

## INLIER PRONE MODELS: A REVEIW

K. Muralidharan

Department of Statistics, Faculty of Science  
The M. S. University of Baroda, Vadodara 390 002, India.  
e-mail: *lmv\_murali@yahoo.com*

*Paper received on 03 September 2009; revised, 13 January 2010; accepted, 03 March 2010.*

**Abstract.** Inliers in a data set are observations which appear to be too small compared to the remaining observations in the usual life testing experiments with failure time distribution (FTD),  $F(x, \theta)$ ,  $x \geq 0$ ,  $\theta \in \Omega$ . These inconsistent observations may be the resultant of the occurrences of instantaneous failures or early failures in life testing experiments. To accommodate such instances the failure time distributions (FTD) are modified by defining early failures as those observations which are very small compared to the other positive observations and are called inliers. We propose various inlier prone models and their statistical significances. The maximum likelihood estimates (MLE) are obtained for parameter  $\theta$  of the target distribution  $F$  and also  $\phi$  of the contaminating population  $\mathcal{G}$ . Assuming that there are  $k$  inliers, we obtained the MLE of  $k$  and proposed tests of hypothesis for  $k = 0$  (no inliers) when target population is exponential with mean life  $\theta$ . We present different methods of detecting inliers when  $k$  unknown and provide test statistics of one and multiple inliers.

**Keywords.** failure time distribution; instantaneous failures; early failures; exchangeable inliers model; labeled slippage inliers model; order statistics; outliers.

## 1 Introduction

The occurrence of instantaneous or early failures in life testing experiment is observed in electronic parts as well as in clinical trials. These occurrences may be due to inferior quality or faulty construction or due to no response to the treatments. These situations can be modeled by modifying commonly used parametric models such as exponential, gamma, Weibull and lognormal distribution among others. The modified model is then a non-standard mixture of distribution by mixing a singular distribution at zero to accommodate such failures.

Consider the data on Schedule 2 of Experiment 3 conducted by [19] on drying of woods (for more details see the numerical example section). Out of

37 observations, there are seventeen instantaneous failures and twenty positive observations. Further from inspection, it is seen that the first five observations among the positive observations are very nominal and closed to zeroes. The observations corresponds to instantaneous failures together with early failures may be termed as inliers here. We can contemplate similar situations in the following examples also: (a) in auditing, the procedure is to determine fraudulent claims of higher value than the actual expenses incurred. The receipts are classified as correct and suspicious. The correct receipts are like instantaneous failures and given value  $X = 0$  and the suspect receipts are measured according to size of the fraud  $X$ , (b). In the mass production of technological components of hardware, intended to function over a period of time, some components may fail on installation and therefore have zero life lengths. A component that does not fail on installation will have a life length that is a positive random variable whose distribution may take different forms, (c). Consider measurements of physical performance scores of patients with a debilitating disease such as multiple sclerosis. There will be frequent zero measurements from those giving no performance and many observations with graded positive performance, (d). In a study of tooth decay, the numbers of surfaces in a mouth which are filled, missing or decayed are scored to produce a decay index. Healthy teeth are scored 0 for no evidence of decay and is therefore a mixture of a mass point at 0 and a nontrivial continuous distribution of decay score, (e). Time until remission is of interest in studies of drug effectiveness for treatment of certain diseases. Some patients respond and some do not. The distribution is a mixture of a mass point at 0, which corresponds to instantaneous remission and a nontrivial continuous distribution of positive remission times, (f). The rainfall measurement at a place recorded during a season is modeled as a continuous distribution with a nonsingular distribution at zero, where zero measures those days having no rainfall etc. For more such examples see Statistical models and analysis in Auditing: Panel on nonstandard mixtures of distributions, Statistical science, 1989.

Thus inliers in a data set are a subset of observations not necessarily all zeroes, which are relatively small as compared with rest of the observations and appear to be inconsistent with the remaining data. In all the above examples, inliers is a natural occurrence and the model  $\mathcal{I} = \{F(x, \theta), x \geq 0, \theta \in \Omega\}$  where  $F(x, \theta)$  is a continuous failure time distribution function (*df*) with  $F(0) = 0$  is to be suitably modified to account inliers. In some of the above examples say (d), (f) and the example based on [19] etc. we see that the inliers are desirable sometimes. [9] have first introduced the term inliers in connection with the estimation of parameters of early failure model with modified failure time distribution (FTD) being an exponential distribution with mean  $\theta$  and the number of inliers known. [17] discuss various methods of inliers model when the number of inliers are unknown. In the following sections we provide some inlier prone models and their analysis and practical importance.

## 2 Instantaneous failure models

The occurrence of instantaneous failures when some items put on test giving  $X_i = 0$  is quite common in electronic component and some other situations. Note that because of the limited accuracy of measuring failure time it is possible that we record  $X_i = 0$  for some units although  $P[X_i = 0|\theta] = 0$ . To accommodate such instantaneous failures, the model  $\mathcal{I}$  is modified to  $\mathcal{G} = \{G(x, \theta, \alpha), x \geq 0, q \in \Omega, 0 < \alpha < 1\}$  by using a mixture in the proportion  $1 - \alpha$  and  $\alpha$  respectively of a singular random variable  $Z$  at zero and with a random variable  $X$  with  $dfF \in \mathcal{I}$ . Thus the modified failure time distribution is given by

$$G(x, \theta, \alpha) = \begin{cases} 1 - \alpha, & x = 0 \\ 1 - \alpha + \alpha F(x, \theta) & x > 0 \end{cases} \quad (2.1)$$

and the corresponding density function as

$$g(x, \theta, \alpha) = \begin{cases} 1 - \alpha, & x = 0 \\ \alpha f(x, \theta) & x > 0 \end{cases} \quad (2.2)$$

Some of the references which treat the above situation are [2, 6, 12, 19, 14, 15, 9] and references contained therein. [20] and [16] considered the case where  $F$  is a two-parameter Gamma distribution with shape parameter  $\beta$  and scale parameter  $\theta$ . The above model is easy to analyze and provide a general formulation for instantaneous failures.

## 3 Early failure models

To accommodate early failures defined as those units  $X_i$  with observed failure times  $X_i \leq \delta$ , we consider the mixture of  $\mathcal{I}$  with a known FTD  $H(x)$  with  $H(\delta) = 1$ , where  $\delta$  is sufficiently small and assumed known. If early failures are reported nominally as a class with  $X_i = \delta$  then the mixed distribution is

$$g(x, \alpha, \theta) = \begin{cases} 0 & x < \delta \\ 1 - \alpha + \alpha F(\delta, \theta), & x = \delta \\ \alpha f(x, \theta), & x > \delta \end{cases} \quad (3.1)$$

A basic result due to the above formulation is that, if  $\mathcal{I}$  is  $m$ -dimensional Cramer family then the modified family  $\mathcal{G}$  is also a  $(m + 1)$  dimensional Cramer family. If further  $\mathfrak{S}$  is an exponential family of dimension  $m$  then  $\mathcal{G}$  is  $(m + 1)$  dimensional exponential family admitting minimal complete sufficient statistic  $(n_0, T_1, T_2, \dots, T_m)$  where  $n_0$  is the number of observations  $\leq \delta$  and  $T_r = \sum_{x_i > \delta} k_r(x_i)$ ,  $r = 1, 2, \dots, m$  where  $(k_1(x), k_2(x), \dots, k_m(x))$  is minimal complete sufficient statistics for the family  $\mathfrak{S}$ . Similarly, if  $\mathfrak{S}$  is Cramer

family so is the modified family given by (3.1). Some of the references, which treat early failure analysis with exponential distributions, are [8, 9], wherein they treat early failures as inliers using the sample configurations. [17, 18] considered the early failure analysis for Weibull distribution. The above two models are further studied exclusively in respect of inliers by [10, 11]. The authors describe various test procedures and the masking effect for testing number of inliers as the loss of power due to presence of more than anticipated discordant observations in the sample.

The early failures are thus defined normatively as those units which fail before  $d$  assumed known. However, early failures can also be defined by the configuration of observed data. *i.e.* those observations which are very small compared to the other observations. Note that observations are all non-negative and those observations which belong to the left tail of the distribution are to be regarded as early failures. Thus the problem of early failures is akin to the problem of outliers where observations which are very large as compared to the other observations are suspected as outliers. A very important distinction between inliers and outliers is that the inliers form a group of observations and are not excluded from further analysis. For the outlier generating models we refer to the text of [3] and a review by [5].

## 4 Nearly instantaneous failure models

It is seen that the models (2.2) and (3.1) are represented as a mixture of a singular distribution at zero and a suitable FTD,  $\mathfrak{F}$  in different proportion. Because of the singular nature of the distribution, it is unable to define the failure rate function meaningfully. [13] have integrated the above two models to a single model as a mixture of two continuous distributions. This modification allows establishing and studying the failure rate function via mixture distribution.

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  $h(x) = \frac{f(x)}{R(x)}$  provided the density exists.

Instead of assuming an instant or an early failure to occur at a particular time point as in the original model, we now represent this model as a mixture of a generalized Dirac delta function and a pdf  $f(x, \theta)$ . Thus the resulting modification gives rise to a density function:

$$f(x) = \alpha \Delta_{\delta}(x - x_0) + (1 - \alpha)f(x, \theta), \quad 0 < \alpha < 1, \quad (4.1)$$

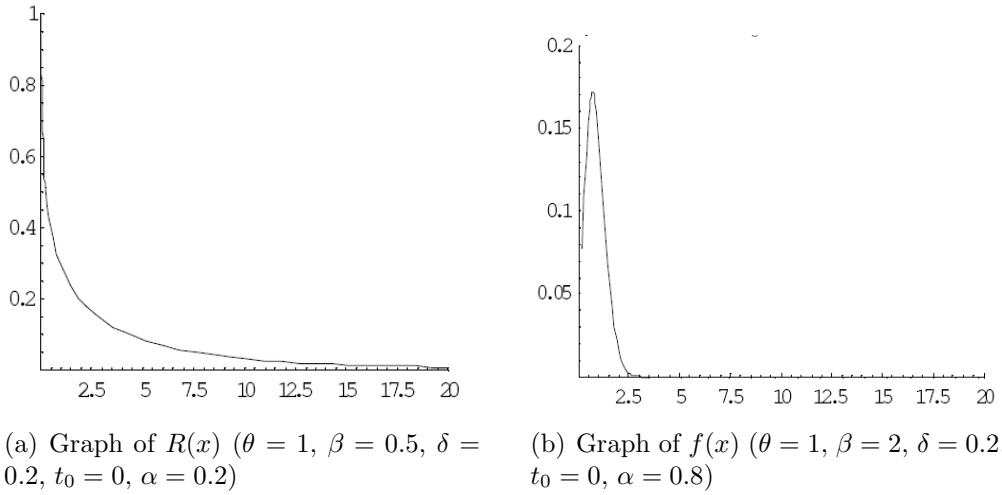


Figure 1:

where

$$\Delta_\delta(x - x_0) = \begin{cases} \frac{1}{\delta}, & x_0 \leq x \leq x_0 + \delta \\ 0, & \text{otherwise} \end{cases} \quad (4.2)$$

for sufficiently small  $\delta$ . We note that

$$\Delta(t - t_0) = \lim_{\delta \rightarrow 0} \Delta_\delta(t - t_0)$$

where  $\Delta(\cdot)$  is the Dirac delta function that is well known in mathematical analysis. We may view the Dirac delta function as a normal distribution having a zero mean and standard deviation that tends to 0. For a fixed value of  $\delta$ , equation (4.2) denotes a uniform distribution over an interval  $[x_0, x_0 + \delta]$  so the modified model is now effectively a mixture of a Weibull with a uniform distribution. For  $x_0 = 0$ , it corresponds to the case with instantaneous failures and for (small)  $x_0 \neq 0$  it corresponds to the case with early failures. Noting from (4.1) and (4.2), we see that the mixture density function is not continuous at  $x_0$  and  $x_0 + \delta$ . However, both the distribution and survival functions are continuous. A typical survival function and density function for the above model is presented in figures 1(a) and 1(b) respectively. For different characterizations and properties of this model with  $f(x, \theta)$  being Weibull  $(\theta, \beta)$ , we refer to the authors' paper.

## 5 $M_k$ inliers models and $L_k$ inliers models

Suppose that  $n$  units are put on test and  $n_0$  units fail instantaneously and  $(n - n_0)$  failure times are available. Out of these positive observations we have

to determine which are inliers or early failures. Before the start of the experiment we do not know which units will fail instantaneously or will produce inliers. These experimental conditions are to be modeled in  $M_k$  inlier model for given  $k$ . Let us reliable failure times of these  $(n - n_0)$  units as  $(X_1, X_2, \dots, X_{n-n_0})$ . Then in  $M_k$  inlier model, we assume that  $(n - n_0 - k)$  are from target population with pdf  $f \in \mathfrak{F}$  and  $k$  are from the inlier population  $g \in \mathcal{G}$ . Thus the joint pdf of  $(X_1, X_2, \dots, X_{n-n_0})$  can be written as

$$L(x_1, x_2, \dots, x_{n-n_0} | f, g, v) = \left\{ \prod_{i \in v} g(x_i) \right\} \left\{ \prod_{i \notin v} f(x_i, \theta) \right\}, \quad f \in \mathfrak{F}, g \in \mathcal{G}, v \in \nu. \quad (5.1)$$

where  $v$  is the new parameter representing set of inliers and ranges over  $\nu$ , the set of integers  $(i_1, i_2, \dots, i_k)$  chosen out of  $(1, 2, \dots, (n - n_0))$  with cardinality  $\binom{n-n_0}{k}$ . This is so far similar to the model  $M_k$  for  $k$  outliers. The main difference in  $M_k$  inlier model is that  $\psi(x) = \frac{\partial G}{\partial F} = \frac{g(x)}{f(x)}$  is assumed to be strictly decreasing function of  $x$ . The following theorem will help us to write the likelihood function under  $M_k$  and  $L_k$ .

**Theorem 5.1** *Let  $X_{(1)} < X_{(2)} < \dots < X_{(n-n_0)}$  be the order statistics and  $(R_1, R_2, \dots, R_{n-n_0})$  be the corresponding rank order statistics, then  $\text{Max } \varphi(r_1, r_2, \dots, r_k) = \varphi(1, 2, \dots, k)$  and  $(x_{(1)}, x_{(2)}, \dots, x_{(k)})$  have the maximum probability of being inliers.*

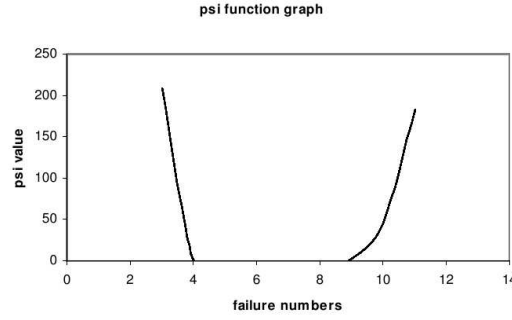
For proof, we refer to [11]. As a consequence of this theorem,  $\hat{v} = k$  and therefore, the likelihood under  $M_k$  inlier model is

$$L(x | g, f, \hat{v}) = \prod_{i=1}^k g(x_{(i)}) \prod_{i=k+1}^{n-n_0} f(x_{(i)}), \quad f \in \mathfrak{F}, g \in \mathcal{G}. \quad (5.2)$$

But  $L(x | g, f, \hat{v})$  is likelihood and not the joint pdf of  $(x_{(1)}, x_{(2)}, \dots, x_{(n-n_0)})$ . Making it a pdf, the model for  $L_k$  inliers is therefore

$$L_k(\underline{x} | g, f) = \frac{(n - n_0)! k!}{\varphi(1, 2, \dots, k)} \prod_{i=1}^k g(x_{(i)}) \prod_{i=k+1}^{n-n_0} f(x_{(i)}), \quad f \in \mathfrak{F}, g \in \mathcal{G}, \quad (5.3)$$

where  $\varphi(1, 2, \dots, k)$  is as defined above and is a norming constant to make  $L_k$  a pdf. The model  $L_k$  is called as the labeled slippage model and it can also be derived as model from  $M_k$  with  $(Y_1, Y_2, \dots, Y_k)$  are iid distributed as  $\mathcal{G}$  and  $(V_1, V_2, \dots, V_{n-n_0})$  as iid  $F$  and with the additional condition  $\text{Max}(Y_1, Y_2, \dots, Y_k) \leq \text{Min}(V_1, V_2, \dots, V_{n-n_0})$ . The object of the experiment is to make inferences about the target population  $F \in \mathfrak{F}$  and the parameter  $v$  in  $M_k$  and parameters of  $g \in \mathcal{G}$  are nuisance parameters. The model  $M_k$  or  $L_k$  for inliers assumes  $k$  known. In practice  $k$  is not known and is to be estimated from

Figure 2: Graph of  $\varphi(1, 2, \dots, k)$ 

the data  $(x_{(1)}, x_{(2)}, \dots, x_{(n-n_0)})$ . The possible values of  $k$  are  $\{0, 1, 2, \dots, n - n_0\}$  but usually  $(k + n_0) \leq [n/2]$  otherwise the conclusions drawn from the experiment are suspect. Also too large number of instantaneous failures or early failures may indicate a factor not taken into consideration in modeling  $F \in \mathfrak{F}$ . One can introduce  $G_1, G_2, \dots \in \mathcal{G}$  for inliers but that will increase the nuisance parameters of  $\mathcal{G}$ . If we assume  $g(x)$  and  $f(x)$  are exponential with respective parameters  $\phi$  and  $\theta$  such that  $\phi < \theta$ , then the likelihood corresponds to  $M_k$  inlier model is given by

$$L(\underline{x}, k, \phi, \theta) = \frac{(n - n_0)!}{k!(n - n_0 - k)!} \frac{1}{\phi^k \theta^{n - n_0 - k}} \exp \left\{ - \sum_{i \in v} \frac{x_i}{\phi} - \sum_{i \notin v} \frac{x_i}{\theta} \right\} \quad (5.4)$$

And the likelihood corresponds to  $L_k$  inlier model is given by

$$L_k(\underline{x}|n_0, f, g) = \frac{k!(n - n_0 - k)}{\varphi(1, 2, \dots, k)} \frac{1}{\phi^k \theta^{n - n_0 - k}} e^{-\sum_{i=1}^k x_{(i)}/\phi} e^{-\sum_{i=k+1}^{n-n_0} x_{(i)}/\theta}$$

where

$$\varphi(1, 2, \dots, k) = \frac{(n - n_0 - k)\phi}{\theta} B\left(k + 1, \frac{(n - n_0 - k)\phi}{\theta}\right).$$

To estimate the value of  $k$ , we recommend that maximize  $\ln L_k(\underline{x}|n_0, \hat{\theta}, \hat{\phi}, \hat{v})$  for fixed  $k$  and then determine  $\text{Max}_{0 \leq k \leq N} \ln L_k(\underline{x}|n_0, \hat{\theta}, \hat{\phi}, \hat{v})$  and take  $\hat{k} = k_0$  at which this maximum is attained.

As an example we generated 5 observations from  $g$  as exp (0.04) and 10 observations from  $f$  as exp (5) thus having  $n = 15$  observations as 0.01339, 0.02679, 0.03442, 0.05519, 0.09459, 0.32254, 0.64367, 1.19427, 3.00276, 3.14612, 3.15643, 3.94635, 5.17659, 9.79405 and 12.52736. The graph of  $\varphi(1, 2, \dots, k)$  for the above data set is presented in Figure 1(b). The estimates are respectively  $\hat{\phi} = 0.05364$ ,  $\hat{\theta} = 4.6411$  and  $\hat{k} = 5$ .

## 6 Inliers as instantaneous and early failures

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 models of inliers. Assuming that the data is usually consisting of  $n_0$  instantaneous failures,  $k$  early failures as indicated by sample configuration and the rest  $n - n_0 - k$  observations belong to the target population. In the identified inliers model with  $r$  and  $v$  known the likelihood of the sample is

$$L = \binom{n}{n_0} (1 - \alpha)^{n_0} \alpha^{n-n_0} \left\{ \prod_{i=1}^k g(x_i, \phi) \prod_{i=k+1}^n f(x_i, \theta) \right\} \quad (6.1)$$

and the likelihood under the labeled slippage model is

$$L = \binom{n}{n_0} (1 - \alpha)^{n_0} \alpha^{n-n_0} \left\{ \frac{k!(n - n_0 - k)!}{\varphi(1, 2, \dots, k)} \prod_{i=1}^k g(x_{(i)}, \phi) \prod_{i=k+1}^n f(x_{(i)}, \theta) \right\} \quad (6.2)$$

where  $\varphi(1, 2, \dots, k)$  is as defined earlier.

The above likelihood of the sample assumes that between the experiments when units are placed on test we do not know which of the units fail instantaneously. i.e.  $X_{i_1} = 0, X = 0, \dots, X_{i_{r_0}} = 0$ , 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.

Since  $\varphi(1, 2, \dots, k)$  is a function of  $\varphi$  and  $\theta$ , for simplicity, we assume  $\varphi(1, 2, \dots, k) = \varphi_k(\phi, \theta)$  and  $g$  and  $f$  are exponential with parameters  $\phi$  and  $\theta$ , then the likelihood equation of the parameters under the model (6.2) are

$$\frac{\partial \ln L}{\partial \alpha} = -\frac{n_0}{1 - \alpha} + \frac{n - n_0}{\alpha} = 0 \quad (6.3)$$

$$\frac{\partial \ln L}{\partial \phi} = \frac{\partial \ln \varphi_k(\phi, \theta)}{\partial \phi} - \frac{k}{\phi} + \frac{1}{\phi^2} \sum_{i=1}^k x_{(i)} = 0 \quad (6.4)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \ln \varphi_k(\phi, \theta)}{\partial \theta} - \frac{(n - n_0 - k)}{\theta} + \frac{1}{\theta^2} \sum_{i=k+1}^{n-n_0} x_{(i)} = 0 \quad (6.5)$$

(6.3) can be solved to get the estimate of  $p$  as  $\hat{p} = \frac{n-n_0}{n}$ . Solving (6.4) and (6.5) simultaneously we get the estimate of  $\phi$  and  $\theta$ . Since

$$\varphi_k(\phi, \theta) = \frac{(n - n_0 - k)\phi \Gamma((k + 1)\Gamma(n - n_0 - k)\frac{\phi}{\theta})}{\theta \Gamma((n - n_0 - k)\frac{\phi}{\theta} + k + 1)}$$

and

$$\ln \varphi_k(\phi, \theta) = C + \ln \theta - \ln \Gamma(z) - \ln \Gamma(z + k + 1),$$



where  $z = \frac{(n-n_0-k)\phi}{\theta}$ . Therefore,

$$\frac{\partial}{\partial \phi} \ln \varphi_k(\phi, \theta) = \frac{1}{\phi} + [\Psi(z) - \Psi(z+k+1)] \left[ \frac{(n-n_0-k)}{\theta} \right]$$

and

$$\frac{\partial}{\partial \theta} \ln \varphi_k(\phi, \theta) = -\frac{1}{\theta} + [\Psi(z) - \Psi(z+k+1)] \left[ -\frac{(n-n_0-k)\phi}{\theta^2} \right]$$

where  $\Psi(t) = \frac{d}{dt} \log \Gamma(t)$ . Using [1], we get

$$\Psi(z) - \Psi(z+k+1) = -\sum_{j=1}^k \frac{1}{z+j}. \quad (6.6)$$

Thus the likelihood equations (6.4) and (6.5) can be written as

$$\frac{\partial \ln L}{\partial \phi} = \frac{k+1}{\phi} - \frac{(n-n_0-k)}{\theta} \sum_{j=1}^k \frac{1}{z+j} - \frac{1}{\phi^2} \sum_{i=1}^k x_{(i)} = 0 \quad (6.7)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{(n-n_0-k-1)}{\theta} + \frac{(n-n_0-k)\phi}{\theta^2} \sum_{j=1}^k \frac{1}{z+j} - \frac{1}{\theta^2} \sum_{k+1}^{n-r_0} x_{(i)} = 0 \quad (6.8)$$

Equations (6.7) and (6.8) may be solved simultaneously to get the estimates of  $\phi$  and  $\theta$ .

## 7 Testing of hypothesis about $k$

Consider the problem of testing  $H_0 : k = 0$  (i.e. no inliers) versus  $H_1 : k = 1$  (i.e one inlier). The joint pdf under  $H_0$  is given by  $M_0(n_0, \underline{x}|\theta)$  and under  $H_1$ , it is given by  $M_1(n_0, \underline{x}|\theta, \phi, v)$ .  $n_0$  in both cases is  $b(n, 1 - \alpha)$  and

$$M_1(n_0, \underline{x}|\theta, \phi, v) = \text{constant} \frac{1}{\phi} \exp\left\{-\frac{1}{\phi} x_{(1)}\right\} \frac{1}{\theta^{n-n_0-1}} \exp\left\{-\frac{1}{\theta} \sum_{i \notin v} x_{(i)}\right\}$$

and

$$M_0(n_0, \underline{x}|\theta) = \text{constant} \frac{1}{\theta^{n-n_0}} \exp\left\{-\frac{1}{\theta} \sum_{i=1}^{n-n_0} x_{(i)}\right\}.$$

The MLE of  $\theta$  under  $M_0$  is

$$\hat{\theta} = \frac{1}{n-n_0} \sum_{i=1}^{n-n_0} x_{(i)}$$

and

$$\max_{\theta} M_0(n_0, x|\theta) = \hat{M}_0 = \left( \frac{n - n_0}{\sum x_{(i)}} \right)^{n-n_0} \exp(-n - n_0).$$

Under  $H_1$ , the MLE of  $v$  is  $x_{\hat{v}} = x_{(1)}$ ,  $\hat{\phi} = x_{(1)}$  and  $\hat{\theta} = \frac{1}{n-n_0-1} \sum_{i=2}^{n-n_0} x_{(i)}$ . Therefore,

$$\text{Max}_{v, \theta_1, \theta_2} M_1(\underline{x}, \theta, \phi, v, n_0) = \hat{M}_1 = \left( \frac{n - n_0 - 1}{\sum_{i=2}^{n-n_0} x_{(i)}} \right)^{n-n_0} \frac{1}{x_{(1)}} \exp\{-n - n_0\}.$$

The likelihood ratio test is equivalent to rejecting  $H_0$  if  $\left(1 - \frac{x_{(1)}}{T}\right)^{n-n_0-1} \frac{x_{(1)}}{T} <$  constant, where  $T = \sum_{i=2}^{n-n_0} x_{(i)}$ . This is equivalent to rejecting  $H_0$  if  $\frac{x_{(1)}}{T} < c$ ,  $c$  is chosen such that

$$\text{Sup}_{\theta_1} P_{H_0} \left[ \frac{x_{(1)}}{\sum_{i=1}^{n-n_0} x_{(i)}} < c \right] = \alpha.$$

But the distribution of  $\frac{x_{(1)}}{\sum_{i=1}^{n-n_0} x_{(i)}}$  is independent of  $\theta$  under  $H_0$ . [11] have called this test as Cochran's test as it is analogous to the test based on  $\frac{x_{(1)}}{\sum_{i=1}^{n-n_0} x_{(i)}}$  derived by [4] to test the largest of set of variances as a fraction of that total in analysis of variance problems. [11] have derived distribution of the test statistic both under  $H_0$  and  $H_1$  and have shown that the test is equivalent to  $\frac{\sum_{i=1}^{n-n_0} x_{(i)}}{x_{(1)}} > c$ , where  $c = \frac{n-n_0}{1-(1-\alpha)^{\frac{1}{n-n_0-1}}} - 1$ . While the determination of  $c$  under  $H_0$  presented no problem, but the calculation of power was more difficult. Assuming that  $x_{(1)}$  is an inlier, i.e. using  $L_1$ , it was proved that the power of the test for one inlier is

$$P_1(\lambda) = 1 - \left( \frac{c - (n - n_0) + 1}{c + \lambda} \right)^{(n-n_0-1)}, \quad \text{where } \lambda = \frac{\theta}{\phi}.$$

In a similar way we can test for  $M_k$  ( $k$  inliers) versus no inliers. In this case the likelihood ratio test is obtained as  $\lambda(\underline{x}) \approx \left(1 - \frac{T_1}{T}\right)^{n-n_0-1} \frac{T_1}{T}$ , where  $T_1 = \sum_{i=1}^k x_{(i)}$  and  $T = \sum_{i=1}^{n-n_0} x_{(i)}$  and we reject  $H_0$  if  $\frac{T_1}{T} < C$ , where  $C$  is such that  $\text{Sup}_{\theta_1} P_{H_0} \left[ \frac{T_1}{T} < C \right] = \alpha$ . In Table 1 we present the simulated values of  $P_1(\lambda)$  and  $P_k(\lambda)$  for various values of  $k$  and  $\lambda$ . Each value in Table 3 is corresponding to 5000 simulations. It is seen that for  $\lambda = 1$ , the test attains the level.

## 8 Numerical computation

Here we present a method of estimating the inliers through various models described above on [20] data on drying of woods under different experiments

Table 1: The values of  $P_1(\lambda)$  and  $P_k(\lambda)$  for  $\alpha = 0.05$  (Cochran test)

$N$	$k$	$\lambda$					
		1	5	10	15	20	30
10	1	.050	.069	.093	.116	.138	.179
	2	.050	.052	.079	.093	.105	.148
	3	.050	.046	.067	.078	.088	.089
20	1	.050	.059	.072	.083	.095	.118
	2	.050	.057	.067	.078	.079	.095
	4	.050	.055	.057	.068	.068	.070
	6	.050	.052	.054	.054	.057	.066
30	1	.049	.056	.065	.073	.080	.098
	2	.049	.055	.060	.065	.077	.082
	4	.049	.054	.058	.060	.069	.074
	6	.049	.053	.056	.058	.060	.065
	8	.049	.053	.055	.056	.057	.059

Table 2: Estimates and their variances

Schedule	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\mu}$	$AV(\hat{\alpha})$	$AV(\hat{\theta})$	$AV(\hat{\alpha}\hat{\theta})$
1	0.6487	4.8686	3.158	0.006159	0.987568	0.23627
2	0.5405	2.8207	1.296	0.006712	0.397847	0.03581

and schedules. We reproduce Vanmann's data on Experiment 3 on two batches of 37 boards by using two different schedules.

**Schedule 1**  $x_i = 0$ ,  $i = 1, 2, \dots, 13$  and the other 24 positive observations arranged in increasing order of their magnitude are 0.08, 0.32, 0.38, 0.46, 0.71, 0.82, 1.15, 1.23, 1.40, 3.00, 3.23, 4.03, 4.20, 5.04, 5.36, 6.12, 6.79, 7.90, 8.27, 8.62, 9.50, 10.15, 10.58 and 17.49.

**Schedule 2**  $x_i = 0$ ,  $i = 1, 2, \dots, 17$  and the other 20 positive observations arranged in increasing order of their magnitude are 0.02, 0.02, 0.02, 0.04, 0.09, 0.23, 0.26, 0.37, 0.93, 0.94, 1.02, 2.23, 2.79, 3.93, 4.47, 5.12, 5.19, 5.39, 6.83 and 8.22.

[11] analyzed the above data using models with instantaneous failures only and exponential distribution with mean  $\theta$ . Now  $E(X) = \mu = \alpha\theta$  and lesser the value of  $\mu$  better is the process and  $\hat{\mu} = \hat{\alpha}\hat{\theta}$  and  $AV(\hat{\mu}) = \frac{\hat{\alpha}\hat{\theta}^2(2-\hat{\alpha})}{n}$ . The estimates and their estimated variances for the two Schedules above are given in Table 2.

We next consider the  $M_k$  inliers model. Then  $\ln L$  is given by (5.4) and MLE's are  $\hat{\alpha} = \frac{n-n_0}{n}$ ,  $\hat{\phi} = \frac{1}{k} \sum_{i=1}^k x_{(i)}$ ,  $\hat{\theta} = \frac{1}{n-n_0-k} \sum_{i=k+1}^{n-n_0} x_{(i)}$ . To determine  $k$  we maximize  $\ln L(\underline{x}|n_0, \hat{\alpha}, \hat{\theta}_1, \hat{\theta}, k)$  over  $k$ . Accordingly we find that for Schedule 1  $\hat{k} = 9$  and for Schedule 2  $\hat{k} = 7$ . For  $L_k$  inliers model also we obtain similar estimates for  $k$ . The other estimates are given in Table 3. We

Table 3:

Schedule	Model	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\theta}$	$AV(\hat{\alpha})$	$AV(\hat{\phi})$	$AV(\hat{\theta})$
1	$M_k$	0.6487	07278	7.3520	0.0061	0.0589	6.0058
	$L_k$	0.6487	05881	5.9148	0.0059	0.0565	5.0067
2	$M_k$	0.5405	0971	3.6485	0.0067	0.0013	1.9017
	$L_k$	0.5405	0916	3.0458	0.0064	0.0010	1.8876

Table 4:

Schedule	Model	$\hat{\alpha}$	$\hat{\theta}$	$AV(\hat{\alpha})$	$AV(\hat{\theta})$
1	Early failure	0.5129	5.9520	0.0067	4.8796
	Lai et al	0.5673	6.2122	0.0056	4.3890
2	Early failure	0.4187	3.0458	0.0058	1.4567
	Lai et al	0.4343	3.7685	0.0057	1.0346

also observe that  $n_0 + \hat{k}$  is 22 in Schedule 1 and 26 in Schedule 2 indicating that it is more than  $\lceil 37/2 \rceil = 18$  and the inferences drawn from the experiment are suspect. However estimates of both  $\hat{\phi}$  and  $\hat{\theta}$  are larger in Schedule 1 than Schedule 2. Also  $\mu = \alpha\theta$  is estimated at 4.7692 for Schedule 1 which is larger than 1.972 for Schedule 2 and hence Schedule 2 is preferred.

The estimates correspond to early failures model are  $\hat{\alpha} = \frac{n-n_0}{n} \exp(\hat{\delta}/\hat{\theta})$  and  $\hat{\theta} = \frac{\sum_{i=1}^{n-n_0}}{n-n_0} - \hat{\delta}$ . For [13] model, we do not have any closed form expressions for the estimates and are numerically estimated through likelihood procedure. See the appendix of [13] paper for the R-code for estimation. In table 4 we present the estimates where a value of  $\delta = 1.4$  is assumed for Schedule 1 and  $\delta = 0.10$  is assumed for Schedule 2.

## Acknowledgements

The author thanks the referee and editor for their comments and suggestions.

## References

- [1] Abramovitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical functions.*, General publishing Company Ltd, Canada.
- [2] Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability Mass at the origin. *J. Amer. Stat. Assn.* Vol. 50, 901–908.
- [3] Barnett, V. and Lewis, T. (1984). *Outliers in Statistical data*, John Wiley & Sons, New York.

- [4] Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Annals of Eugenics*, 11, 47–52.
- [5] Gather, U. and Kale, B. K. (1992). Outlier generating models: A review. In contributions to Stochastics. Ed. N. Venugopal. Wiley Eastern, 59–85.
- [6] Jayade, V.P. and Prasad, M.S. (1990). Estimation of parameters of mixed failure time distribution. *Comm. Statist. Theory and Methods*, 19(12), 4667–4677.
- [7] Kale, B.K. (1998). Optimal estimating equations for discrete data with higher frequencies at a point. *J. Indian. Statist. Asso.*, 36, 125–136.
- [8] Kale, B.K. (2003). Modified failure time distributions to accommodate instantaneous and early failures. In Industrial Mathematics and Statistics. Ed. J. C. Misra, pp. 623–648. Narosa Publishing House, New Delhi.
- [9] Kale, B.K. and Muralidharan, K. (2000). Optimal estimating equations in mixture distributions accommodating instantaneous or early failures. *J.Indian .Statist. Assoc.*, 38, 317–329.
- [10] Kale, B.K. and Muralidharan, K. (2006). Maximum likelihood estimation in presence of Inliers. *Journal of Indian Society for Prob. and Statist.* 10, 65–80.
- [11] Kale, B.K. and Muralidharan, K. (2007). Masking effect of inliers. *Journal of Indian Statist. Assoc.* 45(1), 33–49.
- [12] Kleyle , R.M. and Dahiya, R.L. (1975). Estimation of parameters of mixed failure time distribution from censored data. *Comm. Statist. Theory and Methods*, 4(9),873–882.
- [13] Lai, C. D., Khoo, B. C., Muralidharan, K. and Xie, M. (2007). Weibull model allowing nearly instantaneous failures. *J. Applied Mathematics and Decision Sciences*, Article ID 90842, 11 pages.
- [14] Muralidharan, K. (1999). Tests for the mixing proportion in the mixture of a degenerate and exponential distribution. *J. Indian Stat. Assn.*, Vol. 37(2), 105–119.
- [15] Muralidharan, K. (2000). The UMVUE and Bayes estimate of reliability of mixed Failure time distribution., *Comm.Statist-Simulations & computations*, 29(2), 603–619.
- [16] Muralidharan, K. and Kale, B.K. (2002). Modified Gamma distribution with Singularity at zero. *Comm. Statist-Simulations & computations*, 31(1), 143–158.

- [17] Muralidharan, K. and Lathika, P. (2004). The concept of inliers. Proceedings of the First Sino-International Symposium on Probability, Statistics and Quantitative Management., Taiwan, October, 77–92.
- [18] Muralidharan, K and Lathika, P. (2006). Analysis of instantaneous and early failures in weibull distribution. *Metrika*, 64, 305–316.
- [19] Vannman, K. (1991). Comparing samples from nonstandard mixtures of distributions with Applications to quality comparison of wood. Research report 1991:2 submitted to Division of Quality Technology, Lukea University, Lukea, Swedon.
- [20] Vannman, K. (1995). On the distribution of the estimated mean from the nonstandard Mixtures Of distribution. *Comm. Statist. Theory and Methods*, 24(6), 1569–1584.
- [21] Statistical models and analysis in Auditing: Panel on nonstandard mixtures of distributions, *Statistical Science*, vol.4(1), 2–33, 1989.

ProbStat Forum is an e-journal. For details please visit; <http://www.probstat.org.in>.