









## Bayesian estimation of Rayleigh distribution in the presence of outliers using progressive censoring

Farzana Noor<sup>\*1</sup> , Ahthasham Sajid<sup>2</sup> , Maha Ghazal<sup>1</sup> , Imranullah Khan<sup>3</sup> ,  
Mehwish Zaman<sup>1</sup> , Imran Baig<sup>4</sup> 

<sup>1</sup>Department of Mathematics and Statistics, International Islamic University, Islamabad-Pakistan

<sup>2</sup>Department of Computer Science, Faculty of ICT Baluchistan University of Information Technology Engineering and Management Sciences, Quetta, Baluchistan, Pakistan

<sup>3</sup>College of Underwater Acoustic Communication Engineering, Harbin Engineering University, 150001, Heilongjiang, Harbin, China

<sup>4</sup>Department of Electrical Engineering, College of Engineering, Dhofar University, Oman

### Abstract

In this article, Maximum likelihood estimation (MLE) and Bayesian estimation for Rayleigh distribution using progressive type-II censoring in the presence of outliers is considered. Inverse Gamma prior and Jeffreys prior are used for Bayesian estimation. Squared error loss function (SELF), precautionary loss function (PLF) and K-loss function (KLF) are used for obtaining the expressions of Bayes estimators and posterior risks. Credible intervals are also derived. A simulation study is presented to discuss the behavior of Bayes estimators. Applicability of the undertaken study is highlighted using three real data sets.

**Mathematics Subject Classification (2010).** 62N05, 62F10, 62F12, 62F15

**Keywords.** Rayleigh distribution, outliers, progressive type-II censoring, life testing experiments, prior distribution, loss function

### 1. Introduction

An observation markedly away from the measurements in the data set is referred to as an outlier. Though outliers generally represent random error but may contain valuable information. Data especially from industrial processes may found to be contaminated with outliers. These outliers are often eliminated or discarded from the data set before analysis. But it is of immense importance to know the nature of outliers present in the data sets and handle these nuisance quantities carefully rather than to discard them. Methods and tools exist for finding and testing the outliers contained in the data set. For detailed study on outliers we refer to Anscombe [6], Grubbs [15], Barnett [10], Kale [17], Hawkins [16], Barnett [11], and Kale and Kale [18].

\*Corresponding Author.

Email addresses: farzana.akhtar@iiu.edu.pk (F. Noor), ahthasham.sajid@buitms.edu.pk (A. Sajid), ghazalmaha4@gmail.com (M. Ghazal), khan@hrbeu.edu.cn (I. Khan), mehwishbalouch008@gmail.com (M. Zaman), ibaig@du.edu.om (I. Baig)

Received: 22.10.2019; Accepted: 19.09.2020

In life testing experiments and survival analysis, censoring is a commonly and frequently used technique. It is introduced in situations when a portion of lifetime distribution is known and the remaining portion is known to exceed some value. A huge literature is available on the study of censored sampling spanning at about 50 years and inference on the basis of censored samples for a range of life time distributions such as Normal, Exponential, Gamma, Rayleigh, Weibull, Extreme value, Log-normal, Inverse Gaussian, Laplace have been conducted by many authors.

A more generalized version of the type-II censoring scheme known as the progressive type-II censoring scheme is available that provides the facility to remove the units during the experiment and it also reduces cost and consumption of time during the experiment. A substantial description on progressive censoring has been given by Balakrishnan and Aggarwala [9]. Studies on different lifetime distributions using progressive censoring schemes have been conducted such as by Aslam et al. [7], Mann [20], Childa and Balakrishnan [12], and Soliman [24]. Maximum likelihood estimation and Bayesian estimation under progressive type-II censoring scheme is given by Yuen and Tse [26], Wu and Chang [25], and Sarhan and Abuammoh [22].

Rayleigh distribution is a widely used distribution in life testing experiments, survival as well as reliability analysis. Rayleigh distribution is analyzed by Dyer and Whisenand [14]. Bayesian predictive intervals are derived for the two-component Rayleigh mixture model by Saleem and Aslam [21]. Classical estimation using type-I progressive censoring is presented on mixed Rayleigh distribution by Afify [2]. Bayesian estimation of Rayleigh distribution using progressive type-II censoring is conducted by Azimi and Yaghmaei [8]. In this study, our focus is to obtain parameter estimates of Rayleigh distribution when data from the considered model contain outliers. Outliers are generally discarded so it deems important to utilize them efficiently.

## 2. Materials and methods

In this section, we introduce the model, construct its likelihood function, obtain posterior density and conduct Bayesian estimation of the proposed model.

### 2.1. Rayleigh probability model and its likelihood function

Consider a random sample  $X_1, X_2, \dots, X_n$  such that  $m$  of them are distributed as:

$$f(x, \alpha, \beta) = \frac{2x}{\alpha\beta} e^{-\frac{x^2}{\alpha\beta}}, \quad 0 < x < \infty; \quad \alpha, \beta > 0 \quad (2.1)$$

and remaining  $n - m$  of them can be modeled by:

$$f(x, \alpha) = \frac{2x}{\alpha} e^{-\frac{x^2}{\alpha}}, \quad 0 < x < \infty; \quad \alpha > 0 \quad (2.2)$$

The joint probability density function of  $X_1, X_2, \dots, X_n$  in the presence of outliers according to Dixit [13] is:

$$f(x_1, x_2, \dots, x_n; \alpha, \beta) = \prod_{i=1}^n f(x_i; \alpha) \sum_{\underline{A}} \prod_{j=1}^m \left[ \frac{f(x_{A_j}; \alpha, \beta)}{f(x_{A_j}; \alpha)} \right] \quad (2.3)$$

where

$$\sum_{\underline{A}} = \sum_{A_1=1}^{n-m+1} \sum_{A_2=A_1+1}^{n-m+2}, \dots, \sum_{A_m=A_{m-1}+1}^n$$

By substituting expressions from (2.1), and (2.2) in (2.3), we get

$$f(x_1, x_2, \dots, x_n, \alpha, \beta) = \frac{2^n}{\alpha^n \beta^n} \prod_{i=1}^n x_i e^{-\frac{1}{\alpha} \left( \sum_{i=1}^n x_i^2 \right)} \sum_{\underline{A}} \prod_{j=1}^m e^{-\frac{1}{\alpha} \left( \frac{1}{\beta} - 1 \right) (x_{A_j})^2} \quad (2.4)$$

and marginal distribution of  $X$ 's can be obtained as:

$$f(x_i, \alpha, \beta) = \frac{m}{n} \left( \frac{2x_i}{\alpha\beta} e^{-\frac{x_i^2}{\alpha\beta}} \right) + \left( \frac{n-m}{n} \right) \left( \frac{2x_i}{\alpha} e^{-\frac{x_i^2}{\alpha}} \right) \quad x > 0; \alpha, \beta > 0 \quad (2.5)$$

The likelihood function under the progressive type-II censoring scheme in the presence of outliers is given by

$$L(\alpha, \beta|x) = \frac{C_1}{C(r, s)} \prod_{i=1}^r [f(x_i, \alpha) S(x_i, \alpha)^{R_i}] \sum_A \sum_{j=1}^s \left[ \frac{f(x_{A_j}; \alpha, \beta) S(x_{A_j}; \alpha, \beta)^{R_i}}{f(x_{A_j}; \alpha) S(x_{A_j}; \alpha)^{R_i}} \right] \quad (2.6)$$

substitution of respective functions and simplification leads to:

$$L(x|\alpha, \beta) = \frac{2^r}{\alpha^r \beta^s} \prod_{i=1}^r (x_i) e^{-\frac{1}{\alpha} \sum_{i=1}^r (x_i^2)(R_i+1)} \sum_A e^{-\frac{1}{\beta} \left( \frac{1}{\beta} - 1 \right) \sum_{j=1}^s (x_{A_j}^2)(R_{A_j}+1)} \quad 0 < x < \infty; \alpha, \beta > 0 \quad (2.7)$$

Here,  $s$  is the number of outliers out of  $r$ . where

$$\sum_A = \sum_{A_1=1}^{r-s+1} \sum_{A_2=A_1+1}^{r-s+2}, \dots, \sum_{A_r=A_{r-1}+1}^r$$

### 2.2. Maximum likelihood estimation

To get Maximum Likelihood Estimators (MLEs) for the Rayleigh probability model in the presence of outliers, we take logarithm of the likelihood function (2.7) and differentiate it with respect to parameters which yields the following equations.

$$\frac{\partial(L(\alpha, \beta|x))}{\partial\alpha} = -\frac{r}{\alpha} + \frac{1}{\alpha^2} \sum_{i=1}^r (x_i^2)(R_i + 1) + \sum_A \frac{1}{\alpha^2} \left( \frac{1}{\beta} - 1 \right) \sum_{j=1}^s (x_{A_j}^2)(R_{A_j} + 1) \quad (2.8)$$

$$\frac{\partial(L(\alpha, \beta|x))}{\partial\beta} = -\frac{s}{\beta} + \frac{1}{\alpha\beta^2} \sum_{j=1}^s (x_{A_j}^2)(R_{A_j} + 1) \quad (2.9)$$

Solving simultaneously (2.8) and (2.9) for  $\alpha, \beta$  and equating to zero, we can get MLEs for the parameters of interest. Assuming asymptotic normality 100(1 -  $\alpha$ )% confidence intervals for  $\alpha$  and  $\beta$  can be obtained as  $\hat{\alpha} \pm Z_{\alpha/2} SE(\hat{\alpha})$  and  $\hat{\beta} \pm Z_{\alpha/2} SE(\hat{\beta})$  where  $Var(\hat{\alpha})$  and  $Var(\hat{\beta})$  are the diagonal entities obtained by inverting the Fisher Information Matrix which may be obtained as:

$$I(\alpha, \beta) = \begin{bmatrix} E\left(-\frac{\partial^2}{\partial\alpha^2} \text{Log}L(\alpha, \beta)\right) & E\left(-\frac{\partial^2}{\partial\alpha\partial\beta} \text{Log}L(\alpha, \beta)\right) \\ E\left(-\frac{\partial^2}{\partial\alpha\partial\beta} \text{Log}L(\alpha, \beta)\right) & E\left(-\frac{\partial^2}{\partial\beta^2} \text{Log}L(\alpha, \beta)\right) \end{bmatrix} \quad (2.10)$$

### 2.3. Bayesian estimation

In this section, we obtain Bayes estimators, posterior risks and credible intervals for Rayleigh distribution. We need to specify a loss function to arrive at Bayes estimator and its posterior risk which is the expected value of a given loss function Ali [5]. Three loss functions, Squared error loss function (SELF), precautionary loss function (PLF) and K-loss Function (KLF) is used to derive the expressions of Bayes estimators and their posterior risks. We refer to Ali [3] and Ali [4] for derivation and details on loss functions. These three loss functions are used to deal with different situations so employed with the purpose to choose efficient Bayes estimators.

**2.3.1. Posterior distribution assuming informative prior (IP).** The inverse gamma distribution is used as prior distribution for both parameters. The joint prior density function of  $\alpha$  and  $\beta$  may be written as:

$$p(\alpha, \beta) = \frac{b^a}{\Gamma a} \frac{d^c}{\Gamma c} \alpha^{-(a-1)} \beta^{-(c-1)} e^{-\frac{b}{\alpha} - \frac{d}{\beta}} \quad \alpha, \beta > 0 \quad (2.11)$$

and posterior distribution in the presence of  $s$  outliers under progressive type-II censoring is:

$$p(\alpha, \beta | x) = \sum_{\underline{A}} \left[ \frac{\alpha^{-(r+a+1)} e^{-\frac{1}{\alpha}[b+D_1-D_2]} \sum_{\underline{A}} \left[ \beta^{-(c+s+1)-1} \exp^{-\frac{d}{\beta}} \right]}{\Gamma(r+a+1)\Gamma(c+s+1)/(b+D_1-D_2)^{(r+a+1)} d^{(c+s+1)}} \right] \quad (2.12)$$

$$\alpha, \beta > 0$$

Here  $\sum_{\underline{A}} \left[ \frac{\Gamma(r+a+1)\Gamma(c+s+1)}{(b+D_1-D_2)^{(r+a+1)} d^{(c+s+1)}} \right]$  is the normalizing constant.

$$D_1 = \sum_{i=1}^r (x_i^2)(R_i + 1) \quad D_2 = \sum_{j=1}^s (x_{A_j}^2)(R_{A_j} + 1)$$

Marginal posterior distributions of  $\alpha$  and  $\beta$  having  $s$  outliers which are used to obtain Bayes estimators are:

$$p(\alpha | x) = \sum_{\underline{A}} \left[ \frac{\alpha^{-(r+a+1)-1} e^{-\frac{1}{\alpha}[b+D_1-D_2]} \Gamma(c+s+1)}{d^{(c+s+1)}} \right] \quad (2.13)$$

$$p(\beta | x) = \sum_{\underline{A}} \left[ \frac{\beta^{-(c+s+1)-1} e^{-\frac{d}{\beta}} \Gamma(r+a+1)}{(b+D_1-D_2)^{(r+a+1)}} \right] \quad (2.14)$$

**2.3.2. Posterior distribution assuming non-informative prior (NIP).** Non-informative priors are also the important part of Bayesian statistics. We use Jeffrey's prior as NIP to carry out Bayesian analysis. The joint Jefferys prior for the parameters of Rayleigh distribution is:

$$g(\alpha, \beta) \propto \frac{1}{\alpha\beta} \quad (2.15)$$

Hence the posterior distribution in the presence of  $s$  outliers under progressive type-II censoring is:

$$p(\alpha, \beta | x) = \sum_{\underline{A}} \left[ \frac{\alpha^{-(r+1)-1} e^{-\frac{1}{\alpha}[D_1-D_2]} \sum_{\underline{A}} \left[ \beta^{-(s+1)-1} \exp^{-\frac{1}{\beta}} \right]}{\Gamma(r+1)\Gamma(s+1)/(D_1-D_2)^{(r+1)}} \right] \quad (2.16)$$

## 2.4. Bayes estimators and their posterior risks under different loss functions

In this section we present Bayes estimates and posterior risk under SELF, PLF, KLF.

**2.4.1. Square error loss function (SELF).** The SELF, defined as  $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$  equally weights the under and overestimation hence regarded as a symmetric loss function. Bayes estimators obtained under SELF are posterior expectations while risk is obtained as posterior variance. Bayes estimators and their posteriors risk of  $\alpha$  and  $\beta$  are:

$$\hat{\alpha} = \sum_{\underline{A}} \left[ \frac{\Gamma(c+s+1)\Gamma(r+a)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a)}} \right] \tag{2.17}$$

$$\hat{\beta} = \sum_{\underline{A}} \left[ \frac{\Gamma(r+a+1)\Gamma(c+s)}{(b+D_1-D_2)^{(r+a+1)}d^{(c+s)}} \right] \tag{2.18}$$

and

$$R(\hat{\alpha}) = \sum_{\underline{A}} \left[ \frac{\Gamma(c+s+1)\Gamma(r+a-1)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a-1)}} \right] - \hat{\alpha} \tag{2.19}$$

$$R(\hat{\beta}) = \sum_{\underline{A}} \left[ \frac{\Gamma(c+s-1)\Gamma(r+a+1)}{d^{(c+s-1)}(b+D_1-D_2)^{(r+a+1)}} \right] - \hat{\beta} \tag{2.20}$$

**2.4.2. Precautionary loss function (PLF).** Everything in life is not symmetric. This fact invites us to use loss functions that are asymmetric by nature. PLF which is an asymmetric loss function and is particularly applied when underestimation is of serious concern is defined as.  $L(\theta, d) = \frac{(\theta-d)^2}{d}$ . Bayes estimators and posteriors risk under PLF are obtained as:

$$\hat{d} = [E_{\theta|x}(\theta^2)^{\frac{1}{2}}], \quad \rho(\hat{d}) = 2[E_{\theta|x}(\theta^2)^{\frac{1}{2}}] - 2E_{\theta|x}(\theta)$$

Bayes estimators and their posteriors risk of  $\alpha$  and  $\beta$  when PLF is used are;

$$\hat{\alpha} = \left[ \sum_{\underline{A}} \frac{\Gamma(c+s+1)\Gamma(r+a-1)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a-1)}} \right]^{\frac{1}{2}} \tag{2.21}$$

$$\hat{\beta} = \left[ \sum_{\underline{A}} \frac{\Gamma(c+s-1)\Gamma(r+a+1)}{d^{(c+s-1)}(b+D_1-D_2)^{(r+a+1)}} \right]^{\frac{1}{2}} \tag{2.22}$$

$$R(\hat{\alpha}) = 2 \left[ \hat{\alpha} - \sum_{\underline{A}} \frac{\Gamma(c+s+1)\Gamma(r+a)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a)}} \right] \tag{2.23}$$

$$R(\hat{\beta}) = 2 \left[ \hat{\beta} - \sum_{\underline{A}} \frac{\Gamma(c+s)\Gamma(r+a+1)}{d^{(c+s)}(b+D_1-D_2)^{(r+a+1)}} \right] \tag{2.24}$$

**2.4.3. K-loss function.** The K-loss function defined as  $k(\theta, \hat{\theta}) = \left( \sqrt{\frac{\hat{\theta}}{\theta}} - \sqrt{\frac{\theta}{\hat{\theta}}} \right)^2$  was introduced to cope with the measure of the inaccuracy of an estimator of a scale parameter specifically when distribution is defined on positive real line. Bayes estimators and their posteriors risk of  $\alpha$  and  $\beta$  by using K-loss function are:

$$\hat{\alpha} = \left[ \frac{\sum_{\underline{A}} \frac{\Gamma(c+s+1)\Gamma(r+a)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a)}}}{\sum_{\underline{A}} \frac{\Gamma(c+s+1)\Gamma(r+a+2)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a+2)}}} \right]^{\frac{1}{2}} \tag{2.25}$$

$$\hat{\beta} = \left[ \frac{\sum_{\underline{A}} \frac{\Gamma(c+s+)\Gamma(r+a+1)}{d^{(c+s+)}(b+D_1-D_2)^{(r+a+1)}}}{\sum_{\underline{A}} \frac{\Gamma(c+s+2)\Gamma(r+a+1)}{d^{(c+s+2)}(b+D_1-D_2)^{(r+a+1)}}} \right]^{\frac{1}{2}} \tag{2.26}$$

$$R(\hat{\alpha}) = 2 \left[ \left( \sum_{\underline{A}} \frac{\Gamma(c+s+1)\Gamma(r+a)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a)}} \right) \left( \sum_{\underline{A}} \frac{\Gamma(c+s+1)\Gamma(r+a+2)}{d^{(c+s+1)}(b+D_1-D_2)^{(r+a+2)}} - 1 \right) \right] \quad (2.27)$$

$$R(\hat{\beta}) = 2 \left[ \left( \sum_{\underline{A}} \frac{\Gamma(c+s)\Gamma(r+a+1)}{d^{(c+s)}(b+D_1-D_2)^{(r+a+1)}} \right) \left( \sum_{\underline{A}} \frac{\Gamma(c+s+2)\Gamma(r+a+1)}{d^{(c+s+2)}(b+D_1-D_2)^{(r+a+1)}} - 1 \right) \right] \quad (2.28)$$

## 2.5. Credible intervals

Lower and upper bounds of  $\alpha$  using Mathematica 12 Software are obtained as:

$$\sum_{\underline{A}} \left[ \frac{(b+D_1-D_2)^{(r+a+1)}(L)^{(-r-a-1)} \left( \frac{L}{(b+D_1-D_2)} \right)^{(r+a+1)} \Gamma\left(r+a+1, \frac{b+D_1-D_2}{L}\right)}{\Gamma(r+a+1)} \right] \quad (2.29)$$

and

$$\sum_{\underline{A}} \left[ \frac{(b+D_1-D_2)^{(r+a+1)} \left[ \left( \frac{1}{b+D_1-D_2} \right)^{r+a+1} \Gamma(r+a+1) - U^{(-r-a-1)} \left( \frac{U}{b+D_1-D_2} \right)^{r+a+1} \Gamma\left(r+a+1, \frac{b+D_1-D_2}{L}\right) \right]}{\Gamma(r+a+1)} \right] \quad (2.30)$$

Similarly, expressions of lower and upper bound of  $\beta$  are obtained.

$$\sum_{\underline{A}} \left[ \frac{(d)^{(c+s+1)}(L)^{(-c-s-1)} \left( \frac{L}{d} \right)^{(c+s+1)} \Gamma\left(c+s+1, \frac{d}{L}\right)}{\Gamma(c+s+1)} \right] \quad (2.31)$$

and

$$\sum_{\underline{A}} \left[ \frac{(d)^{(c+s+1)} \left[ \left( \frac{1}{d} \right)^{c+s+1} \Gamma(c+s+1) - U^{(-c-s-1)} \left( \frac{U}{d} \right)^{c+s+1} \Gamma\left(c+s+1, \frac{d}{U}\right) \right]}{\Gamma(c+s+1)} \right] \quad (2.32)$$

## 3. Results and discussions

In this section, we provide a numerical analysis comprising of simulation analysis and real data analysis to illustrate the properties and behavior of derived estimators.

### 3.1. Simulation study

A simulation study is conducted for different values of  $n$ ,  $m$ ,  $s$  and  $r$ , where  $r$  is fixed. We firstly obtain simulated values for  $\alpha$ ,  $\beta$  by taking  $a = 3$ ,  $b = 1$ ,  $c = 5$  and  $d = 2.4$  from independently distributed prior distribution given in (2.11) using Mathematica program. Values of hyper parameters are obtained equating the value of a parameter to the prior mean. The simulated values for parameters using 1000 replications are obtained as:  $\alpha = 1.0827$  and  $\beta = 1.3024$ . After the simulation of prior parameters, data is generated from the Rayleigh distribution in the presence of outliers and without outliers using simulated parameters. Further data is generated for dropout units that are  $R_i$  and  $R_{Aj}$  for the case of outliers and without outliers respectively from a binomial distribution given as.

$$R_1 \sim \text{Bin}[n - m - (r - s), 0.05]$$

$$R_1 |_{R_1, R_2, \dots, R_{r-1}} \sim \text{Bin}[n - m - (r - s) - \sum_{i=1}^{r-1} R_i, 0.05]$$

$$R_{A_1} \sim \text{Bin}[m - s, 0.05]$$

$$R_{A_j} |_{R_{A_1}, R_{A_2}, \dots, R_{A_{r-1}}} \sim \text{Bin}[m - s - \sum_{j=1}^{r-1} R_j, 0.05]$$

Simulated Bayes estimates with 1000 replications are obtained using the informative inverse Gamma prior utilizing the expressions given in (2.17)-(2.28). The same procedure is repeated for non-informative Jeffreys prior to obtain Bayes estimates and posterior risks for  $\alpha$  and  $\beta$  by setting the prior parameters  $a = b = c = d = 0$ . Credible intervals are obtained for  $\alpha$  and  $\beta$  using IP as well as NIP.

**Table 1.** ML estimates with MSEs in parenthesis and Bayes estimates with posterior risks using IP when n=50 and r=9

		Estimates							
m	s	$\alpha_{ML}$	$\beta_{ML}$	$\alpha_{SELF}$	$\beta_{SELF}$	$\alpha_{PLF}$	$\beta_{PLF}$	$\alpha_{KLF}$	$\beta_{KLF}$
28	1	1.5034 (0.5234)	0.5655 (0.1721)	1.8533 (0.3901)	0.1644 (0.0021)	1.9581 (0.1994)	0.1707 (0.0211)	1.8326 (0.1667)	0.1589 (0.1420)
	2	1.5344 (0.5180)	0.5647 (0.1502)	1.5340 (0.2717)	0.1647 (0.0020)	1.6558 (0.1898)	0.1709 (0.0213)	1.5234 (0.1666)	0.1591 (0.1424)
	3	1.5632 (1.4190)	0.5648 (0.3645)	1.8636 (0.3913)	0.1648 (0.0021)	1.9974 (0.2088)	0.1711 (0.0215)	1.8341 (0.1666)	0.1593 (0.1426)
30	1	1.4190 (0.5803)	0.3645 (0.1635)	1.9409 (0.4082)	0.1645 (0.0021)	1.9984 (0.1987)	0.1707 (0.0212)	1.8398 (0.1666)	0.1589 (0.1420)
	2	1.4952 (0.5935)	0.3647 (0.1621)	1.8950 (0.3998)	0.1647 (0.0021)	1.9670 (0.2012)	0.1709 (0.0213)	1.8316 (0.1666)	0.1591 (0.1424)
	3	1.5244 (0.5556)	0.3648 (0.1225)	1.9245 (0.4055)	0.1648 (0.0021)	1.9558 (0.1926)	0.1711 (0.0215)	1.7934 (0.1666)	0.1592 (0.1426)
33	1	1.5890 (0.5807)	0.3645 (0.2633)	1.889 (0.3855)	0.1644 (0.0021)	1.9758 (0.1875)	0.1707 (0.0211)	1.8435 (0.1666)	0.1589 (0.1420)
	2	1.5975 (0.5782)	0.3021 (0.3132)	1.8578 (0.3728)	0.0021 (0.0020)	1.9699 (0.1801)	0.1709 (0.0213)	1.8170 (0.1666)	0.1591 (0.1424)
	3	1.6170 (0.4953)	0.3648 (0.2031)	1.9073 (0.3932)	0.1648 (0.0021)	1.9697 (0.1809)	0.1711 (0.0215)	1.8004 (0.1666)	0.1592 (0.1426)
36	1	2.3442 (0.6552)	0.4021 (0.2001)	2.5442 (0.6372)	0.00208 (0.0021)	2.6876 (0.2291)	0.1707 (0.0212)	2.3917 (0.1666)	0.1589 (0.1420)
	2	2.4901 (0.6397)	0.4647 (0.4901)	2.4951 (0.6191)	0.1647 (0.0021)	2.6841 (0.2286)	0.1709 (0.0213)	2.4141 (0.1667)	0.1591 (0.1424)
	3	2.5450 (0.6373)	0.1648 (0.1905)	2.5403 (0.6343)	0.1648 (0.0021)	2.6510 (0.2258)	0.1710 (0.0215)	2.4437 (0.1666)	0.1592 (0.1426)

**Table 2.** Credible intervals using IP when n=50 and r=9

m	s	$\alpha_{L(ML)}$	$\alpha_{U(ML)}$	$\beta_L$	$\beta_U$	$\alpha_L$	$\alpha_U$	$\beta_L$	$\beta_U$
28	1	0.4775	2.5292	0.2281	0.9028	0.5744	3.6333	0.0979	0.9955
	2	0.5191	2.5496	0.2703	0.8590	0.5795	3.6435	0.0984	0.9922
	3	0.5191	2.5496	0.2703	0.8590	0.4569	2.8377	0.0993	0.9855
30	1	0.2816	2.5563	0.0440	0.6849	0.6153	3.8920	0.0979	0.9955
	2	0.3319	2.6584	0.0469	0.6824	0.4567	2.8713	0.0984	0.9922
	3	0.4354	2.6133	0.1247	0.6049	0.6273	3.8963	0.0993	0.9855
33	1	0.4508	2.7271	-0.1515	0.8805	0.4665	2.9507	0.0979	0.9955
	2	0.4642	2.7307	-0.1157	0.7199	0.4945	3.1090	0.0984	0.9922
	3	0.6462	2.5877	-0.0332	0.7628	0.6384	3.9649	0.0993	0.9855
36	1	1.0600	3.6283	0.0099	0.7942	0.6899	4.3641	0.0979	0.9951
	2	1.2362	3.7439	0.0921	0.8372	0.7792	4.8989	0.0984	0.9922
	3	1.2958	3.7941	0.0914	0.8381	0.6273	3.8960	0.0993	0.9855

**Table 3.** Bayes estimates and posterior risks using NIP

		Estimates n=50 r=9					
m	s	$\alpha_{SELF}$	$\beta_{SELF}$	$\alpha_{PLF}$	$\beta_{PLF}$	$\alpha_{KLF}$	$\beta_{KLF}$
28	1	2.3950 (0.8948)	0.1025 (0.0012)	2.5686 (0.3615)	0.1085 (0.0119)	2.2663 (0.2222)	0.0975 (0.2113)
	2	1.9173 (0.6022)	0.1026 (0.0012)	2.1016 (0.3123)	0.1086 (0.0119)	1.8947 (0.2222)	0.0976 (0.2119)
	3	2.3436 (0.8578)	0.1027 (0.0012)	2.5487 (0.3227)	0.1087 (0.0120)	2.3372 (0.2222)	0.0976 (0.2121)
30	1	2.4099 (0.8892)	0.1025 (0.0012)	2.6559 (0.3495)	0.1085 (0.0119)	2.3018 (0.2222)	0.0975 (0.2113)
	2	2.4309 (0.9166)	0.1026 (0.0012)	2.5915 (0.3429)	0.1086 (0.0119)	2.2656 (0.2222)	0.0976 (0.2119)
	3	2.4404 (0.9049)	0.1027 (0.0012)	2.5692 (0.3380)	1.7552 (0.0120)	0.0976 (0.1666)	0.0976 (0.2121)
33	1	2.4155 (0.8779)	0.1025 (0.0012)	2.5514 (0.3126)	0.1085 (0.0119)	0.0975 (0.2222)	0.0975 (0.2113)
	2	2.3676 (0.8635)	0.1026 (0.0012)	2.6002 (0.3086)	0.1086 (0.0119)	2.3361 (0.2222)	0.0976 (0.2119)
	3	2.4068 (0.8690)	0.1027 (0.0012)	2.5352 (0.3466)	0.1087 (0.0120)	2.3506 (0.2222)	0.0976 (0.2121)
36	1	3.2249 (1.4126)	0.1025 (0.0012)	3.4308 (0.3954)	0.1085 (0.0119)	3.0703 (0.2222)	0.0975 (0.2113)
	2	3.2388 (1.4345)	0.1026 (0.0012)	3.4379 (0.3957)	0.1086 (0.0119)	3.1009 (0.2222)	0.0976 (0.2119)
	3	3.2500 (1.4455)	0.1027 (0.0012)	3.4691 (0.4008)	0.1087 (0.0120)	3.0533 (0.2222)	0.0976 (0.2121)

From the results reported in Tables 1–4, it is observed that parameter  $\alpha$  is estimated a bit higher when non-informative prior is used while the parameter  $\beta$  seems quite consistent under both priors. However, the performance of informative prior is found more impressive as compare to non-informative prior because posterior risks of both parameters are smaller under informative prior. It is also noted that KLF worked well for  $\alpha$







for first parameter while the 2nd parameter is estimated to be 0.15. Posterior risks obtained for these estimates are smaller for IP than NIP which establish the superiority of IP employed. When obtained Bayes estimates are compared with ML estimates, we get estimates with MSEs that are higher than posterior risks. Credible intervals for informative prior found pretty narrow while these are broader for non-informative prior. But ML confidence intervals are narrower than Bayes credible intervals. When we compare the performance of loss functions, KLF performs better than PLF and PLF proved to be better than SELF on the basis of respective posterior risks for first parameter. For the case of the second parameter, SELF outperformed the two other loss functions.

**Data set 2.** The data set for analysis is taken from Abushal [1] that represents the survival times (in years) of a group of patients having the treatment of Chemotherapy. Survival of patients may largely depend upon different types of treatments and on other factors as well. Hence, such data is found suitable to be analyzed by mixture model containing two types of data, with and without outliers. The data is divided into two parts to extract information for a mixture model and is given in Table 10.

**Table 10.** Data set of survival times of patients

X observations						Y observations					
0.260*	0.282	0.296	0.334	0.395	0.458	0.507	0.529	0.534	0.540	0.570	0.641
0.466	0.501	0.841	0.863	1.099	1.219	0.644	0.696	1.553	1.581	1.589*	2.178*
1.271	1.326	1.447	1.458	2.830	3.578	2.343*	2.416*	2.444*	2.825*		
3.658	3.743	3.978	4.003	4.033	0.047*						
0.115*	0.121*	0.132*	0.164*	0.197							
0.203											

Here, in this data set  $n_1 = 30, n_2 = 16, n = n_1 + n_2, m = 25, r = 9$  and the total number of removals are “ $n - m - r$ ” and we take the removals from the scheme as;  $R_1 = \{1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0\}$  and  $R_2 = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1\}$  respectively from no outliers and outliers observations. The results can be viewed from Tables 11–14.

**Table 11.** ML estimates with MSEs in parenthesis and Bayes estimates with posterior risks for IP

Estimates n=46, m=25, r=9								
s	$\alpha_{ML}$	$\beta_{ML}$	$\alpha_{SELF}$	$\beta_{SELF}$	$\alpha_{PLF}$	$\beta_{PLF}$	$\alpha_{KLF}$	$\beta_{KLF}$
1	3.5974 (0.1420)	0.1819 (0.3800)	3.1543 (0.0265)	0.1645 (0.9045)	3.2947 (0.0021)	0.1707 (0.2805)	3.0306 (0.0125)	0.1589 (0.1667)
2	3.4957 (0.2985)	0.1801 (0.0250)	3.1544 (0.9045)	0.1647 (0.0021)	3.2947 (0.2805)	0.1710 (0.0125)	3.0306 (0.1667)	0.1592 (0.1424)
3	3.3945 (0.2955)	0.1801 (0.0250)	3.1544 (0.9045)	0.1648 (0.0021)	3.2947 (0.2805)	0.1711 (0.0125)	3.0306 (0.1667)	0.1592 (0.1427)

**Table 12.** Confidence interval and credible intervals using IP

Credible Intervals n=46, m=25, r=9						
s	$\alpha_{L(ML)}$	$\alpha_{U(ML)}$	$\alpha_L$	$\alpha_U$	$\beta_L$	$\beta_U$
1	2.8526	6.4500	1.5401	9.7412	0.0979	0.9955
2	2.9106	6.4063	1.5440	9.7064	0.0984	0.9923
3	2.8153	6.2098	1.5518	9.6380	0.0993	0.9856

**Table 13.** Bayes estimators and posterior risks for NIP

Estimates n=46,m=25, r=9						
s	$\alpha_{SELF}$	$\beta_{SELF}$	$\alpha_{PLF}$	$\beta_{PLF}$	$\alpha_{KLF}$	$\beta_{KLF}$
1	4.0947 (2.0959)	0.1025 (0.0013)	4.3431 (0.4967)	0.1085 (0.0119)	3.8846 (0.2222)	0.0975 (0.2113)
2	4.0947 (2.0958)	0.1026 (0.0012)	4.3431 (0.4967)	0.1086 (0.0119)	3.8846 (0.2222)	0.0976 (0.2119)
3	4.0947 (2.0958)	0.1027 (0.0013)	4.3431 (0.4967)	0.1087 (0.0120)	3.8846 (0.2222)	0.0976 (0.2121)

**Table 14.** Credible intervals using NIP

Credible Intervals n=46, m=25, r=9				
s	$\alpha_L$	$\alpha_U$	$\beta_L$	$\beta_U$
1	1.80675	15.2907	0.05622	2.30547
2	1.81187	15.2248	0.05662	2.59624
3	1.82206	15.0955	0.05728	2.26215

From results presented in Tables 11–14, we come to conclude that the average survival of patients measured through informative prior is nearly 3.0 to 3.15 years for first parameter which is slightly higher when estimated under non-informative prior. Also informative prior was found better than non-informative prior by producing smaller risks, which is desirable in Bayesian estimation. Comparing ML and Bayes estimates it is concluded that Bayesian estimation is preferred over classical as PRs are much smaller than the MSEs obtained under use of the classical technique. Conclusion about loss functions remains the same as discussed earlier for simulation analysis and data set 1.

**Data set 3.** The data given in Kohansal [19] has been used in estimation of the stress-strength reliability for inverted exponentiated Rayleigh distribution which is about the strength measured in GPA for single carbon fiber. The data set modified for the mixture model (modeled by two different density functions) is given in Table 15.

**Table 15.** Data on strength of carbon fiber

X observations						Y observations					
1.865*	2.382	2.809	2.675	3.377	2.490	1.966*	2.434	2.818	2.257	3.220	4.024
3.537	2.616	2.514	1.944*	2.426	1.901*	2.301*	3.294	2.454	1.997*	2.435	2.821
2.738	3.871	2.648	2.027*	2.937	3.012	2.350	3.223	4.027	2.535	3.628	2.659
1.958*	2.566	2.132*	2.740	3.408	2.880	2.006*	2.478	2.848	2.532	3.235	2.977
2.274	3.264	2.397	2.140*	2.570	2.203	2.726	2.098*	3.030	2.240	2.629	2.228*
3.125	3.886	2.361	2.511	3.554	2.618	2.575	3.435	3.272	2.445	2.359	3.332
2.179*	2.586	2.474	3.139	3.971	2.614	2.253*	2.633	2.525	2.917	3.493	3.562
2.684	2.055*	2.224	2.773	2.518	3.145	2.624	2.554	3.852	2.270	2.642	2.856
2.021*	2.937	2.954	2.301	2.382	2.800	2.928	3.501	2.063*	2.996	2.770	
2.522	3.346	2.272	3.243	2.396	2.697						



#### 4. Conclusion

In this study, an attempt has been made to estimate parameters of Rayleigh distribution when data are contaminated with outliers using the Bayesian approach. ML estimation is also conducted to establish a comparison between the two techniques. Inverse Gamma prior and Jeffreys prior are used for Bayesian estimation. While three loss functions SELF, PLF, and KLF are used. Prior information which is the basis of Bayesian estimation whenever available enhances the utility of analysis. In this study informative prior throughout got superiority by providing better estimates along with smaller posterior risks and pretty good credible intervals. When performance of loss functions is compared, it is found that KLF is better for  $\alpha$  (shape parameter) and SELF is better for  $\beta$  (outlying parameter). When comparing the results of the simulation study for informative prior, it is found that KLF is better for  $\alpha$  (shape parameter) and SELF is better for  $\beta$  (outlying parameter). When sample size increases, posterior risks decrease continuously for  $\alpha$  under PLF. For non-informative prior, KLF provides better results for  $\alpha$  and SELF provides better results for  $\beta$  for each sample size. For data sets 1 and 2, KLF performs better for  $\alpha$  and gives less posterior risks as compared to SELF and PLF for informative and non-informative prior as well. Further, it is also observed that  $\beta$  gives more efficient results under SELF, PLF and KLF when informative prior is used. It is also observed that all the Bayes estimates lie in the range of credible intervals. It is also observed that Bayesian approach gives better results as PRs are found lesser than MSEs which we obtain by using the classical approach in all the cases.

**Acknowledgment.** We are thankful to the Editor concerned for her timely response and guidance. The authors are thankful to anonymous referees for their useful comments and suggestions that improved the quality of the paper.

#### References

- [1] T.A. Abushal, *Estimation of the unknown parameters for the compound Rayleigh distribution based on progressive first-failure-censored sampling*, OJS **1** (03), 161–171, 2011.
- [2] W.M. Afify, *Classical estimation of mixed Rayleigh distribution in type-I progressive censored*, Stat. Theory Appl. **10** (4), 619–632, 2011.
- [3] S. Ali, *Mixture of the inverse Rayleigh distribution: Properties and estimation in a Bayesian framework*, Appl. Math. Model. **39**, 515–530, 2015.
- [4] S. Ali, *On the Bayesian estimation of the weighted Lindley distribution*, J. Stat. Comput. Simul. **85** (5), 855–880, 2015.
- [5] S. Ali, M. Aslam, A. Nasir and S.M.A. Kazmi, *Scale parameter estimation of the Laplace model using different asymmetric loss functions*, International Journal of Statistics and Probability **1** (1), 105–127, 2012.
- [6] F.J. Anscombe, *Residuals*, in: Time series and Statistics, 244–250, Palgrave Macmillan, London, 1990.
- [7] M. Aslam, F. Noor and S. Ali, *Shifted Exponential distribution: Bayesian estimation, prediction and expected test time under progressive censoring*, J. Test. Eval. **48** (2), 1576–1593, 2020.
- [8] R. Azimi and F. Yaghmaei, *Bayesian estimation based on Rayleigh progressive type II censored data with Binomial removals*, Journal of Quality and Reliability Engineering, 1–6, 2013.
- [9] N. Balakrishnan and R. Aggarwala, *Progressive Censoring: Theory, Methods, and Applications*, Springer Science and Business Media, 2000.
- [10] V. Barnett, *The study of outliers: Purpose and Model*, J. R. Stat. Soc. Ser. C. Appl. Stat. **27** (3), 242–250, 1978.

- [11] A. Childs and N. Balakrishnan, *Conditional inference procedures for The Laplace distribution when the observed samples are progressively censored*, *Metrika* **52** (3), 253–265, 2000.
- [12] U.J. Dixit and M.J. Nooghabi, *Bayesian inference for the Pareto lifetime model in the presence of outliers under progressive censoring with binomial removals*, *Hacet. J. Math. Stat.* **46** (5), 887–906, 2017.
- [13] D.D. Dyer and C.W. Whisenand, *Best linear unbiased estimator of the parameter of the Rayleigh distribution-Part I: Small sample theory for censored order statistics*, *IEEE Transactions on Reliability* **46** (1), 27–34, 1973.
- [14] F.E. Grubbs, *Procedures for detecting outlying observations in samples*, *Technometrics* **11** (1), 1–21, 1969.
- [15] D.M. Hawkins, *Identification of outliers* **11**, Chapman and Hall, London, 1980.
- [16] B.K. Kale, *Outliers-A review*, *J. Indian Statist. Assoc.* **17**, 51–67, 1979.
- [17] V.B. Kale and B.K. Kale, *Outliers in exponential sample—A Bayesian approach*, *Gujarat Statistical Review*, 1992.
- [18] A. Kohansal, *Large Estimation of the stress-strength reliability of progressively censored inverted exponentiated Rayleigh distributions*, *J. Appl. Math. Stat. Inform.* **13** (1), 49–76, 2017.
- [19] N.R. Mann, *Best linear invariant estimation for Weibull parameters under progressive censoring*, *Technometrics* **13** (3), 521–533, 1971.
- [20] M. Saleem and M. Aslam, *On prior selection for the mixture of Rayleigh distribution using predictive Intervals*, *Pakistan J. Statist.* **24** (1), 21–35, 2008.
- [21] M. Saleem and M. Aslam, *On Bayesian analysis of the Rayleigh survival time assuming the random censor time*, *Pakistan J. Statist.* **25** (2), 71–82, 2009.
- [22] A.M. Sarhan and A. Abuammoh, *Statistical inference using progressively type-II censored data with random scheme*, *Int. Math. Forum (Online)* **35** (3), 1713–1725, 2008.
- [23] R. Shanker, F. Hagos and S. Sujatha, *On modeling of Lifetimes data using exponential and Lindley distributions*, *Biom Biostat Int J* **2** (5), 1-9, 2015.
- [24] A.A. Soliman, *Estimation of parameters of life from progressively censored data using Burr-XII model*, *IEEE Transactions on Reliability* **54** (1), 34–42, 2005.
- [25] S.J. Wu and C.T. Chang, *Parameter estimations based on exponential progressive type II censored data with binomial removals*, *Int. J. Inf. Manag. Sci.* **13** (3), 37–46, 2002.
- [26] H.K. Yuen and S.K. Tse, *Parameters estimation for Weibull distributed lifetimes under progressive censoring with random removals*, *J. Stat. Comput. Simul.* **55** (1-2), 57–71, 1996.