Chapter 4

Inferences on inliers in Lindley distribution

4.1. Introduction

While investigating the suitable models for inliers, it has come to our notice that, the failure time distribution (FTD) used for modeling the positive observations (lifetime greater than zero) itself can be mixture of distributions. Such mixtures can be a good model for describing the system characteristics of the failure mechanism. Towards this, in this chapter, we study a model based on Lindley distribution to study the occurrence of inliers.

Lindley distribution is a mixture of exponential (θ) and gamma $(2, \theta)$ distributions with mixing proportion $\frac{\theta}{1+\theta}$, and is proposed by Lindley (1958) in the context of Bayesian statistics as a counter example of fiducial statistics. A random variable X is said to have the Lindley distribution with parameter θ if its probability density function (pdf) is defined as

$$f(x;\theta) = \begin{cases} \frac{\theta^2}{1+\theta}(1+x)\, e^{-\theta x}, & x > 0;\ \ \theta > 0 \\ 0, & o.w. \end{cases} \tag{4.1.1}$$

The corresponding cumulative distribution function (CDF) is

$$F(x;\theta) = \begin{cases} 1 - e^{-\theta x}\left[1 + \frac{\theta x}{1+\theta}\right], & x > 0;\ \ \theta > 0 \\ 0, & o.w. \end{cases} \tag{4.1.2}$$

Ghitany et al. (2008) have discussed the various statistical properties of Lindley distribution and shown its applicability over the exponential distribution. They have found that the Lindley distribution performs better than the exponential model. One of the main reasons to consider the Lindley distribution over the exponential distribution is its time-dependent/increasing hazard rate. Since last decade, Lindley distribution has been widely used in different setup by many authors. Of course, Lindley is more flexible than exponential but exponential has some advantage over Lindley due to its simplicity.

The rest of the chapter has been organized as follows. In Section 4.2, we present the Lindley's inliers model and demonstrated its flexibility by showing the wide variety of shapes of the density, and survival functions. The estimation procedures, tests of hypothesis for the parameters are given in subsequent sections. The applicability of the model through practical example is given along with the simulation of parameter estimates in last two sections.

4.2. The Lindley's Inliers Model

The instantaneous failure model given in (1.3.1) with FTD as Lindley distribution as given above is

$$g(x; p, \theta) = \begin{cases} 1-p, & x = 0 \\ p\,\frac{\theta^2}{1+\theta}(1+x)\,e^{-\theta x}, & x > 0 \end{cases} \qquad (4.2.1)$$

The plot of $g(x)$ for various mixing proportions are presented in Figures 4.1. The Figure 4.2 shows the plot of survival function graphs for different value of θ.

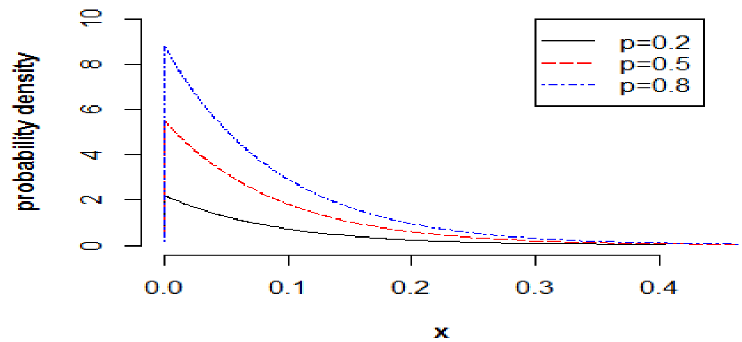

Figure 4.1: Plot of g(x): (θ=12)

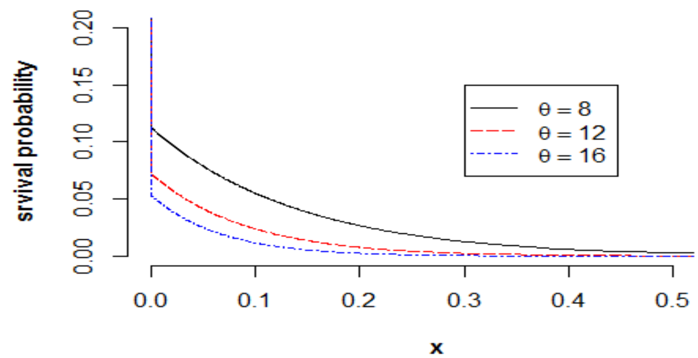

Figure 4.2: Plot of S(t): (p=0.8)

4.3. Estimation

The model in (4.2.1) has two parameters p and θ. To estimate the model parameters, maximum likelihood estimation and minimum variance unbiased estimation method is proposed.

4.3.1. Maximum Likelihood Estimation

Let $X_1, X_2, \ldots, X_n$ a random sample of size n form $g \in G$ given (4.2.1), then the likelihood function is

$$L(p,\theta|\underline{x}) = \prod_{i=1}^{n}(1-p)^{I(x_i)}\left[p\frac{\theta^2}{1+\theta}(1+x_i)\,e^{-\theta x_i}\right]^{(1-I(x_i))}$$

where $I(x_i) = \begin{cases} 1, & if\ x_i = 0 \\ 0, & o.w. \end{cases}$

Let $n - r = \sum_{i=1}^{n} I(x_i)$, where r is number of observations greater than 0 and $p = P(x > 0)$. Then,

$$L(p,\theta|\underline{x}) = (1-p)^{(n-r)}\,p^r\left(\frac{\theta^2}{1+\theta}\right)^r\prod_{x_i>0}(1+x_i)\,e^{-\theta\sum_{x_i>0}\ x_i}$$

The log-likelihood function is

$$\ln L = (n-r)\ln(1-p) + r\ln p + r\ln\theta^2 - r\ln(1+\theta)$$

$$+\sum_{x_i>0}\ln(1+x_i) - \theta\sum_{x_i>0}\ x_i$$

The likelihood estimates are obtained by solving the equations

$$\frac{\partial lnL(p,\theta|\underline{x})}{\partial p} = -\frac{(n-r)}{1-p} + \frac{r}{p} = 0 \qquad (4.3.1)$$

and

$$\frac{\partial lnL(p,\theta|\underline{x})}{\partial \theta} = \frac{2r}{\theta} - \frac{r}{1+\theta} - \sum_{x_i>0}\ x_i = 0. \qquad (4.3.2)$$

The estimator $\hat{p}_{MLE}$ from equation (4.3.1) is

$$\hat{p}_{MLE} = \frac{r}{n} \qquad (4.3.3)$$

and from (4.3.2), the we get the estimate of θ as

$$\hat{\theta}_{MLE} = \frac{r - \sum_{x_i>0} x_i + \sqrt{\left(\sum_{x_i>0} x_i\right)^2 + 6\, r \sum_{x_i>0} x_i + r^2}}{2 \sum_{x_i>0} x_i} \qquad (4.3.4)$$

which is free from the parameter *p*.

4.3.2. Asymptotic distribution of MLE

For inlier prone Lindley distribution $g(x; p, \theta)$ as given in (4.2.1),

$$\frac{\partial ln\ g(p,\theta|x)}{\partial p} = \begin{cases} -\dfrac{1}{1-p}, & x = 0 \\ \dfrac{1}{p}, & x > 0 \end{cases}$$

and

$$\frac{\partial ln\ g(p,\theta|x)}{\partial \theta} = \begin{cases} 0, & x = 0 \\ \dfrac{2}{\theta} - \dfrac{1}{1+\theta} - x, & x > 0 \end{cases}$$

One can verify that $E\left(\frac{\partial ln\ g(p,\theta|x)}{\partial p}\right) = 0$ and $E\left(\frac{\partial ln\ g(p,\theta|x)}{\partial \theta}\right) = 0$. Also,

$$\frac{\partial^2 ln\ g(p,\theta|x)}{\partial p^2} = \begin{cases} -\dfrac{1}{(1-p)^2}, & x = 0 \\ -\dfrac{1}{p^2}, & x > 0 \end{cases}$$

$$\frac{\partial^2 ln\ g(p,\theta|x)}{\partial \theta^2} = \begin{cases} 0, & x = 0 \\ -\dfrac{2}{\theta^2} + \dfrac{1}{(1+\theta)^2}, & x > 0 \end{cases}$$

and

$$\frac{\partial^2 ln\ g(p,\theta|x)}{\partial p\ \partial\theta} = 0\ \ \forall\ x.$$

Hence, the Fisher information's are

$$I_{pp} = E\left(-\frac{\partial^2\ ln\ g(p,\theta|x)}{\partial p^2}\right) = \frac{1}{p(1-p)}$$

$$I_{\theta\theta} = E\left(-\frac{\partial^2\ ln\ g(p,\theta|x)}{\partial\theta^2}\right) = \frac{p(2+4\theta+p^2)}{\theta^2\ (1+\theta)^2}$$

and

$$I_{p\theta} = E\left(-\frac{\partial^2 ln\ g(p,\theta|x)}{\partial p\ \partial\theta}\right) = 0.$$

Therefore, the Fisher information matrix $I_g(p,\theta)$ is given by

$$I_g(p,\theta) = \begin{bmatrix} I_{pp} & I_{p\theta} \\ I_{\theta p} & I_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \frac{1}{p(1-p)} & 0 \\ 0 & \frac{p(2+4\theta+p^2)}{\theta^2\ (1+\theta)^2} \end{bmatrix}. \qquad (4.3.5)$$

The inverse matrix $I_g^{-1}(p,\theta)$ is given by

$$I_g^{-1}(p,\theta) = \begin{bmatrix} p(1-p) & 0 \\ 0 & \frac{\theta^2\ (1+\theta)^2}{p(2+4\theta+p^2)} \end{bmatrix}.$$

Using the standard result of MLE, we have asymptotic distribution of MLE as

$$(\hat{p}_{MLE}, \hat{\theta}_{MLE})' \sim AN^{(2)}\left[(p,\theta)', \frac{1}{n}\ I_g^{-1}(p,\theta)\ \right]. \qquad (4.3.6)$$

Using the estimated variances, one can also propose large sample tests for p and θ. The approximate $(1-\alpha)\%$ confidence interval for p and θ are respectively given by

$$\hat{p}_{MLE} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_{MLE}(1-\hat{p}_{MLE})}{n}} \tag{4.3.7}$$

and

$$\hat{\theta}_{MLE} \pm z_{\alpha/2} \sqrt{\frac{\hat{\theta}^2_{MLE}(1+\hat{\theta}_{MLE})^2}{n\,\hat{p}_{MLE}\left(2+4\hat{\theta}_{MLE}+\hat{\theta}^2_{MLE}\right)}}\,. \tag{4.3.8}$$

4.3.3. Uniformly Minimum Variance Unbiased Estimation

Referring to section 1.5.5, with d=0, we obtain UMVUE of the mixture density in the following way: Writing (4.2.1), in the form of (1.5.9), we have

$$g(x;p,\theta) = \frac{(1+x)^{1-I(x)}\,[e^{-\theta}]^{(1-I(x))x}\left(\frac{(1+\theta)}{\theta^2}\frac{(1-p)}{p}\right)^{I(x)}}{\left(\frac{p(1+\theta)}{\theta^2}\right)} \tag{4.3.9}$$

$$= (a(x))^{(1-I(x))}\,\frac{[h(\theta)]^{d(x)(1-I(x))}}{\left[\frac{g(\theta)}{p}\right]}\,\left[g(\theta)\left(\frac{p}{1-p}\right)\right]^{I(x)}$$

where, $I(x)$ is as in (4.3.1), $a(x) = 1 + x$, $h(\theta) = e^{-\theta}$, $d(x) = x$, $g(\theta) = \frac{1+\theta}{\theta^2}$ and $g(\theta) = \int_{x>0} a(x)[h(\theta)]^{d(x)}\,dx$. The density in (4.3.9) so obtained is defined with respect to a measure $\mu(x)$which is the sum of Lebesgue measure over $(0,\infty)$ and a singular measure at 0, is a well-known form of a two-parameter exponential family with natural parameters $(\eta_1,\eta_2) = \left(\log\left(\frac{(1+\theta)}{\theta^2}\;\frac{(1-P)}{p}\right),\,\log\left(e^{-\theta}\right)\right)$ generated by underlying indexing parameters (p,θ). Hence $\left(I(x),\left(1-I(x)\right)x\right)$ is jointly minimal sufficient for (p,θ), as $I(x)$ and $(1-I(x))x$ do not satisfy any linear restriction.

Let $X_1, X_2, \ldots, X_n$ be n independent random sample from mixture density (4.3.9) and if we denote $z = \sum_{i=1}^{n}(1-I(x_i))x_i$. Then Joint pdf of the sample is given by

$$g(\underline{x};\theta,p)=\binom{n}{r}\ p^{r}(1-p)^{n-r}\ \frac{1}{\binom{n}{r}}\prod\nolimits_{x_{i>0}} a(x_i)\ \frac{[e^{-\theta}]^{z}}{\left(\frac{1+\theta}{\theta^{2}}\right)^{r}} \tag{4.3.10}$$

$$=P(n-R=n-r)\ g(z;\theta\ |(n-r).$$

Therefore, by Neyman's factorization theorem $(n-R,Z)$ are jointly sufficient for (p,θ). Also, $n-R$ is binomial which is the same as that of R with parameter (n,p), and is a complete family. The conditional pdf of variable $(Z|R=r,r>0)$ is

$$g(z;\theta|r)=\begin{cases}\left(\sum_{u=0}^{r}\frac{(r)_u}{(r)^u}\frac{z^{r+u-1}}{(r-1)!\,u!}\right)\frac{\left[e^{-\theta}\right]^{z}}{\left(\frac{1+\theta}{\theta^{2}}\right)^{r}}, z>0\\ 1, \qquad z=0\end{cases}$$

where, $(r)_u=r(r-1)\ldots(r-u+1)$ and $(r)^u=r(r+1)\ldots(r+u-1)$. Here $g(z;\theta|r)$ depends only on θ, and belongs to complete family. Hence $(n-R,Z)$ are jointly complete sufficient for (p,θ). The joint distribution of $(n-R,Z)$ can be written as

$$g(z;\theta,p)=\begin{cases}B(z,r,n)\frac{[e^{-\theta}]^{z}\left(\frac{1-p}{p}\frac{1+\theta}{\theta^{2}}\right)^{n-r}}{\left(\frac{1+\theta}{\theta^{2}p}\right)^{n}}, z>0\\ (1-p)^{n}, \qquad z=0\end{cases} \tag{4.3.11}$$

where

$$B(z,r,n)=\begin{cases}1, & z=0;\ r=0\\ \binom{n}{r}B(z|r), & z>0;\ r=1,2,\ldots,n\end{cases} \tag{4.3.12}$$

and

$$B(z|r)=\frac{1}{(r-1)!}\sum\nolimits_{u=0}^{r}\frac{(r)_u}{(r)^u}\frac{z^{r+u-1}}{u!}.$$

Here $B(z,r,n)$ is such that

$$(1-p)^n + \sum_{r=1}^{n} \int_{z>0} B(z,r,n) \frac{\left[e^{-\theta}\right]^z \left(\frac{1-p}{p}\frac{1+\theta}{\theta^2}\right)^{n-r}}{\left(\frac{1+\theta}{\theta^2 p}\right)^n} dz = 1.$$

Again, using Lemma 1.5.1, the UMVUEs of parametric function $\phi(p,\theta)$ exists if and only if $\phi(p,\theta)$ can be expressed in the form

$$\phi(p,\theta) = \alpha(0,0,\mathrm{n})(p)^n + \sum_{r=1}^{n} \int_{z>0} \alpha(z,r,n) \frac{\left[e^{-\theta}\right]^z \left(\frac{1-p}{p}\frac{1+\theta}{\theta^2}\right)^{n-r}}{\left(\frac{1+\theta}{\theta^2 p}\right)^n} dz.$$

Thus, using (1.5.15) the UMVUE of function $\phi(p,\theta)$ in $g(x;p,\theta)$ is given by

$$\psi(Z,R,n) = \frac{\alpha(Z,R,n)}{B(Z,R,n)}, B(Z,R,n) \neq 0 \tag{4.3.13}$$

Like any other model, here also it is difficult to find an unbiased estimate for the parameter θ alone. Therefore, we obtain the UMVUE for parametric functions as in line with the theorems and corollary of section 1.5.5.

Result 4.1. *Suppose $X_1, X_2, \ldots, X_n$ is a random sample from (4.3.9). For $m \leq n$ and using Result 1.5.1, the UMVUE of $(1-p)^m$ is $G_m(Z,R,n)$ as given by (1.5.16).*

Result 4.2. *If $m = 1$ in Result 4.1, the UMVUE of $(1-p)$, the probability of observing 'zero' in mixture distribution given in (4.3.9) is given by*

$$G_1(z,r,n) = \begin{cases} \dfrac{n-r}{n}, & r > 0; z > 0 \\ 1, & r = 0; z = 0 \end{cases}$$

Result 4.3. *For $m \leq \frac{n}{2}$ and using Result 1.5.2, the UMVUE of the variance of $G_m(Z,R,n)$ is given by (1.5.17).*

Result 4.4. *For $m = 1$ in Result 4.3, the UMVUE of the variance of UMVUE of $(1-p)$ is given by.*

$$\widehat{var}[G_1(z,r,n)] = \begin{cases} \frac{r(n-r)}{n^2(n-1)}, & r = 1, 2, \dots, (n-1) \\ 0, & o.w. \end{cases}$$

Result 4.5. *For $k > 0$ and using Result 1.5.3, the UMVUE of parametric function*

$$(1-p)^n + \left(\frac{p\,\theta^2}{1+\theta}\right)^k [1-(1-p)^{n-k}],\ \theta > 0; 0 < p < 1$$

is obtained as

$$H_k(z,r,n) = \begin{cases} \frac{(r-1)_k (r)_k}{(n)_k} \frac{\sum_{u=0}^{r-k} \frac{(r-k)_u}{(r-k)^u} \frac{z^{u-k}}{u!}}{\sum_{u=0}^{r} \frac{(r)_u}{(r)^u} \frac{z^u}{u!}}, & z > 0;\ r = 1,2,\dots,n \\ 1, & z = 0;\ r = 0 \end{cases}$$

where $(r)_k = r(r-1)(r-2)\dots(r-k+1)$ *and*

$(r)^k = r(r+1)(r+2)\dots(r+k-1).$

Result 4.6. *The UMVUE of the variance of $H_k(Z,R,n)$, according to Result 1.5.4 is obtained as*

$$\widehat{var}[H_k(z,r,n)]$$
$$= \begin{cases} \left(\frac{(r-1)_k (r)_k}{(n)_k} \frac{\sum_{u=0}^{r-k} \frac{(r-k)_u}{(r-k)^u} \frac{z^{u-k}}{u!}}{\sum_{u=0}^{r} \frac{(r)_u}{(r)^u} \frac{z^u}{u!}}\right)^2 - \frac{(r-1)_{2k} (r)_{2k}}{(n)_{2k}} \frac{\sum_{u=0}^{r-2k} \frac{(r-2k)_u}{(r-2k)^u} \frac{z^{u-2k}}{u!}}{\sum_{u=0}^{r} \frac{(r)_u}{(r)^u} \frac{z^u}{u!}}, & z > 0; r = 1,\dots,n \\ 0, & o.w. \end{cases}$$

Result 4.7. *For fixed x, according to Result 1.5.5 the UMVUE of pdf $g(x;p,\theta)$ given in (4.3.9) is obtained as*

$$\phi_x(z,r,n) = \begin{cases} \frac{r(r-1)}{n}(1+x)(1+{}^{x}/_{z})^{r-1}\frac{\sum_{u=0}^{r-1}\frac{(r-1)_u}{(r-1)^u}\frac{(z-x)^{u-1}}{u!}}{\sum_{u=0}^{r}\frac{(r)_u}{(r)^u}\frac{z^u}{u!},}, & x>0; z>x; r=1,2,\dots,n \\ \frac{n-r}{n}, & x=0; r=0,1,\dots,n-1 \\ 0, & o.w. \end{cases}$$

Result 4.8. *For $r = n$ that is when all the observations are coming from the pdf in (4.1.1), then the UMVUE of the density $f(x;\theta)$ is simplified as*

$$\phi_x(z,r,n) = \begin{cases} (n-1)(1+x)(1+{}^{x}/_{z})^{n-1}\frac{\sum_{u=0}^{n-1}\frac{(n-1)_u}{(n-1)^u}\frac{(z-x)^{u-1}}{u!}}{\sum_{u=0}^{r}\frac{(n)_u}{(n)^u}\frac{z^u}{u!},}, & x>0; z>x \\ 0, & o.w. \end{cases}$$

Note: This result is matched with the result obtained by Maiti and Mukherjee (2016),

Result 4.9. *Using Result 1.5.6, the UMVUE of the variance of $\phi_x(Z,R,n)$ obtained as*

$$\widehat{var}[\phi_x(z,r,n)] = \begin{cases} \frac{r^2(r+1)^2}{n^2}(1+x)^2(1-{}^{x}/_{z})^{2(r-1)}\frac{\left(\sum_{u=0}^{r-1}\frac{(r-1)_u}{(r-1)^u}\frac{(z-x)^{u-1}}{u!}\right)^2}{\left(\sum_{u=0}^{r}\frac{(r)_u}{(r)^u}\frac{(z)^u}{u!}\right)^2} \\ -\frac{r(r-1)^2(r-2)}{n(n-1)}(1-x)^2(1-{}^{x}/_{z})^{2r-3}\left(\frac{\sum_{u=0}^{r-1}\frac{(r-1)_u}{(r-1)^u}\frac{(z-x)^{u-1}}{u!}}{\sum_{u=0}^{r}\frac{(r)_u}{(r)^u}\frac{(z)^u}{u!}}\right)\left(\frac{\sum_{u=0}^{r-2}\frac{(r-2)_u}{(r-2)^u}\frac{(z-2x)^{u-1}}{u!}}{\sum_{u=0}^{r}\frac{(r-1)_u}{(r-1)^u}\frac{(z-x)^u}{u!}}\right), \\ \qquad x>0;\ z>2x;\ r=2,3,\dots,n \\ \frac{r^2(r+1)^2}{n^2}(1+x)^2(1-{}^{x}/_{z})^{2(r-1)}\frac{\left(\sum_{u=0}^{r-1}\frac{(r-1)_u}{(r-1)^u}\frac{(z-x)^{u-1}}{u!}\right)^2}{\left(\sum_{u=0}^{r}\frac{(r)_u}{(r)^u}\frac{(z)^u}{u!}\right)^2}, & x>0; x<z<2x \\ \frac{r(n-r)}{n^2(n-1)}, & x=0; r=0,1,\dots,n-1 \end{cases}$$

For $r = n$ the above result reduce to the case of- Lindley distribution, with no inliers.

Result 4.10. *For fixed z and r, using corollary 1.5.1 the UMVUE of the survival function $S(t) = P(X > t)$, $t \geq 0$ obtained as*

$$\hat{S}(t) = \int_{x>t} \phi_x(z, r, n)\, dx$$

$$= \begin{cases} \dfrac{r(r-1)}{n \sum_{u=0}^{r} \frac{(r)_u}{(r)^u} \frac{z^u}{u!}} \displaystyle\sum_{u=0}^{r-1} \frac{(r-1)_u}{(r-1)^u} \frac{1}{u!} \left[(1+z) \frac{(z-t)^{r+u-1}}{r+u-1} - \frac{(z-t)^{r+u}}{r+u}\right], & t < z \\ 0, & o.w. \end{cases}$$

Result 4.11. *Using Result 1.5.7, the UMVUE of the variance of $\hat{S}(t)$ obtained as*

$$\widehat{var}[\hat{S}(t)]`$$

$$= \begin{cases} \left[\dfrac{r(r-1)}{n \sum_{u=0}^{r} \frac{(r)_u}{(r)^u} \frac{z^u}{u!}} \displaystyle\sum_{u=0}^{r-1} \frac{(r-1)_u}{(r-1)^u} \frac{1}{u!} \left[(1+z) \frac{(z-t)^{r+u-1}}{r+u-1} - \frac{(z-t)^{r+u}}{r+u}\right]\right]^2 \\ \qquad - \dfrac{r(r-1)^2(r-2)}{n(n-1) \sum_{u=0}^{r} \frac{(r)_u}{(r)^u} \frac{z^{r+u-1}}{u!}} S^*, & z > 2t;\ r = 3, \ldots, n \\ \left[\dfrac{r(r-1)}{n \sum_{u=0}^{r} \frac{(r)_u}{(r)^u} \frac{z^u}{u!}} \displaystyle\sum_{u=0}^{r-1} \frac{(r-1)_u}{(r-1)^u} \frac{1}{u!} \left[(1+z) \frac{(z-t)^{r+u-1}}{r+u-1} - \frac{(z-t)^{r+u}}{r+u}\right]\right]^2, & t < z < 2t; r = 2,3, \ldots, n \\ 0, & o.w. \end{cases}$$

where,

$$S^* = \sum_{u=0}^{r-2} \frac{(r-2)_u}{(r-2)^u} \frac{1}{u!} \frac{(z-2t)^{r+u-1}}{(r+u-2)} \left\{(1+t) \left[\frac{1+z-t}{r+u-1} - \frac{z-2t}{r+u}\right] + \frac{z-2t}{r+u-1} \left[\frac{1+z-t}{r+u} - \frac{z-2t}{r+u+1}\right]\right\}$$

For $r = n$, both the above results reduce to the case of Lindley distribution, with no inliers.

4.4. Testing of hypothesis for parameters

Since p is the mixing parameter and involve only the number of observations correspond to the positive part, the most powerful test for testing $H_0: p = 1$ against $H_1: p < 1$ of size α is similar to the one obtained in (1.6.1), with power function $\beta(p) = 1 - (1-\alpha)p^n$. Whereas, the locally most powerful (LMP) test for p of size α for testing $H_0: p = 1$ against $H_1: p < 1$ for θ known according to (1.6.2) is given by

$$\Phi_2(x) = \begin{cases} 1, & r < p[n + (1-p)c_\alpha] \\ \gamma, & r = p[n + (1-p)c_\alpha] \\ 0, & r > p[n + (1-p)c_\alpha] \end{cases}$$

where c_α and γ are such that $E_{H_0}[\Phi_2(x)] = \alpha$.

Similarly, the most powerful test for $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$, $(\theta_1 > \theta_0)$ for p known according to (1.6.3) is given by

$$\Phi_3(x) = \begin{cases} 1, & \sum_{x_i > 0} x_i > \dfrac{c_\alpha + 2r(\log\theta_0 - \log\theta_1) - r[\log(1+\theta_0) - \log(1+\theta_1)]}{(\theta_0 - \theta_1)} \\ \gamma, & \sum_{x_i > 0} x_i = \dfrac{c_\alpha + 2r(\log\theta_0 - \log\theta_1) - r[\log(1+\theta_0) - \log(1+\theta_1)]}{(\theta_0 - \theta_1)} \\ 0, & \sum_{x_i > 0} x_i < \dfrac{c_\alpha + 2r(\log\theta_0 - \log\theta_1) - r[\log(1+\theta_0) - \log(1+\theta_1)]}{(\theta_0 - \theta_1)} \end{cases}$$

where c_α and γ are such that $E_{H_0}[\Phi_3(x)] = \alpha$.

In a similar way, the LMP test of size α for testing $H_0: \theta \leq \theta_0$ against $H_1: \theta > \theta_0$ for p known based on n iid observations from the density $g(x; p, \theta)$ according to (1.6.4) is given by

$$\Phi_4(x) = \begin{cases} 1, & \sum_{x_i>0} x_i < \frac{r(2+\theta_0)}{\theta_0(1+\theta_0)} - c_\alpha \\ \gamma, & \sum_{x_i>0} x_i = \frac{r(2+\theta_0)}{\theta_0(1+\theta_0)} - c_\alpha \\ 0, & \sum_{x_i>0} x_i > \frac{r(2+\theta_0)}{\theta_0(1+\theta_0)} - c_\alpha \end{cases}$$

where c_α and γ are such that $E_{H_0}[\Phi_4(x)] = \alpha$.

4.5. Simulation study

A simulation study is conducted to check the performance of estimators of the parameters and parametric functions for various choices of inliers. Each estimate is based on a simulation of 1000 random samples of size $n = 50$ with different choices of r. The value of $\theta = 0.5$ is assumed for the target Lindley distribution. Table 4.1 presents the various estimates of parameters along with their standard error of estimate which is shown in brackets. Note that the estimate of p and is θ is comparable in all cases. It is also seen that the standard error is very small for every combination of (n, r).

Table 4.1. *Summary of maximum likelihood estimates of the model*

(n,r)	Parameter	Estimates(se)	95% confidence interval
(50,15)	$\hat{p}_{MLE}$	0.30108 (0.06409)	(0.17547, 0.42669)
	$\hat{\theta}_{MLE}$	0.52079 (0.09991)	(0.32496, 0.71661)
(50,25)	$\hat{p}_{MLE}$	0.49990 (0.07003)	(0.36264, 0.63716)
	$\hat{\theta}_{MLE}$	0.51042 (0.07493)	(0.36357, 0.65728)
(50,40)	$\hat{p}_{MLE}$	0.80108 (0.05540)	(0.69327, 0.91045)
	$\hat{\theta}_{MLE}$	0.50552 (0.05853)	(0.39304, 0.62249)
(50,50)	$\hat{p}_{MLE}$	1.00000 (0.01414)	(0.97228, 1.0000)
	$\hat{\theta}_{MLE}$	0.50705 (0.05223)	(0.40469, 0.60941)

Table 4.2 presents the UMVU estimates of parametric functions and UMVUE of its variance for model for different values of k and different combination of n and r. It may be observed from Table 4.2 that for each combination of n and r, the UMVU

estimates are decreasing function of k. It is also noted that as the number of inliers decreases the UMVU estimate of the parametric function increases and UMVUE of the variance of the estimate decreases.

Table 4.2. *Summary of estimates of parametric functions and its estimate of the variance*

(n,r)	k	UMVUE of $(1-p)^n + \left(\frac{p\,\theta^2}{1+\theta}\right)^k [1-(1-p)^{n-k}]$ (UMVUE of its variance)
(50,15)	1	0.05033 (0.00040)
	2	0.00253 (4.830e-06)
	3	0.00013 (3.503e-08)
(50,25)	1	0.08319 (0.00058)
	2	0.00694 (1.817e-05)
	3	0.00058 (3.557e-07)
(50,40)	1	0.13395 (0.00078)
	2	0.01791 (5.938e-05)
	3	0.00239 (2.613e-06)
(50,50)	1	0.16742 (0.00085)
	2	0.02796 (0.00010)
	3	0.00467 (6.802e-06)

The UMVU estimate of probability density function and survival function for the above model for various combinations are shown in Table 4.3. The entries in brackets are the UMVUE of variances of estimates.

Table 4.3. *Summary of estimates of pdf and survival functions*

(n,r)	k	UMVUE of PDF	UMVUE of survival function
(50,15)	2	0.05545 (0.00224)	0.18435 (0.00226)
	4	0.03394 (0.00098)	0.09478 (0.00126)
	6	0.01745 (0.00030)	0.04473 (0.00056)
	8	0.00823 (7.626e-05)	0.02006 (0.00020)
(50,25)	2	0.09192 (0.00588)	0.30668 (0.00299)
	4	0.05642 (0.00262)	0.15800 (0.00189)
	6	0.02908 (0.00077)	0.07471 (0.00089)
	8	0.01375 (0.00019)	0.03355 (0.00032)
(50,40)	2	0.14768 (0.01478)	0.49108 (0.00294)
	4	0.09040 (0.00660)	0.25251 (0.00254)
	6	0.04647 (0.00191)	0.11921 (0.00130)
	8	0.02194 (0.00046)	0.05350 (0.00049)
(50,50)	2	0.18440 (0.02282)	0.61175 (0.00218)
	4	0.11274 (0.01019)	0.31403 (0.00277)
	6	0.05786 (0.00293)	0.14793 (0.00153)
	8	0.02725 (0.00070)	0.06620 (0.00058)

Table 4.3 shows that as the value of k increases, the estimate of the pdf and survival function decreases and UMVUE of the variance of the estimate become very small for every combination of n and r. The UMVU estimates of the pdf and survival function increases as the number of inliers decreases.

4.6. Data analysis

Here we discuss two examples from Appendix. The first example is based on the dataset A.5 of SO_2 concentration $\left(\mu_g/m^3\right)$ in air for industrial area of Amritsar collected from 1 April 2017 to 30 April 2017 at 5:00 A.M. It is observed that SO_2 concentration in air is zero for three days of month April 2017, industrial area of Amritsar. The second example is based on 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. Both the example was fitted for Lindley's inlier model. The density and survival plots along with the parameter estimates are presented below.

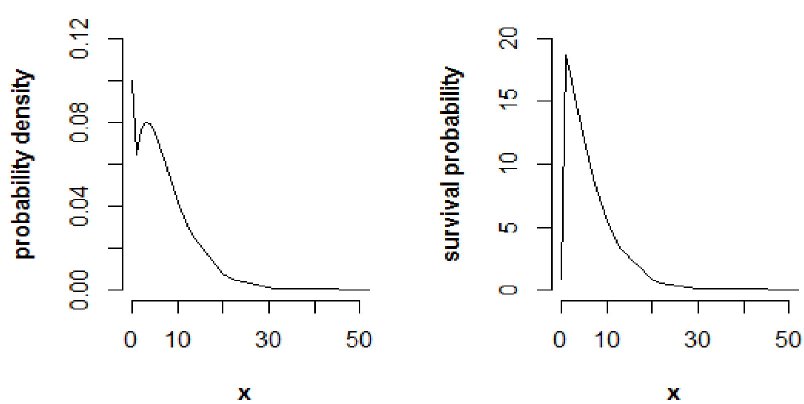

Figure 4.3. *pdf and survival function of SO_2 concentration in air data (p =0.9, θ=0.2365)*

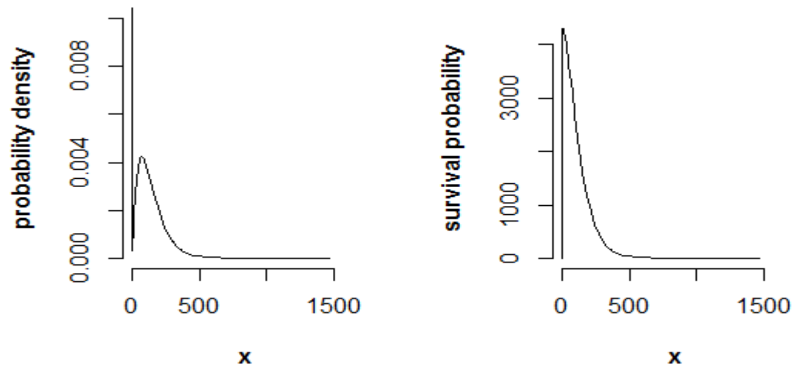

Figure 4.4. *pdf and survival function of NFHS-3 data (p =0.82759, θ=0.01396)*

Figures 4.3 shows the density function and the survival function curve for dataset in A.5 and Figures 4.4 show the density function and the survival function curve for dataset in A.6.

The summary of estimates of the model for some selected values of parameters of the above two examples is shown in Table 4.4. The entry in brackets is the UMVUE of standard error of the estimate.

Table 4.4. *Summary of estimates of the models*

Estimator	Estimates (SE)	
	SO_2 concentration in air data	NFHS-3 data
$\hat{p}_{MLE}$ & $\widehat{SE}(\hat{p}_{MLE})$	0.90000 (0.05477)	0.82759 (0.04050)
MLE of $\hat{\theta}_{MLE}$ & $\widehat{SE}(\hat{\theta}_{MLE})$	0.23650 (0.03248)	0.01396 (0.00116)
95% CI for p	(0.79265, 1.00735)	(0.74821, 0.90696)
95% CI for θ	(0.17284 0.30017)	(0.01168, 0.01624)
UMVUE of $p(x=0)$, $G_1(z,r,n)$	0.10000 (0.05571)	0.17241 (0.04073)
UMVUE of parametric function, $H_1(z,r,n)$	0.11014 (0.02320)	0.00571 (0.00074)
UMVUE of pdf x = 2, $\phi_2(z,r,n)$	0.05991 (0.05909)	0.01689 (0.01385)
UMVUE of survival function at time t, $\hat{S}(t)$		
$\hat{S}(1)$	0.84850 (0.05389)	0.05988 (0.01844)
$\hat{S}(2)$	0.77960 (0.05487)	0.01422 (0.00650)
$\hat{S}(3)$	0.70248 (0.05870)	0.00318 (0.00189)