Chapter 7

Inferences on inliers in Pareto II distribution with Type II censored data

7.1. Introduction

In this chapter, we model the inliers situation using the Type II censored lifetime data from a Pareto II distribution. As discussed in the previous chapter, here also, we follow the same censoring scheme and sample discussed therein. The Pareto Type II distribution (also called Lomax distribution with location parameter zero) has the probability distribution function (pdf)

$$f(x;\underline{\alpha}) = \frac{\theta\beta^{\theta}}{(x+\beta)^{(1+\theta)}},\ x > 0; \beta > 0, \theta > 0 \tag{7.1.1}$$

where $\underline{\alpha} = (\beta, \theta)$, β is a scale parameter and θ 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 Aban et al. (2006) 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 Reed and Jorgensen (2004) for details.

This chapter is published as research article in *Reliability: Theory and Applications*. Vol. 13, No. 3 (50): 60 –89, 2018.

When β =1, the Pareto Type II inliers distribution has the density function as

$$h(x;\underline{\theta}) = \begin{cases} p_1, & x = 0 \\ p_2, & x = 1 \\ (1 - p_1 - p_2)\frac{\theta}{(1+x)}\left(\frac{2}{(1+x)}\right)^{\theta}, & x > 1 \end{cases} \tag{7.1.2}$$

The parameter vector to be estimated are $\underline{\theta} = (p_1, p_2, \theta)$. In various sections below, we suggest product moment estimator along with MLE and UMVU estimator of the unknown parameters in the model. We consider four real datasets for implementing the proposed model in the last section.

7.2. Maximum Likelihood Estimation

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 $X_1, X_2, \ldots, X_{n-r_1-r_2}$. 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 $X_{(1)}, X_{(2)}, \ldots,\ X_{(n-r_1-r_2-c^*)}$ denote ordered observed failure time of these $n - r_1 - r_2 - c^*$ items from $h \in \mathcal{H}$ as given in (7.1.2). Then the likelihood equation can be written as

$$L(\underline{\theta}|\underline{x}) = \prod_{i=1}^{n-r_1-r_2-c^*} h(x_{(i)};\underline{\theta})\left[1 - H(x_{(n-r_1-r_2-c^*)};\underline{\theta})\right]^{c^*}$$

If we define $I_1(x) = \begin{cases} 1, & x = 0 \\ 0, & o.w. \end{cases}$ and $I_2(x) = \begin{cases} 1, & x = 1 \\ 0, & o.w. \end{cases}$, then the likelihood equation is

$$L(\underline{\theta}|\underline{x}) = {p_1}^{r_1}{p_2}^{r_2}(1-p_1-p_2)^{(n-r_1-r_2)}\frac{(n-r_1-r_2)!}{c^*!}\,\theta^{n-r_1-r_2-c^*}$$

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

(7.2.1)

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. Let the maximum likelihood estimator of parameter $\underline{\theta} = (p_1, p_2, \theta)$, is $\hat{\underline{\theta}}_{MLE} = (\hat{p}_{1MLE}, \hat{p}_{2MLE}, \hat{\theta}_{MLE})$, say. We now investigate the following four possible cases of likelihood estimates:

Case (*i*). $r_1 = 0$, that is $r_2 = n$. The likelihood function simply reduces to $L(\underline{\theta}|\underline{x}) = p_2^n$. Obviously, this is maximum when $p_2 = 1$. This corresponds to the maximum likelihood estimator $\hat{p}_{2MLE} = \frac{r_2}{n}$. Since $L(\underline{\theta}|\underline{x}) = p_2^n$ is free from the other parameters, the maximum likelihood estimator of other parameters does not exist.

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

Case (*iii*). $r_1 < n, r_2 < n$ but $r_1 + r_2 = n$. The likelihood function simply reduces to $L(\underline{\theta}|\underline{x}) = p_1^{r_1}p_2^{r_2}$. Here $p_1 + p_2 < n$ and the likelihood function $L(\underline{\theta}|\underline{x}) < \left(\frac{x_1}{n}\right)^{r_1}\left(\frac{x_2}{n}\right)^{r_2}$. So $\hat{p}_{1MLE} = \frac{r_1}{n}$ and $\hat{p}_{2MLE} = \frac{r_2}{n}$. The maximum likelihood of other parameters does not exist.

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

$$log\, L(\underline{\theta}|\underline{x}) = r_1\, log\, p_1 + r_2\, log\, p_2 + (n - r_1 - r_2)\, log(1 - p_1 - p_2)$$
$$+ log(n - r_1 - r_2)! - log\, c^*! + (n - r_1 - r_2 - c^*)\, log\, \theta$$
$$- \sum_{i=1}^{n-r_1-r_2-c^*} log\left(1 + x_{(i)}\right)$$
$$-\theta \left[\sum_{i=1}^{n-r_1-r_2-c^*} log\left(\frac{1+x_{(i)}}{2}\right) + c^*\, log\left(\frac{1+x_{(n-r_1-r_2-c^*)}}{2}\right)\right] \quad (7.2.2)$$

The maximum likelihood estimator of parameter $\underline{\theta}$ is obtained by solving the following likelihood equations:

$$\frac{\partial\, log\, L(\underline{\theta}|\underline{x})}{\partial p_1} = \frac{r_1}{p_1} - \frac{n-r_1-r_2}{1-p_1-p_2} = 0 \quad (7.2.3)$$

$$\frac{\partial\, log\, L(\underline{\theta}|\underline{x})}{\partial p_2} = \frac{r_2}{p_2} - \frac{n-r_1-r_2}{1-p_1-p_2} = 0 \quad (7.2.4)$$

and

$$\frac{\partial\, log\, L(\underline{\theta}|\underline{x})}{\partial \theta} = \frac{n - r_1 - r_2 - c^*}{\theta}$$
$$- \left[\sum_{i=1}^{n-r_1-r_2-c^*} log\left(\frac{1 + x_{(i)}}{2}\right) + c^*\, log\left(\frac{1 + x_{(n-r_1-r_2-c^*)}}{2}\right)\right] = 0 \quad (7.2.5)$$

Solving (7.2.3) and (7.2.4) simultaneously, we get

$$\hat{p}_{1MLE} = \frac{r_1}{n} \quad (7.2.6)$$

and

$$\hat{p}_{2MLE} = \frac{r_2}{n} \quad (7.2.7)$$

From (7.2.5), the estimate of θ is

$$\hat{\theta}_{MLE} = \frac{n-r_1-r_2-c^*}{\sum_{i=1}^{n-r_1-r_2-c^*} log\left(\frac{1+x_{(i)}}{2}\right) + c^* \, log\left(\frac{1+x_{(n-r_1-r_2-c^*)}}{2}\right)} \tag{7.2.8}$$

7.3. Asymptotic distribution of MLE

For density function $h(x;\underline{\theta})$ where $\underline{\theta} = (p_1, p_2, \theta)$ given in (7.1.2),

$$\frac{\partial \, log \, h(x;\underline{\theta})}{\partial p_1} = \begin{cases} \dfrac{1}{p_1}, & x = 0 \\ 0, & x = 1 \\ -\dfrac{1}{(1-p_1-p_2)}, & x > 1 \end{cases}$$

$$\frac{\partial \, log \, h(x;\underline{\theta})}{\partial p_2} = \begin{cases} 0, & x = 0 \\ \dfrac{1}{p_2}, & x = 1 \\ -\dfrac{1}{(1-p_1-p_2)}, & x > 1 \end{cases}$$

and

$$\frac{\partial \, log \, h(x;\underline{\theta})}{\partial \theta} = \begin{cases} 0, & x = 0 \\ 0, & x = 1 \\ \dfrac{1}{\theta} - [log(1+x) - log \, 2], & x > 1 \end{cases}$$

One can verify that $E\left(\frac{\partial \, log \, h(x;\underline{\theta})}{\partial p_1}\right) = 0$, $E\left(\frac{\partial \, log \, h(x;\underline{\theta})}{\partial p_2}\right) = 0$ and $E\left(\frac{\partial \, log \, h(x;\underline{\theta})}{\partial \theta}\right) = 0$.

Also,

$$\frac{\partial^2 \, log \, h(x;\underline{\theta})}{\partial p_1^2} = \begin{cases} -\dfrac{1}{p_1^2}, & x = 0 \\ 0, & x = 1 \\ -\dfrac{1}{(1-p_1-p_2)^2}, & x > 1 \end{cases}$$

$$\frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_2^2} = \begin{cases} 0, & x = 0 \\ -\frac{1}{p_2^2}\,, & x = 1 \\ -\frac{1}{(1-p_1-p_2)^2}, & x > 1 \end{cases}$$

$$\frac{\partial^2 \log h(x;\underline{\theta})}{\partial \theta^2} = \begin{cases} 0, & x = 0 \\ 0, & x = 1 \\ -\frac{1}{\theta^2}, & x > 1 \end{cases}$$

$$\frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_1\, \partial p_2} = \begin{cases} 0, & x = 0 \\ 0, & x = 1 \\ -\frac{1}{(1-p_1-p_2)^2}, & x > 1 \end{cases}$$

$$\frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_1\, \partial \theta} = 0 \; = \frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_2\, \partial \theta} \quad \forall\, x$$

Hence, the Fisher information is

$$I_{p_1p_1} = E\left(-\frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_1^2}\right) = \frac{1-p_2}{p_1(1-p_1-p_2)}$$

$$I_{p_2p_2} = E\left(-\frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_2^2}\right) = \frac{1-p_1}{p_2(1-p_1-p_2)}$$

$$I_{\theta\theta} = E\left(-\frac{\partial^2 \log h(x;\underline{\theta})}{\partial \theta^2}\right) = \frac{1-p_1-p_2}{\theta^2}$$

$$I_{p_1p_2} = E\left(-\frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_1\, \partial p_2}\right) = \frac{1}{1-p_1-p_2}$$

$$I_{p_1\theta} = E\left(-\frac{\partial^2 \log h(x;\underline{\theta})}{\partial p_1\,\partial\theta}\right) = 0$$

and

$$I_{p_2\theta} = E\left(-\frac{\partial^2 \log h(x;\,\underline{\theta})}{\partial p_2\,\partial\theta}\right) = 0$$

The Fisher information matrix $I_h(\underline{\theta})$ is given by

$$I_h(\underline{\theta}) = \begin{bmatrix} \frac{1-p_2}{p_1(1-p_1-p_2)} & \frac{1}{1-p_1-p_2} & 0 \\ \frac{1}{1-p_1-p_2} & \frac{1-p_1}{p_2(1-p_1-p_2)} & 0 \\ 0 & 0 & \frac{1-p_1-p_2}{\theta^2} \end{bmatrix} \tag{7.3.1}$$

The determinant of $I_h(\underline{\theta})$ is given by $\frac{1}{p_1p_2\theta^2}$ and hence, the inverse matrix $I_h^{-1}(\underline{\theta})$ is given by

$$I_h^{-1}(\underline{\theta}) = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & 0 \\ -p_1p_2 & p_2(1-p_2) & 0 \\ 0 & 0 & \frac{\theta^2}{1-p_1-p_2} \end{bmatrix} \tag{7.3.2}$$

Using the standard result of MLE, we have $\left(\underline{\hat{\theta}}_{MLE}\right)' \sim AN^{(3)}\left[(\underline{\theta})', \frac{1}{n} I_h^{-1}(\underline{\theta})\right]$. Using the estimated variances, one can also propose large sample tests for p_1, p_2 and θ. The approximate $(1-\alpha)\%$ confidence interval for p_1, p_2, and θ are respectively given by

$$\hat{p}_{1MLE} \pm z\alpha/_2 \sqrt{\frac{\hat{p}_{1MLE}(1-\hat{p}_{1MLE})}{n}} \tag{7.3.3}$$

$$\hat{p}_{2MLE} \pm z\alpha/_2 \sqrt{\frac{\hat{p}_{2MLE}(1-\hat{p}_{2MLE})}{n}} \tag{7.3.4}$$

and

$$\hat{\theta}_{MLE} \pm z\alpha/_{2} \sqrt{\frac{\hat{\theta}^2_{MLE}}{(n-c^*)\,(1-\hat{p}_{1MLE}-\hat{p}_{2MLE})}} \tag{7.3.5}$$

7.4. Maximum Product Spacings Estimation

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 $X_1, X_2, \dots, X_{n-r_1-r_2}$. With same setup of 'Type II censored sample' used in 7.2, let $Y_{(1)}, Y_{(2)}, \dots, Y_{(m)}$ denote the m distinct values (including the observations 0 and 1). Referring equation (1.5.8) and Singh et al. (2016), the product spacing (in presence of ties) with Type II censoring sample can be written as

$$S\left(\underline{\theta}|\underline{y}\right) = \prod_{j=1}^{m}\left[\frac{H\left(y_{(j)};\ \underline{\theta}\right) - H\left(y_{(j-1)};\ \underline{\theta}\right)}{R_j}\right]^{R_j}\left[1 - H\left(y_{(m)};\ \underline{\theta}\right)\right]^{c^*}$$

$$= \left(\frac{p_1}{r_1}\right)^{r_1}\left(\frac{p_2}{r_2}\right)^{r_2}\left[(1-p_1-p_1)\left(\frac{1-e^{-\theta\, log\left(\frac{1+\,y_{(3)}}{2}\right)}}{R_3}\right)\right]^{R_3}$$

$$\prod_{j=4}^{m}\left[(1-p_1-p_1)\left(\frac{e^{-\theta\, log\left(\frac{1+\,y_{(j-1)}}{2}\right)} - e^{-\theta\, log\left(\frac{1+\,y_{(j)}}{2}\right)}}{R_j}\right)\right]^{R_j}$$

$$\left[(1-p_1-p_1)\left(\frac{e^{-\theta\log\left(\frac{1+\,y_{(m)}}{2}\right)}}{R_3}\right)\right]^{c^*} \tag{7.4.1}$$

Let the maximum product spacings estimator of parameter $\underline{\theta} = (p_1, p_2, \theta)$ is $\underline{\hat{\theta}}_{MPSE} = (\hat{p}_{1MPSE}, \hat{p}_{2MPSE}, \hat{\theta}_{MPSE})$, say. We now investigate the following four possible cases of product spacings estimator.

Case (*i*). $r_1 = 0$, that is $r_2 = n$. The product spacings reduces to $S\left(\underline{\theta}|\underline{y}\right) = \left(\frac{p_2}{r_2}\right)^{r_2}$. Obviously, this is maximum when $p_2 = 1$. This corresponds to the maximum product spacings estimator $\hat{p}_{2MPSE} = 1$. Since $S\left(\underline{\theta}|\underline{y}\right) = \left(\frac{p_2}{r_2}\right)^{r_2}$ is free from the other parameters, the maximum product spacings estimator of other parameters does not exist.

Case (*ii*). $r_2 = 0$, that is $r_1 = n$. The product spacings reduces to $S\left(\underline{\theta}|\underline{y}\right) = \left(\frac{p_1}{r_1}\right)^{r_1}$. Obviously, this is maximum when $p_1 = 1$. This corresponds to the maximum product spacings estimator $\hat{p}_{1MPSE} = 1$. Since $S\left(\underline{\theta}|\underline{y}\right) = \left(\frac{p_1}{r_1}\right)^{r_1}$ is free from the other parameters, the maximum product spacings estimator of other parameters does not exist.

Case (*iii*). $r_1 < n, r_2 < n$ but $r_1 + r_2 = n$. The product spacings reduces to $S\left(\underline{\theta}|\underline{y}\right) = \left(\frac{p_1}{r_1}\right)^{r_1}\left(\frac{p_2}{r_2}\right)^{r_2}$. Here, $p_1 + p_2 < n$ and the product spacing function $S\left(\underline{\theta}|\underline{y}\right) < \left(\frac{y_1}{(m-c^*)r_1}\right)^{r_1}\left(\frac{y_2}{m-c^*r_2}\right)^{r_2}$. So, $\hat{p}_{1MPSE} = \frac{r_1}{n}$ and $\hat{p}_{2MPSE} = \frac{r_2}{n}$, the maximum product spacings estimator of other parameters does not exist.

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

$$log\, S\left(\underline{\theta}|\underline{y}\right) = r_1[log\, p_1 - log\, r_1] + r_2[log\, p_2 - log\, r_2] + R_3\, log(1 - p_1 - p_1)$$

$$+R_3\left[log\left(\frac{1 - e^{-\theta\, log\left(\frac{1+ y_{(3)}}{2}\right)}}{R_3}\right) - \, log\, R_3\right]$$

$$+\sum_{j=1}^{m} R_j\left[log(1 - p_1 - p_1) + \, log\left(e^{-\theta\, log\left(\frac{1+ y_{(j-1)}}{2}\right)} - e^{-\theta\, log\left(\frac{1+ y_{(j)}}{2}\right)}\right) - log\, R_j\right]$$

$$+c^*\left[log(1 - p_1 - p_1) - \theta\, log\left(\frac{1+ y_{(m)}}{2}\right)\right] \qquad (7.4.2)$$

The maximum product spacing estimator of parameter $\underline{\theta}$, is obtained by solving the following likelihood equations:

$$\frac{\partial \log S(\underline{\theta}|\underline{y})}{\partial p_1} = \frac{r_1}{p_1} - \frac{R_3}{1-p_1-p_2} - \sum_{j=1}^{m+1} \frac{R_j}{1-p_1-p_2} - \frac{c^*}{1-p_1-p_2} = 0 \qquad (7.4.3)$$

$$\frac{\partial \log S(\underline{\theta}|\underline{y})}{\partial p_2} = \frac{r_2}{p_2} - \frac{R_3}{1-p_1-p_2} - \sum_{j=1}^{m+1} \frac{R_j}{1-p_1-p_2} - \frac{c^*}{1-p_1-p_2} = 0 \qquad (7.4.4)$$

and

$$\frac{\partial \log S(\underline{\theta}|\underline{y})}{\partial \theta} = \frac{R_3\, e^{-\theta \log\left(\frac{1+y_{(3)}}{2}\right)} \log\left(\frac{1+y_{(3)}}{2}\right)}{\left[1 - e^{-\theta \log\left(\frac{1+y_{(3)}}{2}\right)}\right]}$$

$$+\sum_{j=1}^{m} \frac{R_j \left[e^{-\theta \log\left(\frac{1+y_{(j)}}{2}\right)} \log\left(\frac{1+\ y_{(j)}}{2}\right) - \ e^{-\theta \log\left(\frac{1+y_{(j-1)}}{2}\right)} \log\left(\frac{1+\ y_{(j-1)}}{2}\right)\right]}{e^{-\theta \log\left(\frac{1+y_{(j-1)}}{2}\right)} - e^{-\theta \log\left(\frac{1+y_{(j)}}{2}\right)}}$$

$$-c^* \log\left(\frac{1+\ y_{(m)}}{2}\right) = 0$$

That is,

$$\frac{R_3 \log\left(\frac{1+y_{(3)}}{2}\right)}{\left[\left(\frac{1+y_{(3)}}{2}\right)^{\theta} - 1\right]} + \sum_{j=1}^{m} \frac{R_j\left[\left(\frac{1+y_{(j-1)}}{2}\right)^{\theta} \log\left(\frac{1+y_{(j)}}{2}\right) - \left(\frac{1+y_{(j)}}{2}\right)^{\theta} \log\left(\frac{1+y_{(j-1)}}{2}\right)\right]}{\left(\frac{1+y_{(j)}}{2}\right)^{\theta} - \left(\frac{1+y_{(j-1)}}{2}\right)^{\theta}} - \ c^* \log\left(\frac{1+y_{(m)}}{2}\right) = 0$$

(7.4.5)

Solving (7.4.3) and (7.4.4) simultaneously, we get

$\hat{p}_{1MPSE} = \frac{r_1}{n}$ and $\hat{p}_{2MPSE} = \frac{r_2}{n}$.

Equation (7.4.5) cannot solved analytically, therefor for getting solution to Equation (7.4.5) we use Newton-Rapson method iteratively to get, $\hat{\theta}_{MPSE}$.

7.5. Uniformly Minimum Variance Unbiased Estimation

Here, we propose the distributional properties of complete sufficient statistic and study UMVU estimation for parameters and various parametric functions of the model in subsequent section. Referring (1.5.24) the model in (7.1.2) can be expressed as

$$h(x;\underline{\theta}) = \left(\frac{1}{(1+x)}\right)^{(1-I_1(x)-I_2(x))} \left(\frac{p_1}{\theta(1-p_1-p_2)}\right)^{I_1(x)} \left(\frac{p_2}{\theta(1-p_1-p_2)}\right)^{I_2(x)}$$

$$\left(e^{-\theta}\right)^{\left\{log\left(\frac{1+x}{2}\right)(1-I_1(x)-I_2(x))\right\}} \left(\frac{1}{\theta(1-p_1-p_2)}\right)^{-1}$$

$$= (a(x))^{(1-C_1(x)-C_2(x))} \frac{\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{C_i(x)}}{g(\underline{\theta})} \qquad (7.5.1)$$

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 $C_3(X) = log\left(\frac{1+X}{2}\right)\left(1 - I_1(X) - I_2(X)\right)$. 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 θ_i, i=1,2 and 3. Here $g(\underline{\theta}) = \int_{x>1}(a(x))^{(1-C_1(x)-C_2(x))}\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{C_i(x)} dx$. The density in (7.5.1) 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 $(\eta_1,\ \eta_2,\ \eta_3) = \left(log\left(\frac{p_1}{\theta(1-p_1-p_2)}\right), log\left(\frac{p_2}{\theta(1-p_1-p_2)}\right), log\left(e^{-\theta}\right)\right)$ generated by underlying indexing parameters $\underline{\theta} = (p_1, p_2, \theta)$. Hence $C(X) = \left(C_1(X), C_2(X), C_3(X)\right) = \left(I_1(X),\ I_2(X), \log\left(\frac{1+X}{2}\right)\left(1 - I_1(X) - I_2(X)\right)\right)$ is jointly complete sufficient for $\underline{\theta} = (p_1, p_2, \theta)$. Now solving the equation (1.5.24), we get

$$\begin{bmatrix} E\left(C_1(x)\right) \\ E\left(C_2(x)\right) \\ E\left(C_3(x)\right) \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \frac{(1-p_1-p_2)}{\theta} \end{bmatrix} \qquad (7.5.2)$$

and the variance-covariance matrix Σ given equation (1.5.31) is obtained as

$$\Sigma = \left[\sigma_{ij}\right]_{3\times 3} = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -\theta\, p_1(1-p_1-p_2) \\ -p_1p_2 & p_2(1-p_2) & -\theta\, p_2(1-p_1-p_2) \\ -\frac{p_1(1-p_1-p_2)}{\theta} & -\frac{p_2(1-p_1-p_2)}{\theta} & \frac{\left[1-(p_1+p_2)^2\right]}{\theta^2} \end{bmatrix} \qquad (7.5.3)$$

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

We now propose some uniformly minimum variance unbiased estimators for parameters and some parametric function of the model (7.5.1) in various subsections below.

7.5.1. UMVU Estimation of parameters

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

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

and

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

$$i = 2, \dots, n-r_1-r_2-c^* \qquad (7.5.4)$$

It can be seen that

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

and

$$|J| = \frac{c^*! \prod_{i=1}^{n-r_1-r_2-c^*}(1+x_{(i)})}{(n-r_1-r_2)!} \qquad (7.5.5)$$

Using (7.5.4) and (7.5.5),

$$h\left(\underline{y};\ \underline{\theta}\right) = {p_1}^{r_1}{p_2}^{r_2}\ (1-p_1-p_2)^{(n-r_1-r_2)}\ \theta^{(n-r_1-r_2-c^*)}\ \ e^{-\theta \sum_{i=1}^{n-r_1-r_2-c^*} y_i} \qquad (7.5.6)$$

$$= \frac{\left(\frac{p_1}{\theta(1-p_1-p_2)}\right)^{z_1}\left(\frac{p_2}{\theta(1-p_1-p_2)}\right)^{z_2}\left(e^{-\theta}\right)^{z_3}(1-p_1-p_2)^{c^*}}{\left(\frac{1}{\theta(1-p_1-p_2)}\right)^{n-c^*}}$$

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,

$$h\left(\underline{y};\underline{\theta}\right) = \frac{n!}{r_1!\ \ r_2!\ (n-r_1-r_2)!}\,{p_1}^{r_1}{p_2}^{r_2}(1-p_1-p_2)^{\,(n-r_1-r_2)}$$

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

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

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

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

which belongs to the one-parameter exponential family. Hence $Z_3|Z_1, Z_2$ is complete sufficient for θ 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 θ and is also a complete family of distribution. Therefore, $Z = (Z_1, Z_2, Z_3)$ is complete sufficient for $\underline{\theta} = (p_1, p_2, \theta)$. The Joint distribution of $Z = (Z_1, Z_2, Z_3)$ is

$$h_Z\left(z; \underline{\theta}\right) = \frac{n!}{r_1!\ \ r_2!\ (n - r_1 - r_2)!} {p_1}^{r_1} {p_2}^{r_2} (1 - p_1 - p_2)^{\ (n-r_1-r_2)}$$

$$\frac{{z_3}^{(n-r_1-r_2-c^*-1)}}{\Gamma n - r_1 - r_2 - c^*}\ \theta^{n-r_1-r_2-c^*}\ e^{-\theta\, z_3},$$

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

$$= B(z_1, z_2, z_3, c^*, n)\ \frac{\prod_{i=1}^{3} \left(h_i(\underline{\theta})\right)^{z_i}}{g(\underline{\theta})^{n-c^*}} (1 - p_1 - p_2)^{c^*} \tag{7.5.7}$$

where

$$B(z_1, z_2, z_3, c^*, n) = \begin{cases} \dfrac{n!}{r_1!\ \ r_2!\ \ (n-r_1-r_2)!} \dfrac{{z_3}^{(n-r_1-r_2-c^*-1)}}{\Gamma n - r_1 - r_2 - c^*}, & z_3 > 0;\ r_1 + r_2 - 1 < n - c^* \\ 1, & z_3 = 0;\ r_1 = 0\ \ or\ r_2 = 0 \end{cases} \qquad (7.5.8)$$

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

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

Using (7.5.2),

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

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

and

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

which in turn give UMVUE's of p_1, p_2 and θ respectively as

$$\hat{p}_{1UMVUE} = \frac{Z_1}{n-c^*} = \frac{r_1}{n-c^*} \qquad (7.5.9)$$

$$\hat{p}_{2UMVUE} = \frac{Z_2}{n-c^*} = \frac{r_2}{n-c^*} \qquad (7.5.10)$$

and

$$\hat{\theta}_{UMVUE} = \frac{(n-c^*)(1-\hat{p}_1-\hat{p}_2)}{Z_3} \qquad (7.5.11)$$

7.5.2. UMVU Estimation of parametric functions

Let $X_1, X_2, \dots, X_{n-c^*}$ be Type II censored random sample from (7.5.1), then there exists an UMVUE of $\Phi(\underline{\theta})$ if and only if $\Phi(\underline{\theta})\left[g(\underline{\theta})\right]^{n-c^*}$ can be expressed in the form

$$\frac{\Phi(\underline{\theta})\left[g(\underline{\theta})\right]^{n-c^*}}{(1-p_1-p_2)^{c^*}}$$

$$= \int_{z_1 \in T(n-c^*)} \int_{z_2 \in T(n-c^*)} \int_{z_3 \in T(n-c^*)} \alpha(z_1, z_2, z_3, c^*, n) \prod_{i=1}^{3} \left(h_i(\underline{\theta})\right)^{z_i} dz_1\, dz_2\, dz_3$$

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

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

The following results are now obvious.

Result 7.1. *The UMVUE of* $\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{k_i} = \left(\frac{1}{\theta(1-p_1-p_2)}\right)^{k_1+k_2} {p_1}^{k_1}\, {p_2}^{k_2} e^{-\theta k_3}$ *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 $k_1 \le r_1;\ k_2 \le r_2;\ k_3 \le z_3;\ k_1 + k_2 \le n - r_1 - r_2 - c^*;\ r_1 + r_2 - 1 < n - c^*$, *and* $(r)_k = \frac{r!}{(r-k)!}$, $[r]_k = \frac{\Gamma r + k}{\Gamma r}$.

Corollary 7.1. If $k_1 \neq 0,\ k_2 = 0$ and $k_3 = 0$, then UMVUE of $\left(h_1(\underline{\theta})\right)^{k_1} = \left(\frac{p_1}{\theta(1-p_1-p_2)}\right)^{k_1}$ 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 7.2. If $k_1 = 0, k_2 \neq 0$ and $k_3 = 0$, then UMVUE of $\left(h_2(\underline{\theta})\right)^{k_2} = \left(\frac{p_2}{\theta(1-p_1-p_2)}\right)^{k_2}$ 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 7.3. If $k_1 = 0, k_2 = 0$ and $k_3 \neq 0$, then UMVUE of $\left(h_3(\underline{\theta})\right)^{k_3} = e^{-\theta k_3}$ 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 7.2. *The UMVUE of the variance of* $H_{k_1,k_2,k_3}(Z_1, Z_2, Z_3, c^*, n)$*, is given by*

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

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

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

Corollary 7.4. The UMVUE of the variance of $H_{k_1}(Z_1,Z_2,Z_3,c^*,n)$, is given by

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

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

Corollary 7.5. The UMVUE of the variance of $H_{k_2}(Z_1,Z_2,Z_3,c^*,n)$, is given by

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

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

Corollary 7.6. The UMVUE of the variance of $H_{k_3}(Z_1,Z_2,Z_3,c^*,n)$, is given by

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

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

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

Result 7.3. *The UMVUE of* $\left[g(\underline{\theta})\right]^k = \left(\frac{1}{\theta(1-p_1-p_2)}\right)^k, k \neq 0$ *as per the model given in (7.5.1) is*

$$G_k(z_1,z_2,z_3,c^*,n) = \frac{B(z_1,z_2,z_3,c^*,n+k)}{B(z_1,z_2,z_3,c^*,n)}$$

$$= \frac{[n+1]_k \quad {z_3}^k}{[n-r_1-r_2+1]_k \quad [n-r_1-r_2-c^*]_k},$$

$$k \le n-r_1-r_2-c^*;\ r_1+r_2-1 < n-c^*$$

Result 7.4. *The UMVUE of the variance of* $G_k(Z_1,Z_2,Z_3,c^*,n)$ *is given by*

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

$$=\left[\frac{[n+1]_k \quad {z_3}^k}{[n-r_1-r_2+1]_k \quad [n-r_1-r_2-c^*]_k}\right]^2 - \frac{[n+1]_{2k} \quad {z_3}^{2k}}{[n-r_1-r_2+1]_{2k} \quad [n-r_1-r_2-c^*]_{2k}},$$

$$2k \le n-r_1-r_2-c^*;\ r_1+r_2-1 < n-c^*$$

Result 7.5. *For fixed x, the UMVUE of the density given in (7.5.1) is*

$$\phi_x(z_1,z_2,z_3,c^*,n) = a(x)\,\frac{B(z_1-C_1(x), z_2-C_2(x), z_3-C_3(x),c^*,n-1)}{B(z_1,z_2,z_3,c^*,n)}$$

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

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

Result 7.6. *The UMVUE of the variance of* $\phi_x(Z_1, Z_2, Z_3, c^*, n)$ *is given by*

$$\widehat{var}[\phi_x(z_1, z_2, z_3, c^*, n)] = \phi_x^2(z_1, z_2, z_3, c^*, n)$$
$$-\phi_x(z_1, z_2, z_3, c^*, n)\ \phi_x(z_1 - C_1(x),\ z_2 - C_2(x),\ z_3 - C_3(x), c^*, n-1)$$

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

Result 7.7. *For a fixed* $z = (z_1, z_2, z_3, c^*, n)$, *the UMVUE of the survival function* $S(t) = P(X > t),\ t \geq 0$ *is obtained as*

$$\hat{S}(t) = \left(\frac{(r_1)_{I_1(t)}(r_2)_{I_2(t)}(n - r_1 - r_2)_{(1-I_1(t)-I_2(t))}(n - r_1 - r_2 - c^* - 1)_{(1-I_1(t)-I_2(t))}}{n\left[(n - r_1 - r_2 - c^*) - \left(1 - I_1(t) - I_2(t)\right)\right]}\right)$$
$$(n - r_1 - r_2 - c^* - 1)_{(1-I_1(t)-I_2(t))}\left(Z_3 - \log\left(\frac{1+t}{2}\right)\left(1 - I_1(t) - I_2(t)\right)\right)^{(I_1(t)+I_2(t))}$$
$$\left(1 - \frac{\log\left(\frac{1+t}{2}\right)\left(1 - I_1(t) - I_2(t)\right)}{Z_3}\right)^{(n-r_1-r_2-c^*-1)},$$
$$Z_3 > \log\left(\frac{1+t}{2}\right);\ r_1 + r_2 - 1 < n - c^*$$

Result 7.8. *For the fixed* $z = (z_1, z_2, z_3, c^*, n)$, *the UMVUE of the* $var\left(\hat{S}(t)\right)$, *is obtained as*

$$\widehat{var}\left(\hat{S}(t)\right) = \left[\hat{S}(t)\right]^2 - \frac{1}{n(n-1)}\left(1 - \frac{2\,log\left(\frac{1+t}{2}\right)\left(1 - I_1(t) - I_2(t)\right)}{z_3}\right)^{(n-r1-r2-c^*-1)}$$

$$\left(\frac{(r_1)_{2I_1(t)}(r_2)_{2I_2(t)}(n-r_1-r_2)_{2(1-I_1(t)-I_2(t))}\;(n-r_1-r_2-c-1^*)_{2(1-I_1(t)-I_2(t))}}{\left[(n-r_1-r_2-c^*)-2\left(1-I_1(t)-I_2(t)\right)\right]\left[(n-r_1-r_2-c^*+1)-2\left(1-I_1(t)-I_2(t)\right)\right]}\right)$$

$$\left(z_3 - 2\,log\left(\tfrac{1+t}{2}\right)(1 - I_1(t) - I_2(t)\right)^{2\left(I_1(t)+I_2(t)\right)}, Z_3 > 2\,log\left(\tfrac{1+t}{2}\right); r_1 + r_2 - 1 < n - c^*$$

7.6. Illustrative examples

We consider four dataset from an appendix to illustrate the proposed model. 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:

Example 7.1. Consider Dataset A.8, where Table A.2. shows the loss ratios (yearly data) for earthquake insurance in California from 1971 through 1993. 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.

Example 7.2. We consider dataset A.6, is on child's age at death from the woman's questionnaire of NFHS-3 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. 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.

Example 7.3. We use dataset A.9. on 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 having percentage burnt area less than 1 sq. Km. conveniently considered here as observation 1

Example 7.4. This example based on dataset A.10, about the amount of snowfall in all 50 states of US. It is observed that there are three states having decade's average amount of snowfall zero and for four states having decade’s average amount of snowfall less than 1 inches (coded as observation1).

For all the datasets above we have calculated parameter estimates, goodness-of-fit criteria values, goodness-of-fit statistics and corresponding p-values (see Table 7.1 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 7.1. *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)
Snowfall	$\hat{\beta}$=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)

The plot of pdf, $h(x)$ and survival function, $S(x)$ for all four datasets under study, is displayed in Figure 7.1 and Figure 7.2 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 7.2. Whereas Table 7.3 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.

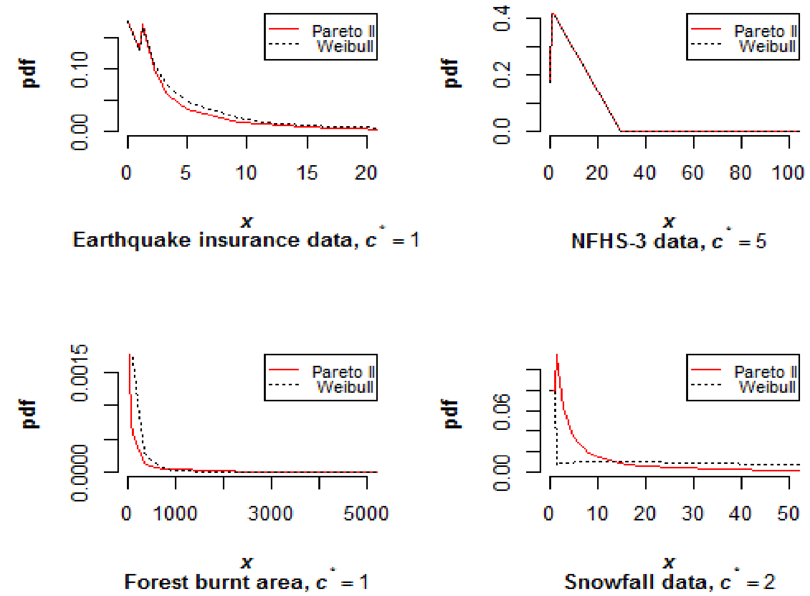

Fig. 7.1 *Density plot to various data sets censored at value* c^*

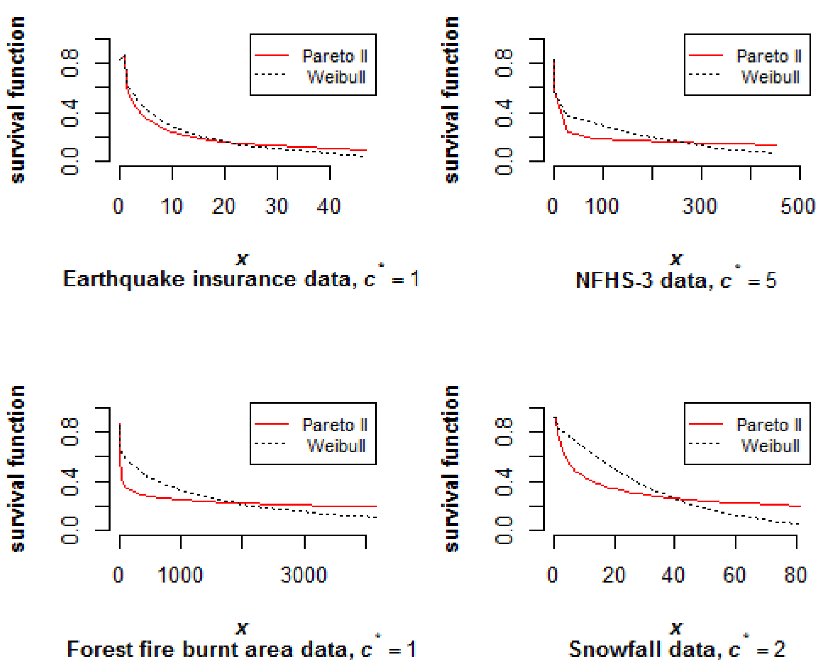

Fig. 7.2 *Survival function plot to various data sets censored at value* c^*

Table 7.2. *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
$\hat{p}_{1MLE}$	0.17391 (0.07904)	0.17241 (0.04050)	0.16667 (0.06212)	0.08000 (0.03837)
$\hat{p}_{2MLE}$	0.13043 (0.07022)	0.42529 (0.05300)	0.13889 (0.05764)	0.08000 (0.03837)
$\hat{\theta}_{MLE}$	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 θ	(0.30648, 0.92191)	(0.12871, 0.26206)	(0.10041, 0.23292)	(0.26651, 0.50449)
$\hat{p}_{1MPSE}$	0.17391 (0.07904)	0.17241 (0.04050)	0.16667 (0.06212)	0.08000 (0.03837)
$\hat{p}_{2MPSE}$	0.13043 (0.07022)	0.42529 (0.05300)	0.13889 (0.05764)	0.08000 (0.03837)
$\hat{\theta}_{MPSE}$	0.32998 (0.02744)	0.11016 (0.00140)	0.12313 (0.00421)	0.17832 (0.00503)
$\hat{p}_{1UMVUE}$	0.18182 (0.08223)	0.18293 (0.04269)	0.17143 (0.06370)	0.08333 (0.03989)
$\hat{p}_{2UMVUE}$	0.13636 (0.07317)	0.45122 (0.05495)	0.14286 (0.05915)	0.08333 (0.03989)
$\hat{\theta}_{UMVUE}$	0.61420 (0.13812)	0.19539 (0.02791)	0.16667 (0.02967)	0.38550 (0.05643)
95% CI of p_1	(0.02065, 0.34299)	(0.09925, 0.26660)	(0.04657, 0.29629)	(0.00515, 0.16152)
95% CI of p_2	(0.00000, 0.27976)	(0.34351, 0.55892)	(0.02693, 0.25879)	(0.00515, 0.16152)
95% CI of θ	(0.34348 0.88492)	(0.14069, 0.25008)	(0.10850, 0.22483)	(0.27489, 0.49610)
$\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{k_i} = \left(\frac{\theta}{1-p_1-p_2}\right)^2 p_1 p_2 e^{-\frac{1}{\theta}}$ $k_1 = 1, k_2 = 1, k_3 = 1$	0.04992 (0.04299)	8.62541 (4.73007)	1.240801 (0.89610)	3.07667 (0.02775)
$h_1(\underline{\theta}) = \frac{\theta\, p_1}{1-p_1-p_2}, k_1 = 1, k_2 = 0, k_3 = 0$	0.38309 (0.22204)	2.13250 (0.74236)	1.38463 (0.66351)	0.24131 (0.12880)
$h_2(\underline{\theta}) = \frac{\theta\, p_2}{1-p_1-p_2}, k_1 = 0, k_2 = 1, k_3 = 0$	0.28732 (0.18391)	5.26018 (1.52329)	1.15386 (0.58886)	0.24131 (0.12880)
$h_3(\underline{\theta}) = e^{-\frac{1}{\theta}}, k_1 = 0, k_2 = 0, k_3 = 1$	0.55693 (0.08844)	0.82738 (0.00109)	0.85191 (0.02856)	0.68545 (0.04162)
$g(\underline{\theta}) = \frac{\theta}{1-p_1-p_2}, k = 1$	2.29856 (0.64068)	12.51070 (2.73276)	8.53853 (1.92013)	3.07667 (0.51361)

Table 7.3. *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	ϕ_{10} = 0.01415 (0.00185)	ϕ_{10} = 0.02091 (0.00295)	ϕ_{100} = 0.00030 (5.043e-05)	ϕ_{100} = 0.00110 (1.968e-04)	ϕ_{650} = 6.912e-05 (7.524e-06)	ϕ_{650} = 0.00022 (3.235e-05)	ϕ_{25} = 0.00469 (0.00028)	ϕ_{25} = 0.01383 (0.00114)
	ϕ_{15} = 0.00784 (0.00112)	ϕ_{15} = 0.01317 (0.00113)	ϕ_{500} = 5.473e-05 (7.032e-06)	ϕ_{500} = 0.00030 (6.096e-05)	ϕ_{1350} = 2.966e-05 (3.199e-06)	ϕ_{1350} = 0.00012 (8.101e-06)	ϕ_{50} = 0.00186 (0.00013)	ϕ_{50} = 0.00676 (0.00068)
	ϕ_{40} = 0.00175 (0.00046)	ϕ_{40} = 0.00273 (0.00046)	ϕ_{1000} = 2.387e-05 (3.185e-06)	ϕ_{1000} = 1.851e-05 (1.282e-05)	ϕ_{2500} = 1.451e-05 (1.624e-06)	ϕ_{2500} = 6.421e-05 (3.185e-06)	ϕ_{100} = 0.00072 (7.193e-05)	ϕ_{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)
	$\hat{S}_{15}$ = 0.20013 (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)