Chapter-6

Inliers estimation in generalized failure distributions

6.1 Introduction

Generalized distributions are not frequently used for modeling life data as the life testing distribution but they have the ability to mimic the attributes of other distributions such as the exponential, Weibull or lognormal, based on the values of the distribution's parameters. Generalized exponential distribution has a right skewed unimodal density function and monotone hazard function similar to the density functions and hazard functions of the gamma and Weibull distributions. It is observed that the bivariate generalized exponential distribution provides a better fit than the bivariate exponential distribution. While the generalized gamma distribution is not often used to model life data by itself, its ability to behave like other more commonly-used life distributions is sometimes used to determine which of those life distributions should be used to model a particular set of data. It is observed that it can be used quite effectively to analyze lifetime data in place of gamma, Weibull and log-normal distributions. The genesis of this model is different estimation procedures and their properties, estimation of the stress-strength parameter, closeness of this distribution to some of the well known distribution functions, etc will be studied in this chapter from the inlier observations perspective.

6.2 Instantaneous Failures

As usual to accommodate the possibility of instantaneous failures, the class of generalized failure time distribution (GFTD) $\Im=\{F(x,\theta),\theta\in\Omega\}$ is modified to a new distribution $\mathcal{G}=\{G(x,\theta,p)=(1-p)+pF(x,\theta),F\in\Im,x\geq 0,0<p<1\}$ where $f(x,\theta)$ is of the form

$$f(x,\theta,\beta)=\left(\frac{\phi(x)}{\theta}\right)^{\beta}\left[\frac{\phi'(x)}{\phi(x)}\right]\frac{1}{\overline{|\beta}}\exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right),\quad \phi(x)>0,\theta,\beta>0 \qquad (6.2.1)$$

One may refer to Johnson and Kotz, Johnson and Balakrishnan (1970) etc. for other version of generalized densities. The above density is studied by Chaturvedi and Usha (2008).

6.2.1 Maximum likelihood estimation in instantaneous failures

The modified general failure time density function is given as

$$g(x,p,\theta,\beta)=\begin{cases}1-p, & \phi(x)=0\\ \frac{p}{\overline{|\beta}}\left(\frac{\phi(x)}{\theta}\right)^{\beta}\left[\frac{\phi'(x)}{\phi(x)}\right]\exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right), & \phi(x)>0\end{cases} \qquad (6.2.2)$$

Let $X_1,X_2,......X_n$, be a random sample of size n from $g\in\mathcal{G}$.

$$L(x,p,\theta,\beta)=\prod_{i=1}^{n}g(x_i,p,\theta,\beta)$$

Define

$$Z(x)=\begin{cases}1, & x=0\\ 0, & x>0\end{cases}$$

Then the likelihood is given by

$$L(x;p,\theta,\beta)=\prod_{i=1}^{n}(1-p)^{z(x_i)}\left[pf(x_i,\theta,\beta)\right]^{1-z(x_i)}$$

$$=(1-p)^{\sum z(x_i)}\left(\frac{p}{\theta^{\beta}\lceil\beta}\right)^{n-\sum z(x_i)}\prod_{x_i>0}\left[\{\phi(x)\}^{\beta}\left(\frac{\phi'(x)}{\phi(x)}\right)\exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right)\right]^{1-z(x_i)}$$

It is possible to show that (6.2.2) is a member of three parameter exponential family with $\left(\sum_{i=1}^{n}z(x_i),\sum[1-z(x_i)]\ln\{\phi(x_i)\},\sum[1-z(x_i)]\{\phi(x_i)\}\right)$ are jointly complete sufficient for (p,β,θ), provided $\phi(x)$ is real valued and strictly increasing function of *x* with $\phi(0)=0$ and its inverse function exists.

The estimating equations are constructed from the log likelihood and are given by

$$\frac{\partial\ln L}{\partial p}=-\frac{\sum z(x_i)}{1-p}+\frac{n-\sum z(x_i)}{p}=0 \qquad (6.2.3)$$

$$\frac{\partial\ln L}{\partial\beta}=-\left[n-\sum z(x_i)\right]\left(\frac{\partial\ln(\lceil\beta)}{\partial\beta}+\ln\theta\right)+\sum[1-z(x_i)]\ln\{\phi(x_i)\}=0 \qquad (6.2.4)$$

$$\frac{\partial\ln L}{\partial\theta}=\frac{\beta}{\theta}\sum z(x_i)-\frac{n\beta}{\theta}+\frac{1}{\theta^2}\sum[1-z(x_i)]\phi(x_i)=0 \qquad (6.2.5)$$

Since the equation (6.2.3) is independent of θ and β, one can solve and get $\hat{p}=\frac{n-r}{n}$, if $\sum z(x_i)=r$. The estimates $\hat{\theta}$ and $\hat{\beta}$ are obtained by solving (6.2.4) and (6.2.5) which are the conditional likelihood equations given (*n-r*) positive observations. One can also obtain the Fisher information as the expectation of second derivative of the likelihood equations above once the form of $\phi(x)$ is known.

6.3 Early failures

To accommodate the possibility of instantaneous and early failures the class of generalized failure time distribution (GFTD) $\Im=\{F(x,\theta),\theta\in\Omega\}$ is modified to distribution $\mathcal{G}_1=\{G_1(x,\theta,p)=(1-p)+pF(\delta,\theta)+pf(x,\theta),F\in\Im,x\geq 0,0<p<1\}$. The failure time correspond to early failures which are reported as δ which is very very small and hence the modified model will be a mixture in the proportion 1-*p* and *p*. The estimation procedure for the parameters involved in the model. The modified generalized failure time distribution is given by distribution function

$$\mathcal{G}_1(x,p,\theta)=\begin{cases}0 & x<\delta\\ 1-p+pF(\delta,\theta), & x=\delta\\ pf(x,\theta), & x>\delta\end{cases} \tag{6.3.1}$$

which can be simplified as

$$g_1(x,p,\theta)=\begin{cases}0 & x<\delta\\ 1-p\bar{F}(\delta,\theta), & x=\delta\\ pf(x,\theta), & x>\delta\end{cases} \tag{6.3.2}$$

6.3.2 Maximum likelihood estimation in early failures

On substituting the modified general failure time distribution is given as

$$g_1(x,p,\theta,\beta)=\begin{cases}1-p\bar{F}(\delta,\theta,\beta), & \phi(x)=\delta\\ \frac{p}{\Gamma\beta}\left(\frac{\phi(x)}{\theta}\right)^{\beta}\left[\frac{\phi'(x)}{\phi(x)}\right]\exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right), & \phi(x)>\delta\end{cases} \tag{6.3.3}$$

Let $X_1,X_2,......X_n$, be a random sample of size *n* from $g_1\in\mathcal{G}_1$

$$L(x,p,\theta,\beta)=\prod_{i=1}^{n}g_1(x_i,p,\theta,\beta)$$

then define

$$Z(x)=\begin{cases}1, & x=\delta \\ 0, & x>\delta\end{cases}$$

$$L(x,p,\theta,\beta)=\prod_{i=1}^{n}\left(1-p\bar{F}\{\delta,\theta,\beta\}\right)^{z(x_i)}\left[pf(x_i,\theta,\beta)\right]^{1-z(x_i)}$$

$$=\left(1-p\bar{F}(\delta,\theta,\beta)\right)^{\sum z(x_i)}\left(\frac{p}{\theta^{\beta}\overline{|\beta}}\right)^{n-\sum z(x_i)}\prod_{x_i>\delta}\left[\{\phi(x)\}^{\beta}\left(\frac{\phi'(x)}{\phi(x)}\right)\exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right)\right]^{1-z(x_i)}$$

Here again it is possible to show that (6.3.3) is a member of three parameter exponential family with $\left(\sum_{i=1}^{n}z(x_i),\sum\left[1-z(x_i)\right]\ln\{\phi(x_i)\},\sum\left[1-z(x_i)\right]\{\phi(x_i)\}\right)$ are jointly complete sufficient for (p,β,θ), provided $\phi(x)$ is real valued and strictly increasing function of x with $\phi(\delta)=0$ and its inverse function exists. The estimating equations are constructed from the log likelihood and are given by

$$\frac{\partial \ln L}{\partial p}=-\frac{\sum z(x_i)\bar{F}(\delta,\theta,\beta)}{(1-p)\bar{F}(\delta,\theta,\beta)}+\frac{n-\sum z(x_i)}{p}=0 \tag{6.3.4}$$

$$\frac{\partial \ln L}{\partial \beta}=-\frac{p\sum z(x_i)}{(1-p)\bar{F}(\delta,\theta,\beta)}\frac{\partial \bar{F}(\delta,\theta,\beta)}{\partial \beta}-\left[n-\sum z(x_i)\right]\frac{\partial \ln\left(\overline{|\beta}\right)}{\partial \beta}$$
$$+\ln\theta+\sum\left[1-z(x_i)\right]\ln\{\phi(x_i)\}=0 \tag{6.3.5}$$

$$\frac{\partial \ln L}{\partial \theta}=-\frac{p\sum z(x_i)}{(1-p)\bar{F}(\delta,\theta,\beta)}\frac{\partial \bar{F}(\delta,\theta,\beta)}{\partial \theta}+\frac{\beta}{\theta}\sum z(x_i)-\frac{n\beta}{\theta}$$
$$+\frac{1}{\theta^2}\sum\left[1-z(x_i)\right]\phi(x_i)=0 \tag{6.3.6}$$

Solving (6.3.4) one gets

$$\hat{p}=\frac{(n-r)}{n\bar{F}(\delta,\theta,\beta)} \tag{6.3.7}$$

Equation (6.3.5) and (6.3.6) have to be solved simultaneously using numerical iterative method to obtain the estimation of parameters under study .

6.4 Nearly instantaneous failure

Let $F(x)$ and $R(x)=1-F(x)$ denote the cumulative distribution function and the survival function of the mixture, respectively. *F* is continuous and its density be given by $f(x)=F'(x)$.The component distribution functions and their survival functions are $F_i(x)$ and $R_i(x)=1-F_i(x)$ respectively, i=1,2. The hazard rate of a lifetime distribution is defined as $h(x)=f(x)/R(x)$ provided the density exists. Instead of assuming an instant or an early failures to occur at a particular point, as in the original model as above, we now represent this model as a mixture of the generalized Dirac delta function and the generalized failure time. Thus the resulting modification gives rise to a density function:

$$f(x)=(1-p)\delta_d(x-x_0)+\frac{p}{\Gamma\beta}\left(\frac{\phi(x)}{\theta}\right)^{\beta}\left[\frac{\phi'(x)}{\phi(x)}\right]\exp\left(-\left[\frac{\phi(x)}{x}\right]\right) ,$$

where

$$p+q=1,\ 0<p<1,\ \phi(x)>0,\ \beta>0. \qquad (6.4.1)$$

and

$$\delta_d(x-x_0)=\begin{cases}\frac{1}{d}, & x_0\le x<x_0+d\\ 0, & otherwise\end{cases} , \qquad (6.4.2)$$

for sufficiently small *d*. Here $p>0$ is the mixing proportion. Also note that

$$\delta(x-x_0)=\lim_{d\to 0}\delta_d(x-x_0) \qquad (6.4.3)$$

where $\delta(\cdot)$ is the Dirac delta function as given in section (2.4) of chapter 2. Both the distribution and survival functions are continuous.

Writing

$$f_1(x)=\delta_d(x-x_0) \text{ and } f_2(x)=\frac{1}{\lceil\beta}\left(\frac{\phi(x)}{\theta}\right)^{\beta}\left[\frac{\phi'(x)}{\phi(x)}\right]\exp\left(-\left[\frac{\phi(x)}{x}\right]\right)$$

Then (6.4.1) can be written as

$$f(x)=q f_1(x)+p f_2(x) \quad \text{where} \quad p+q=1,\ 0<p<1 \tag{6.4.4}$$

so that

$$F(x)=qF_1(x)+pF_2(x) \tag{6.4.5}$$

the corresponding survival function is

$$R(x)=1-F(x)=q+p-qF_1(x)-p F_2(x)=qR_1(x)+pR_2(x) \tag{6.4.6}$$

and the hazard function of the mixture distribution is

$$h(x)=\frac{qf_1(x)+pf_2(x)}{qR_1(x)+pR_2(x)} \tag{6.4.7}$$

Now using above results, in terms of density function of particular distribution, given in equation (6.2.2) one can obtain various characteristics.

6.4.1 Characteristics of the model

The life time models are generally characterised in terms of its hazard rate function, survival function and the mean residual life functions. Below we obtain these characteristics and obtain some useful relationship between them. The reliability (survival) functions of the respective component distributions are given by

$$R_1(x)=\begin{cases}1, & 0\le x<x_0\\ \dfrac{d+x_0-x}{d}, & x_0\le x\le x_0+d\\ undefined, & t\ge t_0+d\end{cases} \tag{6.4.8}$$

and

$$R_2(x)=\bar{F}_2(x) \tag{6.4.9}$$

The hazard rates are, respectively,

$$h_1(x)=\begin{cases}0, & 0\le x<x_0\\ \dfrac{1}{d+x_0-x}, & x_0\le x\le x_0+d\\ \infty, & x\ge x_0+d\end{cases} \tag{6.4.10}$$

and

$$h_2(x)=\frac{f_2(x)}{\bar{F}_2(x)} \tag{6.4.11}$$

It can be shown (6.4.10) and (6.4.11) that for any mixture of two continuous distributions the hazard rate function can be expressed as

$$h(x)=\frac{f(x)}{R(x)}=w(x)h_1(x)+[1-w(x)]h_2(x) \tag{6.4.12}$$

where $w(x)=q\,R_1(x)/R(x)$ for all $x\ge 0$. In our case,

$$w(x)=\begin{cases}\dfrac{q}{R(t)}, & 0\le x<x_0\\ \dfrac{qR_1(x)}{R(x)}, & x_0\le x\le x_0+d\\ 0, & x\ge x_0+d\end{cases} \tag{6.4.13}$$

Establishing some interesting relationship between the survival function and hazard function through $w(x)$ as follows:

Since

$$w(x)=q\,R_1(x)/R(x)$$

$$w'(x)=\frac{q\left[R_1'(x)R(x)-R_1(x)R'(x)\right]}{\left[R(x)\right]^2}$$

upon substituting the value of $R(x)$ from above and simplifying, we get

$$w'(x)=\frac{pq\left[R_1'(x)R_2(x)-R_1(x)R_2'(x)\right]}{\left[R(x)\right]^2}$$

If the terms, are rearranged one gets

$$w'(x)=\frac{pqR_1(x)R_2(x)\left[\frac{R_1'(x)}{R_1(x)}-\frac{R_2'(x)}{R_2(x)}\right]}{\left[R(x)\right]^2}.$$

Now recall,

$$w(x)=\frac{qR_1(x)}{R(x)},1-w(x)=\frac{pR_2(x)}{R(x)},h_1(x)=-\frac{R_1'(x)}{R_1(x)} \text{ and } h_2(x)=-\frac{R_2'(x)}{R_2(x)}$$

hence

$$w'(x)=w(x)\left[1-w(x)\right]\left\{h_2(x)-h_1(x)\right\} \tag{6.4.14}$$

in a similar way, one can show that

$$h'(x)=w'(x)h_1(x)+w(x)h_1'(x)-w'(x)h_2(x)+\left[1-w(x)\right]h_2'(x) \tag{6.4.15}$$

also, since $f_1(x)=-R_1(x)$, one gets

$$w(x)\,h_1(x)=q\frac{R_1(x)}{R(x)}\frac{f_1(x)}{R_1(x)}=\frac{qf_1(x)}{R(x)}$$

which shows that, (6.4.12) is well defined for all $x > 0$. Thus the summarized expression for $R(x), h(x)$ and $m(x)$, are respectively, given as

$$R(x)=\begin{cases} q+p\bar{F}(x), & 0<x<x_0 \\ \frac{q[d+x_0-x]}{d}+p\bar{F}(x), & x_0\le x\le x_0+d \\ p\bar{F}(x), & x>x_0+d \end{cases} \tag{6.4.16}$$

$$h(x)=\begin{cases}\left[\dfrac{p\bar{F}(x)f_2(x)}{q+p\bar{F}(x)}\right], & 0\le x\le x_0\\ \dfrac{q+dp\bar{F}(x)f_2(x)}{q(d-x)+dp\bar{F}(x)}, & x_0\le x\le x_0+d\\ f_2(x), & x>x_0+d\end{cases} \tag{6.4.17}$$

The mean residual life (MRL) of a random variable X defined for all x as

$$m_X(x)=E(X-x/X>x)=\frac{\int_x^{\infty} R_X(y)dy}{R_X(x)}$$

This is the expected additional time to failure given survival to x, which can also be expressed in terms of mixture of two MRL's as

$$m(x)=qm_1(x)+pm_2(x) \tag{6.4.18}$$

where

$$m_1(x)=\begin{cases}\dfrac{x_0-x}{2}, & 0\le x<x_0\\ \dfrac{x_0+d-x}{2}, & x_0\le x<x_0+d\\ 0, & x>x_0+d\end{cases} \tag{6.4.19}$$

and

$$m_2(x)=\frac{\int_x^{\infty}\bar{F}_2(y)dy}{\bar{F}_2(x)}, \qquad y>d \tag{6.4.20}$$

6.4.2 Particular Case When ($X_0 = 0$)

Consider a special case of model (6.4.1) whereby $x_0 = 0$. The model may be called the model with "nearly instantaneous failure". In this case, (6.4.3) is simplified giving the hazard rate of the uniform distribution as

$$h_1(x)=\begin{cases}\frac{1}{d-x}, & 0\le x\le d\\ \infty, & x>d\end{cases} \tag{6.4.21}$$

and its survival rate function is given as

$$R_1(x)=\begin{cases}\frac{d-x}{d}, & 0\le x\le d\\ 0, & x>d\end{cases} \tag{6.4.22}$$

Thus the generalized model with "nearly instantaneous failure" occurring uniformly over [0, *d*] has the survival function

$$R(x)=\begin{cases}\frac{q(d-x)}{d}+p\bar{F}_2(x), & 0\le x\le d\\ p\bar{F}_2(x), & x>d\end{cases} \tag{6.4.23}$$

and the hazard function as

$$h(x)=\begin{cases}\frac{q+dpf_2(x)}{q(d-x)+dp}, & 0\le x\le x_0+d\\ f_2(x), & x>x_0+d\end{cases} \tag{6.4.24}$$

One can study the above characteristics by plotting graphs, with various combinations of values of parameters.

6.5 Testing of hypothesis

Here the interest is to test the hypothesis, whether sample observations belong to inliers population against hypothesis that it belongs to target population. Refer equation (1.6.1), the hypothesis can be written as

$$H_0:\xi=\phi \text{ versus } H_0:\xi\ne\phi. \tag{6.5.1}$$

where ξ is the common population parameter under study. Below we discuss various computationally simple test procedures to detect inliers in a model.

6.5.1 Sequential Probability Ratio Test (SPRT) to detect inliers in the model

SPRT to test the hypothesis whether a observation belongs to inlier population with p.d.f. $g(x,\phi)$ against hypothesis that it belongs to target population with p.d.f. $f(x,\theta)$. i.e. equation (6.5.1).

The likelihood when H_1 is true , is given by

$$L_1 = \prod_{i=1}^{r} f(x_i,\theta)$$

and under H_0, it is

$$L_0 = \prod_{i=1}^{r} g(x_i,\phi)$$

And likelihood ratio λ_r is given by $\lambda_r = \frac{L_1}{L_0}$ or equivalently

$$\ln\lambda_r = \sum_{i=1}^{r} \ln\frac{f(x_i,\theta)}{g(x_i,\phi)} = \sum_{i=1}^{r} z_i \qquad (6.5.2)$$

For deciding number of inliers we continue to take ordered observations one by one till we reject H_0. That is

if $\sum_{i=1}^{r} z_i \le \log B$ accept H_0 and take the next observation.

and

if $\sum_{i=1}^{r} z_i \ge \log A$ reject H_0 and stop.

The corresponding m represents the first observation from target observation and hence $\hat{r} = r-1$ are the number of inliers. And $B=\frac{\gamma}{1-\alpha}$, $A=\frac{1-\gamma}{\alpha}$, where α represents probability of type I error and γ represents probability of type II error. Now the SPRT procedure is investigated for following special cases.

Case 1: Testing for scale parameter when shape parameter $\beta_0 = \beta_1 = b$.

To test: $H_0:\theta=\theta_0$ against $H_1:\theta=\theta_1$

θ_0 and θ_1 are the scale parameters of inlier distribution and target distribution respectively.

The test statistics is

$$\ln\lambda_{r\theta}=r\beta_0\left[\ln\theta_0-\ln\theta_1\right]+\sum_{i=1}^{r}\phi(x)\left[\frac{1}{\theta_0}-\frac{1}{\theta_1}\right] \tag{6.5.3}$$

Reject H_0 when

$$\sum_{i=1}^{r}\phi(x)>\frac{\ln A-r\beta_0\left[\ln\theta_0-\ln\theta_1\right]}{\left[\frac{1}{\theta_0}-\frac{1}{\theta_1}\right]} \tag{6.5.4}$$

Case 2: Testing for shape parameter when scale parameter $\theta_0=\theta_1=\theta$.

To test : $H_0:\beta=\beta_0$ against $H_1:\beta=\beta_1$

β_0 and β_1 are the shape parameters of inliers distribution and target distributions respectively.

The test statistics is

$$\ln\lambda_{r\beta}=\left[r\beta_0-r\beta_1\right]\left\{\ln\theta-\sum_{i=1}^{r}\ln\phi(x)\right\}+r\left[\ln\overline{\beta_0}-\ln\overline{\beta_1}\right] \tag{6.5.5}$$

Reject H_0 when

$$\sum_{i=1}^{r}\ln\phi(x)>\frac{\ln A-r\left[\ln\overline{\beta_0}-\ln\overline{\beta_1}\right]}{r\left[\beta_1-\beta_0\right]}+\ln\theta \tag{6.5.6}$$

6.5.2 Most Powerful Test

For the hypothesis as defined in equation (6.5.1), the most powerful test to reject H_0 is given by

$$\psi(x)=\begin{cases}1, & \dfrac{P_1(x)}{P_0(x)}>C\alpha \\ 0, & \dfrac{P_1(x)}{P_0(x)}<C\alpha\end{cases} \tag{6.5.7}$$

Where $P_1(x)$ and $P_0(x)$ are likelihood functions under distribution of target population $\mathfrak{I}$ and inlier population $\mathcal{G}$ respectively $C\alpha$ is such that test attains level of the test when H_0 is true. We reject H_0 for large values of the ratio $\dfrac{P_1(x)}{P_0(x)}$.

Case 1: Testing for scale parameter when shape parameter $\beta_0=\beta_1=b$.

To test $H_0:\theta=\theta_0$ against $H_1:\theta=\theta_1$ where the parameters are as defined before in section (6.5.1).

The most powerful test is given by

$$\psi(x)=\begin{cases}1, & \sum_{i=1}^{n}\phi(x)>\dfrac{C_\alpha-n\beta[\log\theta_0-\log\theta_1]}{\left[\dfrac{1}{\theta_0}-\dfrac{1}{\theta_1}\right]} \\ 0, & otherwise.\end{cases} \tag{6.5.8}$$

Case 2: Testing for shape parameter when scale parameter $\theta_0=\theta_1=\theta$.

To test : $H_0:\beta=\beta_0$ against $H_1:\beta=\beta_1$

The most powerful test is given by

$$\psi(x)=\begin{cases}1, & \sum_{i=1}^{n}\log\phi(x)>\dfrac{C_\alpha-n\left[\log\lceil\beta_0-\log\lceil\beta_1\right]}{n[\beta_1-\beta_0]}+\log\theta \\ 0, & otherwise.\end{cases} \tag{6.5.9}$$

$C\alpha$ is such that test attains level of the test when H_0 is true. Once $C\alpha$ is obtained we can find power function under alternative hypothesis.

6.6 Information Criterion

Again three information criteria which are already discussed in section (2.5) are used such as Schwarz's Information criterion (*SIC= -2lnL(Θ) + p ln n),* Schawarz's Bayesian Information criterion (*BIC= -lnL(Θ) + 0.5(p ln n)/n*) and Hannan-Quinn criterion (*HQ= -lnL(Θ) + p ln[ln(n)]*) to detect number of inliers. *L*(*Θ*) represents the maximum likelihood function and *p* is the number of free parameters that need to be estimated under the model.

Below we develop the procedure for *SIC* scheme. We consider the model of no inliers as Model *SIC*(0), where all the observations are from target population. Model *SIC*(*r*) will denote *r* observations are inlier and remaining (*n-r*) observations are from target population. Our aim is to obtain number of inliers in the sample. For density in equation (6.2.2) the model with zero inlier is given by

$$SIC(0)=2n\beta_1\ln\theta_1-2(\beta_1-1)\sum_{i=1}^{n}\ln\phi(x_i)+2n\ln\overline{|\beta_1}+\frac{2\sum_{i=1}^{n}\phi(x_i)}{\theta_1}+2\ln n \qquad (6.6.1)$$

and the model with *r* inliers is as

$$SIC(r)=2\beta_0 r\ln\theta_0-2(\beta_0-1)\sum_{i=1}^{r}\ln\phi(x_i)+2n\ln\overline{|\beta_0}+\frac{2\sum_{i=1}^{r}\phi(x_i)}{\theta_1}+2\beta_1(n-r)\ln\theta_1$$

$$-2(\beta_1-1)\sum_{i=r+1}^{n}\ln\phi(x_i)+2(n-r)\ln\overline{|\beta_1}+\frac{2\sum_{i=r+1}^{n}\phi(x_i)}{\theta_1}+4\ln n \qquad (6.6.2)$$

According to the procedure, the model(0) is selected with no inliers if $SIC(0)<\min_{1\le r\le n-1} SIC(r)$. And the model(r) is selected if $SIC(0)>\min_{1\le r\le n-1} SIC(r)$. Similarly we can find criteria for *BIC* and *HQ*.

6.7 Estimation and test for specific distributions

One can obtain, life distribution, such as, exponential, gamma, Weibull and Rayleigh distribution by substituting appropriate form of the parameters.

6.7.1 Exponential model

If $\phi(x)=x$ and $\beta=1$ then (6.2.2) reduces to a one parameter exponential distribution with the MLE under instantaneous failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i}{n-r}. \tag{6.7.1}$$

The MLE under early failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{ne^{-\frac{x}{\theta}}}, \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>\delta} x_i}{n-r}-\delta. \tag{6.7.2}$$

The test criteria to test H_0: There are no inliers in data set against a single inlier is present in the data from exponential distribution is given by

Reject H_0

$$\text{if} \quad \frac{x_{(1)}}{\sum_{i=1}^{n} x_{(i)}}<C \tag{6.7.3}$$

where C is to be chosen such that

$$P\left(\frac{x_{(1)}}{\sum_{i=1}^{n} x_{(i)}}<C\right)=\alpha,$$

where α is the size of the test.

6.7.2 Rayleigh model

If $\phi(x)=x^2$ and $\beta=1$ then (6.2.2) reduces to a one parameter Rayleigh distribution , and the MLE under instantaneous failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i^2}{n-r}. \tag{6.7.4}$$

The MLE under early failure model for $\hat{p}$ and $\hat{\theta}$ is given as

$$\hat{p}=\frac{n-r}{n}e^{\frac{x^2}{\theta}} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>\delta} x_i^2}{n-r}-\delta^2 \tag{6.7.5}$$

To test H_0 : all observations are from Rayleigh distribution with parameter θ against a single inlier is present in the data, is given by

Reject H_0

if $$\frac{x_{(1)}^2}{\sum_{i=1}^{n} x_{(i)}^2}<C \tag{6.7.6}$$

where C is to be chosen such that we attain the size of the test under null hypothesis.

6.7.3 Weibull model

If $\phi(x)=x^b$ and $\beta=1$ then (6.2.2) reduces to a two parameter Weibull distribution. The MLE under instantaneous failure model for p, b and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i^b}{n-r}. \tag{6.7.7}$$

and for the estimate of b one has to solve the following equation

$$\frac{(n-r)}{b}+\sum_{x_i>0}\ln x_i-\frac{(n-r)\sum_{x_i>0} x_i^b \ln x_i}{\sum_{x_i>0} x_i^b}=0 \tag{6.7.8}$$

similarly the MLE under early failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{n}e^{\frac{x^b}{\theta}} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i^b}{n-r}-\delta^b \tag{6.7.9}$$

and for parameter b one has to solve the following equation

$$\frac{(n-r)\delta-\sum_{x_i>\delta} x_i^b \ln x_i}{\sum_{x_i>\delta} x_i^b-(n-r)\delta^b}+\frac{1}{b}+\frac{\sum_{x_i>\delta}\ln x_i}{(n-r)}=0 \tag{6.7.10}$$

The test for presence of single inlier in Weibull family is derived in section (6.7.4.1).

6.7.4 Gamma model

If $\phi(x)=x$ then (6.2.2) reduces to a two parameter Gamma distribution with the MLE under instantaneous failure model for p, β and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i}{(n-r)\beta} \tag{6.7.11}$$

where for β one has to solve the equation

$$\frac{(n-r)}{\beta}-(n-r)\frac{\partial \ln(\lceil\beta)}{\partial\beta}+\ln\theta+\sum_{x_i>0}\ln x_i=0 \tag{6.7.12}$$

The MLE under early failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{n\bar{F}(\delta,\theta,\beta)} \tag{6.7.13}$$

For getting the estimates of θ,β the following two equations are obtained, which are to be solved simultaneously

$$\frac{rp}{1-p\bar{F}(\delta,\theta,\beta)}\frac{\partial\bar{F}(\delta,\theta,\beta)}{\partial\beta}-(n-r)\frac{\partial\ln(\lceil\beta)}{\partial\beta}+\ln\theta+\sum_{x_i>0}\ln x_i=0 \tag{6.7.14}$$

$$\frac{rp}{1-p\bar{F}(\delta,\theta,\beta)}\frac{\partial\bar{F}(\delta,\theta,\beta)}{\partial\theta}-(n-r)\frac{\beta}{\theta}+\frac{\sum_{x_i>\delta}x_i}{\theta^2}=0 \quad (6.7.15)$$

where

$$\bar{F}(\delta,\theta,\beta)=1-\frac{x^{\beta}}{\theta^{\beta}}\frac{\left(\Gamma\beta-\Gamma\left(\beta,\frac{x}{\theta}\right)\right)}{\Gamma\beta}.$$

The test for single inlier in Gamma distribution is equivalent to that of exponential distribution.

6.7.4.1 Testing for one inlier in Weibull family

Consider the problem of $H_0: r=0$ (i.e. no inliers) versus $H_0: r=1$ (i.e. one inlier) in data with Weibull distribution. The joint pdf under H_0 is given by L_0 and under H_1 is given by L_1. Hence

$$L_1=c\phi\theta^{n-1}\exp\left(-\left[\frac{x_{(1)}}{\phi}+\frac{\sum_{i=2}^{n}x_{(i)}}{\theta}\right]\right) \quad (6.7.12)$$

and

$$L_0=c\theta^{n}\exp\left(-\left[\frac{\sum_{i=2}^{n}x_{(i)}}{\theta}\right]\right) \quad (6.7.13)$$

We already know that $\hat{\theta}=\frac{\sum_{1=1}^{n}x_{(i)}}{n}$ under H_0 where as under H_1 $\hat{\theta}=\frac{\sum_{1=2}^{n}x_{(i)}}{(n-1)}$ and $\hat{\phi}=x_{(1)}$.

Substituting the above values in equations (6.7.12) and (6.7.13), one can obtain likelihood ratio test as

$$\frac{L_0}{L_1} < C \quad \Rightarrow \quad \frac{\left(\sum_{i=2}^{n} x_{(i)}^{\beta}\right)^{n-1}}{\left(\sum_{i=1}^{n} x_{(i)}^{\beta}\right)^{n}} x_{(1)}^{\beta} < C \tag{6.7.14}$$

$$= \frac{\left(\sum_{i=1}^{n} x_{(i)}^{\beta} - x_{(1)}^{\beta}\right)^{n-1}}{\left(\sum_{i=1}^{n} x_{(i)}^{\beta}\right)^{n}} x_{(1)} < C$$

$$= \frac{\left(T - x_{(1)}^{\beta}\right)^{n-1}}{(T)^{n-1}} \frac{x_{(1)}^{\beta}}{T} < C \quad \text{where} \quad T = \sum_{1=1}^{n} x_{(i)}^{\beta}.$$

The test is to reject H_0 if

$$\frac{x_{(1)}^{\beta}}{T} < C \tag{6.7.15}$$

where C is to be chosen such that

$$P\left(\frac{x_{(1)}^{\beta}}{T} < C\right) = \alpha,$$

and α is the size of the test. For the simulated data from Weibull (0.02, 5) the values of C for various size of the test are obtained in table (6.7.1). Using these C values power of the test have been obtained in table (6.7.2). For computation of power the data is simulated from Weibull (0.001, 1).

Table 6.7.1. Values of C

α	n			
	10	30	50	100
0.01	0.00065	0.00023	0.00015	7.64E-05
0.025	0.00095	0.00027	0.00016	8.14E-05
0.05	0.00149	0.00029	0.00017	8.44E-05
0.10	0.00290	0.00051	0.0002	9.09E-05
0.25	0.00670	0.00089	0.00038	0.0001
0.95	0.02343	0.00324	0.00119	0.00036
0.99	0.03555	0.00499	0.00189	0.00051

Table 6.7.2. Power of the test

n			
10	30	50	100
0.004	0.005	0.007	0.008
0.036	0.035	0.046	0.047
0.083	0.073	0.095	0.098
0.197	0.234	0.242	0.245
0.423	0.467	0.492	0.494
0.888	0.934	0.945	0.947
0.978	0.991	0.989	0.99

The power in above table are found using C values obtained in table (6.7.1).

6.8 Application

The data, collected by Amutha and Porchelvan (2009), represents monthly rainfall (in mm) during year 2004 and 2006 for the estimation of surface runoff in

Malattar Sub-watershed in Andhra Pradesh. The watershed experiences tropical monsoon climate with normal temperature, humidity and evaporation throughout the year. Runoff is one of the important hydrologic variables used in water resources applications and management planning. For gauged watershed accuracy of estimation of runoff on land surface and river requires much time and effort.

Set 1 (2004) : 3.40, 0.00, 0.00, 15.80, 232.80, 8.80, 123.20, 47.00, 154.00, 103.20, 89.80 and 12.20.

Set 2 (2006) : 0.00, 0.00, 21.40, 60.20, 53.86, 93.20, 27.80, 45.40, 205.40, 101.20, 128.20 and 0.00.

Using Kolmogorov-Smirnov test, we have come to the conclusion that exponential distribution fits well to above set 1 and set 2. Hence the analysis for the data set 1 and 2 is conducted for Exponential and Rayleigh distribution. Estimates of parameters with their standard error are calculated for instantaneous failure, early failures and nearly instantaneous model shown in the tables below.

Table 6.8.1. Instantaneous Failures

Distribution	Parameter	Set 1		Set 2	
		Estimates	Standard Error	Estimates	Standard Error
	$\hat{p}$	0.83333	2.68328	0.75	0.012217
Exponential	$\hat{\theta}$	0.01265	0.00400	2.30941	0.003863
Rayleigh	$\hat{\theta}$	0.000166	5.24224E-05	0.000103011	3.43E-05

Table 6.8.2. Early Failures

Distribution	Parameter	Set 1 $\delta = 20$		Set 2 $\delta = 30$	
		Estimates	Standard Error	Estimates	Standard Error
	$\hat{p}$	0.50000	2.00000	0.58333	2.02837
Exponential	$\hat{\phi}$	0.09950	0.04062	0.04065	0.016595
	$\hat{\theta}$	0.00799	0.00325	0.010182	0.004157
Rayleigh	$\hat{\phi}$	0.008205	0.00259	0.001624	0.001149
	$\hat{\theta}$	0.00010	3.170E-05	8.126E-05	3.071E-05

Table 6.8.3. Nearly Instantaneous Failures

Distribution	Parameter	Set 1 δ=20		Set 2 δ=21.4	
		Estimates	Standard Error	Estimates	Standard Error
Exponential	$\hat{\theta}$	0.01481	0.00428	0.01583	0.00457
Rayleigh	$\hat{\theta}$	0.000104	3.3E-05	0.000136	3.95E-05

From above table we can clearly observe that Rayleigh distribution fits better to above data sets.

6.9 Future Prospects

We have considered the Bayesian approach to inliers problem only for exponential model in this chapter. Also considered in this chapter is the inlier estimation of mixture of two different distributions from exponential family. This is further extended for mixture of any two life testing distributions when inliers are encountered. Bayesian method for estimation of parameters of mixture distribution of inliers and target population, assuming distribution other than exponential is also explored. It is possible to have observations as inliers, target and outliers, thus leading to mixture of three densities. The estimation procedure for such a model is challenging. We will be pursuing this study in future.