

# A Pareto II Model With Inliers at Zero and One Based on TYPE-II Censored Samples

Bavagosai Pratima<sup>1</sup> and K. Muralidharan<sup>2</sup>

Department of Statistics, Faculty of Science  
The Maharaja Sayajirao University of Baroda, Vadodara 390002, India  
Email: <sup>1</sup>Pratimabava@outlook.com, <sup>2</sup>lmv\_murali@yahoo.com

## Abstract

Inliers (instantaneous or early failures) are natural occurrences of a life test, where some of the items fail immediately or within a short time of the life test due to mechanical failure, inferior quality or faulty construction of items and components. The inconsistency of such life data is modeled using a nonstandard mixture of distributions; where degeneracy can happen at discrete points at zero and one. In this paper, parameters estimation based on Type-II censored sample from a Pareto type II distribution with a discrete mass at zero and one is study. The Maximum Likelihood Estimators (MLE) are developed for estimating the unknown parameters. The Fisher information matrix, as well as the asymptotic variance-covariance matrix of the MLEs, are derived. Uniformly Minimum Variance Unbiased Estimate (UMVUE) of model parameters as well as UMVUE of the density function, reliability function, and some other parametric function are obtained along with the standard error of estimators. The model is implemented on various real data sets and compared with Weibull inliers model.

**Keywords:** early failures; failure time distribution; infant mortality rate; inliers; instantaneous failures; type-II censored sample.

## I. Introduction

There are a plethora of examples of phenomena concerning nature, life and human activities where the real data do not conform to the standard distributions. In such cases, we either use mixtures of standard distributions of similar types or non-standard mixtures of degenerate distribution and a standard distribution, which may be again a discrete or continuous one. Since *inliers* are inconsistent observations, which are generally the results of instantaneous and early failures, modeling with inliers involve non-standard mixtures of distributions. In the former case, the random variable will have a discrete probability mass at the origin (that is life will be zero) and some positive lifetimes, and in the latter case, the failure times may be smaller in relation to other lifetimes. These occurrences may be due to mechanical failure, inferior quality or faulty construction or defective parts of items and components. Such failures usually discard the assumption of a single mode distribution and hence the usual method of modeling and inference procedures may not be accurate in practice. [2] was the first to discuss the inference problem of instantaneous failures in life testing. The author has provided the efficient estimation of parametric functions under various probability models. [13] have introduced the term *inliers* in connection with the estimation of  $(p, \theta)$  of early failure model with modified failure time distribution (FTD) being an exponential distribution with mean  $\theta$  assuming  $p$  known. Later on, many authors have

studied these kinds of models (see [18], [13] and [15]).

There are many practical contexts, where inliers can be natural occurrences of the specific situations involved and degeneracy can happen at two discrete points and a positive distribution for the remaining lifetimes. Some of the situations are as follows:

1. The size of tumor lesions is of interest to treat *Hematologic malignancy* patients. The measurement effect is zero who have lesions absent (or due to disappearance of tumor during treatment), though who have lesions present at baseline that are measurable but do not meet the definitions of measurable disease may be considered as measurement 1, otherwise lesions can be accurately measured as longest diameter to be recorded in at least one dimension by chest x-ray, with CT scan or with calipers by clinical examination. Similarly, in studies like Bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusions, lymphangitis cutis/pulmonitis, inflammatory breast disease, and abdominal masses, either the effect is absent or present but not followed by CT or MRI, are considered as non-measurable otherwise accurately measurable on a continuous scale.
2. In the mass production of technological components of hardware, intended to function over a period of time, some components may fail on installation and therefore have zero life lengths, some component that does not fail on installation but fails with negligible life (may be coded as one for simplicity), and others that will have a life length a positive random variable whose distribution may take different forms.
3. In a clinical trial laboratory, a particular drug is designed and given to certain species of hens so that the new chicks have a weight greater than usual. The possible weight of chicks may be modeled as a continuous distribution, with discrete mass at 'zero' and 'one', where zero measures those chicks having no gain of weight, and one measures those chicks with negligible gain of weight than usual, and the remaining chicks having weight gain in some continuous measurement.
4. The rainfall measurement at a place recorded during a season is modeled as a continuous distribution, with a discrete mass at 'zero' where zero measures those days having no rainfall, and at 'one', one measures those days with no rain but humid and cloudy conditions, and a continuous variable having some positive amount of rain.
5. In the studies of genetic birth defects, children can be characterized by three variables: first, a discrete variable to indicate whether a child is affected and born dead; second, a child is affected and has a neonatal death; and third, a continuous variable measuring the survival time of affected children born alive. We may consider this as a nonstandard mixture of the mass point at "zero" (for children born dead), at "one" (for children born and neonatal death), and a nontrivial continuous distribution for other surviving children.

Similarly, one can contemplate many such examples in practical situations involving degeneracy at two or more points and positive configurations of observations. Authors [16] and [17], have modeled the above situation using exponential distribution and Weibull distribution respectively. In this article, we model the inliers situation using the type-II censored lifetime data from a Pareto II distribution. As per the scheme, if  $n$  units are placed on the test and the experiment is terminated after a prefixed number of failures say,  $c < n$ , then the observed failure times are  $X_{(1)}, X_{(2)}, \dots, X_{(c)}$  where  $X_{(c)} < X_{(n)}$ . The remaining  $n - c$  items are regarded as censored data. The family of the Pareto distribution is well known in the literature for its capability in modeling the heavy-tailed distributions. The Pareto Type II distribution (also called Lomax distribution with location parameter zero) has the probability distribution function (pdf)

$$f(x, \underline{\alpha}) = \frac{\theta \beta^\theta}{(x + \beta)^{(1+\theta)}}, \quad x > 0, \beta > 0, \theta > 0 \tag{1}$$

where  $\underline{\alpha} = (\beta, \theta)$ ,  $\beta > 0$  is a scale parameter and  $\theta > 0$  is a shape parameter. The Pareto distribution has been used in connection with studies of income, property values, insurance risk, migration, size of cities and firms, word frequencies, business mortality, service time in queuing

systems, etc. The paper by [1] contains a detailed list of important areas where heavy-tailed distributions are found applicable. There are also recent applications of the Pareto distribution in data sets on earthquakes, forest fire areas, fault lengths on Earth and Venus, and on oil and gas fields sizes, see [22] for details.

The presentation of the paper is as follows: The model description is given in Section II. In Section III, we derive the MLE of the unknown parameters along with the interval estimation of parameters. The UMVU estimation of model parameters and various parametric functions are given in Section IV. For illustration, we consider four real datasets for implementing the proposed model in Section V.

## II. Model description

If 0 and 1 are natural occurrence of a life test as described above with other positive observations, then the distribution function of such a inliers model can be written as:

$$H(x; p_1, p_2, \underline{\alpha}) = \begin{cases} 0, & x < 0 \\ p_1, & 0 \leq x < 1 \\ p_1 + p_2, & x = 1 \\ p_1 + p_2 + (1 - p_1 - p_2) \frac{F(x; \underline{\alpha}) - F(1; \underline{\alpha})}{1 - F(1; \underline{\alpha})}, & x \geq 1 \end{cases} \quad (2)$$

The fact is that the probability measure generated by  $H(\cdot)$  is composed of three measures, say  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$ , where  $\mu_3$  is absolutely continuous with respect to the Lebesgue measure on  $R$  and  $\mu_1$  and  $\mu_2$  are singular with respect to the Lebesgue measure on  $R$ . The corresponding likelihood function of the model is

$$h(x; p_1, p_2, \underline{\alpha}) = \begin{cases} p_1, & x = 0 \\ p_2, & x = 1 \\ (1 - p_1 - p_2) \frac{f(x; \underline{\alpha})}{1 - F(1; \underline{\alpha})}, & x > 1 \end{cases} \quad (3)$$

where  $p_1$  and  $p_2$  are the proportion of 0 and 1 observations respectively. For  $\beta = 1$ , the Pareto Type II inliers distribution has the likelihood function

$$h(x; p_1, p_2, \theta) = \begin{cases} p_1, & x = 0 \\ p_2, & x = 1 \\ (1 - p_1 - p_2) \frac{\theta}{(1+x)} \left(\frac{2}{(1+x)}\right)^\theta, & x > 1 \end{cases} \quad (4)$$

The parameter estimates are obtained in the next section.

## III. The Maximum Likelihood Estimation of $\underline{\theta} = (p_1, p_2, \theta)$

Suppose  $n$  items placed on life test, where  $r_1$  items have life zero where as  $r_2$  items have life 1 and remaining  $n - r_1 - r_2$  items have life greater than 1, is denoted by  $X_1, X_2, \dots, X_{n-r_1-r_2}$ . By applying the technique of 'Type-II censored sample', the experiment terminates after prefixed number of failures  $n - r_1 - r_2 - c$  out of  $n - r_1 - r_2$  items, where,  $n - r_1 - r_2 - c < n - r_1 - r_2$ . Clearly, if  $n - r_1 - r_2 - c = n - r_1 - r_2$ , then the experiment is not terminated and all  $n - r_1 - r_2$  lifetimes are observed. Let  $n - r_1 - r_2 - c^* = \min(n - r_1 - r_2 - c, n - r_1 - r_2)$  and  $X_{(1)}, X_{(2)}, \dots, X_{(n-r_1-r_2-c^*)}$  denote ordered observed failure time of these  $n - r_1 - r_2 - c^*$  items from  $h \in \mathcal{H}$  as given in (4). Then the likelihood equation can be written as

$$L(\underline{x}; \underline{\theta}) = \prod_{i=1}^n h(x_i; \underline{\theta})$$

If we define

$$I_1(x) = \begin{cases} 1, & x = 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$I_2(x) = \begin{cases} 1, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Then the likelihood equation can be written as

$$= p_1^{r_1} p_2^{r_2} (1 - p_1 - p_2)^{(n-r_1-r_2)} \frac{(n-r_1-r_2)!}{c^*!} \theta^{n-r_1-r_2-c^*} \prod_{i=1}^{n-r_1-r_2-c^*} \frac{1}{1+x_{(i)}} e^{-\theta \left\{ \sum_{i=1}^{n-r_1-r_2-c^*} \left[ \log\left(\frac{1+x_{(i)}}{2}\right) \right] + c^* \left[ \log\left(\frac{1+x_{(n-r_1-r_2-c^*)}}{2}\right) \right] \right\}} \quad (5)$$

where  $r_1 = \sum_{i=1}^n I_1(x_{(i)})$  and  $r_2 = \sum_{i=1}^n I_2(x_{(i)})$ , denotes the number of zero and one observations respectively. We now investigate the following four possible cases of likelihood estimates:

*Case (i).*  $r_2 = 0$ , that is  $r_1 = n$ . The likelihood function simply reduces to  $L(\underline{x}; \underline{\theta}) = p_1^n$ . Obviously, this is maximum when  $p_1 = 1$ . This corresponds to the maximum likelihood estimator  $\hat{p}_1 = \frac{r_1}{n}$ . Since  $L(\underline{x}; \underline{\theta}) = p_1^n$  is free from the other parameters, the maximum likelihood estimator of other parameters do not exist.

*Case (ii).*  $r_1 = 0$ , that is  $r_2 = n$ . The likelihood function simply reduces to  $L(\underline{x}; \underline{\theta}) = p_2^n$ . Obviously, this is maximum when  $p_2 = 1$ . This corresponds to the maximum likelihood estimator  $\hat{p}_2 = \frac{r_2}{n}$ . Since  $L(\underline{x}; \underline{\theta}) = p_2^n$  is free from the other parameters, the maximum likelihood estimator of other parameters do not exist.

*Case (iii).*  $r_1 < n, r_2 < n$  but  $r_1 + r_2 = n$ . The likelihood function simply reduces to  $L(\underline{x}; \underline{\theta}) = p_1^{r_1} p_2^{r_2}$ . Here  $p_1 + p_2 < n$ . Then the likelihood function  $L(\underline{x}; \underline{\theta}) < \left(\frac{x_1}{n}\right)^{r_1} \left(\frac{x_2}{n}\right)^{r_2}$ . So  $\hat{p}_1 = \frac{r_1}{n}$  and  $\hat{p}_2 = \frac{r_2}{n}$ . The maximum likelihood of other parameters do not exist.

*Case (iv).*  $r_1 + r_2 < n$ . The log-likelihood function is given by

$$\begin{aligned} \log L(\underline{x}; \underline{\theta}) &= r_1 \log p_1 + r_2 \log p_2 + (n - r_1 - r_2) \log(1 - p_1 - p_2) + \log(n - r_1 - r_2)! \\ &\quad - \log c^*! + (n - r_1 - r_2 - c^*) \log \theta - \sum_{i=1}^{n-r_1-r_2-c^*} \log(1 + x_{(i)}) \\ &\quad - \theta \left\{ \sum_{i=1}^{n-r_1-r_2-c^*} \left[ \log\left(\frac{1+x_{(i)}}{2}\right) \right] + c^* \left[ \log\left(\frac{1+x_{(n-r_1-r_2-c^*)}}{2}\right) \right] \right\} \end{aligned} \quad (6)$$

The maximum likelihood estimator of parameter  $\underline{\theta} = (p_1, p_2, \theta)$  is obtained by solving the following likelihood equations:

$$\frac{\partial \log L(\underline{x}; \underline{\theta})}{\partial p_1} = \frac{r_1}{p_1} - \frac{n-r_1-r_2}{1-p_1-p_2} = 0 \quad (7)$$

$$\frac{\partial \log L(\underline{x}; \underline{\theta})}{\partial p_2} = \frac{r_2}{p_2} - \frac{n-r_1-r_2}{1-p_1-p_2} = 0 \quad (8)$$

and

$$\frac{\partial \log L(\underline{x}; \underline{\theta})}{\partial \theta} = \frac{n-r_1-r_2-c^*}{\theta} - \left\{ \sum_{i=1}^{n-r_1-r_2-c^*} \left[ \log\left(\frac{1+x_{(i)}}{2}\right) \right] + c^* \left[ \log\left(\frac{1+x_{(n-r_1-r_2-c^*)}}{2}\right) \right] \right\} = 0 \quad (9)$$

Solving (7) and (8) simultaneously, we get

$$\hat{p}_1 = \frac{r_1}{n} \quad (10)$$

$$\hat{p}_2 = \frac{r_2}{n} \quad (11)$$

From (9), the estimate of  $\theta$  is

$$\hat{\theta} = \frac{n-r_1-r_2-c^*}{\sum_{i=1}^{n-r_1-r_2-c^*} [\log(1+x_{(i)}) - \log 2] + c^* [\log(1+x_{(n-r_1-r_2-c^*)}) - \log 2]} \quad (12)$$

The approximate  $(1 - \alpha)\%$  confidence interval for  $p_1$ ,  $p_2$  and  $\theta$  are respectively given by

$$\hat{p}_1 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}}, \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n}} \quad \text{and} \quad \hat{\theta} \pm z_{\alpha/2} \sqrt{\frac{\hat{\theta}^2}{(n-c^*)\hat{p}^*}} \quad \text{where, } \hat{p}^* = 1 - \hat{p}_1 - \hat{p}_2.$$

#### IV. Unbiased estimation

Many authors have studied the problem of minimum variance unbiased estimation for different classes of distributions. [23], [12] and [5] have studied the estimation problem for power series distribution, [20] has studied the same for generalized power series distribution, [7] and [5] have studied for modified power series distribution. [19] has studied the UMVUE of parameters for the multivariate modified power series distribution. All these studies include discrete distributions only. [9] has studied the problem of MVU estimation in one parameter exponential family of distributions which includes power series distribution, modified power series distribution and univariate continuous distributions. Further, a characterization property of power series distribution using one and two moments was given by [14]. [8] extended this for the one-parameter exponential family of distribution which includes all earlier cases. [10] have further studied MVU estimation in the multi-parameter exponential family of distributions. Here, we propose the distributional properties of complete sufficient statistic and study UMVU estimation for various parametric functions of the model.

The model in (4) can be expressed as

$$h(x; \underline{\theta}) = \frac{\left(\frac{1}{1+x}\right)^{(1-I_1(x)-I_2(x))} \left(\frac{p_1}{\theta(1-p_1-p_2)}\right)^{I_1(x)} \left(\frac{p_2}{\theta(1-p_1-p_2)}\right)^{I_2(x)} (e^{-\theta})^{\left\{\log\left(\frac{1+x}{2}\right)\right\}(1-I_1(x)-I_2(x))}}{\left(\frac{1}{\theta(1-p_1-p_2)}\right)^{C_1(x)}} = \frac{(a(x))^{(1-C_1(x)-C_2(x))} \prod_{i=1}^3 (h_i(\underline{\theta}))^{C_i(x)}}{g(\underline{\theta})} \quad (13)$$

where,  $a(X) = \frac{1}{(1+X)}$ ;  $h_1(\underline{\theta}) = \frac{p_1}{\theta(1-p_1-p_2)}$ ;  $h_2(\underline{\theta}) = \frac{p_2}{\theta(1-p_1-p_2)}$ ;  $h_3(\underline{\theta}) = e^{-\theta}$ ;  $g(\underline{\theta}) = \frac{1}{\theta(1-p_1-p_2)}$ ;  $C_1(X) = I_1(X)$ ;  $C_2(X) = I_2(X)$  and  $C_3(X) = \left[\log\left(\frac{1+X}{2}\right)\right](1 - I_1(X) - I_2(X))$ . Also  $a(X) > 0$ ,  $C_i(X)$ ,  $i = 1, 2$  and  $3$  are nontrivial real-valued statistics,  $g(\underline{\theta})$  and  $h_i(\underline{\theta})$  are at least twice differentiable functions of  $\theta_i$ ,  $i=1,2$  and  $3$ . Here  $g(\underline{\theta}) = \int_{x>1} (a(x))^{(1-C_1(x)-C_2(x))} \prod_{i=1}^3 (h_i(\underline{\theta}))^{C_i(x)} dx$ . The density in (13) so obtained is defined with respect to a measure  $\mu(x)$  which is the sum of Lebesgue measure over  $(1, \infty)$  a well-known form of a three parameter exponential family with natural parameters  $(\eta_1, \eta_2, \eta_3) = \left(\log\left(\frac{p_1}{\theta(1-p_1-p_2)}\right), \log\left(\frac{p_2}{\theta(1-p_1-p_2)}\right), \log(e^{-\theta})\right)$  generated by underlying indexing parameters  $\underline{\theta} = (p_1, p_2, \theta)$ . Hence  $C(X) = (C_1(X), C_2(X), C_3(X)) = \left(I_1(X), I_2(X), \left[\log\left(\frac{1+X}{2}\right)\right](1 - I_1(X) - I_2(X))\right)$  is jointly complete sufficient for  $\underline{\theta} = (p_1, p_2, \theta)$ . The distributional properties of  $C(X) = (C_1(X), C_2(X), C_3(X))$  are presented in appendix A. We now propose some uniformly minimum variance unbiased estimators for parameters and some parametric function of the model (13) in various subsections below.

##### I. Uniformly Minimum Variance Unbiased Estimation of parameters

For the Type-II censored sample discussed in the previous section, consider the following transformation

$$Y_1 = (n - r_1 - r_2) \left( \left[ \log\left(\frac{1+x_{(i)}}{2}\right) \right] \right),$$

and

$$Y_i = (n - r_1 - r_2 - i + 1) \left\{ \left[ \log \left( \frac{1+x(i)}{2} \right) \right] - \left[ \log \left( \frac{1+x(i-1)}{2} \right) \right] \right\};$$

$$i = 2, 3, \dots, n - r_1 - r_2 - c^* \quad (14)$$

It can be seen that

$$\sum_{i=1}^{n-r_1-r_2-c^*} Y_i = \sum_{i=1}^{n-r_1-r_2-c^*} \left[ \log \left( \frac{1+x(i)}{2} \right) \right] + c^* \left[ \log \left( \frac{1+x(n-r_1-r_2-c^*)}{2} \right) \right]$$

and

$$|J| = \frac{c^*! \prod_{i=1}^{n-r_1-r_2-c^*} (1+x(i))}{(n-r_1-r_2)!} \quad (15)$$

Using (14) and (15),

$$h(\underline{y}; \underline{\theta}) = p_1^{r_1} p_2^{r_2} (1 - p_1 - p_2)^{(n-r_1-r_2)} \theta^{(n-r_1-r_2-c^*)} e^{-\theta \sum_{i=1}^{n-r_1-r_2-c^*} y_i} \quad (16)$$

$$= \frac{\left( \frac{p_1}{\theta(1-p_1-p_2)} \right)^{Z_1} \left( \frac{p_2}{\theta(1-p_1-p_2)} \right)^{Z_2} (e^{-\theta})^{Z_3} (1-p_1-p_2)^{c^*}}{\left( \frac{1}{\theta(1-p_1-p_2)} \right)^{n-c^*}}$$

where

$$Z_1 = \sum_{i=1}^n C_1(X_i) = \sum_{i=1}^{n-c^*} I_1(Y_i) = r_1$$

$$Z_2 = \sum_{i=1}^n C_2(X_i) = \sum_{i=1}^{n-c^*} I_2(Y_i) = r_2$$

and

$$Z_3 = \sum_{i=1}^n C_3(X_i) = \sum_{i=1}^{n-r_1-r_2-c^*} Y_i$$

Hence by Neyman Factorization theorem  $Z = (Z_1, Z_2, Z_3)$  is jointly sufficient for  $\underline{\theta} = (p_1, p_2, \theta)$ . Also,

$$h(\underline{y}; \underline{\theta}) = \frac{n!}{r_1! r_2! (n-r_1-r_2)!} p_1^{r_1} p_2^{r_2} (1 - p_1 - p_2)^{(n-r_1-r_2)} \frac{\theta^{(n-r_1-r_2-c^*)}}{\left( \frac{n!}{r_1! r_2! (n-r_1-r_2)!} \right)} e^{-\theta \sum_{i=1}^{n-r_1-r_2-c^*} y_i}$$

$$= P(Z_1 = r_1, Z_2 = r_2) h(\underline{y}; \theta | Z_1 = r_1, Z_2 = r_2)$$

Here distribution of  $(Z_1, Z_2)$  is trinomial and is a complete family of distribution and

$$h(\underline{y}; \theta | Z_1 = r_1, Z_2 = r_2) = \frac{\theta^{(n-r_1-r_2-c^*)} e^{-\theta \sum_{i=1}^{n-r_1-r_2-c^*} y_i}}{\left( \frac{n!}{r_1! r_2! (n-r_1-r_2)!} \right)}$$

which belongs to the one-parameter exponential family. Hence  $Z_3 | Z_1, Z_2$  is complete sufficient for  $\theta$  and also a member of the exponential family. The distribution of  $Z_3 | Z_1, Z_2$  is Gamma with parameter  $(n - r_1 - r_2 - c^*, \theta)$  with pdf

$$h(z_3; \theta | n - r_1 - r_2 - c^*) = \frac{z_3^{(n-r_1-r_2-c^*-1)} \theta^{n-r_1-r_2-c^*} e^{-\theta z_3}}{\Gamma_{n-r_1-r_2-c^*}}, z_3 > 0; \theta > 0$$

which depends only on  $\theta$  and is also a complete family of distribution. Therefore, using result of [11]  $Z = (Z_1, Z_2, Z_3)$  is complete sufficient for  $\underline{\theta} = (p_1, p_2, \theta)$ . The Joint distribution of  $Z = (Z_1, Z_2, Z_3)$  is

$$h_z(z; \underline{\theta}) = \frac{n!}{r_1! r_2! (n - r_1 - r_2)!} p_1^{r_1} p_2^{r_2} (1 - p_1 - p_2)^{(n-r_1-r_2)} \frac{z_3^{(n-r_1-r_2-c^*-1)}}{\Gamma n - r_1 - r_2 - c^*} \theta^{n-r_1-r_2-c^*} e^{-\theta z_3},$$

$$0 \leq r_1, r_2 \leq n - c^*; z_3 > 0; 0 \leq p_1, p_2 \leq 1; \theta > 0$$

$$= B(z_1, z_2, z_3, c^*, n) \frac{\prod_{i=1}^3 (h_i(\underline{\theta}))^{z_i}}{g(\underline{\theta})^{n-c^*}} (1 - p_1 - p_2)^{c^*}$$

where

$$B(z_1, z_2, z_3, c^*, n) = \begin{cases} \frac{n!}{r_1! r_2! (n-r_1-r_2)!} \frac{z_3^{(n-r_1-r_2-c^*-1)}}{\Gamma n - r_1 - r_2 - c^*}, & z_3 > 0; r_1 + r_2 - 1 < n - c^* \\ 1, & z_3 = 0; r_1 = 0 \text{ or } r_2 = 0 \end{cases} \quad (17)$$

$z_i \in T(n - c^*) \subseteq \mathbb{R}, \underline{\theta} \in \Omega$ . Here  $z = (z_1, z_2, z_3, c^*, n)$  and  $B(z_1, z_2, z_3, c^*, n)$  are such that

$$\frac{g(\underline{\theta})^{n-c^*}}{(1-p_1-p_2)^{c^*}} = \int_{z_1 \in T(n-c^*)} \int_{z_2 \in T(n-c^*)} \int_{z_3 \in T(n-c^*)} B(z_1, z_2, z_3, c^*, n) \prod_{i=1}^3 (h_i(\underline{\theta}))^{z_i} dz_1 dz_2 dz_3$$

Since  $(C_1(x)) = p_1, E(C_2(x)) = p_2$  and  $E(C_3(x)) = \frac{(1-p_1-p_2)}{\theta}$  (see Appendix A for details). Hence,

$$E(Z_1) = E(\sum_{j=1}^n C_1(x_j)) = \sum_{j=1}^{n-c^*} E(I_1(y_j)) = (n - c^*) p_1,$$

$$E(Z_2) = E(\sum_{j=1}^n C_2(x_j)) = \sum_{j=1}^{n-c^*} E(I_2(y_j)) = (n - c^*) p_2,$$

and

$$E(Z_3) = E(\sum_{j=1}^n C_3(x_j)) = \sum_{i=1}^{n-r_1-r_2-c^*} E(Y_i) = (n - c^*) \frac{(1-p_1-p_2)}{\theta},$$

which in turn give UMVUE's of  $p_1, p_2$  and  $\theta$  as

$$\hat{p}_1 = \frac{Z_1}{n-c^*} = \frac{r_1}{n-c^*} \quad (18)$$

$$\hat{p}_2 = \frac{Z_2}{n-c^*} = \frac{r_2}{n-c^*} \quad (19)$$

and

$$\hat{\theta} = \frac{(n-c^*)(1-\hat{p}_1-\hat{p}_2)}{Z_3} \quad (20)$$

For variance computation, see Appendix A. Note that, the likelihood estimate and minimum variance unbiased estimate of the parameters coincides everywhere when  $c^*=0$ .

## II. Uniformly Minimum Variance Unbiased Estimation of parametric functions

Let  $X_1, X_2, \dots, X_{n-c^*}$  be Type-II censored random sample from (13), then there exists an UMVUE of  $\Phi(\underline{\theta})$  if and only if  $\Phi(\underline{\theta})[g(\underline{\theta})]^{n-c^*}$  can be expressed in the form

$$\frac{\Phi(\underline{\theta})[g(\underline{\theta})]^{n-c^*}}{(1 - p_1 - p_2)^{c^*}} = \int_{z_1 \in T(n-c^*)} \int_{z_2 \in T(n-c^*)} \int_{z_3 \in T(n-c^*)} \alpha(z_1, z_2, z_3, c^*, n) \prod_{i=1}^3 (h_i(\underline{\theta}))^{z_i} dz_1 dz_2 dz_3$$

Thus, the UMVUE of a function  $\Phi(\underline{\theta})$  of  $\underline{\theta}$  in  $h(x; \underline{\theta})$  is given by

$$\psi(Z_1, Z_2, Z_3, c^*, n) = \frac{\alpha(Z_1, Z_2, Z_3, c^*, n)}{B(Z_1, Z_2, Z_3, c^*, n)}, \quad B(Z_1, Z_2, Z_3, c^*, n) \neq 0$$

The following results are now obvious.

**Result 1** The UMVUE of  $\prod_{i=1}^3 (h_i(\underline{\theta}))^{k_i} = \left(\frac{1}{\theta(1-p_1-p_2)}\right)^{k_1+k_2} p_1^{k_1} p_2^{k_2} e^{-\theta k_3}$  is given by

$$H_{k_1, k_2, k_3}(z_1, z_2, z_3, c^*, n) = \frac{B(z_1 - k_1, z_2 - k_2, z_3 - k_3, c^*, n)}{B(z_1, z_2, z_3, c^*, n)}$$

$$= \frac{(r_1)_{k_1} (r_2)_{k_2} \left(1 - \frac{k_3}{z_3}\right)^{(n-r_1-r_2-c^*-1)} (z_3 - k_3)^{k_1+k_2}}{[n-r_1-r_2+1]_{k_1+k_2} [n-r_1-r_2-c^*]_{k_1+k_2}},$$

where  $k_1 \leq r_1; k_2 \leq r_2; k_3 \leq z_3; k_1 + k_2 \leq n - r_1 - r_2 - c^*$ ;  $r_1 + r_2 - 1 < n - c^*$ , and  $(r)_k = \frac{r!}{(r-k)!}$ ,  $[r]_k = \frac{\Gamma r+k}{\Gamma r}$ .

**Corollary 1** If  $k_1 \neq 0, k_2 = 0$  and  $k_3 = 0$ , then UMVUE of  $(h_1(\theta))^{k_1} = \left(\frac{p_1}{\theta(1-p_1-p_2)}\right)^{k_1}$  is given by

$$H_{k_1}(z_1, z_2, z_3, c^*, n) = \frac{B(z_1-k_1, z_2, z_3, c^*, n)}{B(z_1, z_2, z_3, c^*, n)} = \frac{(r_1)_{k_1} z_3^{k_1}}{[n-r_1-r_2+1]_{k_1} [n-r_1-r_2-c^*]_{k_1}},$$

$k_1 \leq r_1; k_1 \leq n - r_1 - r_2 - c^*; r_1 + r_2 - 1 < n - c^*$

**Corollary 2** If  $k_1 = 0, k_2 \neq 0$  and  $k_3 = 0$ , then UMVUE of  $(h_2(\theta))^{k_2} = \left(\frac{p_2}{\theta(1-p_1-p_2)}\right)^{k_2}$  is given by

$$H_{k_2}(z_1, z_2, z_3, c^*, n) = \frac{B(z_1, z_2-k_2, z_3, c^*, n)}{B(z_1, z_2, z_3, c^*, n)} = \frac{(r_2)_{k_2} z_3^{k_2}}{[n-r_1-r_2+1]_{k_2} [n-r_1-r_2-c^*]_{k_2}},$$

$k_2 \leq r_2; k_2 \leq n - r_1 - r_2 - c^*; r_1 + r_2 - 1 < n - c^*$

**Corollary 3** If  $k_1 = 0, k_2 = 0$  and  $k_3 \neq 0$ , then UMVUE of  $(h_3(\theta))^{k_3} = e^{-\theta k_3}$  is given by

$$H_{k_3}(z_1, z_2, z_3, c^*, n) = \frac{B(z_1, z_2, z_3 - k_3, c^*, n)}{B(z_1, z_2, z_3, c^*, n)} = \left(1 - \frac{k_3}{z_3}\right)^{n-r_1-r_2-c^*-1},$$

$k_3 \leq z_3; r_1 + r_2 - 1 < n - c^*$

**Result 2** The UMVUE of the variance of  $H_{k_1, k_2, k_3}(Z_1, Z_2, Z_3, c^*, n)$ , is given by

$$\widehat{var}[H_{k_1, k_2, k_3}(z_1, z_2, z_3, c^*, n)] = H_{k_1, k_2, k_3}^2(z_1, z_2, z_3, c^*, n) - H_{2k_1, 2k_2, 2k_3}(z_1, z_2, z_3, c^*, n)$$

$$= \left[ \frac{(r_1)_{k_1} (r_2)_{k_2} \left(1 - \frac{k_3}{z_3}\right)^{(n-r_1-r_2-c^*-1)} (z_3 - k_3)^{k_1+k_2}}{[n-r_1-r_2+1]_{k_1+k_2} [n-r_1-r_2-c^*]_{k_1+k_2}} \right]^2 - \frac{(r_1)_{2k_1} (r_2)_{2k_2} \left(1 - \frac{2k_3}{z_3}\right)^{(n-r_1-r_2-c^*-1)} (z_3 - 2k_3)^{2(k_1+k_2)}}{[n-r_1-r_2+1]_{2(k_1+k_2)} [n-r_1-r_2-c^*]_{2(k_1+k_2)}},$$

where  $2k_1 \leq r_1; 2k_2 \leq r_2; 2k_3 \leq z_3; 2(k_1 + k_2) \leq n - r_1 - r_2 - c^*$ ;  $r_1 + r_2 - 1 < n - c^*$ .

**Corollary 4** The UMVUE of the variance of  $H_{k_1}(Z_1, Z_2, Z_3, c^*, n)$ , is given by

$$\widehat{var}[H_{k_1}(z_1, z_2, z_3, c^*, n)] = H_{k_1}^2(z_1, z_2, z_3, c^*, n) - H_{2k_1}(z_1, z_2, z_3, c^*, n)$$

$$= \left[ \frac{(r_1)_{k_1} z_3^{k_1}}{[n-r_1-r_2+1]_{k_1} [n-r_1-r_2-c^*]_{k_1}} \right]^2 - \frac{(r_1)_{2k_1} z_3^{2k_1}}{[n-r_1-r_2+1]_{2k_1} [n-r_1-r_2-c^*]_{2k_1}},$$

$2k_1 \leq r_1; 2k_1 \leq n - r_1 - r_2 - c^*; r_1 + r_2 - 1 < n - c^*$

**Corollary 5** The UMVUE of the variance of  $H_{k_2}(Z_1, Z_2, Z_3, c^*, n)$ , is given by

$$\widehat{var}[H_{k_2}(z_1, z_2, z_3, c^*, n)] = H_{k_2}^2(z_1, z_2, z_3, c^*, n) - H_{2k_2}(z_1, z_2, z_3, c^*, n)$$

$$= \left[ \frac{(r_2)_{k_2} z_3^{k_2}}{[n-r_1-r_2+1]_{k_2} [n-r_1-r_2-c^*]_{k_2}} \right]^2 - \frac{(r_2)_{2k_2} z_3^{2k_2}}{[n-r_1-r_2+1]_{2k_2} [n-r_1-r_2-c^*]_{2k_2}},$$

$2k_2 \leq r_2; 2k_2 \leq n - r_1 - r_2 - c^*; r_1 + r_2 - 1 < n - c^*$

**Corollary 6** The UMVUE of the variance of  $H_{k_3}(Z_1, Z_2, Z_3, c^*, n)$ , is given by

$$\widehat{var}[H_{k_3}(z_1, z_2, z_3, c^*, n)] = H_{k_3}^2(z_1, z_2, z_3, c^*, n) - H_{2k_3}(z_1, z_2, z_3, c^*, n)$$

$$= \left(1 - \frac{k_3}{z_3}\right)^{2(n-r_1-r_2-c^*-1)} - \left(1 - \frac{2k_3}{z_3}\right)^{n-r_1-r_2-c^*-1},$$

$2k_3 \leq z_3; r_1 + r_2 - 1 < n - c^*$

**Result 3** The UMVUE of  $[g(\theta)]^k = \left(\frac{1}{\theta(1-p_1-p_2)}\right)^k, k \neq 0$  as per the model given in (13) is



$$G_k(z_1, z_2, z_3, c^*, n) = \frac{B(z_1, z_2, z_3, c^*, n+k)}{B(z_1, z_2, z_3, c^*, n)} = \frac{[n+1]_k z_3^k}{[n-r_1-r_2+1]_k [n-r_1-r_2-c^*]_k}, k \leq n-r_1-r_2-c^*; r_1+r_2-1 < n-c^*$$

**Result 4** The UMVUE of the variance of  $G_k(Z_1, Z_2, Z_3, c^*, n)$  is given by

$$\widehat{var}[G_k(z_1, z_2, z_3, n)] = G_k^2(z_1, z_2, z_3, c^*, n) - G_{2k}(z_1, z_2, z_3, c^*, n) = \left[ \frac{[n+1]_k z_3^k}{[n-r_1-r_2+1]_k [n-r_1-r_2-c^*]_k} \right]^2 - \frac{[n+1]_{2k} z_3^{2k}}{[n-r_1-r_2+1]_{2k} [n-r_1-r_2-c^*]_{2k}}, 2k \leq n-r_1-r_2-c^*; r_1+r_2-1 < n-c^*$$

**Result 5** For fixed  $x$ , the UMVUE of the density given in (13) is

$$\phi_x(z_1, z_2, z_3, c^*, n) = a(x) \frac{B(z_1-c_1(x), z_2-c_2(x), z_3-c_3(x), c^*, n-1)}{B(z_1, z_2, z_3, c^*, n)} = \left( \frac{1}{1+x} \right) \frac{(r_1)_{I_1(x)} (r_2)_{I_2(x)} (n-r_1-r_2)_{(1-I_1(x)-I_2(x))} (n-r_1-r_2-c^*-1)_{(1-I_1(x)-I_2(x))}}{n \left[ z_3 - \left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x)) \right]^{(1-I_1(x)-I_2(x))}} \left( 1 - \frac{\left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x))}{z_3} \right)^{(n-r_1-r_2-c^*-1)}, z_3 > \left[ \log\left(\frac{1+x}{2}\right) \right]; r_1+r_2-1 < n-c^*$$

**Result 6** The UMVUE of the variance of  $\phi_x(Z_1, Z_2, Z_3, c^*, n)$  is given by

$$\widehat{var}[\phi_x(z_1, z_2, z_3, c^*, n)] = \phi_x^2(z_1, z_2, z_3, c^*, n) - \phi_x(z_1, z_2, z_3, c^*, n) \phi_x(z_1 - c_1(x), z_2 - c_2(x), z_3 - c_3(x), c^*, n-1) = \phi_x^2(z_1, z_2, z_3, c^*, n) - \left( \frac{1}{1+x} \right)^2 \frac{(r_1)_{2I_1(x)} (r_2)_{2I_2(x)} (n-r_1-r_2)_{2(1-I_1(x)-I_2(x))} (n-r_1-r_2-c^*-1)_{2(1-I_1(x)-I_2(x))}}{n(n-1) \left[ z_3 - 2 \left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x)) \right]^{2(1-I_1(x)-I_2(x))}} \left( 1 - \frac{2 \left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x))}{z_3} \right)^{(n-r_1-r_2-c^*-1)}, z_3 > 2 \left[ \log\left(\frac{1+x}{2}\right) \right]; r_1+r_2-1 < n-c^*$$

**Result 7** For a fixed  $z = (z_1, z_2, z_3, c^*, n)$ , the UMVUE of the survival function  $S(x) = p(X > x)$ ,  $x \geq 0$  is obtained as

$$\hat{S}(x) = \left( \frac{(r_1)_{I_1(x)} (r_2)_{I_2(x)} (n-r_1-r_2)_{(1-I_1(x)-I_2(x))} (n-r_1-r_2-c^*-1)_{(1-I_1(x)-I_2(x))}}{n [(n-r_1-r_2-c^*) - (1-I_1(x)-I_2(x))]} \right)^{(n-r_1-r_2-c^*-1)} \left( z_3 - \left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x)) \right)^{(I_1(x)+I_2(x))} \left( 1 - \frac{\left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x))}{z_3} \right)^{(n-r_1-r_2-c^*-1)}, z_3 > \left[ \log\left(\frac{1+x}{2}\right) \right]; r_1+r_2-1 < n-c^*$$

**Result 8** For the fixed  $z = (z_1, z_2, z_3, c^*, n)$ , the UMVUE of the  $var(\hat{S}(x))$ , is obtained as

$$\widehat{var}(\hat{S}(x)) = [\hat{S}(x)]^2 - \frac{1}{n(n-1)} \left( 1 - \frac{2 \left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x))}{z_3} \right)^{(n-r_1-r_2-c^*-1)} \left( \frac{(r_1)_{2I_1(x)} (r_2)_{2I_2(x)} (n-r_1-r_2)_{2(1-I_1(x)-I_2(x))} (n-r_1-r_2-c^*-1)_{2(1-I_1(x)-I_2(x))}}{[(n-r_1-r_2-c^*) - 2(1-I_1(x)-I_2(x))] [(n-r_1-r_2-c^*+1) - 2(1-I_1(x)-I_2(x))]} \right) \left( z_3 - 2 \left[ \log\left(\frac{1+x}{2}\right) \right] (1-I_1(x)-I_2(x)) \right)^{2(I_1(x)+I_2(x))}, z_3 > 2 \left[ \log\left(\frac{1+x}{2}\right) \right]; r_1+r_2-1 < n-c^*$$

### III. Real data illustration

In this section, we have considered four inliers prone data set to illustrate our proposed work. The motivation behind considering a different variety of data sets is to show the flexibility of the proposed model in different situations. The detailed description regarding the data sets is given below:

**Dataset 1:** The data in Table 1 shows the loss ratios (yearly data) for earthquake insurance in

California from 1971 through 1993. The data are taken from [6] and also used by [4] for their study. Note that, for four years there was no loss for earthquake insurance and the information where loss of less than 1 billion dollars per year is considered as 1, for simplicity. The analysis of this data is carried out at the end of this section.

**Table 1.** California earthquake insurance data

Year	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
Loss ratios	17.4	0.0	0.6	3.4	0.0	0.0	0.7	1.5	2.2	9.2	0.9	0.0
Year	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	
Loss ratios	2.9	5.0	1.3	9.3	22.8	11.5	129.8	47.0	17.2	12.8	3.2	

**Dataset 2:** The National Family Health Survey (NFHS) is a large-scale, multi-round survey conducted in a representative sample of households throughout India. The First National Family Health Survey (NFHS-1) was conducted in 1992-93, the Second National Family Health Survey (NFHS-2) was conducted in 1998-99 and the Third National Family Health Survey (NFHS-3) was carried out in 2005-06. The survey is based on a sample of households that is representative at the national and state levels. The NFHS-3 fieldwork, conducted by 18 research organizations between December 2005 and August 2006, interviewed women at age 15-49. We consider the data on child's age at death from the woman's questionnaire of NFHS-3. For comprehensive data, one may visit [24]. For Gujarat state, there are 15 stillbirths (the death of a baby before or during the birth after 28 weeks of gestation) considered as observation 0, 37 neonatal deaths (the death of a baby within the first 28 days of life) considered as observation 1 and other observations of age at death in days as: 30, 30, 30, 31, 31, 60, 62, 62, 62, 90, 90, 90, 92, 93, 150, 182, 213, 242, 272, 273, 300, 303, 333, 334, 335, 356, 360, 365, 366, 450, 730, 731, 732, 732 and 1462. This is a perfect data for inliers model with two discrete point at zero and one. Authors of this paper had already modeled this data using exponential and Weibull distribution. The analysis based on Pareto Type II distribution is presented below.

**Dataset 3:** [23] have analyzed and quantified forest burnt area in India using AWiFS data for the year 2014. The burnt area map from AWiFS data involves Forest type map of 2013 at 56 m resolution prepared as part of the national carbon project. India has a geographical area of about 3,287,263 sq. km. It comprises 29 states and 7 union territories. The country has 21% of the geographical area under forest cover. Forest fires occur in India mainly between January and June. They are more frequent between February and May in different biogeographic zones of India. State/Union Territory-wise analysis of the percentage of forest burnt area (area in sq. km) is available in [23], page 1531. We consider State/Union Territory burnt area from February to May 2014. There are six State/Union Territory (Delhi, Andaman and Nicobar, Chandigarh, Daman and Diu, Lakshadweep and Pondicherry) having burnt area zero, five State/Union Territory (Goa(0.04), Jammu and Kashmir (0.11), Dadra and Nagar Haveli (0.23), Punjab (0.85) and Himachal Pradesh (0.91)) having percentage burnt area less than 1 sq. Km. conveniently considered here as observation 1, and the remaining 25 State/Union Territory burnt area in sq. Km. are: 6611.86, 102.70, 941.11, 1773.22, 4606.69, 487.81, 1.84, 2587.40, 1920.35, 82.01, 3342.66, 5066.66, 1974.23, 457.50, 421.03, 975.79, 8186.46, 364.17, 2.50, 4275.64, 2955.23, 739.00, 459.07, 42.01 and 386.37. The analysis is reported below.

**Dataset 4:** This data is about the amount of snowfall in all 50 states of US. According to the National Climatic Data Center, the data were populated considering the average snowfall for almost three decades from 1981 to 2010, available at [25]. The average amount snowfall per year (in inches) for 50 states of US are: 5.2, 0.5, 1.6, 74.5, 0.3, 0.0, 19.1, 40.5, 20.2, 0.0, 0.7, 0.0, 19.2, 24.6, 25.9, 34.9, 14.7, 12.5, 0.0, 61.8, 20.2, 43.8, 51.1, 54, 0.9, 17, 38.1, 25.9, 21.8, 60.8, 16.5, 9.6, 123.8, 7.6, 51.2, 27.5, 7.8, 3, 28.2, 33.8, 43.9, 6.3, 1.5, 56.2, 81.2, 10.3, 5.0, 62.0, 50.9 and 91.4. It is observed that there

are three decades having an average amount of snowfall zero and for four states having decades average amount of snowfall less than 1 inches (coded as observation1).

For all the data sets above we have calculated parameter estimates, goodness-of-fit criteria values, goodness-of-fit statistics and corresponding  $p$ -values (see Table 2 for details) for positive observations only. It may be noted from the table that for all the considered data sets, the Pareto Distribution fits well (see  $p$ -values).

**Table 2.** The parameter estimates, goodness-of-fit criteria and corresponding  $p$ -value for various datasets (Pareto distribution).

Data	MLE (SE)	AIC	BIC	K-S ( $p$ -value)	CVM ( $p$ -value)	AD ( $p$ -value)
Earthquake insurance	$\hat{\beta}$ = 19.5743 (19.2742) $\hat{\theta}$ = 2.0113 (1.4153)	124.7323	126.2778	0.1213 (0.9498)	0.0362 (0.9563)	0.2901 (0.9448)
NFHS-3	$\hat{\beta}$ =18557.4806 (34321.4861) $\hat{\theta}$ = 65.5015 (119.8512)	470.3576	473.4683	0.1210 (0.6848)	0.0898 (0.6400)	0.6150 (0.6327)
Forest burnt area	$\hat{\beta}$ =3418.3510 (4828.3362) $\hat{\theta}$ = 2.6249 (2.7363)	431.6623	434.1000	0.1446 (0.6214)	0.0984 (0.5964)	1.0663 (0.3236)
Snow fall	$\hat{\beta}$ =2907.8650 (8293.9850) $\hat{\theta}$ = 87.5320 (247.1416)	383.029	386.5043	0.1049 (0.7447)	0.0933 (0.6208)	0.5532 (0.6922)

(\* Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov (K-S) Statistic, Cramer-Von Mises (CVM) statistics, Anderson-Darling (AD) statistic).

The plot of pdf,  $h(x)$  and survival function,  $S(x)$  for all four datasets under study, is displayed in Figure 1 and Figure 2 respectively for varying censoring schemes under Pareto II and the Weibull distribution. For the data sets under study, the summary of the various estimates of parameters and parametric functions along with their standard error (shown in bracket) and 95 % confidence interval considering censoring schemes at value  $c^*$  is given in Table 3. Whereas Table 4 shows, the UMVU estimate of pdf and survival function with Pareto II and the Weibull distribution for varying censoring schemes. It is observed that Pareto distribution has a heavier tail than Weibull.

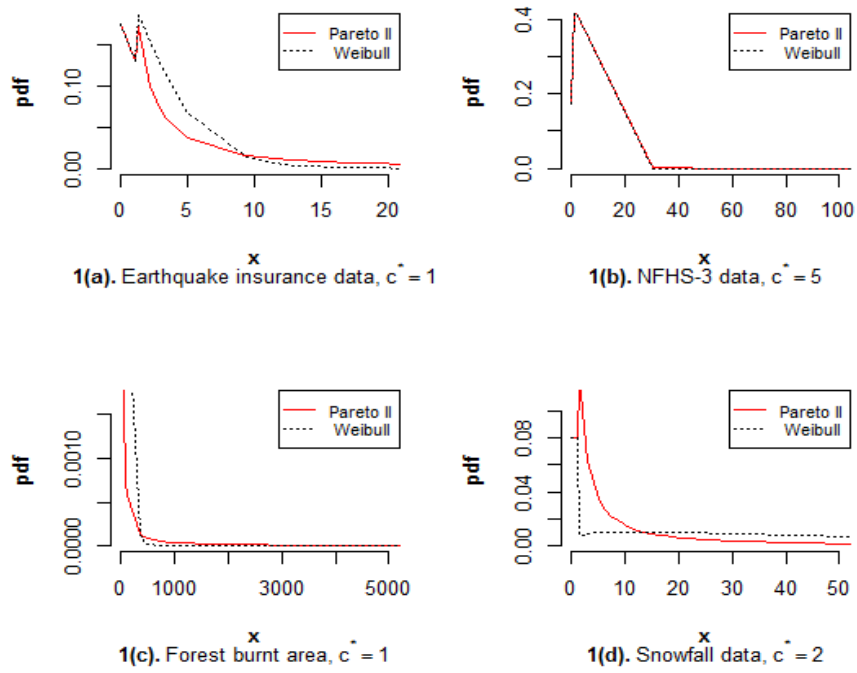


Fig. 1 Density plot to various data sets censored at value  $c^*$

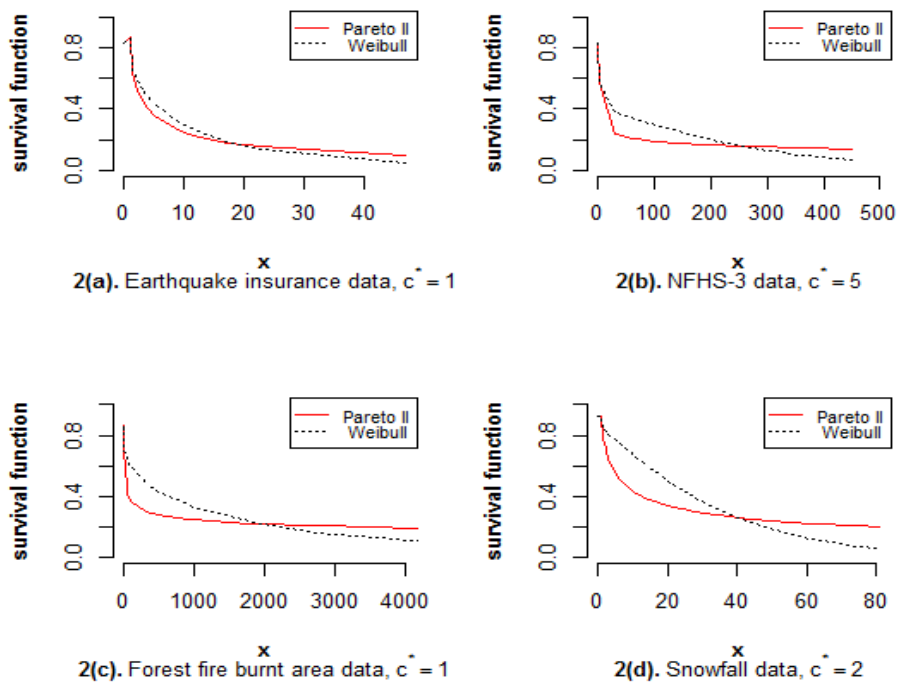


Fig. 2 Survival function plot to various data sets censored at value  $c^*$

**Table 3.** Summary of estimates of parameters/parametric functions of Pareto II distribution censored at  $c^*$ .

Parameter/Parametric function	Earthquake insurance data	NFHS-3 data	Forest fire burnt area data	Snowfall data
	$c^*=1$	$c^*=5$	$c^*=1$	$c^*=2$
MLE (SE) of $p_1$	0.17391 (0.07904)	0.17241 (0.04050)	0.16667 (0.06212)	0.08000 (0.03837)
MLE (SE) of $p_2$	0.13043 (0.07022)	0.42529 (0.05300)	0.13889 (0.05764)	0.08000 (0.03837)
MLE (SE) of $\theta$	0.61420 (0.15857)	0.19539 (0.03402)	0.16667 (0.03380)	0.38550 (0.06095)
95% CI of $p_1$	(0.01901, 0.32882)	(0.09304, 0.25179)	(0.04493, 0.28841)	(0.00480, 0.15520)
95% CI of $p_2$	(0.00000, 0.26807)	(0.32140, 0.52917)	(0.02592, 0.25186)	(0.00480, 0.15520)
95% CI of $\theta$	(0.30648, 0.92191)	(0.12871, 0.26206)	(0.10041, 0.23292)	(0.26651, 0.50449)
UMVUE (SE) of $p_1$	0.18182 (0.08223)	0.18293 (0.04269)	0.17143 (0.06370)	0.08333 (0.03989)
UMVUE (SE) of $p_2$	0.13636 (0.07317)	0.45122 (0.05495)	0.14286 (0.05915)	0.08333 (0.03989)
UMVUE (SE) of $\theta$	0.61420 (0.13812)	0.19539 (0.02791)	0.16667 (0.02967)	0.38550 (0.05643)
95% CI of UMVUE $p_1$	(0.02065, 0.34299)	(0.09925, 0.26660)	(0.04657, 0.29629)	(0.00515, 0.16152)
95% CI of UMVUE of $p_2$	(0.00000, 0.27976)	(0.34351, 0.55892)	(0.02693, 0.25879)	(0.00515, 0.16152)
95% CI of UMVUE $\theta$	(0.34348, 0.88492)	(0.14069, 0.25008)	(0.10850, 0.22483)	(0.27489, 0.49610)
$\prod_{i=1}^3 (h_i(\theta))^{k_i} = \left(\frac{\theta}{1-p_1-p_2}\right)^2 p_1 p_2 e^{-\frac{1}{\theta}}$ $k_1 = 1, k_2 = 1, k_3 = 1$	0.04992 (0.04299)	8.62541 (4.73007)	1.240801 (0.89610)	3.07667 (0.02775)
$h_1(\theta) = \frac{\theta p_1}{1-p_1-p_2}, k_1 = 1, k_2 = 0, k_3 = 0$	0.38309 (0.22204)	2.13250 (0.74236)	1.38463 (0.66351)	0.24131 (0.12880)
$h_2(\theta) = \frac{\theta p_2}{1-p_1-p_2}, k_1 = 0, k_2 = 1, k_3 = 0$	0.28732 (0.18391)	5.26018 (1.52329)	1.15386 (0.58886)	0.24131 (0.12880)
$h_3(\theta) = e^{-\frac{1}{\theta}}, k_1 = 0, k_2 = 0, k_3 = 1$	0.55693 (0.08844)	0.82738 (0.00109)	0.85191 (0.02856)	0.68545 (0.04162)
$g(\theta) = \frac{\theta}{1-p_1-p_2}, k = 1$	2.29856 (0.64068)	12.51070 (2.73276)	8.53853 (1.92013)	3.07667 (0.51361)

**Table 4.** Summary of estimates of pdf and reliability function of the various data sets censored at  $c^*$ .

Function	Earthquake insurance data		NFHS-3 data		Forest fire burnt area data		Snowfall data	
	$c^*=1$		$c^*=5$		$c^*=1$		$c^*=2$	
	Pareto-II	Weibull	Pareto-II	Weibull	Pareto-II	Weibull	Pareto-II	Weibull
pdf	$\phi_{10} = 0.01415$ (0.00185)	$\phi_{10} = 0.02091$ (0.00295)	$\phi_{100} = 0.00030$ (5.043e-05)	$\phi_{100} = 0.0011$ (1.968e-04)	$\phi_{650} = 6.912e-05$ (7.524e-06)	$\phi_{650} = 0.00022$ (3.235e-05)	$\phi_{25} = 0.00469$ (0.00028)	$\phi_{25} = 0.01383$ (0.00114)
	$\phi_{15} = 0.00784$ (0.00112)	$\phi_{15} = 0.01317$ (0.00113)	$\phi_{500} = 5.473e-05$ (7.032e-06)	$\phi_{500} = 0.0003$ (6.096e-05)	$\phi_{1350} = 2.966e-05$ (3.199e-06)	$\phi_{1350} = 0.00012$ (8.101e-06)	$\phi_{50} = 0.00186$ (0.00013)	$\phi_{50} = 0.00676$ (0.00068)
	$\phi_{40} = 0.00175$ (0.00046)	$\phi_{40} = 0.00273$ (0.00046)	$\phi_{1000} = 2.387e-05$ (3.185e-06)	$\phi_{1000} = 1.851e-05$ (1.282e-05)	$\phi_{2500} = 1.451e-05$ (1.624e-06)	$\phi_{2500} = 6.421e-05$ (3.185e-06)	$\phi_{100} = 0.00072$ (7.193e-05)	$\phi_{100} = 0.00108$ (0.00044)
Survival function	$\hat{S}_{10} = 0.25585$ (0.07038)	$\hat{S}_{10} = 0.17398$ (0.05400)	$\hat{S}_{100} = 0.18996$ (0.03638)	$\hat{S}_{100} = 0.29777$ (0.04249)	$\hat{S}_{650} = 0.27042$ (0.07114)	$\hat{S}_{650} = 0.40038$ (0.06434)	$\hat{S}_{25} = 0.31647$ (0.05347)	$\hat{S}_{25} = 0.43451$ (0.05291)
	$\hat{S}_{15} = 20013$ (0.07279)	$\hat{S}_{15} = 0.00183$ (0.04218)	$\hat{S}_{500} = 0.13903$ (0.03296)	$\hat{S}_{500} = 0.05063$ (0.02039)	$\hat{S}_{1350} = 0.23941$ (0.06868)	$\hat{S}_{1350} = 0.05022$ (0.06206)	$\hat{S}_{50} = 0.24389$ (0.05082)	$\hat{S}_{50} = 0.18368$ (0.04572)
	$\hat{S}_{40} = 0.10957$ (0.05592)	$\hat{S}_{40} = 2.139e-06$ (0.00953)	$\hat{S}_{1000} = 0.12136$ (0.03136)	$\hat{S}_{1000} = 0.00293$ (0.02039)	$\hat{S}_{2500} = 0.21592$ (0.05811)	$\hat{S}_{2500} = 0.01212$ (0.05510)	$\hat{S}_{100} = 0.18692$ (0.04659)	$\hat{S}_{100} = 0.02462$ (0.01331)

### Appendix A. Distributional properties of $C(X)$

Since the moments of  $C(X) = (C_1(X), C_2(X), C_3(X))$  are functions of  $\underline{\theta} = (p_1, p_2, \theta)$ , and  $\beta$  assumed known, they are MVUE's of these functions. Hence, in order to find the moments, differentiating  $g(\underline{\theta})$  partially with respect to  $p_1, p_2$  and  $\theta$  under the regularity conditions, we get

$$\underline{G} = A \underline{\mu}, |A| \neq 0 \tag{i}$$

where

$$\underline{G} = \begin{bmatrix} \frac{\partial \log g(\underline{\theta})}{\partial p_1} \\ \frac{\partial \log g(\underline{\theta})}{\partial p_2} \\ \frac{\partial \log g(\underline{\theta})}{\partial \theta} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 - p_1 - p_2 \\ 1 \\ -\frac{1}{\theta} \end{bmatrix}$$

$$\underline{\mu} = \begin{bmatrix} E(C_1(x)) \\ E(C_2(x)) \\ E(C_3(x)) \end{bmatrix} = \begin{bmatrix} E(I_1(x)) \\ E(I_2(x)) \\ E([\log(1+x) - \log 2](1 - I_1(x) - I_2(x))) \end{bmatrix}$$

and

$$A = \begin{bmatrix} \frac{\partial \log h_1(\theta)}{\partial p_1} & \frac{\partial \log h_2(\theta)}{\partial p_1} & \frac{\partial \log h_3(\theta)}{\partial p_1} \\ \frac{\partial \log h_1(\theta)}{\partial p_2} & \frac{\partial \log h_2(\theta)}{\partial p_2} & \frac{\partial \log h_3(\theta)}{\partial p_2} \\ \frac{\partial \log h_1(\theta)}{\partial \theta} & \frac{\partial \log h_2(\theta)}{\partial \theta} & \frac{\partial \log h_3(\theta)}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{1}{p_1} + \frac{1}{1-p_1-p_2} & \frac{1}{1-p_1-p_2} & 0 \\ \frac{1}{1-p_1-p_2} & \frac{1}{p_2} + \frac{1}{1-p_1-p_2} & 0 \\ -\frac{1}{\theta} & -\frac{1}{\theta} & -1 \end{bmatrix}$$

Equation (i) gives

$$E(C_i(x)) = \frac{|A_i|}{|A|}, i = 1, 2 \text{ and } 3$$

where  $A_i$  is obtained by replacing  $i^{\text{th}}$  column of A by the elements of  $\underline{G}$ . Hence,

$$\underline{\mu} = \begin{bmatrix} E(C_1(x)) \\ E(C_2(x)) \\ E(C_3(x)) \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \frac{(1-p_1-p_2)}{\theta} \end{bmatrix} \tag{ii}$$

Now joint moments of  $C_1^{k_1}(x), C_2^{k_2}(x)$  and  $C_3^{k_3}(x)$  are given as

$$E(C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x)) = \int_x C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x) a(x) \frac{\prod_{i=1}^3 (h_i(\theta))^{C_i(x)}}{g(\theta)} dx$$

which on differentiating with respect to  $p_1, p_2$  and  $\theta$  and using (iv), gives a system of three non-homogeneous equations

$$\underline{G}_1 = A \underline{V}, |A| \neq 0 \tag{iii}$$

where

$$\underline{G}_1 = \begin{bmatrix} \frac{\partial \log E \left( C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x) \right)}{\partial p_1} \\ \frac{\partial \log E \left( C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x) \right)}{\partial p_2} \\ \frac{\partial \log E \left( C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x) \right)}{\partial \theta} \end{bmatrix}$$

$$\underline{V} = \begin{bmatrix} E \left( C_1^{k_1+1}(x) C_2^{k_2}(x) C_3^{k_3}(x) \right) - E(C_1(x))E \left( C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x) \right) \\ E \left( C_1^{k_1}(x) C_2^{k_2+1}(x) C_3^{k_3}(x) \right) - E(C_2(x))E \left( C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x) \right) \\ E \left( C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3+1}(x) \right) - E(C_3(x))E \left( C_1^{k_1}(x) C_2^{k_2}(x) C_3^{k_3}(x) \right) \end{bmatrix} = \begin{bmatrix} \sigma_{1(1,2,3)} \\ \sigma_{2(1,2,3)} \\ \sigma_{3(1,2,3)} \end{bmatrix}, \text{ (say).}$$

Using Cramer’s rule for the solution of a system of linear equations (iii) gives

$$\sigma_{i(1,2,3)} = \frac{|A_i|}{|A|}, i = 1,2 \text{ and } 3$$

where  $A_i$  is obtained by replacing  $i^{\text{th}}$  column of  $A$  by the elements of  $\underline{G}_1$ . For  $k_i = 1$  and  $k_j = 0 \forall i \neq j = 1,2 \text{ and } 3$ , we get covariance between  $C_i(x)$  and  $C_j(x)$  as

$$\sigma_{i(1,2,3)} = \frac{|A_i|_{(k_i=1; k_j=0), i \neq j}}{|A|}.$$

Thus, we have the variance-covariance matrix  $V$  as

$$V = [\sigma_{ij}]_{3 \times 3} = \frac{(|A_i|_{(k_i=1; k_j=0), i \neq j})}{|A|}$$

If  $A_{ij}$  is the cofactor of the element  $a_{ij}$  of  $A$ , then

$$|A_i|_{(k_i=1; k_j=0), i \neq j=1,2,3} = A_{1i} \frac{\partial}{\partial p_1} E(C_i(x)) + A_{2i} \frac{\partial}{\partial p_2} E(C_i(x)) + A_{3i} \frac{\partial}{\partial \theta} E(C_i(x))$$

and hence

$$V = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -\theta p_1(1-p_1-p_2) \\ -p_1p_2 & p_2(1-p_2) & -\theta p_2(1-p_1-p_2) \\ -\frac{p_1(1-p_1-p_2)}{\theta} & -\frac{p_2(1-p_1-p_2)}{\theta} & \frac{[1-(p_1+p_2)]^2}{\theta^2} \end{bmatrix} \quad \text{(iv)}$$

where  $|A| = \frac{1}{p_1 p_2 (1-p_1-p_2)}$ .

### Acknowledgements

This work is supported by DST-FIST No. SR/FST/MSI-104/2015. Authors also thank the reviewers for their useful comments.

### References

- [1] Aban, I. B., Meerschaert, M. M. and Panorska, A. K. (2006). Parameter estimation for the truncated Pareto distribution, *Journal of American Statistical Association*, 101(473): 270–277.
- [2] Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of American Statistical Association*, 50: 901–908.
- [3] Charalambides, C. H. (1974). Minimum variance unbiased estimation for a class of left truncated distributions. *Sankhya A*, 36: 392–418.

- [4] Embrechts, P., Resnick, S. I. and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3 (2): 30–41.
- [5] Gupta, R. C. (1977). Minimum variance unbiased estimation in modified power series distribution and some of its applications. *Communication in Statistics*, 6: 977–991.
- [6] Jaffe, D. M. and Russell, T. (1996). *Catastrophe Insurance, Capital Markets and Uninsurable Risk*. Philadelphia: Financial Institutions Center. The Wharton School, pp. 96–112.
- [7] Jani, P.N. (1977). Minimum variance unbiased estimation for some left-truncated modified power series distributions. *Sankhya*, 39: 258–278.
- [8] Jani, P. N. (1993). A characterization of one-parameter exponential family of distributions. *Calcutta Statistical Association Bulletin*, 43 (3–4): 253–256.
- [9] Jani, P. N. and Dave, H. P. (1990). Minimum variance unbiased estimation in a class of exponential family of distributions and some of its applications. *Metron*, 48: 493–507.
- [10] Jani, P. N. and Singh, A. K. (1995). Minimum variance unbiased estimation in multi-parameter exponential family of distributions. *Metron*, 53: 93–106.
- [11] Jayade, V. P. and Prasad, M. S. (1990). Estimation of parameters of mixed failure time distribution. *Communication in Statistics – Theory Methods*, 19(12): 4667–4677.
- [12] Joshi, S. W. and Park, C. J. (1974). Minimum variance unbiased estimation for truncated power series distributions. *Sankhya A*, 36: 305–314.
- [13] Kale, B. K. and Muralidharan, K. (2000). Optimal estimating equations in mixture distributions accommodating instantaneous or early failures. *Journal of the Indian Statistical Association*, 38: 317–329.
- [14] Khatri, C. G. (1959). On certain properties of power series distributions. *Biometrika*, 46: 486–490.
- [15] Muralidharan, K. (2010). Inlier prone models: A review. *ProbStat Forum*, 3: 38–51.
- [16] Muralidharan K. and Bavagosai, P. (2017) Analysis of lifetime model with discrete mass at zero and one. *Journal of Statistical Theory Practice*, 11(4), 670–692.
- [17] Muralidharan K. and Bavagosai, P. (2018). A new Weibull model with inliers at zero and one based on type-II censored samples. *Journal of Indian Society of Probability and Statistics*, 19: 121–151.
- [18] Muralidharan, K. and Lathika, P. (2006). Analysis of instantaneous and early failures in Weibull distribution. *Metrika*, 64(3): 305–316.
- [19] Patel, S. R. (1978). Minimum variance unbiased estimation of multivariate modified power series distribution. *Metrika*, 25: 155–161.
- [20] Patil, G. P. (1963). Minimum variance unbiased estimation and certain problem of additive number theory. *Annals of Mathematics and Statistics*, 34: 1050–1056.
- [21] Reed, W. J. and Jorgensen, M. (2004). The double Pareto–lognormal distribution—A new parametric model for size distributions. *Communication in Statistics– Theory Methods* 33(8): 1733–1753.
- [22] Reddy, C. S., Jha, C. S., Manaswini, G., Alekhya, V. V. L. P., Pasha, S. V., Satish, K. V., Diwakar, P. G. and Dadhwal, V. K. (2017). Nationwide assessment of forest burnt area in India using Resourcesat-2 AWiFS data. *Current Science*, 112(7): 1521–1532.
- [23] Roy, J. and Mitra, S. K. (1957). Unbiased minimum variance estimation in a class of discrete distributions. *Sankhya*, 18: 371–378.
- [24] [http://www.dhsprogram.com/data/dataset/India\\_Standard-DHS\\_2006.cfm?flag=0](http://www.dhsprogram.com/data/dataset/India_Standard-DHS_2006.cfm?flag=0)
- [25] <https://thetoolboss.com/average-snowfall-us-states>.