Statistics and Applications {**ISSN 2452-7395 (online)**}
Volume 17, No. 1, 2019 (New Series), pp 193–207

UMVU Estimation in Gamma Model with Inliers at Zero and One

K. Muralidharan and Pratima Bavagosai
Department of Statistics, Faculty of Science,
The Maharaja Sayajirao University of Baroda, Vadodara 390002, India

Received: March 09, 2019; Revised: March 16, 2019; Accepted: March 20, 2019

Abstract

Inliers (instantaneous or early failures) are natural occurrences of a life test, where some of the items fail immediately or within a short time of the life test. These failures are either due to mechanical failures, accelerated pressures and interventions, or faulty items. The inconsistency of such life data is modeled using a nonstandard mixture of distributions; as there can be degeneracy in the model with discrete and continuous measures. The model discussed here is a nonstandard mixture of Gamma distribution with degeneracy happens at two discrete points at zero and one. In this paper, we provide the Uniformly Minimum Variance Unbiased (UMVU) estimation of parameters from a Gamma distribution with inliers at zero and one. Along with various other characteristics, the model was numerically illustrated on a real life example.

Key words: Early failures; Failure time distribution; Inliers; Instantaneous failures.

1 Introduction

A Non-standard mixture of distribution is a mixture of degenerate (discrete masses at some points) and continuous observations measured on a particular life time characteristics. Because there are large mass of discrete points and continuous points on a measurable scale, they generally contains inliers and outliers. Inliers is an observation (or a group of observations) are sufficiently small relative to the rest of the observations, which appears to be inconsistent with the remaining dataset, whereas, outliers are large inconsistent observations in relation to the remaining observations. Inliers are either the result of the instantaneous failures or the early failures or both, experienced in life testing experiments, clinical trials, weather predictions, geographic information systems, athlete performance analysis, and many other such applications. The test items that fail at time 0 are called the *instantaneous failures* and the test items that fail prematurely are called the *early failures*. Kale and Muralidharan (2000) was the first to introduce the term *inliers* in connection with the estimation of (p, θ) of early failure model with modified failure time distribution (FTD) as an exponential with mean θ. Some of the practical contexts, where degeneracy can happen at two discrete points with a mix of positive and continuous observations are the following:

1. The size of tumor lesions is of interest to treat *Hematologic malignancy* patients. The measurement effect is zero who have lesions absent (or due to disappearance of tumor during treatment), though who have lesions present at baseline that are evaluable but do not meet the definitions of measurable disease may be considered as measurement

Corresponding Author: K. Muralidharan
Email: lmv_murali@yahoo.com

1, otherwise lesions can be accurately measured as longest diameter to be recorded in at least one dimension by chest x-ray, with CT scan or with calipers by clinical exam. Similarly, in studies like Bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusions, lymphangitis cutis/pulmonitis, inflammatory breast disease, and abdominal masses, either the effect is absent or present but not followed by CT or MRI, are considered as non-measurable otherwise accurately measurable on continuous scale.

2. 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, some component that does not fail on installation but fails with negligible life (may be coded as one for simplicity), and others that will have a life length that is a positive random variable whose distribution may take different forms.
3. In a clinical trial laboratory, a particular drug is designed and given to certain species of hens so that the new chicks have weight greater than usual. The possible weight of chicks may be modeled as a continuous distribution, with discrete mass at ‘zero’ and ‘one’ where zero measures those chicks having no gain of weight, where one measures those chicks with negligible gain of weight than usual, and a continuous variable having target gain in weight.
4. The rainfall measurement at a place recorded during a season is modeled as a continuous distribution, with a discrete mass at ‘zero’ where zero measures those days having no rainfall, at ‘one’ where one measures those days with no rain but humid and cloudy conditions, and a continuous variable having some positive amount of rain.

Aitchison (1955) was the first to discuss the inference problem of instantaneous failures in life testing. The author has provided an efficient estimation of parametric functions under various probability models. Some earlier studies on these type of models are done by Kleyle and Dahiya (1975), Jayade and Prasad (1990), Vannman (1991, 1995), Kale (1998, 2003), Muralidharan (1999) and Shinde and Shanubhougue (2000), Dixit (2003). The inferences on inliers was studied in detail by Kale and Muralidharan (2000, 2006), Muralidharan and Kale (2002), Muralidharan and Lathika (2004, 2006), Muralidharan (2010), Adlouni et al. (2011), Muralidharan and Arti (2008, 2013), Muralidharan and Pratima (2016a,b), Bavagosai and Muralidharan (2016) and the references contained therein.

In this paper, we consider distribution function (DF) of the model with inliers at 0 and 1 as

$$H(x;p_1,p_2,\underline{\tau}) = \begin{cases} 0, & x<0 \\ p_1, & 0\le x<1 \\ p_1+p_2, & x=1 \\ p_1+p_2+(1-p_1-p_2)\frac{F(x;\underline{\tau})-F(1;\underline{\tau})}{1-F(1;\underline{\tau})}, & x\ge 1 \end{cases} . \quad (1)$$

The fact is that the probability measure generated by $H(.)$ is composed of three measures, say μ_1, μ_2 and μ_3, where μ_3 is absolutely continuous with respect to the Lebesgue measure on R and μ_1 and μ_2 are singular with respect to the Lebesgue measure on R. Here p_1 and p_2 is the proportion of 0 and 1 observations and $F(x;\ \underline{\tau})$ is the Failure time distribution (FTD) with the vector of parameters $\underline{\tau}$ involved in it. If we consider FTD as Gamma distribution with parameters $\underline{\tau}=(\alpha,\theta)$ having pdf

$$f(x;\alpha,\theta)=\frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\theta^{\alpha}\,\Gamma\alpha},x>0;\alpha>0,\theta>0 \tag{2}$$

Then the density function of the model (1) reduces to

$$h(x;p_1,p_2,\alpha,\theta)=\begin{cases}p_1, & x=0\\ p_2, & x=1\\ (1-p_1-p_2)\dfrac{x^{\alpha-1}e^{-\frac{(x-1)}{\theta}}}{\theta^{\alpha}\,\Gamma\alpha\left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}\right)}, & x>1\end{cases} \tag{3}$$

This kind of model was first studied by Muralidharan and Bavagosai (2017, 2018) and Bavagosai and Muralidharan (2018) with FTD as exponential, Weibull and Pareto.

In Section 2, we propose the unbiased estimation for model parameters along with the distributional properties of complete sufficient statistics, in Section 3, UMVUE of various parametric functions including pdf and survival function along with the UMVUE of their variance studies. As an illustration, we consider the breast cancer tumor size data from cancer genomic studies for implementing the proposed model in Section 4 and some conclusions are given in Section 5.

2 Unbiased estimation

Define

$$I_1(x)=\begin{cases}1, & x=0\\ 0, & otherwise\end{cases}$$

and

$$I_2(x)=\begin{cases}1, & x=1\\ 0, & otherwise\end{cases}\quad .$$

Then, the model in (3) can be expressed as

$$h(x;p_1,p_2,\alpha,\theta)={p_1}^{I_1(x_i)}{p_2}^{I_2(x_i)}\left((1-p_1-p_2)\frac{{x_i}^{(\alpha-1)}e^{-\frac{(x_i-1)}{\theta}}}{\theta^{\alpha}\,\Gamma\alpha\left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}\right)}\right)^{1-I_1(x_i)-I_2(x_i)} \tag{4}$$

here $\in T\subseteq R,\ \underline{\theta}=(p_1,p_2,\theta)\in\Omega$, and assuming α known integer, which can also be written as

$$h(x;\underline{\theta})=\left(\frac{x^{\alpha-1}}{\Gamma\alpha}\right)^{(1-I_1(x)-I_2(x))}\frac{\left(\dfrac{p_1\theta^{\alpha}\ \sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}}{1-p_1-p_2}\right)^{I_1(x)}\left(\dfrac{p_2\ \theta^{\alpha}\ \sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}}{1-p_1-p_2}\right)^{I_2(x)}\left(e^{-\frac{1}{\theta}}\right)^{(x-1)(1-I_1(x)-I_2(x))}}{\dfrac{\theta^{\alpha}\ \sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}}{1-p_1-p_2}}$$

$$= (a(x))^{(1-C_1(x)-C_2(x))} \frac{\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{C_i(x)}}{g(\underline{\theta})} \tag{5}$$

where, $a(X) = \frac{X^{\alpha-1}}{\Gamma\alpha}$; $h_1(\underline{\theta}) = \frac{p_1\theta^{\alpha}\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}$; $h_2(\underline{\theta}) = \left(\frac{p_2\,\theta^{\alpha}\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)$; $h_3(\underline{\theta}) = e^{-\frac{1}{\theta}}$; $g(\underline{\theta}) = \frac{\theta^{\alpha}\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}$; $C_1(X) = I_1(X)$; $C_2(X) = I_2(X)$ and $C_3(X) = (X-1)\left(1 - I_1(X) - I_2(X)\right)$. Also $a(X) > 0$, $C_i(X), i$ = 1, 2 and 3 are nontrivial real-valued statistics, $g(\underline{\theta})$ and $h_i(\underline{\theta})$ are at least twice differentiable functions of θ_i, i = 1, 2 and 3. The $g(\underline{\theta})$ is such that

$$g(\underline{\theta}) = \int_{x>1}(a(x))^{(1-C_1(x)-C_2(x))}\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{C_i(x)} dx. \tag{6}$$

The density in (5) so obtained is defined with respect to a measure $\mu(x)$ which is the sum of Lebesgue measure over $(1,\infty)$ a well-known form of a three parameter exponential family with natural parameters $(\eta_1,\eta_2,\eta_3) = \left(\log\left(\frac{p_1\theta^{\alpha}\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right), \log\left(\frac{p_2\,\theta^{\alpha}\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right), \log\left(e^{-\frac{1}{\theta}}\right)\right)$ generated by underlying indexing parameters $\underline{\theta} = (p_1,p_2,\theta)$. Here $C(X) = \left(C_1(X), C_2(X), C_3(X)\right)$ is minimal sufficient for $\underline{\theta} = (p_1,p_2,\theta)$ as $C_1(X)$, $C_2(X)$ and $C_3(X)$ do not satisfy any linear restriction. The (η_1,η_2,η_3) too do not satisfy any linear constraint and hence the natural parameter space is a parameter space in E_3 containing a three-dimensional rectangle making (5) a full rank family. The statistics $C(X) = \left(C_1(X), C_2(X), C_3(X)\right) = \left(I_1(X),\ I_2(X), (X-1)\left(1 - I_1(X) - I_2(X)\right)\right)$ is jointly complete sufficient for $\underline{\theta} = (p_1,p_2,\theta)$ and the distribution of $C(X)$ is also a three parameter exponential family.

2.1 Distributional properties of $\boldsymbol{C(X) = \left(C_1(x), C_2(x), C_3(x)\right)}$

Referring Jani and Singh (1996), differentiating $g(\underline{\theta})$ in (6) with respect to p_1, p_2, and θ, and since the range T is independent of $\underline{\theta}$ under the condition of interchangeability of differentiation and integration, we get

$$\underline{G} = A\,\underline{\mu}\,,\ |A| \neq 0 \tag{7}$$

where

$$\underline{G}=\begin{bmatrix}\frac{\partial\log g(\underline{\theta})}{\partial p_1}\\ \frac{\partial\log g(\underline{\theta})}{\partial p_2}\\ \frac{\partial\log g(\underline{\theta})}{\partial\theta}\end{bmatrix}=\begin{bmatrix}\frac{1}{1-p_1-p_2}\\ \frac{1}{1-p_1-p_2}\\ \frac{\alpha}{\theta}+\frac{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i-1)\theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}}\end{bmatrix}$$

$$\underline{\mu}=\begin{bmatrix}E(C_1(x))\\ E(C_2(x))\\ E(C_3(x))\end{bmatrix}=\begin{bmatrix}E(I_1(x))\\ E(I_2(x))\\ E\left((x-1)(1-I_1(x)-I_2(x))\right)\end{bmatrix}$$

and

$$A=\begin{bmatrix}\frac{\partial\log h_1(\underline{\theta})}{\partial p_1} & \frac{\partial\log h_2(\underline{\theta})}{\partial p_1} & \frac{\partial\log h_3(\underline{\theta})}{\partial p_1}\\ \frac{\partial\log h_1(\underline{\theta})}{\partial p_2} & \frac{\partial\log h_2(\underline{\theta})}{\partial p_2} & \frac{\partial\log h_3(\underline{\theta})}{\partial p_2}\\ \frac{\partial\log h_1(\underline{\theta})}{\partial\theta} & \frac{\partial\log h_2(\underline{\theta})}{\partial\theta} & \frac{\partial\log h_3(\underline{\theta})}{\partial\theta}\end{bmatrix}=\begin{bmatrix}\frac{1}{p_1}+\frac{1}{1-p_1-p_2} & \frac{1}{1-p_1-p_2} & 0\\ \frac{1}{1-p_1-p_2} & \frac{1}{p_2}+\frac{1}{1-p_1-p_2} & 0\\ \frac{\alpha}{\theta}+\frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma(i-1)\theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}} & \frac{\alpha}{\theta}+\frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma(i-1)\theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}} & \frac{1}{\theta^2}\end{bmatrix}$$

Equation (7) gives

$$E(C_i(x))=\frac{|A_i|}{|A|}\ , i=1,2\ and\ 3$$

where A_i is obtained by replacing i[th] column of A by the elements of $\underline{G}$. Hence,

$$\underline{\mu}=\begin{bmatrix}E(C_1(x))\\ E(C_2(x))\\ E(C_3(x))\end{bmatrix}=\begin{bmatrix}p_1\\ p_2\\ (1-p_1-p_2)\theta^2\left(\frac{\alpha}{\theta}-\frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma(i-1)\theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}}\right)\end{bmatrix} \tag{8}$$

Now joint moments of $C_1^{k_1}(\boldsymbol{x})$, $C_2^{k_2}(x)$ and $C_3^{k_3}(x)$ are given as

$$E\left(C_1^{k_1}(\boldsymbol{x})\,C_2^{k_2}(\boldsymbol{x})\,C_3^{k_3}(\boldsymbol{x})\right)=\int_x C_1^{k_1}(\boldsymbol{x})\,C_2^{k_2}(\boldsymbol{x})\,C_3^{k_3}(\boldsymbol{x})\,a(x)\frac{\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{C_i(x)}}{g(\underline{\theta})}dx$$

which on differentiating with respect to p_1, p_2 and θ and using (7), gives a system of three non-homogeneous equations

$$\underline{G_1}=A\,\underline{V}\ , |A|\neq 0 \tag{9}$$

where

$$\underline{G_1} = \begin{bmatrix} \frac{\partial \log E\left(C_1^{k_1}(x)\, C_2^{k_2}(x)\, C_3^{k_3}(x)\right)}{\partial p_1} \\ \frac{\partial \log E\left(C_1^{k_1}(x)\, C_2^{k_2}(x)\, C_3^{k_3}(x)\right)}{\partial p_2} \\ \frac{\partial \log E\left(C_1^{k_1}(x)\, C_2^{k_2}(x)\, C_3^{k_3}(x)\right)}{\partial \theta} \end{bmatrix}$$

$$\underline{V} = \begin{bmatrix} E\left(C_1^{k_1+1}(\boldsymbol{x})\, C_2^{k_2}(\boldsymbol{x})\, C_3^{k_3}(\boldsymbol{x})\right) - E\big(C_1(x)\big)E\left(C_1^{k_1}(\boldsymbol{x})\, C_2^{k_2}(\boldsymbol{x})\, C_3^{k_3}(\boldsymbol{x})\right) \\ E\left(C_1^{k_1}(\boldsymbol{x})\, C_2^{k_2+1}(\boldsymbol{x})\, C_3^{k_3}(\boldsymbol{x})\right) - E\big(C_2(x)\big)E\left(C_1^{k_1}(\boldsymbol{x})\, C_2^{k_2}(\boldsymbol{x})\, C_3^{k_3}(\boldsymbol{x})\right) \\ E\left(C_1^{k_1}(\boldsymbol{x})\, C_2^{k_2}(\boldsymbol{x})\, C_3^{k_3+1}(\boldsymbol{x})\right) - E\big(C_3(x)\big)E\left(C_1^{k_1}(\boldsymbol{x})\, C_2^{k_2}(\boldsymbol{x})\, C_3^{k_3}(\boldsymbol{x})\right) \end{bmatrix} = \begin{bmatrix} \sigma_{1(1,2,3)} \\ \sigma_{2(1,2,3)} \\ \sigma_{3(1,2,3)} \end{bmatrix}, \text{ (say).}$$

Using Cramer's rule for the solution of a system of linear equations (9) gives

$$\sigma_{i(1,2,3)} = \frac{|A_i|}{|A|}\ , i = 1,2 \ and\ 3$$

where A_i is obtained by replacing i^th^ column of A by the elements of $\underline{G_1}$. For $k_i = 1$ and $k_j = 0\ \forall\ i \neq j = 1,2\ and\ 3,$ we get covariance between $C_i(x)$ and $C_j(x)$ as

$$\sigma_{i(1,2,3)} = \frac{|A_i\ |_{(k_i=1;k_j=0),i\neq j}}{|A|}.$$

Thus, we have the variance-covariance matrix Σ as

$$\Sigma = \left[\sigma_{ij}\right]_{3\times 3} = \frac{\left(|A_i\ |_{(k_i=1;k_j=0),i\neq j}\right)}{|A|}$$

If A_{ij} is the cofactor of the element a_{ij} of A, then

$$|A_i\ |_{(k_i=1;k_j=0),i\neq j=1,2,3} = A_{1i}\frac{\partial}{\partial p_1}E\big(C_i(x)\big) + A_{2i}\frac{\partial}{\partial p_2}E\big(C_i(x)\big) + A_{3i}\frac{\partial}{\partial \theta}E\big(C_i(x)\big)$$

and hence

$$\Sigma = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -\theta^2\, p_1(1-p_1-p_2)\omega \\ -p_1p_2 & p_2(1-p_2) & -\theta^2\, p_2(1-p_1-p_2)\omega \\ -\theta^2\, p_1(1-p_1-p_2)\omega & -\theta^2\, p_2(1-p_1-p_2)\omega & \theta^2(1-p_1-p_2)\omega_1 \end{bmatrix} \tag{10}$$

where $\omega = \left(\frac{\alpha}{\theta} - \frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma(i-1)\theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}}\right)$ and $\omega_1 = \left(\alpha + \frac{\sum_{i=2}^{\alpha-1}\frac{1}{\Gamma(i-1)\theta^{i}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}} + \left(\frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma i\ \theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^{i}}}\right)^2\right).$

2.2 Uniformly Minimum Variance Unbiased Estimation of parameters

Suppose n items placed on life test, where r_1 items have life zero whereas r_2 items have life 1 and remaining $n-r_1-r_2$ items have life greater than 1 is denoted by $X_1, X_2, \dots, X_{n-r_1-r_2}$ with pdf $h \in \mathcal{H}$ as given in (3), the joint density function is

$$h(\underline{x};\, \underline{\theta}) = \left(\frac{x^{\alpha-1}}{\Gamma\alpha}\right)^{n-r_1-r_2} p_1{}^{r_1} p_2{}^{r_2}\, (1-p_1-p_2)^{(n-r_1-r_2)} \frac{e^{-\frac{\sum_{i=1}^{n-r_1-r_2}(x_i-1)}{\theta}}}{\theta^{\alpha(n-r_1-r_2)}\left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}\right)^{n-r_1-r_2}} \tag{11}$$

$$= \left(\frac{x^{\alpha-1}}{\Gamma\alpha}\right)^{n-r_1-r_2} \frac{\left(\frac{p_1\theta^\alpha\, \sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{z_1} \left(\frac{p_2\,\theta^\alpha\, \sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{z_2} \left(e^{-\frac{1}{\theta}}\right)^{z_3}}{\left(\frac{\theta^\alpha\, \sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{n}}$$

where $Z_1 = \sum_{i=1}^{n} C_1(X_i) = \sum_{i=1}^{n} I_1(X_i) = r_1$; $Z_2 = \sum_{i=1}^{n} C_2(X_i) = \sum_{i=1}^{n} I_2(X_i) = r_2$; and $Z_3 = \sum_{i=1}^{n} C_3(X_i) = \sum_{i=1}^{n-r_1-r_2}(X_i-1)$

Hence by Neyman Factorization theorem $Z = (Z_1, Z_2, Z_3)$ is jointly sufficient for $\underline{\theta} = (p_1, p_2, \theta)$. Also,

$$h(\underline{x}; \underline{\theta}) = \frac{n!}{r_1!\ r_2!\ (n-r_1-r_2)!} p_1{}^{r_1} p_2{}^{r_2} (1-p_1-p_2)^{\,(n-r_1-r_2)}$$

$$\frac{r_1!\ r_2!\ (n-r_1-r_2)!\,1}{n!} \left(\frac{x^{\alpha-1}}{\Gamma\alpha}\right)^{n-r_1-r_2} \frac{e^{-\frac{\sum_{i=1}^{n-r_1-r_2}(X_i-1)}{\theta}}}{\theta^{\alpha(n-r_1-r_2)}\ \left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}\right)^{n-r_1-r_2}}$$

$$= \ \mathrm{P}(Z_1 = r_1, Z_2 = r_2)\ \ h(\underline{x}; \theta | Z_1 = r_1, Z_2 = r_2)$$

Here distribution of (Z_1, Z_2) is trinomial and is a complete family of distribution and

$$h(\underline{x}; \theta \big| Z_1 = r_1, Z_2 = r_2) = \left(\frac{x^{\alpha-1}}{\Gamma\alpha}\right)^{n-r_1-r_2} \frac{r_1!\ r_2!\,(n-r_1-r_2)!}{n!} \frac{e^{-\frac{\sum_{i=1}^{n-r_1-r_2}(X_i-1)}{\theta}}}{\theta^{(n-r_1-r_2)\alpha}\left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}\right)^{n-r_1-r_2}}$$

which belongs to one-parameter exponential family. Hence $(Z_3|Z_1, Z_2)$ is complete sufficient for θ and also a member of exponential family. The distribution of $(Z_3|Z_1, Z_2)$ is Gamma with parameter $(n-r_1-r_2, \theta)$ with pdf

$$h(z_3; \theta | n-r_1-r_2) = \frac{z_3{}^{(n-r_1-r_2)\alpha-1}}{\Gamma(n-r_1-r_2)\alpha} \frac{e^{-\frac{z_3}{\theta}}}{\theta^{(n-r_1-r_2)\alpha}\left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}\right)^{n-r_1-r_2}}, z_3 > 0;\ \theta > 0$$

which depends only on θ and is also a complete family of distribution. Therefore, using result of Jayade (1993) $Z = (Z_1, Z_2, Z_3)$ is complete sufficient for $\underline{\theta} = (p_1, p_2, \theta)$. The Joint distribution of $Z = (Z_1, Z_2, Z_3)$ is

$$h_Z(z;\underline{\theta}) = \frac{n!}{r_1!\ r_2!\ (n-r_1-r_2)!} p_1{}^{r_1} p_2{}^{r_2} (1-p_1-p_2)^{(n-r_1-r_2)} \frac{z_3{}^{(n-r_1-r_2)\alpha-1}}{\Gamma(n-r_1-r_2)\alpha} \frac{e^{-\frac{z_3}{\theta}}}{\theta^{(n-r_1-r_2)\alpha}\left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}\right)^{n-r_1-r_2}},$$

$$0 \le r_1, r_2 \le n;\ z_3 > 0;\ 0 \le p_1, p_2 \le 1;\ \theta > 0$$

$$= B(z_1, z_2, z_3, n)\ \frac{\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{z_i}}{g(\underline{\theta})^n}$$

where

$$B(z_1, z_2, z_3, n) = \begin{cases} \frac{n!}{r_1!\ r_2!\ (n-r_1-r_2)!} \frac{z_3{}^{(n-r_1-r_2)\alpha-1}}{\Gamma(n-r_1-r_2)\alpha},\ z_3 > 0;\ r_1 + r_2 - 1 < n \\ 1, \qquad z_3 = 0;\ r_1 = 0\ \ or\ r_2 = 0 \end{cases} \tag{12}$$

$z_i \in T(n) \subseteq R, \underline{\theta} \in \Omega$. Here $z = (z_1, z_2, z_3, n)$ and $B(z_1, z_2, z_3, n)$ are such that

$$g(\underline{\theta})^n = \int_{z_1\in T(n)}\int_{z_2\in T(n)}\int_{z_3\in T(n)} B(z_1, z_2, z_3,, n) \prod_{i=1}^{3}\left(h_i(\ \underline{\theta}\ )\right)^{z_i} dz_1\, dz_2\, dz_3$$

Since $E\left(C_1(x)\right) = p_1$, $E\left(C_2(x)\right) = p_2$ and

$$E\left(C_3(x)\right) = (1-p_1-p_2)\theta^2\left(\frac{\alpha}{\theta} - \frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma(i-1)\theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}\right), \text{ then}$$

$$E(Z_1) = E\left(\sum_{j=1}^{n} C_1(x_j)\right) = \sum_{j=1}^{n} E\left(I_1(x_j)\right) = n\, p_1,$$

$$E(Z_2) = E\left(\sum_{j=1}^{n} C_2(x_j)\right) = \sum_{j=1}^{n} E\left(I_2(x_j)\right) = n\, p_2,$$

and

$$E(Z_3) = E\left(\sum_{j=1}^{n} C_3(x_j)\right) = \sum_{i=1}^{n-r_1-r_2} E\left((x_i - 1)\left(1 - I_1(x_j) - I_2(x_j)\right)\right)$$

$$= (n - r_1 - r_2)(1-p_1-p_2)\theta^2\left(\frac{\alpha}{\theta} - \frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma \mathrm{i}\ \theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}\right),$$

which in turn give UMVUE's of p_1, p_2 and θ as

$$\hat{p}_1 = \frac{Z_1}{n} = \frac{r_1}{n} \tag{13}$$

$$\hat{p}_2 = \frac{Z_2}{n} = \frac{r_2}{n} \tag{14}$$

and $\hat{\theta}$ is obtained by solving the non-linear equation

$$\frac{z_3}{(n-r_1-r_2)(1-p_1-p_2)} - \theta^2\left(\frac{\alpha}{\theta} - \frac{\sum_{i=1}^{\alpha-1}\frac{1}{\Gamma \mathrm{i}\ \theta^{i+1}}}{\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}}\right) = 0 \tag{15}$$

3 UMVU Estimation of parametric functions

Let $X_1, X_2, \dots, X_n$ be a random sample from (6), then following Jani and Singh (1995) there exists UMVUE of $\Phi(\underline{\theta})$ if and only if $\Phi(\underline{\theta})[g(\underline{\theta})]^n$ can be expressed in the form

$$\Phi(\underline{\theta})[g(\underline{\theta})]^n = \int_{z_1 \in T(n)} \int_{z_2 \in T(n)} \int_{z_3 \in T(n)} \alpha(z_1, z_2, z_3, n) \prod_{i=1}^{3} \left(h_i(\underline{\theta})\right)^{z_i} dz_1\, dz_2\, dz_3$$

Thus, the UMVUE of a function $\Phi(\underline{\theta})$ of $\underline{\theta}$ in $h(x; \underline{\theta})$ is given by

$$\psi(Z_1, Z_2, Z_3, n) = \frac{\alpha(Z_1, Z_2, Z_3, n)}{B(Z_1, Z_2, Z_3, n)}, \quad B(Z_1, Z_2, Z_3, n) \neq 0$$

Then the following result are obvious.

Result 1 *The UMVUE of* $\prod_{i=1}^{3}\left(h_i(\underline{\theta})\right)^{k_i} = \left(\frac{\theta^\alpha \sum_{i=0}^{\alpha-1} \frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{k_1+k_2} p_1^{k_1}\, p_2^{k_2} e^{-\frac{k_3}{\theta}}$ *is given by*

$$H_{k_1,k_2,k_3}(z_1, z_2, z_3, n) = \frac{B(z_1-k_1, z_2-k_2, z_3-k_3, n)}{B(z_1, z_2, z_3 n)}$$

$$= \frac{(r_1)_{k_1}\, (r_2)_{k_2} \left(1-\frac{k_3}{z_3}\right)^{(n-r_1-r_2)\alpha-1} (z_3-k_3)^{\alpha(k_1+k_2)}}{[n-r_1-r_2+1]_{k_1+k_2} [(n-r_1-r_2)\alpha]_{(k_1+k_2)\alpha}},$$

where $k_1 \le r_1;\ k_2 \le r_2;\ k_3 \le z_3;\ k_1 + k_2 \le n - r_1 - r_2;\ r_1 + r_2 - 1 < n,\quad (r)_k = \frac{r!}{(r-k)!}$ *and* $[r]_k = \frac{\Gamma r+k}{\Gamma r}$.

Corollary 1 If $k_1 \neq 0,\ k_2 = 0$ and $k_3 = 0$, then UMVUE of $\left(h_1(\underline{\theta})\right)^{k_1} = \left(\frac{p_1 \theta^\alpha \sum_{i=0}^{\alpha-1} \frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{k_1}$ is given by

$$H_{k_1}(z_1, z_2, z_3, n) = \frac{B(z_1-k_1, z_2, z_3, n)}{B(z_1, z_2, z_3, n)}$$

$$= \frac{(r_1)_{k_1}\, z_3^{\alpha k_1}}{[n-r_1-r_2+1]_{k_1} [(n-r_1-r_2)\alpha]_{\alpha k_1}},$$

$$k_1 \le r_1;\ k_1 \le n - r_1 - r_2;\ \ r_1 + r_2 - 1 < n$$

Corollary 2 If $k_1 = 0, k_2 \neq 0$ and $k_3 = 0$, then UMVUE of $\left(h_2(\underline{\theta})\right)^{k_2} = \left(\frac{p_2\, \theta^\alpha \sum_{i=0}^{\alpha-1} \frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{k_2}$ is given by

$$H_{k_2}(z_1, z_2, z_3, n) = \frac{B(z_1, z_2-k_2, z_3, n)}{B(z_1, z_2, z_3, n)}$$

$$= \frac{(r_2)_{k_2}\, z_3^{\alpha k_2}}{[n-r_1-r_2+1]_{k_2} [(n-r_1-r_2)\alpha]_{\alpha k_2}},\ k_2 \le r_2;\ k_2 \le n - r_1 - r_2;\ r_1 + r_2 - 1 < n$$

Corollary 3 If $k_1 = 0, k_2 = 0$ and $k_3 \neq 0$, then UMVUE of $\left(h_3(\underline{\theta})\right)^{k_3} = e^{-\frac{k_3}{\theta}}$ is given by

$$H_{k_3}(z_1, z_2, z_3, n) = \frac{B(z_1, z_2, z_3 - k_3, n)}{B(z_1, z_2, z_3, n)}$$

$$= \left(1 - \frac{k_3}{z_3}\right)^{(n - r_1 - r_2)\alpha - 1}, \; k_3 \leq z_3;\; r_1 + r_2 - 1 < n$$

Result 2 *The UMVUE of the variance of* $H_{k_1,k_2,k_3}(Z_1, Z_2, Z_3, n)$, *is given by*

$$\widehat{var}\left[H_{k_1,k_2,k_3}(z_1, z_2, z_3, n)\right] = H^2_{k_1,k_2,k_3}(z_1, z_2, z_3, n) - H_{2k_1,2k_2,2k_3}(z_1, z_2, z_3, n)$$

$$= \left[\frac{(r_1)_{k_1}(r_2)_{k_2}\ \left(1 - \frac{k_3}{z_3}\right)^{((n - r_1 - r_2)\alpha - 1)} (z_3 - k_3)^{\alpha(k_1 + k_2)}}{[n - r_1 - r_2 + 1]_{k_1 + k_2}[(n - r_1 - r_2)\alpha]_{\alpha(k_1 + k_2)}}\right]^2$$

$$- \frac{(r_1)_{2k_1}(r_2)_{2k_2}\ \left(1 - \frac{2k_3}{z_3}\right)^{((n - r_1 - r_2)\alpha - 1)} (z_3 - 2k_3)^{2\alpha(k_1 + k_2)}}{[n - r_1 - r_2 + 1]_{2(k_1 + k_2)}[(n - r_1 - r_2)\alpha]_{2\alpha(k_1 + k_2)}},$$

$2k_1 \leq r_1;\ 2k_2 \leq r_2;\ 2k_3 \leq z_3;\ 2(k_1 + k_2) \leq n - r_1 - r_2;\ \ r_1 + r_2 - 1 < n$

Corollary 4 The UMVUE of the variance of $H_{k_1}(Z_1, Z_2, Z_3, n)$, is given by

$$\widehat{var}\left[H_{k_1}(z_1, z_2, z_3, n)\right] = H^2_{k_1}(z_1, z_2, z_3, n) - H_{2k_1}(z_1, z_2, z_3, n)$$

$$= \left[\frac{(r_1)_{k_1}\ {z_3}^{\alpha k_1}}{[n - r_1 - r_2 + 1]_{k_1}[(n - r_1 - r_2)\alpha]_{\alpha k_1}}\right]^2 - \frac{(r_1)_{2k_1}\ \ {z_3}^{2\alpha k_1}}{[n - r_1 - r_2 + 1]_{2k_1}[(n - r_1 - r_2)\alpha]_{2\alpha k_1}},$$

$$2k_1 \leq r_1;\ 2k_1 \leq n - r_1 - r_2;\ r_1 + r_2 - 1 < n$$

Corollary 5 The UMVUE of the variance of $H_{k_2}(Z_1, Z_2, Z_3, n)$, is given by

$$\widehat{var}\left[H_{k_2}(z_1, z_2, z_3, n)\right] = H^2_{k_2}(z_1, z_2, z_3, n) - H_{2k_2}(z_1, z_2, z_3, n)$$

$$= \left[\frac{(r_2)_{k_2}\ \ {z_3}^{\alpha k_2}}{[n - r_1 - r_2 + 1]_{k_2}\ \ [(n - r_1 - r_2)\alpha]_{\alpha k_2}}\right]^2 - \frac{(r_2)_{2k_2}\ \ {z_3}^{2\alpha k_2}}{[n - r_1 - r_2 + 1]_{2k_2}\ \ [(n - r_1 - r_2)\alpha]_{2\alpha k_2}},$$

$$2k_2 \leq r_2;\ 2k_2 \leq n - r_1 - r_2;\ r_1 + r_2 - 1 < n$$

Corollary 6 The UMVUE of the variance of $H_{k_3}(Z_1, Z_2, Z_3, n)$, is given by

$$\widehat{var}\left[H_{k_3}(z_1, z_2, z_3, n)\right] = H^2_{k_3}(z_1, z_2, z_3, n) - H_{2k_3}(z_1, z_2, z_3, n)$$

$$= \left(1 - \frac{k_3}{z_3}\right)^{2(n - r_1 - r_2 - 1)\alpha} - \left(1 - \frac{2k_3}{z_3}\right)^{(n - r_1 - r_2)\alpha - 1},$$

$$2k_3 \leq z_3;\ r_1 + r_2 - 1 < n$$

Result 3 *The UMVUE of* $\left[g(\underline{\theta})\right]^k = \left(\frac{\theta^{\alpha}\left(\sum_{i=0}^{\alpha-1}\frac{1}{\Gamma(i+1)\theta^i}\right)}{1 - p_1 - p_2}\right)^k$, $k \neq 0$ *as given in the mode (6) is*

$$G_k(z_1,z_2,z_3,n)=\frac{B(z_1,z_2,z_3,n+k)}{B(z_1,z_2,z_3,n)}$$

$$=\frac{[n+1]_k \quad {z_3}^{\alpha k}}{[n-r_1-r_2+1]_k \quad [(n-r_1-r_2)\alpha]_{\alpha k}}\,,\ k\le n-r_1-r_2;\ r_1+r_2-1<n$$

Result 4 *The UMVUE of the variance of* $G_k(Z_1,Z_2,Z_3,n)$ *is given by*

$$\widehat{var}[G_k(z_1,z_2,z_3,n)]=G_k^2(z_1,z_2,z_3,n)-G_{2k}(z_1,z_2,z_3,n)$$

$$=\left[\frac{[n+1]_k \quad {z_3}^{\alpha k}}{[n-r_1-r_2+1]_k \quad [(n-r_1-r_2)\alpha]_k}\right]^2-\frac{[n+1]_{2k} \quad {z_3}^{2\alpha k}}{[n-r_1-r_2+1]_{2k} \quad [(n-r_1-r_2)\alpha]_{2\alpha k}}\,,$$
$$2k\le n-r_1-r_2;\ r_1+r_2-1<n$$

Result 5 *For fixed x, the UMVUE of the density is given by*

$$\phi_x(z_1,z_2,z_3,n)=a(x)\frac{B(z_1-C_1(x),\,z_2-C_2(x),\,z_3-C_3(x),n-1)}{B(z_1,z_2,z_3,n)}$$

$$=\left(\frac{x^{\alpha-1}}{\Gamma\alpha}\right)\frac{(r_1)_{I_1(x)}\,(r_2)_{I_2(x)}\,(n-r_1-r_2-1)_{(1-I_1(x)-I_2(x))}\left((n-r_1-r_2)\alpha\right)_{(1-I_1(x)-I_2(x))}}{n\,[\,z_3-(x-1)(1-I_1(x)-I_2(x))]^{(1-I_1(x)-I_2(x))\alpha}}$$
$$\left(1-\frac{(x-1)(1-I_1(x)-I_2(x))}{z_3}\right)^{((n-r_1-r_2)\alpha-1)},z_3>(x-1);\ r_1+r_2-1<n$$

Result 6 *The UMVUE of the variance of* $\phi_x(Z_1,Z_2,Z_3,n)$ *is given by*

$$\widehat{var}[\phi_x(z_1,z_2,z_3,n)]=\phi_x^2(z_1,z_2,z_3,n)$$
$$-\phi_x(z_1,z_2,z_3,n)\ \phi_x(z_1-C_1(x),\ z_2-C_2(x),\ z_3-C_3(x),n-1)$$

$$=\phi_x^2(z_1,z_2,z_3,n)-\left(\frac{x^{\alpha-1}}{\Gamma\alpha}\right)^2\frac{(r_1)_{2I_1(x)}\,(r_2)_{2I_2(x)}\,(n-r_1-r_2)_{2(1-I_1(x)-I_2(x))}\left((n-r_1-r_2)\alpha-1\right)_{2(1-I_1(x)-I_2(x))}}{n(n-1)[\,z_3-2(x-1)(1-I_1(x)-I_2(x)]^{2\alpha(1-I_1(x)-I_2(x))}}$$
$$\left(1-\frac{2(x-1)(1-I_1(x)-I_2(x))}{z_3}\right)^{((n-r_1-r_2)\alpha-1)},z_3>2(x-1);\ r_1+r_2-1<n$$

Result 7 *For fixed z, the UMVUE of the survival function* $S(t)=P(X>t),t\ge 0$ *is given by*

$$\hat{S}(t)=\int_{x>t}\phi_x(z_1,z_2,\dots,z_r,n)\,dx$$

$$=\left(\frac{(r_1)_{I_1(t)}(r_2)_{I_2(t)}(n-r_1-r_2)_{(1-I_1(t)-I_2(t))}\left(\alpha(n-r_1-r_2-1)\right)_{\alpha(1-I_1(t)-I_2(t))}}{n\,\Gamma\alpha}\right)$$
$$\int_t^\infty x^{(\alpha-1)}\frac{\left(1-\frac{(x-1)\left(1-I_1(x)-I_2(x)\right)}{Z_3}\right)^{\alpha(n-r_1-r_2)-1}}{\left(Z_3-(x-1)\left(1-I_1(x)-I_2(x)\right)\right)^{\alpha\left(I_1(x)+I_2(x)\right)}}dx$$

$$Z_3>(t-1);\ r_1+r_2-1<n$$

Result 8 *For the fixed* $z=(z_1,z_2,z_3,n)$, *the UMVUE of the* $var\left(\hat{S}(t)\right)$, *is obtained as*

$$\widehat{var}\big(\hat{S}(t)\big) = \hat{S}^2(t) - 2\iint_{x>y>t}^{\infty} a(x)a(y)\frac{B(z_1 - C_1(x) - C_1(y), \dots, z_r - C_r(x) - C_r(y)\; n-2)}{B(z_1, \dots, z_r, n)}\,dx\,dy$$

$$= \hat{S}^2(t) - 2\left(\frac{(r_1)_{2I_1(t)}(r_2)_{2I_2(t)}(n - r_1 - r_2)_{2(1-I_1(t)-I_2(t))}\,\big(\alpha(n-r_1-r_2-1)\big)_{2\alpha(1-I_1(t)-I_2(t))}}{n(n-1)(\Gamma\alpha)^2}\right)$$

$$\iint_{x>y>t}^{\infty} x^{(\alpha-1)}y^{(\alpha-1)}\frac{\left(1-\dfrac{(x-1)\big(1-I_1(x)-I_2(x)\big)+(y-1)\big(1-I_1(y)-I_2(y)\big)}{Z_3}\right)^{\alpha(n-r_1-r_2)-1}}{\Big(Z_3-(x-1)\big(1-I_1(x)-I_2(x)\big)+(y-1)\big(1-I_1(y)-I_2(y)\big)\Big)^{\alpha\left((1-I_1(x)+I_2(x))+(1-I_1(y)+I_2(y))\right)}}dxdy$$

$$Z_3 > 2(t-1);\; r_1 + r_2 - 1 < n$$

Remark: For α=1, the above results reduce to the case of exponential distribution (see Muralidharan and Pratima, 2017).

4 Illustrative Example

The example is based on tumor size as one of the most important factors in making a clinical and pathological assessment of breast cancer accessed through the cBioPortal software for cancer genomics and was originally developed at Memorial Sloan Kettering Cancer Center (MSK). The cBioPortal software is now available under an open source license via GifHub. We use this portal for breast tumor size data from cancer genomics studies. For comprehensive data, one may visit http://www.cbioportal.org/study?id=brca_metabric#clinical. We followed up the tumor size of 509 samples from 509 female breast cancer patients with histologically confirmed invasive ductal breast carcinoma (IDC) of cohort 1. The largest tumor diameter was chosen as the sizing reference in each case. In the data set, lesions are absent for six patients, whereas for 22 patients lesions present but non-measurable by CT scan or MRI and remaining 481 samples tumor size in mm with corresponding frequency in brackets are: 11(9), 12(12), 13(11), 14(15), 15(11), 16(20), 17(18), 19(17), 20(33), 21(24), 22(21), 23(17), 25(28), 26(18), 27(11), 28(15), 29(13), 30(25), 31(7), 32(6), 33(6), 34(5), 35(11), 36(3), 37(3), 38(4), 39(5), 40(12), 42(3), 43, 44(2), 45(6), 46(2), 47(5), 48, 50(13), 52(2), 53(2), 55(4), 57, 60(6), 62, 65(4), 67, 70(3), 80(2), 84, 90, 100(3), 150,160, and 180.

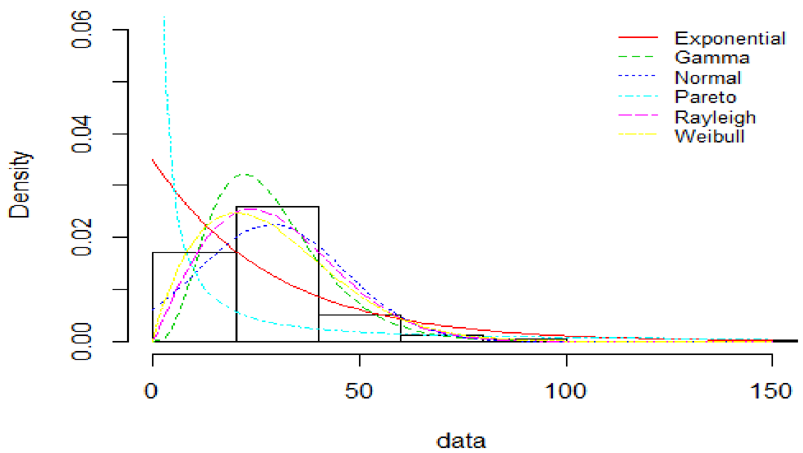

Figure 1. Histogram and theoretical densities

We fitted the number of distributions to the breast tumor size data by using various information criteria. Table 1 shows the fitted values and estimates. The superscripts that indicate the rank obtained by the distribution according to the selection criteria (the smaller the better). The first competing model includes Gamma distribution across the criteria's. The histogram of the data and plots of the fitted densities are displayed in Figure 1. These plots indicate that Gamma distribution provides a good fit to data.

The summary of the UMVU estimates of parameters with their standard error (shown in bracket) along with 95 % confidence interval is given in Table 2 whereas the summary of UMVU estimates of various parametric functions is given in Table 3 for breast tumor size data.

Acknowledgements

Both the authors thank the referee and the Editor for their useful suggestions.

References

Aitchison J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of American Statistical Association*, **50**, 901 – 908.

Adlouni, S. El., Ezzahid, E. and Mouatassim, Y. (2011). Mixed distributions for loss severity modeling with zero in operational risk losses. *International Journal of Applied Mathematics and Statistics*, **21(11)**, 14 pages.

Bavagosai, P. and Muralidharan, K. (2018). A Pareto II model with inliers at zero and one based on Type-II censored samples. *Reliability: Theory and Applications*, **13 No. 3** (**50**), 60 – 89.

Bavagosai, P. and Muralidharan, K. (2016). Some inferential study on inliers in Lindley distribution. *International Journal of Statistics and Reliability Engineering*, **3(2)**, 108 – 129.

Dixit, V.U. (2003) Estimation of parameters of mixed failure time distribution based on an extended modified sampling scheme. *Communication in Statistical Theory Methods*, **32(10)**, 1911 – 1923.

Jani P.N. and Singh A.K. (1995). Minimum variance unbiased estimation in multi-parameter exponential family of distributions. *Metron*, **53**, 93 – 106.

Jayade V.P. (1993). *Statistical Inference for AR-Process With Mixed Error*. Unpublished Ph.D. Thesis, Shivaji University, Kolhapur, Maharashtra.

Jayade, V.P. and Prasad, M.S. (1990). Estimation of parameters of mixed failure time distribution. *Communications in Statistics: Theory and Methods*, **19(12)**, 4667 – 4677.

Kale, B.K. (2003). *Modified Failure Time Distributions to Accommodate Instantaneous and Early Failures.* In J.C. Misra (Ed.) Industrial Mathematics and Statistics, Narosa Publishing House, New Delhi, 623 – 648.

Kale, B. K. (1998). Optimal estimating equations for discrete data with higher frequencies at a point. *Journal of thev Indian Statistical Association*, **36**, 125 – 136.

Kale, B.K. and Muralidharan, K. (2006). Maximum likelihood estimation in presence of inliers. *Journal of the Indian Society for Probability and Statistics*, **10**, 65–80.

Kale, B.K. and Muralidharan, K. (2000). Optimal estimating equations in mixture distributions accommodating instantaneous or early failures. *Journal of the Indian Statistical Association*, **38**(**2**), 317 – 329.

Kleyle, R.M. and Dahiya, R.L. (1975). Estimation of parameters of mixed failure time distribution from censored data. *Communication in Statistics – Theory and Methods*, **4**(**9**), 873 – 882.

Muralidharan, K. (1999). Tests for the mixing proportion in the mixture of a degenerate and exponential distribution. *Journal of the Indian Statistical Association*, **37(2)**, 105 – 119.

Muralidharan K (2010). Inlier prone models: A review. *ProbStat Forum*, 3: 38 – 51.

Muralidharan, K. and Arti, M. (2013). Inlier proness in Normal distribution. *Reliability: Theory and Applications*, **8(1)**, 86 – 99.

Muralidharan, K. and Arti, M. (2008). Analysis of instantaneous and early failures in Pareto distribution. *Journal of Statistical Theory and Applications*, **7**, 187 – 204.

Muralidharan, K. and Bavagosai, P. (2018). A new Weibull model with inliers at zero and one based on type-II censored samples. *Journal of the Indian Society of Probability and Statistics*, **19**, 121 – 151.

Muralidharan, K. and Bavagosai, P. (2017). Analysis of lifetime model with discrete mass at zero and one. *Journal of Statistical Theory Practice*, **11(4)**, 670–692.

Muralidharan, K. and Bavagosai, P. (2016a). Some inferential studies on inliers in Gompertz distribution. *Journal of the Indian Society for Probability and Statistics*, **17**, 35 – 55.

Muralidharan, K. and Bavagosai, P. (2016b). A revisit to early failure analysis in life testing. *Journal of the Indian Statistical Association*, **54(1 & 2)**, 43 – 69.

Muralidharan, K. and Kale, B.K. (2002). Modified gamma distribution with singularity at zero. *Communications in Statistics–Simulation and Computation*, **31(1)**, 143 – 158.

Muralidharan K. and Lathika P. (2006). Analysis of instantaneous and early failures in Weibull distribution. *Metrika*, **64(3)**, 305 – 316.

Muralidharan, K. and Lathika, P. (2004). The concept of inliers. *Proceeding of first Sino-International Symposium on Probability, Statistics and Quantitative Management*, Taiwan, October: 77 – 92.

Shinde, R. L. and Shanubhogue, A. (2000). Estimation of parameters and the mean life of a mixed failure time distribution. *Communications in Statistics–Theory and Methods*, **29(1)**, 2621 – 2642.

Vannman, K. (1995). On the distribution of the estimated mean from the nonstandard mixtures of distribution. *Communication in Statistics–Theory and Methods*, **24(6)**,1569 – 1584.

Vannman, K. (1991). *Comparing Samples From Nonstandard Mixtures of Distributions With Applications to Quality Comparison of Wood*. Research Report: 2, Submitted to Division of Quality Technology, Lulea University, Lulea, Sweden.

http://www.cbioportal.org/study?id=brca_metabric#clinical.

Table 1. Parameter estimates and goodness-of-fit criteria for distributions fitted to the breast tumor size data

Distribution	MLE (SE)	AIC	BIC	K-S	CVM	AD
Exponential	$\hat{\theta}$=0.0349 (0.0016)	4191.4030[5]	4195.5790[5]	0.3236[5]	13.9263[5]	70.7563[4]
Gamma	$\hat{\alpha}$=4.3043 (0.2674) $\hat{\theta}$=0.1503 (0.0099)	3815.4970[1]	3823.8490[1]	0.1301[1]	1.6984[1]	9.5293[1]
Normal	$\hat{\mu}$=28.6424 (0.8076) $\hat{\theta}$=17.7565 (0.5725)	4136.4550[4]	4144.8070[4]	0.2013[4]	5.6140[4]	Inf[6]
Pareto	$\hat{\theta}$=0.3052 (0.0139)	5257.6140[6]	5261.7900[6]	0.5316[6]	36.3271[6]	1685236[5]
Rayleigh	$\hat{\theta}$=23.8300 (0.5433)	3953.5270[3]	3957.7030[2]	0.1846[3]	4.0913[3]	21.8297[3]
Weibull	$\hat{\alpha}$=1.7859 (0.0524) $\hat{\theta}$=32.4110 (0.8802)	3939.0980[2]	3947.4500[3]	0.1503[2]	3.3870[2]	20.6991[2]

Table 2. Summary of UMVU estimates of parameters of breast tumor size data

$\alpha = 2$	UMVUE (SE)	95% CI
$\hat{p}_1$	0.011788 (0.004784)	(0.002412, 0.021164)
$\hat{p}_2$	0.043222 (0.009014)	(0.025556, 0.060888)
$\hat{\theta}$	15.094700 (0.902663)	(13.325510, 16.863890)

Table 3. Summary of estimates of parametric functions of breast tumor size data

parametric function	UMVUE	SE of UMVUE
$h_1(\underline{\theta})\, h_2(\underline{\theta}) h_3(\underline{\theta}) = \left(\dfrac{\theta^\alpha \sum_{i=0}^{\alpha-1} \frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{k_1+k_2} {p_1}^{k_1}\, {p_2}^{k_2} e^{-\frac{k_3}{\theta}}$, $k_1 = 1, k_2 = 1, k_3 = 1$	17.705940	8.472283
$h_1(\underline{\theta}) = \left(\dfrac{p_1 \theta^\alpha \sum_{i=0}^{\alpha-1} \frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{k_1}, k_1 = 1, k_2 = 0, k_3 = 0$	2.287341	0.951184
$h_2(\underline{\theta}) = \left(\dfrac{p_2 \theta^\alpha \sum_{i=0}^{\alpha-1} \frac{1}{\Gamma(i+1)\theta^i}}{1-p_1-p_2}\right)^{k_2}, k_1 = 0, k_2 = 1, k_3 = 0$	8.387144	1.924007
$h_3(\underline{\theta}) = e^{-\frac{1}{\theta}}, k_1 = 0, k_2 = 0, k_3 = 1$	0.930270	0.002169
$g(\underline{\theta}) = \left(\dfrac{\theta^\alpha \left(\sum_{i=0}^{\alpha-1} \frac{1}{\Gamma(i+1)\theta^i}\right)}{1-p_1-p_2}\right)^{k}, k = 1$	201.50210	13.13920
UMVUE of density function at x=20, x=40 x=60	0.025097 0.011820 0.004167	0.000571 0.000337 0.000309
UMVUE of survival function at time x=20, x=40, x=60	0.579584 0.217181 0.069952	0.008231 0.008346 0.009012