Chapter 2

Inlier estimation in Pareto distribution

2.1 Introduction

Pareto distribution has recently been used as a model for file sizes on the internet, insurance losses, and financial behavior of the stock market as well as in telecommunication systems. Many of the empirical studies also use Pareto's law for representing long tail distributions. The proposed study is aimed to look further for the suitability of Pareto distributions in the context of life testing experiments where data involves instantaneous and early failures. The occurrence of instantaneous or early failures in life testing experiment is a phenomenon observed in electronic components as well as in clinical trials. These occurrences may be due to inferior quality or faulty construction or due to no response of the treatments. Such failures usually discard the assumption of a unimodal distribution and hence the usual method of modeling and inference procedures may not be accurate in practice. These situations can be handled by modifying used parametric model Pareto distribution. The modified model is then a non-standard mixture of distribution by mixing a singular distribution at zero to accommodate instantaneous failures.

Consider a model $\Im=\{F(x;\theta),x\geq 0,\theta\in\Omega\}$ where $F(x,\theta)$ is a continuous failure time distribution function (df) with $F(0)=0$. To accommodate a real life situation, where instantaneous failures are observed at the origin, the model $\Im$ is modified to $\mathcal{G}$ = $\{G(x;\alpha,\theta)=(1-\alpha)+\alpha f(x,\theta),0<\alpha<1,F\in\Im\}$ by using a mixture in the proportion 1-α and α respectively of a singular random variable *Z* at zero and with a random variable X with df $F\in\Im$. The df corresponding to G∈ $\mathcal{G}$ is given by

$$\mathrm{G}(x,\theta,\alpha) = (1-\alpha)+\alpha\mathrm{F}(x,\theta) \tag{2.1.1}$$

Thus the modified failure time distribution will have its corresponding probability density function (pdf) as

$$g(x;\theta,\alpha)=\begin{cases}1-\alpha, & x=0\\ \alpha f(x,\theta), & x>0\end{cases} \tag{2.1.2}$$

The problem of inference about (α,θ) has received considerable attention particularly when X is exponential with mean θ. Some of the early works are by Aitchison (1955), Kleyle and Dahiya (1975), Jayade and Prasad (1990), Vannman (1991), Muralidharan (1999, 2000), Kale and Muralidharan (2000) and references contained therein. Vannman (1995) and Muralidharan and Kale (2002) considered the case where *F* is a two parameter Gamma distribution with shape parameter β and scale parameter θ and obtained confidence interval for $\phi=\alpha\beta\theta$ assuming α known and unknown respectively.

To accommodate early failures, the family $\Im$ is modified to $\mathcal{G}_1$ = $\{G_1(x,\theta,\alpha),\ x\geq 0,\ \theta\in\Omega,\ 0<\alpha<1\}$ where the df corresponding to $G_1\in\mathcal{G}_1$ is given by

$$G_1(x,\theta,\alpha)=(1-\alpha)H(x)+\alpha F(x,\theta) \tag{2.1.3}$$

where $H(x)$ is a df with $H(\delta)$ = 1 for δ sufficiently small and assumed to be known and specified in advance. We also assume that the early failures are recorded as a class with notional failure time δ so that the modified family G_1 has a pdf with reference to measure μ which is sum of Lebesgue measure on (δ, ∞) and a singular measure at δ. The corresponding pdf is then given by

$$g_1(x,\alpha,\theta) = \begin{cases} 0, & x < \delta \\ 1 - \alpha + \alpha F(\delta,\theta), & x = \delta \\ \alpha f(x,\theta), & x > \delta \end{cases} \quad (2.1.4)$$

Some of the references which treat early failure analysis with exponential distribution are Kale and Muralidharan (2000), Kale (2001), Kale and Muralidharan (2002), and Muralidharan and Lathika (2006), wherein they treat early failures as inliers using the sample configurations.

The objective is to consider the model G given by (2.1.1), G_1 given by (2.1.3) and nearly instantaneous failures when F is Pareto and study the suitability of Pareto distribution in the context of life testing experiments. The Pareto distribution was originally derived in connection with studying income distribution. The Pareto distribution is a power-tailed distribution which is a special case of a heavy-tailed distribution whose tails go to zero more slowly than exponential. Many of the empirical studies also use Pareto's law for representing long tail distributions. The distribution also comes in various forms and types. Hence modeling differences between one parameter, two parameters and three parameters Pareto will be a point of interest. Fisher, Masi, Gross and Shortle (2005) have studied the modeling difference of such different forms of Pareto distribution in connection with queuing systems. A three parameter Pareto type family has the survival function

$$\bar{F}(x) = \left(\frac{\beta}{x+\beta-\gamma}\right)^{\phi}, x \geq \gamma, \phi > 0, \beta > 0, \gamma > 0 \quad (2.1.5)$$

or the more general form

$$\bar{F}(x) = \frac{\beta^{\phi}}{\beta+(x-\gamma)^{\phi}}, x \geq \gamma, \phi > 0, \beta > 0, \gamma > 0 \quad (2.1.6)$$

A two parameter Pareto can be easily obtained as a particular case of the above distributions for $\gamma = 0$. The other forms of Pareto can be easily obtained for particular cases of β and ϕ.

We study two types of Pareto distribution in the context of instantaneous failures and early failures. From the point of view of estimating equations, Kale and Muralidharan (2000) have shown that $I_g^{(\alpha)}(\theta)$, Fisher information about θ ignoring α in the model G is less than $I_f(\theta)$, the Fisher information about θ in the original model $\Im$. It is also shown that the parameter α is orthogonal to θ in the case of model (2.1.1), whereas, the parameter α is not orthogonal to θ in the case of model (2.1.3). It is possible to show $I_g^{(\alpha)}(\theta) < I_f(\theta)$ although Var (X|g) can be smaller than Var (X|f) in both the models. In the subsequent sections, different types of Pareto distributions have been used with different parameters for analysis. The general theory of estimating equations and Fisher information's for instantaneous failures and early failures have been developed in the next two sections separately. We also discuss the importance of instantaneous and early failures in practical situations through a real life data set obtained by Vannman (1991).

2.2 Analysis for instantaneous failures

In this section we study inference regarding instantaneous failure. We have obtained UMVUE, Fishers information and MLE for the parameters of inlier and target population.

2.2.1 Fisher information

The pdf in (2.1.2) is with respect to the measure $\mu(x)$ which is the sum of Lebesgue measure over $(0,\infty)$ and a singular measure at {0}. If we assume $f\left(x_{(i)},\theta\right)$ as a Cramer family, then ln $[g(x,\theta,\alpha)]$ admits continuous partial derivatives with respect to $(\alpha,\theta)'$ upto order two. Here θ can also be a vector of parameters. Further,

$\int_0^{\infty} g(x,\theta,\alpha)d\mu = 1$, can be differentiated twice under integral sign with respect to $(\alpha,\theta)'$. Hence G satisfies all the regularity conditions of Cramer (1966) and G is a Cramer family. Therefore from (2.1.2),

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

$$\frac{\partial \ln g}{\partial \theta} = \begin{cases} 0, & x = 0 \\ \frac{\partial \ln f(x,\theta)}{\partial \theta}, & x > 0 \end{cases}$$

One can verify that $E\left(\frac{\partial \ln g}{\partial \alpha}\right) = 0$ and $E\left(\frac{\partial \ln g}{\partial \theta}\right) = 0$. The element of the Fisher information matrix, $I_g(\alpha,\theta)$ are

$$I_{\alpha\alpha} = E\left(-\frac{\partial^2 \ln g}{\partial \alpha^2}\right) = \frac{1}{\alpha(1-\alpha)} \tag{2.2.1}$$

$$I_{\theta\theta} = E\left(-\frac{\partial^2 \ln g}{\partial \theta^2}\right) = \alpha I_f(\theta) \tag{2.2.2}$$

and

$$I_{\alpha\theta} = I_{\theta\alpha} = E\left(-\frac{\partial \ln g}{\partial \beta}\frac{\partial \ln g}{\partial \theta}\right) = \int_{S-\{0\}} \frac{\partial \ln f}{\partial \theta}\alpha f(x,\theta)dx = 0 \tag{2.2.3}$$

Hence $I_g(\alpha,\theta)$ = $diag\left(\frac{1}{\alpha(1-\alpha)}, \alpha I_f(\theta)\right)$, which shows that α and θ are orthogonal parameters. Using the definition of Fisher information for θ ignoring α in model $g \in \mathcal{G}$ as given by Liang (1983), denoted by $I_g^{(\alpha)}(\theta)$, we have

$$I_g^{(\alpha)}(\theta) = I_{\theta\theta} - I_{\theta\alpha} I_{\alpha\alpha}^{-1} I_{\alpha\theta} \tag{2.2.4}$$

$$= I_{\theta\theta} = \alpha I_f(\theta) \quad \text{as} \quad I_{\alpha\theta} = I_{\theta\alpha} = 0$$

Since $0 < \alpha < 1$, $I_g^{(\alpha)}(\theta) < I_f(\theta)$ and there is less information about θ ignoring α in the model $\mathcal{G}$ than that in the model $\Im$.

2.2.2 Maximum likelihood estimation.

Now let ($X_1, X_2, \ldots, X_n$) be a random sample of size n from $g \in G$ and define

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

Then the likelihood function can be written as

$$\begin{aligned} L(x;\alpha,\theta) &= \prod_{i=1}^{n} g(x_i,\alpha,\theta) \\ &= \prod_{i=1}^{n} (1-\alpha)^{z(x_i)} [\alpha f(x_i,\theta)]^{1-z(x_i)} \\ &= (1-\alpha)^{\sum z(x_i)} \alpha^{n-\sum z(x_i)} \prod_{x_i>0} f(x_i,\theta) \end{aligned} \qquad (2.2.5)$$

If $\sum_{i=1}^{n} z(x_i) = n_0$, then the likelihood equations are given by

$$\frac{\partial \ln L}{\partial \alpha} = \frac{-n_0}{1-\alpha} + \frac{n-n_0}{\alpha} = 0 \qquad (2.2.6)$$

and

$$\frac{\partial \ln L}{\partial \theta} = \sum_{x_i>0} \frac{\partial \ln f(x_i,\theta)}{\partial \theta} = 0 \qquad (2.2.7)$$

then from (2.2.6), we have $\hat{\alpha} = \frac{n-n_0}{n}$ and $\hat{\theta}$ will be the solution of (2.2.7). Using the standard results on MLE, we have $\frac{1}{n} I_g^{-1}(\alpha,\theta) = diag\left(\frac{\alpha(1-\alpha)}{n}, \frac{1}{n\alpha I_f(\theta)} \right)$.

The pdf of one parameter Pareto is defined as below

$$f(x)=\frac{\beta x^{\beta-1}}{(1+x^{\beta})^2}, x>0, \beta>0$$

$$\int_0^\infty \left[\frac{x^{\beta}(\ln x)^2}{(1+x^{\beta})^2}\right] f(x)dx=\frac{1}{\beta^2}\frac{(\pi^2-6)}{18}$$

where the log likelihood is

$$\ln L = r\ln(1-\alpha)+(n-r)[\ln\alpha+\ln\beta]+(\beta-1)\sum_{x_i>0}\ln x_i-2\sum_{x_i>0}\ln\left(1+x_i^{\beta}\right)$$

and the Fisher information's are

$$I_{\alpha\alpha}=E\left(-\frac{\partial^2 \ln L}{\partial\alpha^2}\right)=\frac{1}{\alpha(1-\alpha)}$$

$$I_{\alpha\beta}=I_{\beta\alpha}=E\left(-\frac{\partial^2 \ln L}{\partial\alpha\partial\beta}\right)=0$$

and

$$I_{\beta\beta}=E\left(-\frac{\partial^2 \ln L}{\partial\beta^2}\right)=\frac{\alpha}{\beta^2}\left[1+\frac{(\pi^2-6)}{9}\right].$$

The pdf for two parameter Pareto is as given below

$$f(x)=\frac{\beta}{\phi}\left(\frac{x}{\phi}\right)^{\beta-1}\frac{1}{\left[1+(x/\phi)^{\beta}\right]^2},\ x\geq 0,\ \phi>0, \beta>0$$

For some computations below, we use the following formulas:

$$\int_0^\infty \left[\frac{(x/\phi)^{\beta}}{\{1+(x/\phi)^{\beta}\}^2}\right] f(x)dx=\frac{1}{6}$$

$$\int_0^\infty \left[\frac{(x/\phi)^{2\beta}}{\{1+(x/\phi)^{\beta}\}^2}\right] f(x)dx=\frac{1}{3}$$

$$\int_0^\infty \left[\frac{(x/\phi)^{\beta}\ln(x/\phi)}{\{1+(x/\phi)^{\beta}\}^2}\right] f(x)dx=-\frac{\phi}{6\beta}[\ln(\beta/\phi)]$$

and

$$\int_0^{\infty}\left[\frac{(x/\phi)^{\beta}\ln(x/\phi)^2}{\{1+(x/\phi)^{\beta}\}^2}\right]f(x)dx=\frac{1}{\beta^2}\frac{(\pi^2-6)}{18}.$$

The log likelihood is

$$\ln L=r\ln(1-\alpha)+(n-r)\left[\ln\alpha+\ln\beta-\beta\ln\phi\right]+(\beta-1)\sum_{x_i>0}\ln x_i-2\sum_{x_i>0}\ln\left[1+\frac{x_i}{\phi}\right]^{\beta}$$

and the Fisher information's are

$$I_{\alpha\alpha}=E\left(-\frac{\partial^2\ln L}{\partial\alpha^2}\right)=\frac{1}{\alpha(1-\alpha)}$$

$$I_{\alpha\beta}=I_{\beta\alpha}=E\left(-\frac{\partial^2\ln L}{\partial\alpha\partial\beta}\right)=0$$

$$I_{\beta\beta}=E\left(-\frac{\partial^2\ln L}{\partial\beta^2}\right)=\frac{\alpha}{\beta^2}\left[1+\frac{(\pi^2-6)}{9}\right]$$

$$I_{\alpha\phi}=I_{\phi\alpha}=E\left(-\frac{\partial^2\ln L}{\partial\alpha\partial\phi}\right)=0$$

$$I_{\phi\beta}=I_{\beta\phi}=E\left(-\frac{\partial^2\ln L}{\partial\phi\partial\beta}\right)=\frac{\alpha}{\phi}-\frac{\alpha}{3\phi}\left[4-\phi\ln\left(\frac{\beta}{\phi}\right)\right]$$

and

$$I_{\phi\phi}=E\left(-\frac{\partial^2\ln L}{\partial\phi^2}\right)=\frac{\alpha\beta^2}{3\phi^2}$$

The above computations for both criteria's are done for Vannman's example given in section (2.8).

2.2.3 Uniformly Minimum Variance Unbiased Estimator (UMVUE)

One can obtain UMVUE of mixture density of instantaneous and positive observation taken from Pareto distribution using the method discussed in Singh (2007). Based on above families we define a new family of df $\Im=\{F(x;\theta,\alpha):x\geq 0,\theta\in\Omega,0<\alpha<1\}$ such that

$$f(x;\theta,\alpha)=\begin{cases}1-\alpha+\alpha f(x;\theta), & x=0\\ \alpha f(x;\theta), & x>0\end{cases}$$

Hence the pdf of mixture family is obtained as

$$f(x;\theta,\alpha)=(1-\alpha)^{\rho}\alpha^{(1-\rho)}\left[\frac{1}{(1+x)}\frac{\theta}{(1+x)^{\theta}}\right]^{(1-\rho)}$$

$$=\frac{\left(\frac{1}{1+x}\right)^{(1-\rho)}\left[\exp(-\theta)\right]^{(1-\rho)\ln(1+x)}\left(\frac{1-\alpha}{\alpha\theta}\right)^{\rho}}{\left(\frac{1}{\alpha\theta}\right)} \tag{2.2.8}$$

which is a member of exponential family with $a(x)=\left(\frac{1}{1+x}\right)$, $h(\theta)=\exp(-\theta)$, $g(\theta)=\frac{1}{\theta}$ and $d(x)=\ln(1+x)$. We have $z=\sum_{x>0}(1-\rho)\ln(1+x)$ and $n-r=\sum_{x>0}\rho_j$ which are jointly complete sufficient statistics for (θ,p). Since $\ln(1+x)$ has exponential distribution with parameter θ.

The UMVUE of mixture density is given by

$$\zeta_x(z,r,n)=\begin{cases}\frac{B(z,r,n-1)}{B(z,r,n)}=\frac{n-r}{n}, & x=0, r=0,1,2....n-1\\ a(x)\frac{B(z-d(x),r-1,n-1)}{B(z,r,n)}, & x>0, z>d(x), r=1,2....n\end{cases}$$

where

$$B(z,r,n)=\begin{cases}\binom{n}{r}B(z|r), & z=r,r+1.....;\ r=1,2....n\\ 1, & z=0,\ r=0\end{cases}$$

and $B(z|r)$ is such that

$$\left[b(\theta)-a(0)\right]^{r}=\sum_{z=r}^{\infty}B(z|r)\theta^{r},\ r=1,2,....n.$$

The above expression simplifies to

$$\zeta_x(z,r,n)=\begin{cases}\dfrac{n-r}{n}, & x=0\\ \dfrac{r(r-1)}{nzx}\left[1-\dfrac{\ln(1+x)}{z}\right]^{r-2}, & x>0\end{cases} \tag{2.2.9}$$

which is UMVUE of mixture density of instantaneous failure and positive observation taken from Pareto distribution.

2.3 Analysis for early failures

If early failures are nominally reported as $X=\delta$ then the df of the modified model G_1 is given as

$$G_1(x,\alpha,\theta)=\begin{cases}0, & x<\delta\\ 1-\alpha+\alpha F(\delta,\theta), & x=\delta\\ 1-\alpha+\alpha F(x,\theta), & x>\delta\end{cases} \tag{2.3.1}$$

The corresponding pdf is given as

$$g_1(x,\alpha,\theta)=\begin{cases}0, & x<\delta\\ 1-\alpha+\alpha F(\delta,\theta), & x=\delta\\ \alpha f(x,\theta), & x>\delta\end{cases} \tag{2.3.2}$$

The Fisher informations can be obtained as

$$I_{\alpha\alpha}=\frac{1-F(\delta,\theta)}{\alpha[1-\alpha+\alpha F(\delta,\theta)]} \tag{2.3.3}$$

$$I_{\theta\theta}=\alpha\left[I_f(\theta)-\int_0^{\delta}\left(\frac{\partial \ln f}{\partial\theta}\right)^2 f(x,\theta)dx\right] \tag{2.3.4}$$

and

$$I_{\alpha\theta}=I_{\theta\alpha}=\frac{-\frac{\partial}{\partial\theta}F(\delta,\theta)}{[1-\alpha+\alpha F(\delta,\theta)]} \tag{2.3.5}$$

where $I_f(\theta)$ is the Fisher information about θ in the original pdf $f(x,\theta)$. Again using (2.3.4), we get the Fisher information about θ ignoring α as

$$I_{g_1}^{(\alpha)}(\theta)=\frac{\alpha\left[I_f(\theta)-\int_0^{\delta}\left(\frac{\partial \ln f}{\partial\theta}\right)^2 f(x,\theta)dx-\left(\frac{\partial F(\delta,\theta)}{\partial\theta}\right)^2\right]}{1-\alpha+\alpha F(\delta,\theta)[1-F(\delta,\theta)]}. \tag{2.3.6}$$

Here one can see that the parameters α and θ are not orthogonal. Also as $0<\alpha<1$, $I_{g_1}^{(\alpha)}(\theta)<I_f(\theta)$. If the n observations $X_1,X_2,\ldots,X_n$ are from $g_1\in\mathcal{G}_1$, then the likelihood is

$$L(x,\alpha,\theta)=[1-\alpha+\alpha F(\delta,\theta)]^{n_0}\,\alpha^{n-n_0}\prod_{x_i>\delta} f(x_i,\theta).$$

Then the ML estimates are the solutions of the following likelihood equations:

$$\frac{\partial \ln L}{\partial\alpha}=\frac{-n_0[1-F(\delta,\theta)]}{1-\alpha+\alpha F(\delta,\theta)}+\frac{n-n_0}{\alpha}=0 \tag{2.3.7}$$

and

$$\frac{\partial \ln L}{\partial\theta}=\frac{-n_0\frac{\partial}{\partial\theta}F(\delta,\theta)}{1-F(\delta,\theta)}+\sum_{x_i>\delta}\frac{\partial \ln f(x_i,\theta)}{\partial\theta}=0 \tag{2.3.8}$$

Equation (2.3.8) does not depend on α and hence one can obtain $\hat{\theta}$ from (2.3.8). Using this $\hat{\theta}$ in (2.3.7) we can obtain $\hat{\alpha}$. Again,

$$L(x,\alpha,\theta)=[1-\alpha+\alpha F(\delta,\theta)]^{n_0}\left(\alpha[1-F(\delta,\theta)]\right)^{n-n_0}\prod_{x_i>\delta}\frac{f(x_i,\theta)}{1-F(\delta,\theta)}$$

That is, the likelihood of the sample under $g_1\in\mathcal{G}_1$ is the product of the likelihoods of n_0 and the conditional likelihood of the sample given n_0 which is same as the likelihood of $(n\text{-}n_0)$ observations coming from the truncated version of $f\in\mathfrak{F}$ or $(g_1\in\mathcal{G}_1)$ restricted to (δ,∞). Now n_0 is binomial with probability of success given by $1\text{-}\alpha\text{+}\alpha F(\delta,\theta)$. For fixed θ and $\alpha\in[0,1]$ this binomial family is complete. Therefore, the optimal estimating equation for θ ignoring α is the conditional score function given n_0

or $\frac{\partial \ln L_{n_0}}{\partial \theta}=0$, where $L_{n_0}=\prod_{x_i>\delta}\frac{f(x_i,\theta)}{1-F(\delta,\theta)}$. Hence optimal estimating equation for θ ignoring α is given by (2.3.8). Thus $\frac{\partial \ln L_{n_0}}{\partial \theta}$ or $\hat{\theta}$ is same as the estimator given by optimal estimating equation for θ ignoring α.

For some computations of one parameter Pareto family defined in section (2.1), we have the following formulas:

$$\int_{\delta}^{\infty}\left[\frac{x^{\beta}\ln x^{2}}{\{1+x^{\beta}\}^{2}}\right]f(x)dx=\frac{1+3u+(1+\ln u)(2\ln u+6u\ln u)}{12(1+u)^{3}},\ u=\delta^{\beta}.$$

The log likelihood of early failure in one parameter Pareto model is

$$\ln \mathrm{L}=\ r\ln\left[1-\frac{\alpha}{(1+\delta^{\beta})}\right]+(n-r)[\ln\alpha+\ln\beta]+(\beta-1)\sum_{x_i>\delta}\ln x_i-2\sum_{x_i>\delta}\ln(1+x_i^{\beta})$$

and the Fisher Information's are

$$I_{\alpha\alpha}=E\left(-\frac{\partial^{2}\ln L}{\partial\alpha^{2}}\right)=\frac{1}{\alpha(1-\alpha+u)},$$

$$I_{\alpha\beta}=I_{\beta\alpha}=E\left(-\frac{\partial^{2}\ln L}{\partial\alpha\partial\beta}\right)=-\frac{u\ln\delta}{(1+u)(1-\alpha+u)},$$

and

$$I_{\beta\beta}=E\left(-\frac{\partial^{2}\ln L}{\partial\beta^{2}}\right)=\frac{\alpha}{\beta^{2}}\left[\frac{1}{1+u}+\left\{\frac{1+3u+(1+\ln u)(2\ln u+6u\ln u)}{6(1+u)^{3}}\right\}\right]$$
$$-\frac{\alpha u(\ln\delta)^{2}}{(1+u)^{3}(1-\alpha+u)}[(1+u)(1-\alpha+u)-u(2+2u-\alpha],$$

where $u=\delta^{\beta}$.

For two parameters Pareto family as defined in section (2.1) we have to use the following formulas:

$$\int_{\delta}^{\infty}\left[\frac{(x/\phi)^{\beta}}{\{1+(x/\phi)^{\beta}\}^{2}}\right]f(x)dx=-\frac{1}{3[1+v]^{3}}+\frac{1}{2[1+v]^{2}},$$

$$\int_{\delta}^{\infty}\left[\frac{(x/\phi)^{2\beta}}{\{1+(x/\phi)^{\beta}\}^{2}}\right]f(x)dx=\frac{1}{[1+v]}-\frac{1}{[1+v]^{2}}+\frac{1}{3[1+v]^{3}},$$

and

$$\int_{\delta}^{\infty}\left[\frac{(x/\phi)^{\beta}\ln(x/\phi)^{2}}{\{1+(x/\phi)^{\beta}\}^{2}}\right]f(x)dx=\frac{1}{6\beta}\left[-\frac{v(1+v)+v(3+v)\ln v}{(1+v)^{3}}+\ln(1+v)\right],$$

where $v=(\delta/\phi)^{\beta}$.

The log likelihood of early failures in two parameters Pareto model is

$$\ln L= r\ln\left[1-\frac{\alpha}{1+(\delta/\phi)^{\beta}}\right]+(n-r)[\ln\alpha+\ln\beta-\beta\ln\phi]+(\beta-1)\sum_{x_i>\delta}\ln x_i$$
$$-2\sum_{x_i>\delta}\ln\left[1+(x_i/\phi)^{\beta}\right]$$

Then Fisher information equations corresponding to two parameters Pareto models are as given below

$$I_{\alpha\alpha}=E\left(-\frac{\partial^2\ln L}{\partial\alpha^2}\right)=\frac{1}{\alpha(1-\alpha+v)}$$

$$I_{\alpha\beta}=I_{\beta\alpha}=E\left(-\frac{\partial^2\ln L}{\partial\alpha\partial\beta}\right)=-\frac{v\ln(\delta/\phi)}{(1+v)(1-\alpha+v)}$$

$$I_{\beta\beta}=E\left(-\frac{\partial^2\ln L}{\partial\beta^2}\right)=\frac{1}{\beta^2}\left[\frac{1}{1+v}+\left\{\frac{1+3v+(1+\ln v)(2\ln v+6v\ln v)}{6(1+v)^3}\right\}\right]$$
$$-\frac{\alpha v[\ln(\delta/\phi)]^2}{(1+v)^3(1+v-\alpha)}[(1+v-\alpha)(1+v)-v(2+2v-\alpha)]$$

$$I_{\alpha\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\alpha\partial\phi} \right) = \frac{\beta v}{\phi(1+v)^2}$$

$$I_{\beta\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\beta\partial\phi} \right) = \frac{1}{3\phi}\left[\frac{v(1+v+v(3+v)\ln v)}{(1+v)^3} - \frac{3(2v^2+3v+1)}{(1+v)^3} - \ln(1+v) \right]$$

$$+ \frac{\alpha v[(1+v-\alpha)(1+v)\{1+\beta\ln(\delta/\phi)\} - \beta v\ln(\delta/\phi)(2+2v-\alpha)]}{\phi(1+v-\alpha)(1+v)^3} - \frac{1}{\phi(1+v)}$$

and

$$I_{\phi\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\phi^2} \right) = \frac{\beta}{\phi^2}\left[-\frac{1}{(1+v)} + \frac{(2+v)(\beta+1)}{6(1+v)^3} + \frac{2\{3(1+v^2)+5v\}}{3(1+v)^3} + \frac{2\alpha\beta v\{(1+v-\alpha)(1+v)(1+\beta) - \beta v(2+2v-\alpha)\}}{\phi^2(1+v-\alpha)(1+v)^3} \right],$$

where $v = (\delta/\phi)^{\beta}$.

The above information is used in illustration given in section (2.8) with comparative study of instantaneous failures and early failures are presented for different situations.

2.4 Nearly instantaneous model

As already discussed in chapter 1 nearly instantaneous model incorporates inliers in better way than the above two models.

2.4.1 Representation of the model

Let $F(x)$ and $R(x) = 1 - F(x)$ denote the cumulative distribution function and the survival function of the mixture, respectively. We assume that 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 failure rate of a lifetime distribution is defined as $h(x) = f(x)/R(x)$ provided the density exists.

We now represent this model as a mixture of the generalized Dirac delta function and the 2-parameter Pareto as opposed to a mixture of a singular distribution with Pareto, as

$$f(x)=p\delta_d(x-x_0)+q\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(x/\beta\right)^{\alpha}\right]^{-2} \quad ,p+q=1, 0<p<1 \ ,x>0 \quad ,\alpha>0,\beta>0 \tag{2.4.1}$$

where

$$\delta_d(x-x_0)=\begin{cases}\frac{1}{d}, & x_0\le x\le x_0+d\\ 0, & otherwise\end{cases} \tag{2.4.2}$$

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

$$\delta(x-x_0)=\lim_{d\to 0}\delta_d(x-x_0) \tag{2.4.3}$$

where $\delta(\cdot)$ is the Dirac delta function. We may view the Dirac delta function as approximately normal distribution having a zero mean and standard deviation that tends to 1. For fixed value of *d*, equation (2.4.2) denotes a uniform distribution over an interval [x_0, x_0+ *d*] so the modified model is now effectively a mixture of a Pareto with a uniform distribution. Instead of including a possible instantaneous failure in the model (2.4.2) is allowed for a possible "nearly instantaneous" failure to occur uniformly over a very small time interval. Note that the case $x_0 = 0$ corresponds to instantaneous failures, whereas $x_0 \neq 0$ (but small) corresponds to the case with early failures. Noting from (2.4.1) and (2.4.2), we see that the mixture density function is not continuous at x_0 and x_0+ *d*. However, both the distribution and survival functions are continuous. Writing

$$f_1(x)=\delta_d(x-x_0) \text{ and } f_2(x)=\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(x/\beta\right)^{\alpha}\right]^{-2} \quad \alpha,\beta>0$$

can be written as

$$f(x)=pf_1(x)+qf_2(x) \quad \text{where } p+q=1,\ 0<p<1 \tag{2.4.4}$$

So

$$F(x)=pF_1(x)+qF_2(x) \tag{2.4.5}$$

and

$$R(x)=1\text{-}F(x)=p+q-[pF_1(x)+qF_2(x)]=pR_1(x)+qR_2(x) \tag{2.4.6}$$

Thus, the failure (hazard) rate function of the mixture distribution is

$$h(x)=\frac{pf_1(x)+qf_2(x)}{pR_1(x)+qR_2(x)} \tag{2.4.7}$$

A mixture distribution involving two 2-parameter Weibull distribution has been thoroughly studied by Lai, Khoo, Murlidharan (2007). The mixture considered was more complex in the sense that one of the mixing distributions has a finite range which poses some challenges. Simulated observations from this model are made by generating uniform variates and Pareto variates with proportions p and $q=1-p$ respectively.

2.4.2 Survival function, failure rate and mean residual life function of the nearly instantaneous model

Recently, failure rates of mixtures are discussed quite extensively. The Reliability (survival) functions of the respective component distributions are given by

$$R_1(x)=\begin{cases}1, & 0\le x<x_0\\ \dfrac{d+x_0-x}{d}, & x_0\le x\le x_0+d\\ 0, & x\ge x_0+d\end{cases} \tag{2.4.8}$$

and

$$R_2(x)=\frac{1}{1+\left(x/\beta\right)^{\alpha}} \tag{2.4.9}$$

The failure rates are, respectively,

$$h_1(x)=\begin{cases} 0, & 0\le x<x_0 \\ \dfrac{1}{d+x_0-x}, & x_0\le x\le x_0+d \\ \infty, & x\ge x_0+d \end{cases} \tag{2.4.10}$$

and

$$h_2(x)=\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(x/\beta\right)^{\alpha}\right]^{-1} \tag{2.4.11}$$

It can be shown (2.4.4) and (2.4.6) that for any mixture of two continuous distributions the failure rate function can be expressed as

$$h(x)=\frac{f(x)}{R(x)}=w(x)h_1(x)+[1-w(x)]h_2(x) \tag{2.4.12}$$

where $w(x)=pR_1(x)/R(x)$ for all $x\ge 0$. In our case,

$$w(x)=\begin{cases} \dfrac{p}{R(x)}, & 0\le x<x_0 \\ \dfrac{pR_1(x)}{R(x)}, & x_0\le x\le x_0+d \\ 0, & x\ge x_0+d \end{cases} \tag{2.4.13}$$

with

$$w'(x)=w(x)\left[1-w(x)\right]\{h_2(x)-h_1(x)\} \tag{2.4.14}$$

Also a simple differentiation shows that

$$h'(x)=w'(x)h_1(x)+w(x)h_1'(x)-w'(x)h_2(x)+\left[1-w(x)\right]h_2'(x) \tag{2.4.15}$$

Now $w(x)h_1(x)=\dfrac{pR_1(x)}{R(x)}\dfrac{f_1(x)}{R_1(x)}=\dfrac{pf_1(x)}{R(x)}$, so (2.4.12) is well defined for all $x > 0$. Summarized expression for $R(x)$, $h(x)$ and $m(x)$ are, respectively, given as

$$R(x)=pR_1(x)+qR_2(x)$$

$$R(x)=\begin{cases} p+q\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}, & 0\le x<x_0 \\ \frac{p[d+x_0-x]}{d}+q\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x_0\le x\le x_0+d \\ q\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x>x_0+d \end{cases} \tag{2.4.16}$$

Recall that *h*(*x*) is discontinuous at both $x=x_0$ and $x=x_0+d$.

$$h(x)=\begin{cases} \left[\frac{q\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}}{p+q\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}}\right]\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}, & 0\le x\le x_0 \\ \frac{p+dq\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-2}}{p(d-x)+dq\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}}, & x_0\le x\le x_0+d \\ \alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x>x_0+d \end{cases} \tag{2.4.17}$$

Then the Mean residual life of an r.v. X is defined for all *x* as

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

This is the expected additional time to failure given survival to *x*.

$$m(x)=pm_1(x)+qm_2(x) \tag{2.4.18}$$

where

$$m_1(x)=\begin{cases}\dfrac{x_0-x}{2}, & 0\le x<x_0\\ \dfrac{x_0+d-x}{2}, & x_0\le x<x_0+d\\ 0, & x>x_0+d\end{cases} \tag{2.4.19}$$

$$m_2(x)=\frac{\displaystyle\int_x^{\infty}\frac{1}{\left[1+\left(y/\beta\right)^{\alpha}\right]}dy}{\dfrac{1}{\left[1+\left(x/\beta\right)^{\alpha}\right]}},\qquad y>x_0+d \tag{2.4.20}$$

2.4.3 Nearly instantaneous failure case ($x_0 = 0$)

Consider a special case of model (2.4.1) whereby $x_0=0$. The model may be called the Pareto with "nearly instantaneous failure" model. In this case, (2.4.10) is simplified giving the failure rate of the uniform distribution as

$$h_1(x)=\begin{cases}\dfrac{1}{d-x}, & o\le x\le d\\ \infty, & x>d\end{cases} \tag{2.4.21}$$

and corresponding to (2.4.8) its survival rate function is given as

$$R_1(x)=\begin{cases}\dfrac{d-x}{d}, & o\le x\le d\\ 0, & x>d\end{cases} \tag{2.4.22}$$

The Pareto model with "nearly instantaneous failure" occurring uniformly over [0, d] has

$$R(x)=\begin{cases}\dfrac{p(d-x)}{d}+q\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & 0\le x\le d\\ q\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x>d\end{cases} \tag{2.4.23}$$

and

$$h(x)=\begin{cases}\dfrac{p+dq\alpha\beta^{\alpha}x^{\alpha-1}\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-2}}{p(d-x)+dq\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}}, & 0\le x\le x_0+d\\ \alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x>x_0+d\end{cases} \tag{2.4.24}$$

We now present some graphical plots of Survival, Density and Failure Rate Functions. Graphical plots are important for ageing distributions. Some graphs are plotted to identify whether the model is useful for specific datasets for which empirical plots are available. All plots are done when $x_0 = 0$, the Pareto with "nearly" instantaneous failure model. A plot of density function, Survival function and MRL functions for various values of *p* are given below.

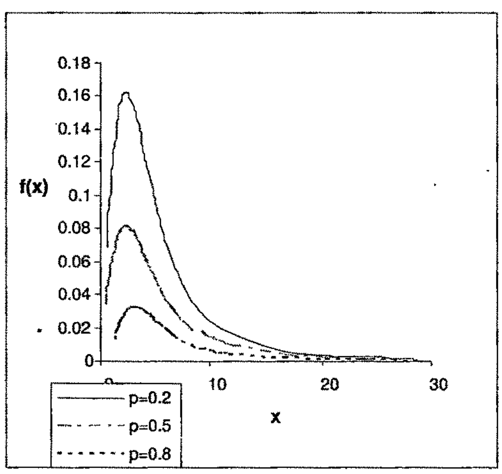


Fig. 2.4.1. Density function f(x): $\beta=3$, $\alpha=2$, $d = 0.5$, $x_0 = 0$.

Failure Rate Functions. The failure rate function is given in fig. (2.4.7). Clearly, its shape is the same as the Pareto distribution after *d*. Thus we focus on the segment from 0 to *d*. The following four figures show that *h(x)* can be increasing, decreasing, or bathtub shaped for $0 \le x \le d$. From the plots, it can be seen that the failure rate function of the model gives rise to several different shapes and bumps; this is expected as mixing with a component distribution that has a finite range often cause some problems. Although the second part can be either increasing or decreasing, the first segment can achieve various shapes. This finding agrees with Block (2003).

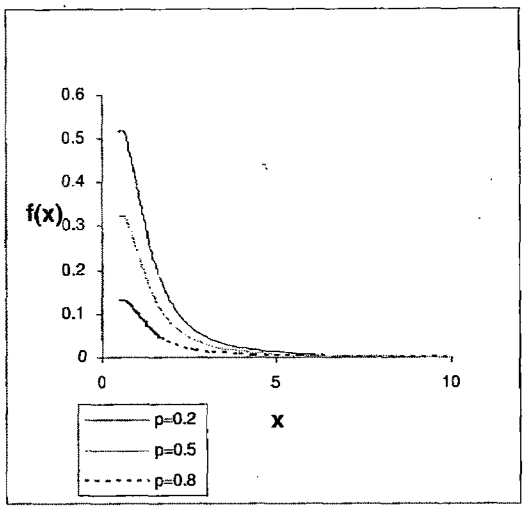


Figure 2.4.2 Density function $f(x)$: $\theta = 1, \alpha = 2, d = 0.2, x_0 = 0$.

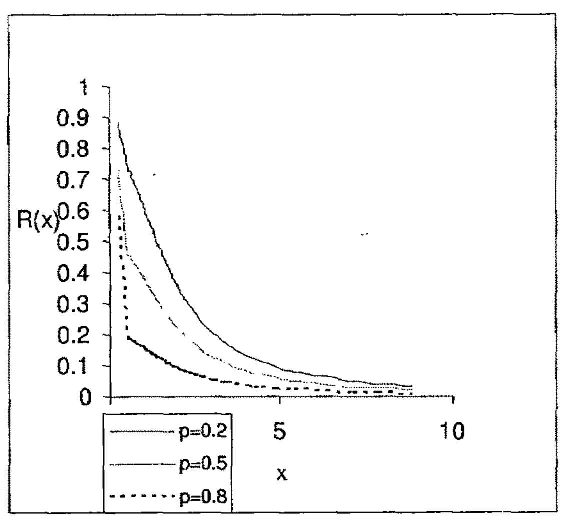


Figure 2.4.3 Reliability function R(x): $\theta = 3, \alpha = 2, d = 0.5, x_0 = 0$

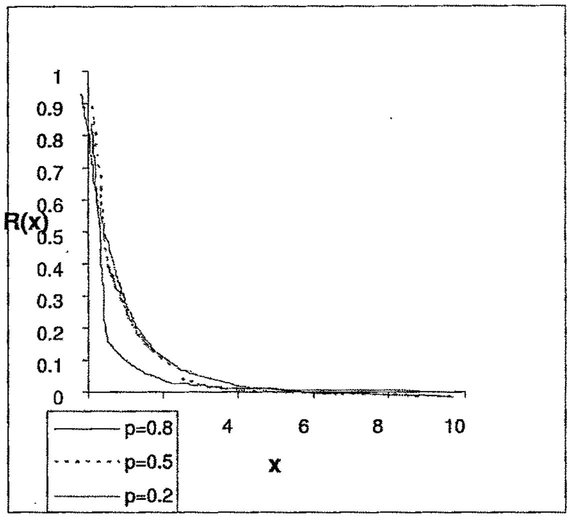


Fig. 2.4.4. Reliability function $R(x)$: $\beta=1$, $\alpha=2$, $d=0.5$, $x_0=0$

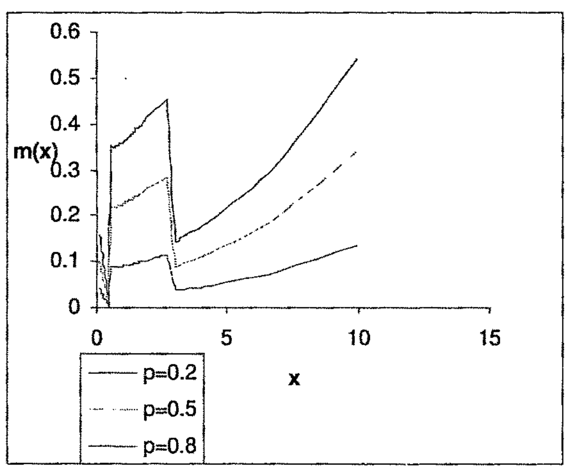


Fig. 2.4.5. Plot of mean residual $m(x)$: $\beta=1$, $\alpha=2$, $d=0.5$, $x_0=0$

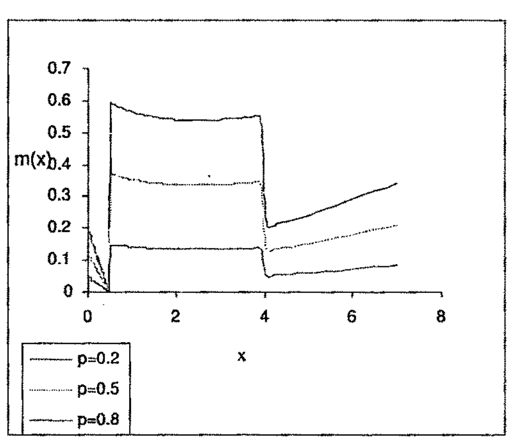
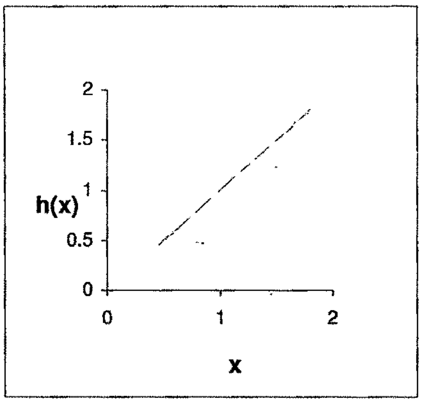
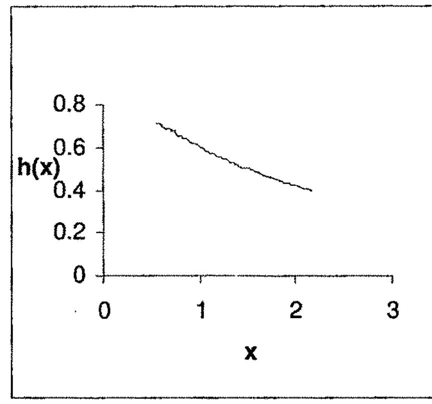


Fig. 2.4.6. Plot of mean residual m(x): $\beta=3$, $\alpha=2$, $d=0.5$, $t_0=0$.

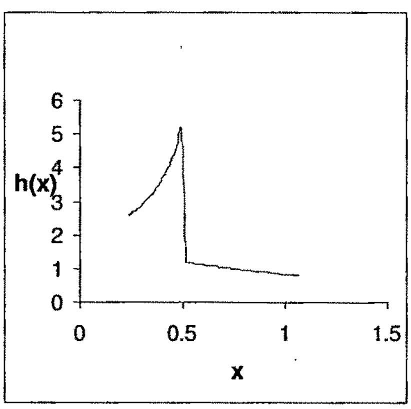
Figure (2.4.6) represents failure rate $h(x)$ for different combinations of α, β, p and d.



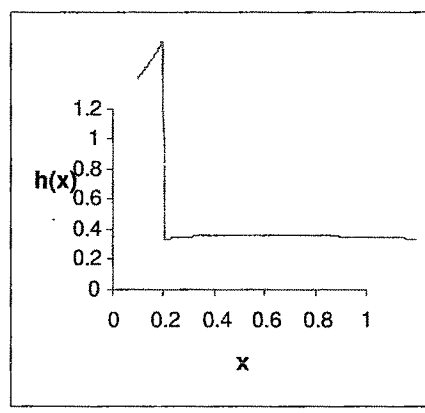
(a) α=1.2, β=1.2, p=0.3, d=0.5



(b) α=1.2, β=1, p=0.08, d=0.5



(c) α=1.2, β=0.5, p=0.5, d=0.5



(d) α=2, β=3, p=0.2, d=0.2

Fig. 2.4.7. Failure rate

2.5 Inlier estimation using L_k and M_k models

In this section we consider the situation where instantaneous (i.e. $X = 0$) failures can also occur by mixing a singular distribution at $X = 0$ with the above model of inliers. Assuming that the data is usually consisting of r_0 instantaneous failures, r_1 early failures as indicated by sample configuration and the rest $n - r_0 - r_1$ observations belong to the target population.

2.5.1 Inlier estimation for labeled slippage (L_k) models

For this model consider observations from inlier with pdf

$$g(x)=\frac{\phi}{(1+x)^{\phi+1}}, x>0,\ \phi>0$$

and those from target population with pdf

$$f(x)=\frac{\theta}{(1+x)^{\theta+1}}, x>0,\ \theta>0$$

Then the likelihood of the sample from population with observations from inliers with pdf $g\left(x_{(i)},\phi\right)$ and target pdf $f\left(x_{(i)},\theta\right)$

$$L=\binom{n}{r_0}(1-\alpha)^{r_0}\alpha^{n-r_0}\left[\frac{r_1!(n-r_0-r_1)!}{\varphi_{r_1}(G,F)}\right]\prod_{i=1}^{r_1} g\left(x_{(i)},\phi\right)\prod_{i=r_1+1}^{n} f\left(x_i,\theta\right) \qquad (2.5.1)$$

where $\varphi_r(G,F)$ is defined as

$$\varphi_r(G,F)=\mathrm{P}\left(X_{(r)}<X_{(r+1)}\mid G,F\right)$$

$$=\int_{-\infty}^{\infty}[G(u)]^r(n-r)[1-F(u)]^{n-r-1}dF(u) \qquad (2.5.2)$$

The likelihood function in (2.5.1) assumes that between the experiments when units are placed on test we do not know which of the units fail instantaneously. Equivalently $X_{i_1}=0,\ X_{i_2}=0,\ \ldots X_{i_{r_0}}=0$ which fail early i.e. those units whose failure time distribution is $g(x_{(i)},\phi)$ with failure rate much larger than that of the failure time distribution of the target population whose failure rate is considerably smaller. The log likelihood of the model is

$$\ln L=r_0\ln(1-\alpha)+(n-r_0)\ln\alpha-\ln\varphi_{r_1}(\phi,\theta)+r_1\ln\phi-(\phi+1)\sum_{i=1}^{r_1}\ln(1+x_{(i)})$$

$$+(n-r_0-r_1)\ln\theta-(\theta+1)\sum_{i=r_1+1}^{n-r_0}\ln(1+x_{(i)})$$

and the likelihood equations are

$$\frac{\partial \ln L}{\partial \alpha}=\frac{-r_0}{(1-\alpha)}+\frac{(n-r_0)}{\alpha}=0 \tag{2.5.3}$$

$$\frac{\partial \ln L}{\partial \phi}=-\frac{\partial}{\partial \phi}\ln\varphi_{r_1}(\phi,\theta)+\frac{r_1}{\phi}-\sum_{i=1}^{r_1}\ln\left(1+x_{(i)}\right) \tag{2.5.4}$$

and

$$\frac{\partial \ln L}{\partial \theta}=-\frac{\partial}{\partial \theta}\ln\varphi_{r_1}(\phi,\theta)+\frac{n-r_0-r_1}{\theta}-\sum_{i=r_1+1}^{n}\ln\left(1+x_{(i)}\right) \tag{2.5.5}$$

Here (2.5.3) can be solved to get the estimate of α as $\hat{\alpha}=(n-r_0)/n$. Solving (2.5.4) and (2.5.5) simultaneously we get the estimate of ϕ and θ. The parameter α is orthogonal to (ϕ,θ). The second order derivatives are

$$\frac{\partial^2 \ln L}{\partial \phi^2}=-\frac{\partial^2}{\partial \phi^2}\ln\varphi_{r_1}(\phi,\theta)-\frac{r_1}{\phi^2} \tag{2.5.6}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2}=-\frac{\partial^2}{\partial \theta^2}\ln\varphi_{r_1}(\phi,\theta)-\frac{(n-r_0-r_1)}{\theta^2} \tag{2.5.7}$$

and

$$\frac{\partial^2 \ln L}{\partial \phi\partial \theta}=-\frac{\partial^2}{\partial \phi\partial \theta}\ln\varphi_{r_1}(\phi,\theta) \tag{2.5.8}$$

where

$$\varphi_{r_1}(\phi,\theta)=(n-r_0-r_1)\int_0^\infty\left\{1-\frac{1}{(1+x)^\phi}\right\}^{r_1}\left[\frac{1}{(1+x)^\theta}\right]^{n-r_0-r_1-1}\frac{\theta}{(1+x)^{\theta+1}}dx$$

$$=\frac{(n-r_0-r_1)\theta}{\phi}\beta\left(r_1+1,\frac{(n-r_0-r_1)\theta}{\phi}\right)$$

$$=\left[\frac{(n-r_0-r_1)\theta}{\phi}\right]\left[\frac{\Gamma(r_1+1)\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}\right)}{\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}+r_1+1\right)}\right] \tag{2.5.9}$$

Taking log on both sides we get

$$\ln\varphi_{r_1}(\phi,\theta)=C+\ln\theta-\ln\phi+\ln\Gamma(z)-\ln\Gamma((z+r_1+1))$$

where

$$z = \{(n - r_0 - r_1)\theta\}/\phi$$

and

$$\frac{\partial}{\partial\phi}\ln\varphi_{r_1}(\phi,\theta) = -\frac{1}{\phi} + \frac{\partial}{\partial\phi}\ln\Gamma z\frac{\partial z}{\partial\phi} - \frac{\partial}{\partial\phi}\ln\Gamma(z + r_1 + 1)\frac{\partial z}{\partial\phi}$$

$$= -\frac{1}{\phi} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(-\frac{(n - r_0 - r_1)\theta}{\phi^2}\right)$$

where

$$\psi(z) = \frac{\partial}{\partial\phi}\Gamma z \quad and\, \Gamma z = \int_0^\infty x^{z-1}e^{-x}dx$$

The second derivative of the likelihood functions are

$$\frac{\partial^2\ln\varphi_{r_1}(\phi,\theta)}{\partial\phi^2} = \frac{1}{\phi^2} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(\frac{2(n - r_0 - r_1)\theta}{\phi^3}\right)$$
$$-\left\{\frac{(n - r_0 - r_1)\theta}{\phi^2}\left[\psi'(z) - \psi'(z + r_1 + 1)\right]\left(-\frac{(n - r_0 - r_1)\theta}{\phi^2}\right)\right\}$$

$$\frac{\partial^2\ln\varphi_{r_1}(\phi,\theta)}{\partial\phi^2} = \frac{1}{\phi^2} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(\frac{2\theta(n - r_0 - r_1)}{\phi^3}\right)$$
$$+\left\{\left(\frac{(n - r_0 - r_1)\theta}{\phi^2}\right)^2\left[\psi'(z) - \psi'(z + r_1 + 1)\right]\right\}$$

where $\psi'(z) = \frac{\partial^2}{\partial\phi^2}\ln\Gamma z$

now

$$\frac{\partial}{\partial\theta}\ln\varphi_{r_1}(\phi,\theta) = \frac{1}{\theta} + \frac{\partial}{\partial\theta}\ln\Gamma z\frac{\partial z}{\partial\theta} - \frac{\partial}{\partial\theta}\ln\Gamma(z + r_1 + 1)\frac{\partial z}{\partial\theta}$$

$$= -\frac{1}{\theta} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(\frac{(n - r_0 - r_1)}{\phi}\right)$$

$$\frac{\partial^2\ln\varphi_{r_1}(\phi,\theta)}{\partial\theta^2} = -\frac{1}{\theta^2} + \left[\psi'(z) - \psi'(z + r_1 + 1)\right]\left(\frac{(n - r_0 - r_1)}{\phi}\right)^2$$

$$\frac{\partial^2 \ln \varphi_{r_1}(\phi,\theta)}{\partial\phi\partial\theta}=\left[\psi(z)-\psi(z+r_1+1)\right]\left(-\frac{(n-r_0-r_1)}{\phi^2}\right)$$
$$-\left(\left\{\frac{(n-r_0-r_1)}{\phi}\right\}\left(\frac{(n-r_0-r_1)\theta}{\phi^2}\right)\left[\psi'(z)-\psi'(z+r_1+1)\right]\right)$$

Using results from Abramovitz and Stegun (1965) we get

$$\left[\psi(z)-\psi(z+r_1+1)\right]=-\sum_{j=1}^{r_1}\frac{1}{z+j} \quad (2.5.10)$$

$$\left[\psi'(z)-\psi'(z+r_1+1)\right]=\sum_{j=1}^{r_1}\frac{1}{(z+j)^2} \quad (2.5.11)$$

Using the above results, we obtain the likelihood equations as

$$\frac{\partial \ln L}{\partial\phi}=0\Rightarrow=\frac{r_1+1}{\phi}-\frac{\theta(n-r_0-r_1)}{\phi^2}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right]-\sum_{i=1}^{r_1}\ln(1+x_i)=0 \quad (2.5.12)$$

$$\frac{\partial \ln L}{\partial\theta}=0\Rightarrow=\frac{(n-r_0-r_1-1)}{\theta}+\frac{(n-r_0-r_1)}{\phi}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right]-\sum_{i=r_1}^{n}\ln(1+x_i)=0 \quad (2.5.13)$$

The above equations may be solved simultaneously to get estimates for ϕ and θ. The Fisher information's are obtained as

$$I_{\phi\phi}=E\left[\frac{-\partial^2 \ln L}{\partial\phi^2}\right]=\frac{r_1+1}{\phi^2}-\frac{2\theta(n-r_0-r_1)}{\phi^3}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right]+\left\{\frac{(n-r_0-r_1)\theta}{\phi^2}\right\}^2\sum_{j=1}^{r_1}\frac{1}{(z+j)^2}$$

$$I_{\theta\theta}=E\left[\frac{-\partial^2 \ln L}{\partial\theta^2}\right]=\frac{n-r_0-r_1-1}{\theta^2}+\left\{\frac{(n-r_0-r_1)}{\phi}\right\}^2\sum_{j=1}^{r_1}\frac{1}{(z+j)^2}$$

and

$$I_{\phi\theta}=E\left[\frac{-\partial^2 \ln L}{\partial\phi\partial\theta}\right]=\frac{\theta(n-r_0-r_1)^2}{\phi^3}\left[\sum_{j=1}^{r_1}\frac{1}{(z+j)^2}\right]+\left\{\frac{(n-r_0-r_1)}{\phi^2}\right\}\sum_{j=1}^{r_1}\frac{1}{(z+j)}$$

The graph of $\varphi_r(G,F)$ to detect inlier is represented on graph (2.5.2).

2.5.2 Identified Inliers Model (M_k model)

Here we assume that the failure times $(X_1, X_2, ...X_n)$ of *n* units put on test are such that *(n-r)* of them are i.i.d. with FTD belonging to $\Im$ characterizing target population and remaining *r* are i.i.d. with FTD from $\mathcal{G}$ causing inlier observations where $G \in \mathcal{G}$ and $F \in \Im$ are such that $\frac{\partial G}{\partial F}$ is decreasing in *X*. As the indexing set *v* and the number of inliers are known we can relate $(X_1, X_2, ...X_r)$ i.i.d. as $G \in \mathcal{G}$ are independently distributed of $(X_{r+1}, X_{r+2}, ...X_n)$ from $F \in \Im$. Then the likelihood of the sample is given by

$$L(x \mid \phi, \theta, v, r) = \prod_{i=1}^{r} g(x_i) \prod_{i=r+1}^{n} f(x_i) \tag{2.5.14}$$

The MLE of parameter of *G* and *F* is a straight forward two sample problem. Suppose that the target population has FTD given by pareto distribution parameter θ and the inliers are given by pareto distribution with parameter ϕ where $\phi > \theta$ and the likelihood in the identified inlier model is given by

$$L(x \mid \phi, \theta, v, r) = \prod_{i=1}^{r} \frac{\phi}{\left(1 + x_{(i)}\right)^{\phi+1}} \prod_{i=r+1}^{n} \frac{\theta}{\left(1 + x_{(i)}\right)^{\theta+1}} \tag{2.5.15}$$

For each $r = 1, 2, ...n$ we find maximum likelihood using equation (2.5.15), and then consider inlier $\hat{r}$ being that value of *r* for which likelihood is maximum.

2.5.3 Simulation study

To illustrate the method of identifying inliers model we have generated 15 independent random samples, where five of them are coming from Pareto with parameter $\phi = 20$ and remaining ten observations from Pareto distribution with parameter $\theta = 0.8$. The samples are 0.01339, 0.02679, 0.03442, 0.05519, 0.09459, 0.32854, 0.64367, 1.19427, 3.00276, 3.14612, 3.15643, 3.94635, 5.17659, 9.79405 and 12.52736. The model under illustration is identified inliers model. The identification is done as follows to evaluate for each fixed r the maximum likelihood equation $\hat{L}_r$ and then consider $\hat{r}$ being that value of r for which likelihood is maximum. The estimates have been presented in table (2.5.1).

It is interesting to note that the maximum likelihood corresponds to $\hat{r}=5$, which was expected. The corresponding estimates of the parameters are $\hat{\phi}$ = 22.96948 and $\hat{\theta}$ =0.704261. The graphical representations of the likelihood plot are given in figure (2.5.1).

Table 2.5.1. The Likelihood and parameter estimates

r	$\hat{\phi}$	$\hat{\theta}$	$\hat{L}_r$
1	75.18149	0.971976	8.46574E-12
2	50.32893	0.904208	1.14685E-10
3	40.77225	0.836623	1.33608E-09
4	31.42176	0.769788	9.1914E-09
5	**22.96948**	**0.704261**	**3.21716E-08**
6	12.06675	0.646565	1.02195E-08
7	7.041070	0.596001	2.28998E-09
8	4.494340	0.553932	4.46583E-10
9	2.841800	0.533335	4.66372E-11
10	2.179040	0.508762	1.37914E-11
11	1.829110	0.476013	6.59953E-12
12	1.576350	0.440886	3.38237E-12
13	1.378100	0.401308	1.74583E-12

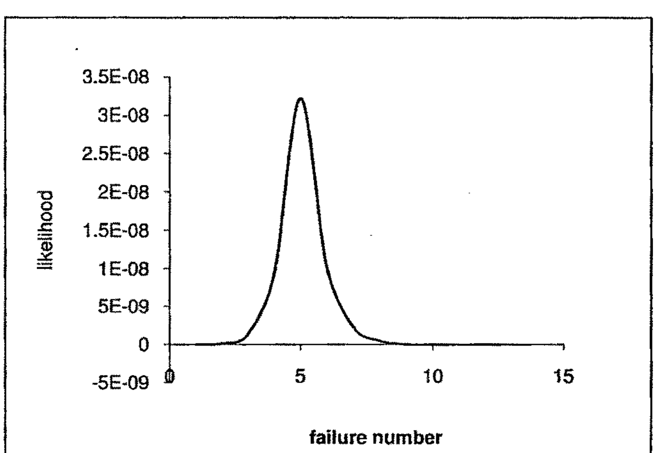


Fig. 2.5.1. Likelihood plot

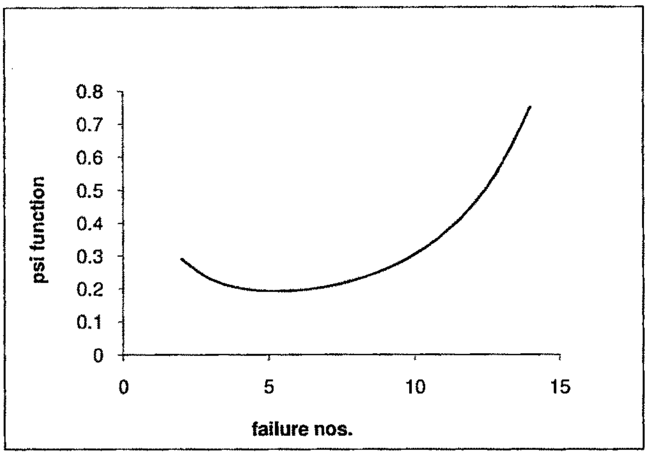


Figure 2.5.2. The graph of $\varphi_r(G,F) = P[X_{(i)} < X_{(i+1)}]$

Clearly the above graph also indicates the number of inliers is 5.

2.6 Inliers detection using information criterion

The most important use of information criterion is, that it helps us in model selection, from the set of different models which all fit the data. These criteria are suitable when the underlying distribution and inlier distribution are available. It is an exploratory data analysis approach as no formal statistical inference is performed. Here three information criteria are discussed, to detect number of inliers in the data set, such as Schwarz's Information criterion ($SIC = 2\ln L(\Theta) + p\ln n$), Schawarz's Bayesian Information criterion ($BIC = -\ln L(\Theta) + \frac{0.5(p\ln n)}{n}$)and Hannan-Quinn criterion : $HQ = -\ln L(\Theta) + p\ln(\ln(n))$ where L(Θ) 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:

Denoting the parameter of X by α_i, i=1,2,...n. The following model of no inliers where X is from one parameter Pareto distribution with pdf

$$f(x) = \frac{\theta}{(1+x)^{\theta+1}}, x > 0, \theta > 0.$$

Now let

$$\text{Model}(0): \alpha_i = \theta \quad i = 1,2,......p \ , \tag{2.6.1}$$

And the model with r inliers as

$$\text{Model}(r): \alpha_i = \begin{cases} \phi, & 1 \le i < r \\ \theta, & r+1 \le i < n \end{cases} \tag{2.6.2}$$

where inliers have pdf $g(x) = \frac{\phi}{(1+x)^{\phi+1}}$, $x > 0$, $\phi > 0$ and r is such that $1 \le r \le n$, is the unknown index of the inliers. Model(0) may also be interpreted as having all observations from the target distribution F with common parameter .

Suppose that the life times of $X_1, X_2, ..., X_n$ is sequence of independent random variables with Pareto distribution having unknown parameter θ. Our aim is to detect those information's(inliers) from the n models given by equation (2.6.2).

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 $SIC(0) > \min_{1 \le r \le n-1} SIC(r)$. For Pareto distribution, the model with 0 inlier is given by

$$SIC(0) = -2n\ln\theta + 2(\theta+1)\sum_{i=1}^{n}\ln(1+x_i) + p\ln n \tag{2.6.3}$$

and

$$SIC(r) = -2r\ln\phi - 2(n-r)\ln\theta + 2(\phi+1)\sum_{i=1}^{r}\ln(1+x_i) + 2(\theta+1)\sum_{i=r+1}^{n}\ln(1+x_i) + p\ln n \tag{2.6.4}$$

where

$$\hat{\phi} = \frac{r}{\sum_{i=1}^{r}\ln(1+x_i)} \quad \text{and} \quad \hat{\theta} = \frac{(n-r)}{\sum_{i=r+1}^{n}\ln(1+x_i)} \tag{2.6.5}$$

The estimate of inliers say r is such that $SIC(r) = \min_{1 \le r \le n} SIC(r)$. The illustration uses this method with the simulated example discussed in the previous section (2.5.1) and Table (2.6.1) presents the parameter estimates and the information criterion values.

Table 2.6.1. Parameter estimates and the information criterion values

$\hat{r}$	$\hat{\theta}$	$\hat{\phi}$	*SIC*(*r*)	*BIC*(*r*)	*HQ*(*r*)
0	1.040442	----------	60.3520	29.82228	58.64077917
1	0.971976	75.18149	53.6980	26.49499	51.98621518
2	0.904208	50.32893	48.4857	23.88883	46.77389739
3	0.836623	40.77225	43.5751	21.43353	41.86328438
4	0.769788	31.42176	39.7180	19.50500	38.00622349
5	**0.704261**	**22.96948**	**37.2124**	**18.25218**	**35.50059008**
6	0.646565	12.06675	39.50599	19.39897	37.79417263
7	0.596001	7.041079	42.4975	20.89472	40.78567634
8	0.553932	4.494343	45.7668	22.52940	44.05502113
9	0.533335	2.841807	50.2853	24.78862	48.57347388
10	0.508762	2.179042	52.7220	26.00698	51.01018362
11	0.476013	1.829118	54.1960	26.74402	52.48427275
12	0.440886	1.576359	55.5329	27.41245	53.82112047
13	0.401308	1.378105	56.8556	28.07379	55.14380810

Clearly $SIC(0) = 60.3526 > SIC(5) = \min_{1 \le r \le n} SIC(r) = 37.21241$. A similar conclusion can be drawn in the case of other information criterions:

$$\text{BIC}(0) = 29.82228 > BIC(5) = \min_{1 \le r \le n} BIC(r) = 18.25218$$

$$HQ(0) = 58.64077917 > HQ(5) = \min_{1 \le r \le n} HQ(r) = 35.50059008.$$

Above table clearly indicates $\hat{r} = 5$ and the corresponding estimates for the parameters are $\hat{\phi}$ = 22.96948 and $\hat{\theta}$ =0.704261.

Next, we carried out an experiment with 1000 samples each of size 15 and number of inliers as 3,4,5 and 6 each with $\theta =$ 0.8 and $\phi = 4,2,1.0,1.33$. The following table entitled power of SIC procedure presents the number of times the SIC procedure correctly identified the number of inliers in proportion to total number of samples. The values clearly indicate the effectiveness of the method in detecting the inliers. One of the important problem while detecting the inliers is the masking effect, where masking effect is defined as the loss of power due to wrong detection of more than one inlier.

Table 2.6.2. Power of SIC procedure

r \ θ/ϕ	0.2	0.4	0.6	0.8
3	0.055	0.083	0.098	0.103
4	0.084	0.116	0.136	0.128
5	0.102	0.153	0.158	0.157
6	0.128	0.168	0.170	0.175

2.7 Inlier estimation through Sequential Probability Ratio Test (SPRT)

To test the hypothesis whether an observation belongs to inliers population against hypothesis that it belongs to target population. The SPRT test is given as follows:

Under H_1 the pdf and likelihood function is given by

$$f(x,\theta)=\theta/(1+x)^{\theta+1}$$

and

$$L_{1m}=\prod_{i=1}^{m} f(x_i,\theta)=\prod_{i=1}^{m}\frac{\theta}{(1+x)^{\theta+1}}$$

Under H_0 the pdf and likelihood function is given by

$$g(x,\phi)=\phi/(1+x)^{\phi+1}$$

and

$$L_{0m}=\prod_{i=1}^{m} g(x_i,\phi)=\prod_{i=1}^{m}\frac{\phi}{(1+x)^{\phi+1}}$$

The likelihood ratio λ_m is given by $\lambda_m=\frac{L_{1m}}{L_{0m}}$ or equivalently

$$\ln\lambda_m=\sum_{i=1}^{m}\ln\frac{f(x_{(i)},\theta)}{g(x_{(i)},\phi)}=\sum_{i=1}^{m} z_{(i)} \qquad m=1,2,\ldots.n \tag{2.7.1}$$

For deciding number of inliers r we continue to take additional observations till we reject H_0. That is,

if $\sum_{i=1}^{m} z_{(i)} \le \ln B$ accept H_0 and take the next observation.

and

if $\sum_{i=1}^{m} z_{(i)} \geq \ln A$ reject H_0 and stop. The corresponding m represents the first observation from $f\left(x_{(i)},\theta\right)$ and number of inliers $r = m$-1.

$$B=\frac{\beta}{1-\alpha}\ ,\ A=\frac{1-\beta}{\alpha} \tag{2.7.2}$$

where α represents probability of type I error and β represents probability of type II error. Hence

$$\ln\lambda_m=\sum_{i=1}^{m}\ln\frac{f\left(x_{(i)},\theta\right)}{g\left(x_{(i)},\phi\right)}=m(\ln\theta-\ln\phi)+\sum_{i=1}^{m}\ln\left(1+x_{(i)}\right)(\theta-\phi) \tag{2.7.3}$$

Arrange $X_{(1)} \leq X_{(2)} \leq ......X_{(n)}$ and apply SPRT process till the hypothesis H_0 is rejected.

Test criteria for rejection of H_0 is

$$\ln\lambda_m>\ln A=\sum_{i=1}^{m}\ln\left(1+x_{(i)}\right)>\frac{\ln A-m(\ln\theta-\ln\phi)}{(\theta-\phi)} \tag{2.7.4}$$

Corresponding value of m for which H_0 was accepted last becomes number of inliers r. The above test is conducted for the example in next section.

2.8 Illustrative Example

The main reason for detecting early failures is that the inclusion of these observations will result in underestimating life expectancy or the reliability of the item or system. This in turn may underestimate the true quality of the product. But there are situations in which instantaneous or early failures may be desirable. For example, consider the following experiment carried by Vannman (1991). A batch of wooden boards is dried by a particular chemical process and the object of the experiment is to compare two processes as regards the extent of deformation of boards due to checking. The measure of damage to the board is the checking area x defined as $x=\frac{l\bar{d}}{hl_0}100$, where l is the length of the check, $\bar{d}$ is the mean depth of the check, h is the thickness of the board area and l_0 is the length of the board. Thus x is the check area measured as percentage of the board area. The boards are dried at the same time under different schedule and under some climatic conditions. When drying boards not all of them will get the checks and a typical sample of wood contain several observations with $x_i = 0$ or $x_i > 0$ but relatively small compared to the rest of the checks. These observations will correspond to instantaneous failures or early failures. Note that the larger the number of instantaneous failures better is the process. Below is the reproduced data of Schedule 1 and 2 of Experiment 3 conducted by Vannman (1991). In both the case n=37. For data refer appendix.

First of all, we justify the Pareto model for the above data using the technique given in Meeker and Escobar(1998) and plotted $log[-log(1-p)]$, $p= F(x_i)$ against $log(x_i)$ and obtained the one parameter Pareto plot and two parameter Pareto plot separately for Schedule 2. For early failure analysis, we assumed δ=0.2. With this the observation 0.08 of Schedule 1 becomes an early failure and the observations 0.02 to 0.09 (total of 5) items of Schedule 2 become early failures.

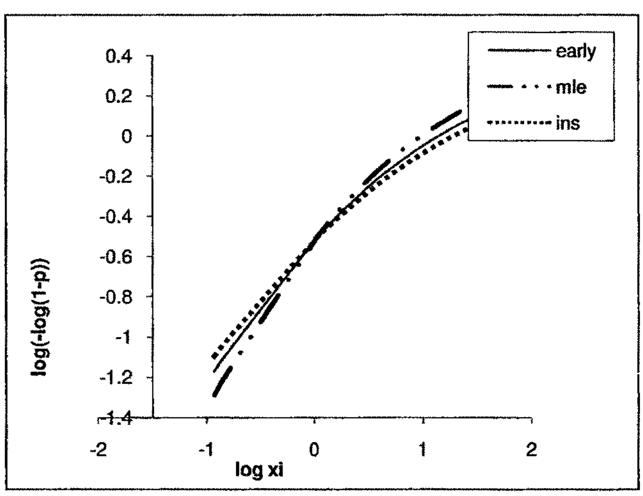


Fig. 2.8.1. One parameter Pareto plot for Schedule-2

The plots are given in figure (2.8.1) and (2.8.2) for one parameter and two parameters Pareto plot respectively for Schedule-2 of experiment 3.

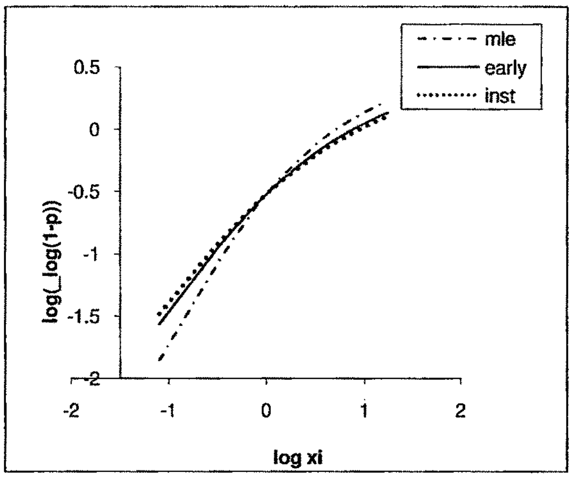


Fig. 2.8.2. Two parameter Pareto plot for Schedule-2

Table 2.8.1. Estimates of one parameter Pareto distribution

Model	Parameter	Estimates	
		Schedule-1	Schedule-2
$\mathcal{G}$ (instantaneous failures)	α	0.64865 (0.006159)	0.54054 (0.006712)
	β	1.00541 (0.029455)	0.803645 (0.022583)
$\mathcal{G}_1$ (early failures)	α	0.86603 (0.012723)	0.65394 (0.017035)
	β	0.577342 (0.011761)	0.305469 (0.003649)

Table 2.8.2. Estimates of two parameter Pareto distribution

Model	Parameter	Estimates	
		Schedule-1	Schedule-2
$\mathcal{G}$ (instantaneous failures)	α	0.64865 (0.006160)	0.54054 (0.0067123)
	β	1.25299 (0.421699)	0.800943 (0.139438)
	ϕ	2.77164 (5.638073)	0.823577 (0.393020)
$\mathcal{G}_1$ (early failures)	α	0.652919 (0.007057)	0.43217 (0.007451)
	β	1.22408 (0.034035)	1.23621 (0.036314)
	ϕ	2.30727 (0.226172)	1.79431 (0.130620)

Note: the values in the parenthesis represents variances of the estimates

The above analysis shows that the results differ in the models $\mathcal{G}$ and $\mathcal{G}_1$. In $\mathcal{G}_1$, even if we keep δ=0.1 or any value in between 0.1 to 0.2 the results are similar. Further, if we ignore the value of α then the information loss of β are 0.064116 for Schedule 1 and 0.048048 for Schedule 2 correspond to the one parameter Pareto distribution. Similarly the information loss for two parameter distributions is 0.0036226 for β and 0.00015074 for ϕ in Schedule 1 and 0.0023976 for β and 0.0004433 for ϕ in Schedule 2, respectively. Thus to retain the complete information the presence of α and δ are very much required. Moreover, from Tables 1 and 2 it is observed that the variance of the estimators of the parameters corresponds to early failures is less than the corresponding variance of instantaneous failures. Also the presence of more parameters makes the model more flexible to use. If in equation (2.1.3), the individual life times $x_i \in (0,\delta)$ are available and are not reported as δ, the problem becomes more complex.

Table 2.8.3. Estimates for instantaneous failures

Schedule		$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
I	Estimates	0.351351	1.25299	2.77164
	Standard Error	2.094713	0.508146	0.154711
II	Estimates	0.459459	0.800943	0.823577
	Standard Error	2.006607	1.02006	0.380592

Table 2.8.4. Uniform spread of "nearly instantaneous" failure times.

Schedule		$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
I	Estimates	0.378378	1.3949	3.10361
	Standard Error	2.06793	0.527328	0.159483
II	Estimates	0.567568	1.23692	1.80127
	Standard Error	2.018516	0.728331	0.298683

If we fit above data to one- parameter Pareto distribution, taking β=1, we get following estimates for the two schedules:

Table:2.8.5. Estimates for instantaneous and nearly instantaneous failure when β=1

Schedule		$\hat{p}$	$\hat{\alpha}$
I	Instantaneous	0.351351 (2.094713)	0.922071 (0.768719)
	Nearly	0.378378 (2.06193)	0.969036 (0.759074)
II	Instantaneous	0.459459 (2.0006607)	0.703233 (1.152618)
	Nearly	0.567568 (2.018516)	1.14447 (1.069799)

Note: Figures in the bracket represents the standard error of the estimates.

Table 2.8.6. Estimates of parameters and r.

r	$\hat{\theta}$	$\hat{\phi}$	$\hat{L}_r$	$SIC(r)$	$Z_{(i)}$
1	0.654913	12.99359	1.5045E-29	135.9111	0.076961
2	0.63143	5.640273	2.51883E-29	134.8804	0.354593
3	0.608352	4.433435	4.99947E-29	133.5093	0.676676
4	0.585805	3.791064	9.15284E-29	132.2999	1.055113
5	0.5654	3.141481	1.17998E-28	131.7918	1.591606
6	0.54536	2.739172	1.50475E-28	131.3056	2.190443
7	0.527292	2.368137	1.5387E-28	131.2609	2.878949
8	0.508935	2.128842	1.67179E-28	131.095	3.403319
9	**0.490793**	**1.942426**	**1.77838E-28**	**130.9714**	3.956704
10	0.479839	1.661219	1.07411E-28	131.9798	4.964562
11	0.468734	1.47416	7.37023E-29	132.7331	
12	0.459438	1.321979	4.93023E-29	133.5372	
13	0.449527	1.212013	3.60868E-29	134.1613	
14	0.441077	1.117822	2.593E-29	134.8224	
15	0.43224	1.043523	1.95208E-29	135.3902	
16	0.424204	0.979354	1.46919E-29	135.9586	
17	0.416518	0.924409	1.11832E-29	136.5043	
18	0.410398	0.874798	8.41417E-30	137.0733	
19	0.403448	0.833225	6.52924E-30	137.5806	
20	0.394893	0.797868	5.19969E-30	138.036	
21	0.385706	0.765915	4.14956E-30	138.4872	
22	0.372682	0.737522	3.35093E-30	138.9147	

For inliers detection based on section (2.5) and (2.6) we have used only schedule 1 data which are shown in table (2.8.6).Clearly, SIC(0) = 139.9487 > SIC(9) = min SIC(r) = 130.9714. Also the likelihood is maximum for r = 9. The corresponding estimates of the parameter are $\hat{\phi}$ = 1.942426 and $\hat{\theta}$ =0.490793. Using SPRT of section (2.7) the hypothesis $H_0:\phi=2$ against $H_0:\theta=0.5$ is also tested, for which $\alpha=0.005$, $\beta=0.065$. Hence $ln\ A$ = -13.4417 and $ln\ B$ = -2.72836 and H_0 is rejected when $Z_{(i)}=\sum_{i=1}^{m}\ln\left(1+x_{(i)}\right)=4.964562$ $>\dfrac{\ln A-m\left(\ln\theta-\ln\phi\right)}{\left(\theta-\phi\right)}=4.340185$. SPRT also gives number of inliers as $\hat{r}=9$.

The Pareto distribution has been used in many reliability fields. However one often finds that it does not fit well in the early part of lifespan for various reasons. In particular, in the cases where initial defects are present causing early failures, the Pareto distribution is found inadequate to model such phenomenon. The proposed model of a modified Pareto mixing with Uniform distribution to model the first phase of lifespan should provide a useful alternative.