply Type-II hybrid censoring scheme is a good description model for experiments that lack the exact registration of the failure times for some units. A prediction study is applied and proven that the predictive sample simulates the observed sample. It is recommended to apply the Multiply hybrid censoring scheme to clinical data especially in cases of pandemics when it is difficult for researchers to register all the failure times exactly.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] E. A. Ahmed, *Bayesian estimation based on progressive Type-II censoring from two-parameter bathtub-shaped lifetime model: a Markov chain Monte Carlo approach*, Journal of Applied Statistics 41 (4) 2014, pp. 752—768.
[2] E. A. Ahmed, *Estimation and prediction for the generalized inverted exponential distribution based on progressively first-failure-censored data with application*, Journal of Applied Statistics 2014. Available at http://dx.doi.org/10.1080/02664763.2016.1214692.
[3] B. Al-zahrani and M. Gindwan, *Parameter estimation of a two-parameter Lindley distribution under hybrid censoring*, International Journal of System Assurance Engineering and Management 5 (2014), pp. 628–636.
[4] A. Azzalini, *Statistical inference based on the likelihood*, Chapman and Hall, London (1996).
[5] A. Banerjee and D. Kundu, *Inference based on Type-II hybrid censored data from a Weibull distribution*, IEEE Transactions on Reliability 57 (2008), pp. 369–378.
[6] S. M. Chen and G. K. Bhattacharyya, *Exact confidence bound for an exponential under hybrid censoring*, Communication in Statistics - Theory and Methods 17 (1988), pp. 1858–1870.
[7] Chien-Wei Wu, Ming-Hung Shu and Yu-Ning Chang, *Variable-sampling plans based on lifetime-performance index under exponential distribution with censoring and its extensions*, Applied Mathematical Modelling 55 (2018), pp. 81–93.
[8] A. Childs, B. Chandrasekhar, N. Balakrishnan and D. Kundu, *Exact likelihood inference based on type-I and type-II hybrid censored samples from the exponential distribution*, Annals of the Institute of Statistical Mathematics 55 (2003), pp. 319–330.
[9] Chin-Diew Lai and Min Xie, *Stochastic Ageing and Dependence for Reliability*, Springer, New York, NY.
[10] M. Y. Danish, I. A. Arshad and M. Aslam, *Bayesian inference for the randomly censored Burr-type XII distribution*, Journal of Applied Statistics 45 (2) 2018, pp. 270—283.
[11] H.A. David and H.N. Nagaraja, *Order Statistics*, 3rd ed., John Wiley & Sons, New York, 2003.
[12] S. Dey and B. Pradhan, *Generalized inverted exponential distribution under hybrid censoring*, Statistical Methodology 18 (2014), pp. 101–114.
[13] S. Dubey, B. Pradhan and D. Kundu (2011), *Parameter estimation of the hybrid censored log-normal distribution*, Journal of Statistical Computation and Simulation 81 (2011), pp. 275–287.
[14] R. M. EL-Sagheer, *Estimation of parameters of Weibull-Gamma distribution based on progressively censored data*, Statistical Papers 59 (2018), pp. 725—757. Available at https://doi.org/10.1007/s00362-016-0787-2.
[15] A. Ganguly, S. Mitra, D. Samanta and D. Kundu, *Exact inference for the two parameter exponential distribution under Type-II hybrid censoring*, Journal of Statistical Planning and Inference 142 (2012), pp. 613–625.
[16] P. K. Gupta and B. Singh, *Parameter estimation of Lindley distribution with hybrid censored data*, International Journal of System Assurance Engineering and Management, 4 (2013), pp. 378–385.
[17] Y. E. Jeon, S. B. Kang, *Bayesian estimation for the exponential distribution based on generalized multiply Type-II hybrid censoring*, Communications for Statistical Applications and Methods 27 (4) 2020, pp. 413—430.
[18] Y. E. Jeon, S. B. Kang, *Estimation for the half-logistic distribution based on multiply Type-II hybrid censoring*, Physica A 550 (2020). Available at https://doi.org/10.1016/j.physa.2020.124501.
[19] N. L. Johnson and S. Kotz, *Encyclopedia of Statistical Sciences*, Wiley, New York, (1983).
[20] J. P. Keating, R. E. Glaser and N. S. Ketchum, *Testing hypotheses about the shape of a gamma distribution*, Technometrics 32 (1990), 67-82.

[21] A. Kohansal, S. Rezakhah and E. Khorram, *Parameter estimation of Type-II hybrid censored weighted exponential distribution*, Communications in Statistics - Simulation and Computation 44 (2015), pp. 1273–1299.

[22] D. Kundu, *On hybrid censored Weibull distribution*, Journal of Statistical Planning and Inference 137 (2007), pp. 2127–2142.

[23] J. F. Lawless, *Statistical models and methods for lifetime data*, John Wiley and Sons, New York (1982).

[24] WH. Lee and KJ. Lee, *Estimating the parameter of an exponential distribution under multiply Type-II hybrid censoring*, Journal of the Korean Data and Information Science Society 29 (2018), pp. 807–814.

[25] M. A. W. Mahmoud , D. A. Ramadan and M. M. M. Mansour, *Estimation of lifetime parameters of the modified extended exponential distribution with application to a mechanical model*, Communications in Statistics - Simulation and Computation 2020. Available at https://doi.org/10.1080/03610918.2020.1821887.

[26] M. M. M. Mansour and M. S. Aboshady, *ASSESSING THE PERFORMANCE OF INSULATING FLUIDS VIA POINT OF STATISTICAL INFERENCE VIEW*, TWMS J. App. and Eng. Math. 12 (2) 2022, pp. 469-480.

[27] M. M. M. Mansour and D. A. Ramadan, *STATISTICAL INFERENCE TO THE PARAMETERS OF THE MODIFIED EXTENDED EXPONENTIAL DISTRIBUTION UNDER THE TYPE-II HYBRID CENSORING SCHEME*, Journal of Applied Probability and Statistics 15 (2) 2020, pp. 19—44.

[28] S. Nadarajaha and S. A. A. Bakar, *An exponentiated geometric distribution*, Applied Mathematical Modelling 40 (2016), pp. 6775–6784.

[29] S. Park and N. Balakrishnan, *A very flexible hybrid censoring scheme and its Fisher information*, Journal of Statistical Computation and Simulation 82 (2012), pp. 41–50.

[30] M. K. Rastogi and Y. M. Tripathi, *Inference on unknown parameters of a Burr distribution under hybrid censoring*, Statistical Papers 54 (2013), pp. 619–643.

[31] R. Royall, *Statistical Evidence: A Likelihood Paradigm*, Chapman and Hall, London (1997).

[32] Tian-Qun Xu and Yue-Peng Chen, *Two-sided M-Bayesian credible limits of reliability parameters in the case of zero-failure data for exponential distribution*, Applied Mathematical Modelling 38 (2014), pp. 2586–2600.