Parameter Estimation of Inverse Rayleigh Distribution under Competing Risk Model for Masked data

M.S. Panwar, Bapat Akanshya Sudhir, Rashmi Bundel and Sanjeev K. Tomer*1
Department of Statistics, Central University of Rajasthan, Kishangarh
*Department of Statistics, Banaras Hindu University, Varanasi
1E-mail: sktomer73@gmail.com

ABSTRACT
This paper tries to derive maximum likelihood estimators (MLEs) for the parameters of the inverse Rayleigh distribution (IRD) when the observed data is masked. MLEs, asymptotic confidence intervals (ACIs) and boot-p confidence intervals (boot-p CIs) for the lifetime parameters have been discussed. The simulation illustrations provided that as the sample size increases the estimated value approaches to the true value, and the mean square error decreases with the increase in sample size, and mean square error increases with increase in level of masking, the ACIs are always symmetric and the boot-p CIs approaches to symmetry as the sample size increases whereas the mean life time due to the local spread of the disease is less than that due to the metastasis spread in case of real data set.

Keywords: Masked data, asymptotic confidence intervals, inverse Rayleigh distribution.

INTRODUCTION
A random variable (rv) $X$ is said to follow inverse Rayleigh distribution (IRD) if its probability density function (pdf) is given by

$$f(x) = \frac{2\theta}{x^3} \exp\left(-\frac{\theta}{x}\right), \quad x > 0, \theta > 0. \quad (1)$$

The reliability function of IRD with pdf (1) is given as follows

$$F(x) = P(X > t) = 1 - \exp\left(-\frac{\theta}{x}\right). \quad (2)$$

Trayer (1964) introduced the IRD and discussed its application in reliability theory. Voda (1972) discussed some properties of the maximum likelihood estimator (MLE) of $\theta$ and the property that the distribution of lifetimes of several types of experimental units can be approximated by the IRD. Mukherjee and Saran (1984) noticed that the hazard rate of IRD is increasing till time $t < 1.069543 \sqrt{\theta}$, then it decreases and then becomes stable after some time. Various methods of estimation for IRD are developed by Gharraph (1993) and Mukherjee and Maiti (1996). A number of researchers have been discussed the IRD (Soliman et al. 2010, Dey 2012).

The theory of competing risk studies the subjects that are exposed to more than one cause of failure, and failure due to one cause excludes the chance of failure due to other causes. An investigator may use this theory to study the reliability characteristics of the components of series system. The failure of one component causes the breakdown of the system, the data can be collected in a pair of the lifetime of the system and the component which causes the failure. Crowder (2001) has presented appropriate competing risk methodology for analysis of such data.

In reliability analysis/competing risk analysis, with multiple causes of failures, there may be cases for which the exact cause of failure may be unknown. It makes the data incomplete. The incompleteness may be due to deficiency of knowledge about the exact cause of failure, identifiability of cause due to time consuming process etc. Such a data is termed as “masked” in literature. Sarhan (2003) discussed the estimations of parameters in Pareto reliability model in the presence of masked data. For some more citations, one may refer to Basu et al. (2003) and Lin et al. (1993).

In this context, this paper is to consider the competing risk analysis of masked data. Assuming the life times of components to be IRD, we obtain MLEs, asymptotic confidence intervals (ACIs) and boot-p confidence intervals (boot-p CIs) for the lifetime parameters of individual components using system lifetime data. We derive expressions to estimate the MLEs of parameter and to compute ACIs and boot-p CIs for parameters.
This paper also attempts the simulation and application of proposed methods using real data sets.

**MODELS**

Some of the used notations and assumptions are given below.

- $n$ 
  #systems under observation
- $k$ 
  #components in each system
- $n_1$ 
  #systems failed due to component 1
- $n_2$ 
  #systems failed due to component 2
- $n_{12}$ 
  #systems with masked causes of failure
- $x_{ij}$ 
  lifetime of $j^{th}$ component in the $i^{th}$ system, $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$.
- $f(x, \theta_j)$ 
  pdf of lifetime of $j^{th}$ component
- $F(x, \theta_j)$ 
  reliability function of $j^{th}$ component
- $z_\alpha$ 
  Upper 100 ($\alpha$)th percentile of standard normal variate
- $\hat{\theta}$ 
  MLE of $\theta$
- $\Theta$ 
  $(\theta_1, \theta_2)$

**Assumptions**

- Each system has 2-independent components connected in series.
- $n$ identical systems are put on the life test.
- Systems, subsystems and components have only two statuses: failure and survival.
- The RV $X_i = \min \{X_{ij}, \ldots, X_{ik}\}$ $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, k$ are independent with $(X_{ij}, X_{j2}, \ldots, X_{jn})$ being identical and having inverse Rayleigh distribution with pdf $f(x, \theta_j)$ and reliability function $F(x, \theta_j)$ of the forms given, respectively, by (1) and (2).
- The parameter $\theta_j$ is unknown for $j = 1, 2, \ldots, k$.
- The observable quantities for the $i^{th}$ system, on the life test, are its life time $x_i$ and a set $S_i$ of system's components that may cause it to fail.
- Masking is $s$-independent of the cause of failure.

**Maximum Likelihood Estimation**

Suppose an experiment is conducted with $n$ identical 2-component systems. Here the system may fail due to a failure of any one component. The experiment terminates when all the systems fail and the life times of all the $n$ systems $X_1, X_2, \ldots, X_n$ along with the cause(s) of failures are obtained i.e., the data $(X_i, S_i), i = 1, 2, \ldots, n$.

With these notations we write the likelihood function of the data as follows.

$$L(\Theta | \mathbf{x}) = \prod_{i=1}^{n} \left( \prod_{j=1}^{k} f(x_i | \theta_j) \right) \prod_{i=1}^{n} F(x_i | \theta_j)$$

(3)

Using (1) and (2), we get the from (3) that

$$L(\Theta | \mathbf{x}) = \prod_{i=1}^{n} \left( \prod_{j=1}^{k} \frac{2\theta_j}{x_i} \exp \left( \frac{\theta_j}{x_i} \right) \left[ 1 - \exp \left( \frac{\theta_j}{x_i} \right) \right] \right) \prod_{i=1}^{n} \left( \frac{2\theta_j}{x_i} \exp \left( \frac{\theta_j}{x_i} \right) \left[ 1 - \exp \left( \frac{\theta_j}{x_i} \right) \right] \right)$$

(4)

Taking logarithm of (4) and differentiating it partially with respect to $\theta_1$ and $\theta_2$, respectively, we get

$$\frac{\partial \log L}{\partial \theta_1} = \frac{n_1}{\theta_1} - \sum_{i=1}^{n} \frac{1}{x_i} \sum_{j=1}^{k} \frac{1}{x_i} \exp \left( - \frac{\theta_j}{x_i} \right) \left[ 1 - \exp \left( - \frac{\theta_j}{x_i} \right) \right]$$

(5)

$$\frac{\partial \log L}{\partial \theta_2} = \frac{n_2}{\theta_2} - \sum_{i=1}^{n} \frac{1}{x_i} \sum_{j=1}^{k} \frac{1}{x_i} \exp \left( - \frac{\theta_j}{x_i} \right) \left[ 1 - \exp \left( - \frac{\theta_j}{x_i} \right) \right]$$

(6)

From (5) and (6), we observe that the likelihood equations of $\theta_1$ and $\theta_2$ cannot be solved analytically. Therefore, we use iterative methods to evaluate these MLEs. Thus, using (5) and (6), we write the following expression for $\theta_1$ and $\theta_2$ in order to obtain their MLEs through numerical procedure.

$$\hat{\theta}_1 = n_1 \left[ \sum_{i=1}^{n} \frac{1}{x_i^2} \sum_{j=1}^{k} \frac{1}{x_i} \exp \left( - \frac{\theta_j}{x_i} \right) \left[ 1 - \exp \left( - \frac{\theta_j}{x_i} \right) \right] \right]$$

(7)
Parameter Estimation of Inverse Rayleigh Distribution under Competing Risk Model for Masked data

\[ \hat{\theta}_2 = n_2 \times \left[ \frac{1}{x_2^2} \exp \left( -\frac{\hat{\theta}_2}{x_2^2} \right) \right] \]

\[ - \sum_{i=1}^{n_2} \hat{\theta}_1 \exp \left( -\frac{\hat{\theta}_2}{x_i^2} \right) \left[ \frac{1}{x_i^2} \exp \left( -\frac{\hat{\theta}_2}{x_i^2} \right) \right] \]

Confidence Interval

Asymptotic Confidence Interval

Here, we evaluate ACI for parameters, using the asymptotic normality of MLEs. Therefore, we first obtain the approximate Fisher's information matrix given by

\[ I(\hat{\Theta}) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \]

\[ A_{11} = -\frac{\partial^2}{\partial \hat{\theta}_1^2} \ln L \]

\[ = -\frac{n_1}{\hat{\theta}_1^3} \sum_{i=1}^{n_1} \left[ 1 - \exp \left( -\frac{\hat{\theta}_1}{x_i^2} \right) \right] \left[ \frac{1}{x_i^2} \exp \left( -\frac{\hat{\theta}_1}{x_i^2} \right) \right] \]

\[ + \sum_{i=1}^{n_1} J(\hat{\theta}_1, \hat{\theta}_1) \left[ \frac{1}{x_i^2} \exp \left( -\frac{\hat{\theta}_1}{x_i^2} \right) \right] \]

\[ A_{22} = -\frac{\partial^2}{\partial \hat{\theta}_2^2} \ln L \]

\[ = -\frac{n_2}{\hat{\theta}_2^3} \sum_{i=1}^{n_2} \left[ 1 - \exp \left( -\frac{\hat{\theta}_2}{x_i^2} \right) \right] \left[ \frac{1}{x_i^2} \exp \left( -\frac{\hat{\theta}_2}{x_i^2} \right) \right] \]

\[ + \sum_{i=1}^{n_2} J(\hat{\theta}_2, \hat{\theta}_2) \left[ \frac{1}{x_i^2} \exp \left( -\frac{\hat{\theta}_2}{x_i^2} \right) \right] \]

Here,

\[ J(\hat{\theta}_1, \hat{\theta}_1) = \left[ 1 - \exp \left( -\frac{\hat{\theta}_1}{x_1^2} \right) \right] \left[ \frac{1}{x_1^2} \exp \left( -\frac{\hat{\theta}_1}{x_1^2} \right) \right] \]

\[ \times \exp \left( -\frac{\hat{\theta}_1}{x_1^2} \right) \left[ 1 - \exp \left( -\frac{\hat{\theta}_1}{x_1^2} \right) \right] \]

\[ K(\hat{\theta}_2, \hat{\theta}_2) = \left[ 1 - \exp \left( -\frac{\hat{\theta}_2}{x_2^2} \right) \right] \left[ \frac{1}{x_2^2} \exp \left( -\frac{\hat{\theta}_2}{x_2^2} \right) \right] \]

\[ \times \exp \left( -\frac{\hat{\theta}_2}{x_2^2} \right) \left[ 1 - \exp \left( -\frac{\hat{\theta}_2}{x_2^2} \right) \right] \]

Thus, we get the 100(1-α)% confidence limits for \( \hat{\theta} ; k = 1, 2 \), by \( \hat{\theta} \pm z_{\alpha/2} \sqrt{I^{-1}(\hat{\theta})} \)

Boot-p Confidence Interval

In this section we define confidence intervals based on bootstrapping. Following are the main steps to calculate boot-p confidence interval:

1. Based on the original sample \( \mathbf{x} = (x_1, x_2, ..., x_n) \) obtain \( \hat{\Theta} \), the MLE of \( \Theta \).
2. Under the same conditions simulate sample, say \((y_1, y_2, ..., y_m)\), from the underlying distribution \( IRD(\hat{\Theta}) \).
3. Compute the MLE of \( \hat{\Theta} \) based on \((y_1, y_2, ..., y_m)\), say \( \hat{\Theta}^* \).
4. Repeat Step 2 and Step 3, B times and obtain \( \hat{\Theta}^1_*, \hat{\Theta}^2_*, ..., \hat{\Theta}^B_* \).
5. Arrange \( \hat{\Theta}^1_*, \hat{\Theta}^2_*, ..., \hat{\Theta}^B_* \) in ascending order.
6. A two-sided 100(1-α)% percentile bootstrap confidence interval of \( \Theta \), say \( [\hat{\Theta}^L_*, \hat{\Theta}^U_*] \), is then given by

\[ [\hat{\Theta}^L_*, \hat{\Theta}^U_*] = [\hat{\Theta}^{*(B_{\alpha/2})}, \hat{\Theta}^{*(B_{1-\alpha/2})}] \] (13)

Fig. 1. Variation of MSE's of \( \hat{\theta}_1 \) with varying sample size and at different levels of masking.
Fig. 2. Variations of MSE’s of $\hat{\theta}_2$ with varying sample size and at different levels of masking.

Simulation study

In this section, we present some numerical illustrations based on simulation study for different sample sizes $n = 10, 15, 20, 30, 40, 50$ and level of masking 10, 15, 20%. Here we consider two cases when $(\theta_1 = \theta_2)$ and $(\theta_1 \neq \theta_2)$.

For the simulation study, we generate $n$ observations from $IRD(\theta_1)$ and $IRD(\theta_2)$. Then the lifetimes of $i^{th}$ system becomes $X_i = \min(X_{i1}, X_{i2})$; $i = 1, 2, ..., n$.

We also note down the component which causes the $i^{th}$ system to fail. Then under different levels of masking we obtain the MLEs of $\theta_1$ and $\theta_2$ from (7) and (8) through iterative methods. We repeat this process 1000 times. For each considered sample size and level of masking, we report the average values of MLEs along with their mean squared errors (MSEs) in Tables 1 and 3 and ACI and boot-p CI in Table 2 and 4.

CONCLUSION

As working on the survival data or competing risk data, the data obtained may not always be complete with known cause of failure. In such a case, we are interested in designing a technique which can help us through the problem. Here, we have obtained the parameters for the data with varying sample sizes, and different levels of masking. The simulation study gives (i) On the basis of sample size- as the sample size increases the estimated value approaches to the true value of parameter, and the mean square error decreases with the increase in sample size, and (ii) On the basis of level of masking- the mean square error increases with increase in level of masking, the ACIs are always symmetric and the boot-p CIs approaches to symmetry as the sample size increases.

REFERENCES


Parameter Estimation of Inverse Rayleigh Distribution under Competing Risk Model for Masked data

Table 1. MLEs and respective MSE's for $\theta_1 = 1.0$ and $\theta_2 = 1.0$ at different masking level

<table>
<thead>
<tr>
<th>$N$</th>
<th>Masking Level</th>
<th>$\hat{\theta}_1$</th>
<th></th>
<th>$\hat{\theta}_2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>10</td>
<td>0.4801 (0.2969)</td>
<td>0.5242 (0.6401)</td>
<td>0.4303 (0.9866)</td>
<td>0.4352 (0.3315)</td>
<td>0.4355 (0.7845)</td>
</tr>
<tr>
<td>15</td>
<td>0.5810 (0.2924)</td>
<td>0.5772 (0.5091)</td>
<td>0.4835 (0.8124)</td>
<td>0.4990 (0.2563)</td>
<td>0.5733 (0.5186)</td>
</tr>
<tr>
<td>20</td>
<td>0.6531 (1.3011)</td>
<td>0.6302 (0.5064)</td>
<td>0.6454 (0.6737)</td>
<td>0.6199 (0.1857)</td>
<td>0.5828 (0.4941)</td>
</tr>
<tr>
<td>30</td>
<td>0.6780 (1.2111)</td>
<td>0.7186 (0.4255)</td>
<td>0.6729 (0.6177)</td>
<td>0.7675 (0.1725)</td>
<td>0.5993 (0.4283)</td>
</tr>
<tr>
<td>50</td>
<td>0.8147 (0.0666)</td>
<td>0.9513 (0.3651)</td>
<td>0.7632 (0.3978)</td>
<td>0.8964 (0.0446)</td>
<td>0.7695 (0.1341)</td>
</tr>
</tbody>
</table>

Table 2. Asymptotic and boot- p confidence intervals for the MLE of $\theta_1 = 1.0$ and $\theta_2 = 1.0$ at different masking level

<table>
<thead>
<tr>
<th>$n$</th>
<th>Masking Level</th>
<th>$\hat{\theta}_1$</th>
<th></th>
<th>$\hat{\theta}_2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>10</td>
<td>(0.1426,0.8176)</td>
<td>(0.1186,0.9297)</td>
<td>(0.1459,0.7147)</td>
<td>(0.1416,0.7289)</td>
<td>(0.1556,0.7155)</td>
</tr>
<tr>
<td></td>
<td>(0.3457,0.7019)</td>
<td>(0.3398,0.8298)</td>
<td>(0.3969,0.4927)</td>
<td>(0.3351,0.6332)</td>
<td>(0.3359,0.5934)</td>
</tr>
<tr>
<td>15</td>
<td>(0.2436,0.9184)</td>
<td>(0.3570,0.7974)</td>
<td>(0.2150,0.7520)</td>
<td>(0.2735,0.7246)</td>
<td>(0.3563,0.7904)</td>
</tr>
<tr>
<td></td>
<td>(0.3761,0.9044)</td>
<td>(0.5039,0.6784)</td>
<td>(0.3834,0.6529)</td>
<td>(0.4377,0.5942)</td>
<td>(0.5394,0.6256)</td>
</tr>
<tr>
<td>20</td>
<td>(0.3474,0.9588)</td>
<td>(0.2217,0.1038)</td>
<td>(0.2139,0.5625)</td>
<td>(0.2619,0.9778)</td>
<td>(0.2281,0.9375)</td>
</tr>
<tr>
<td></td>
<td>(0.4962,0.8625)</td>
<td>(0.5131,0.7809)</td>
<td>(0.3591,0.1487)</td>
<td>(0.4522,0.8521)</td>
<td>(0.4417,0.7821)</td>
</tr>
<tr>
<td>30</td>
<td>(0.4229,0.9330)</td>
<td>(0.4354,0.0018)</td>
<td>(0.3855,0.9604)</td>
<td>(0.4378,0.1097)</td>
<td>(0.3776,0.8209)</td>
</tr>
<tr>
<td></td>
<td>(0.5301,0.8481)</td>
<td>(0.5148,0.9438)</td>
<td>(0.5582,0.8291)</td>
<td>(0.5773,0.9847)</td>
<td>(0.5338,0.6802)</td>
</tr>
<tr>
<td>50</td>
<td>(0.5425,1.0869)</td>
<td>(0.6685,1.2341)</td>
<td>(0.5179,1.0086)</td>
<td>(0.5984,1.1944)</td>
<td>(0.6839,1.2541)</td>
</tr>
<tr>
<td></td>
<td>(0.6660,0.9772)</td>
<td>(0.7814,1.1243)</td>
<td>(0.6529,0.8892)</td>
<td>(0.5859,0.9873)</td>
<td>(0.4865,0.8701)</td>
</tr>
</tbody>
</table>

Table 3. MLEs and respective MSE’s for $\theta_1 = 1.0$ and $\theta_2 = 1.1$ at different masking levels.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Masking Level</th>
<th>$\hat{\theta}_1$</th>
<th></th>
<th>$\hat{\theta}_2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>10</td>
<td>0.5042 (0.7136)</td>
<td>0.1260 (0.7642)</td>
<td>0.0993 (0.8166)</td>
<td>0.7213 (0.6982)</td>
<td>0.2566 (0.7426)</td>
</tr>
<tr>
<td>15</td>
<td>0.5621 (0.6851)</td>
<td>0.3903 (0.7738)</td>
<td>0.2077 (0.7976)</td>
<td>0.7954 (0.5476)</td>
<td>0.4169 (0.6678)</td>
</tr>
<tr>
<td>20</td>
<td>0.6021 (0.6537)</td>
<td>0.5521 (0.7126)</td>
<td>0.2661 (0.7364)</td>
<td>0.8102 (0.5264)</td>
<td>0.5846 (0.6482)</td>
</tr>
<tr>
<td>30</td>
<td>0.6894 (0.1703)</td>
<td>0.5725 (0.3043)</td>
<td>0.4962 (0.4253)</td>
<td>0.8467 (0.2164)</td>
<td>0.6161 (0.2764)</td>
</tr>
<tr>
<td>50</td>
<td>0.7046 (0.0249)</td>
<td>0.6248 (0.1395)</td>
<td>0.5397 (0.2120)</td>
<td>0.9176 (0.1837)</td>
<td>0.7199 (0.1954)</td>
</tr>
</tbody>
</table>
Table 4. Asymptotic and boot-p confidence intervals for the MLEs at $\theta_1 = 1.0$ and $\theta_2 = 1.1$ for different masking level.

<table>
<thead>
<tr>
<th>n</th>
<th>Masking Level</th>
<th>$\hat{\theta}_1$</th>
<th>$\hat{\theta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>ACI boot-p</td>
<td>(0.2235, 0.7849)</td>
<td>(0.2641, 0.9054)</td>
</tr>
<tr>
<td>20</td>
<td>ACI boot-p</td>
<td>(0.3416, 0.7826)</td>
<td>(0.5810, 1.0098)</td>
</tr>
<tr>
<td>30</td>
<td>ACI boot-p</td>
<td>(0.2014, 1.0028)</td>
<td>(0.4028, 1.2176)</td>
</tr>
<tr>
<td>40</td>
<td>ACI boot-p</td>
<td>(0.5106, 0.8682)</td>
<td>(0.4973, 1.1961)</td>
</tr>
<tr>
<td>50</td>
<td>ACI boot-p</td>
<td>(0.5947, 1.1425)</td>
<td>(0.5325, 1.2894)</td>
</tr>
</tbody>
</table>