

APPLICATION OF NON-PARAMETRIC DENSITY ESTIMATION FOR PARAMETERS
OF AN INVERSE GAUSSIAN – INVERSE WEIBULL MIXTURE MODEL

Sultan, Ahmed, M. M¹ and Almubty, Reem Z. M.*²

¹Princess Nora Bint Abdul Rahman University, KSA. E-mail: amsultan_52@yahoo.com

²King Saud University, KSA. E-mail: ralmubty@ksu.edu.sa

(Received on: 14-03-14; Revised & Accepted on: 22-03-14)

ABSTRACT

A technique is introduced for estimating the parameters of a mixture of two distributions. The mixture has a two parameters Inverse Weibull distribution and a two parameters Inverse Gaussian one mixed with a given mixing proportion. The technique is based on a nonparametric density estimator for the mixture using a sample from the given mixture. The minimum distance estimation (MDE) method is used for estimating the parameters of the mixture when the empirical distribution function (EDF) is replaced by the corresponding nonparametric density estimator. A Monte Carlo of size 10000 is used to show the performance of the proposed method. Results from the proposed method show a better performance.

Key Words: Non-parametric density, Inverse Weibull, Inverse Gaussian, Gaussian kernel, hybrid methods, Cramer von Mises statistic.

1. INTRODUCTION

Mixture models are widely used in considerable fields of applications. The fields include among other fields modeling heterogeneity in cluster analysis, evaluation of error rates in data diagnostics and screening as Fonseca [1] showed, latent class analyses, discriminant analyses, survival analyses. McLachlan and Basford [2] pointed the role of mixtures in data analysis as well as inferences on sets of data and introduced suitable models. Priebe [3] found an approximated log-normal probability distribution function using a mixture of 30 normal distributions for 10000 observations using kernel density estimator. Bishop [4] spotted the use of mixture models in neural networks. Dempster *et.al.* [5] introduced algorithm for imputing missing data for general estimation purposes. They an iterative methodology for computing the expectation maximization (EM) for the maximum likelihood estimators of a mixture of distributions. McLachlan[6] explained the use of the EM method to compromise a model for a finite mixture of normal distributions. In this paper a method is introduced for applying the non-parametric density estimation for parameters estimation of a mixture of two components. These components are the two parameters Inverse Weibull distribution and the two parameters Inverse Gaussian distributions. Section 2 surveys the mixture components used and the mathematical formulation for the mixture model. Section 3 discusses the parametric approach for parameters estimation of the mixture model while section 4 introduces a proposed approach based on the minimum distance estimation (MDE) for estimating the mixture parameters. The method involves transforming the problem into an optimization problem over the space of the unknown parameters. Results and conclusion are given in section 4.

2. MODEL COMPONENTS

This section surveys the two parameters Inverse Weibull distribution and the two parameters Inverse Gaussian distribution used as model components for the mixture model under study.

2.1. Inverse Weibull component

Keller and Kmath [7] introduced the two parameter inverse Weibull distribution for life testing analysis. The probability density function for the distribution is given by:

*Corresponding author: Almubty, Reem Z. M.*²*
²King Saud University, KSA. E-mail: ralmubty@ksu.edu.sa

$$f_1(x) = \begin{cases} \beta \alpha^{-\beta} x^{-(\beta+1)} e^{-(\alpha x)^{-\beta}} & x \geq 0, \alpha > 0, \beta > 0 \\ 0 & \text{o.w} \end{cases} \quad (1)$$

where α is the scale parameter and β is the shape parameter. The corresponding cumulative distribution function is given by:

$$F_1(x) = e^{-(\alpha x)^{-\beta}}, \quad x \geq 0, \alpha > 0, \beta > 0 \quad (2)$$

This distribution received its importance from life testing since it can model the failure rate which is very common in reliability and biological research. Pawlas and Szynal [8] gave the characterization conditions for the inverse Weibull distribution. Erto [9] has discussed the properties of this distribution and its potential use as a lifetime model.

2.2. Inverse Gaussian component

The inverse Gaussian probability density function is given by:

$$f_2(x) = \left[\frac{\lambda}{2\pi x^3} \right]^{\frac{1}{2}} \exp \left[-\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right], \quad x > 0, \mu > 0, \lambda > 0 \quad (3)$$

where μ is the expected value and μ^3/λ is the variance and μ/λ is the square of the coefficient of variation. The corresponding cumulative distribution function is given by:

$$F_2(x) = \Phi \left[\left(\frac{\lambda}{x} \right)^{\frac{1}{2}} \left(\frac{x}{\mu} - 1 \right) \right] + e^{\frac{2\lambda}{\mu}} \Phi \left[- \left(\frac{\lambda}{x} \right)^{\frac{1}{2}} \left(\frac{x}{\mu} + 1 \right) \right] \quad (4)$$

with $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$

Johnson and Kotz [10] pointed different areas of application for the distribution. Among these applications are the motion of particles influenced by Brownian motion, sequential analysis, physical applications connected with diffusion processes with boundary conditions, reliability and lifetime data analysis, and many other applications.

Using these two distributions in a mixture gives the following probability density function

$$f(x, \theta) = p f_1(x, \theta_1) + (1 - p) f_2(x, \theta_2) \quad (5)$$

where p is the mixing coefficient. The corresponding cumulative distribution function is given by:

$$F(x, \theta) = p F_1(x, \theta) + (1 - p) F_2(x, \theta) \quad (6)$$

In this paper six different models are taken into account representing various shapes of the mixture. Thus if $(p, \mu, \alpha, \beta, \lambda)$ is a vector of the mixture parameters, then the models considered have the following values: (0.2, 1.5, 2.4, 3.3, 1.5), (0.5, 1.5, 2.4, 3.3, 1.5), (0.9, 1.5, 2.4, 3.3, 1.5), (0.2, 4.0, 0.5, 4.0, 0.6), (0.5, 4.0, 0.5, 4.0, 0.6), and (0.9, 4.0, 0.5, 4.0, 0.6). Figure 1 shows these six different cases respectively.

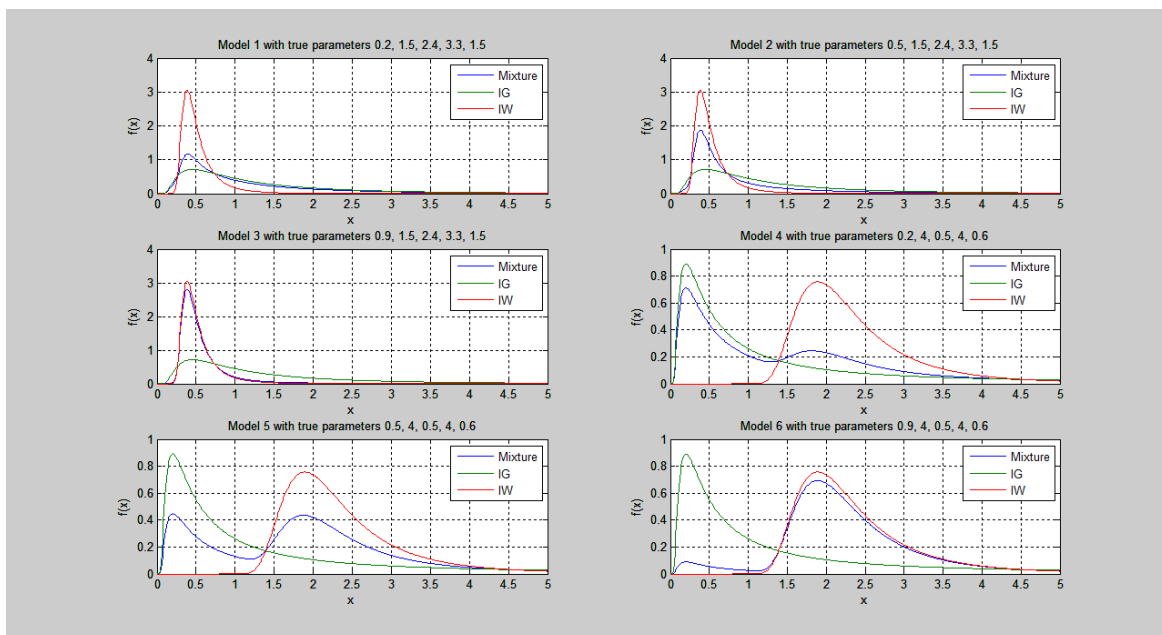


Figure 1: Different shapes of mixture based on parameters values

3. MAXIMIUMLIKELIHOOD APPROCH

The maximum likelihood estimation for the parameters of a mixture of distributions has been studied extensively. Hasselblad [11] was one of the first researchers who used the maximum likelihood estimation for mixtures of two or more components from exponential family. Generally speaking, applying the maximum likelihood method to mixtures involves using numerical iterative technique to find a solution and applying the expectation maximization (EM) algorithm. Hence this finds approximate values for the parameters.

Let L be the likelihood function and L^* be the log-likelihood function, then differentiating L^* with respect to the five parameters $(p, \mu, \alpha, \beta, \lambda)$ of the mixture gives:

$$g_1 = \frac{\partial L^*}{\partial p} = \sum_{i=1}^n \frac{f_1(x_i) - f_2(x_i)}{f(x_i)} \quad (7)$$

$$g_2 = \frac{\partial L^*}{\partial \mu} = \sum_{i=1}^n \frac{1}{f(x_i)} (1 - p) \frac{\partial f_2(x_i)}{\partial \mu} \quad (8)$$

$$g_3 = \frac{\partial L^*}{\partial \alpha} = \sum_{i=1}^n \frac{1}{f(x_i)} p \frac{\partial f_1(x_i)}{\partial \alpha} \quad (9)$$

$$g_4 = \frac{\partial L^*}{\partial \beta} = \sum_{i=1}^n \frac{1}{f(x_i)} p \frac{\partial f_1(x_i)}{\partial \beta} \quad (10)$$

$$g_5 = \frac{\partial L^*}{\partial \lambda} = \sum_{i=1}^n \frac{1}{f(x_i)} (1 - p) \frac{\partial f_2(x_i)}{\partial \lambda} \quad (11)$$

Solving the previous system of nonlinear equations in the five parameters gives estimators for those parameters. This of course involves more details that can be summarized in the following steps:

- Generate 10000 samples of sizes 30, 50, 70, 90 for every model of the six models representing various shapes of the mixture distribution f_t .
- Estimate the five parameters for each sample in (a)
- Find the integrated square error $ISE_{mle} = \int (f_t - f_{mle})^2 dx$ between the model used for generating the sample f and the model with estimated parameters f_{mle} .
- Compute the mean integrated square error MSE_{mle} for the 10000 runs from each sample size with the corresponding standard deviation.

3. NON-PARAMETRIC DENSITY BASED APPROACH

In this section the proposed method of parameter estimation for the mixture model under study will be surveyed. The method is a modification of the general minimum distance estimation (MDE). The method is based on finding the parameters of the distribution that makes the distance between the distribution and the empirical distribution function minimum. MDE uses a goodness of fit statistic as the function to be minimized over the parameters space which consists of five components $(p, \mu, \alpha, \beta, \lambda)$. Among these goodness of fit statistics are chi-square statistic, Cramer-von-Mises statistic, Kolmogorov Smirnov statistic, and Anderson Darling statistic. Hobbs et. al. [12] used the MDE method for the three parameter Weibull distribution. Sultan [13] described a method for the estimation of the three-parameter Weibull distribution function from censored samples. In the current research paper Cramer-von-Mises statistic is to be used. This statistic has the form:

$$W^2 = n \int_{-\infty}^{\infty} (F_t(x) - \hat{F}(x))^2 dF_0(x) \quad (12)$$

where $\hat{F}(x)$ is the empirical distribution function and $F_t(x)$ is the cumulative distribution function of the underlying distribution. A computational form is:

$$W^2 = \sum_{i=1}^n \left(F_t(x_i) - \frac{i - 0.5}{n} \right)^2 + \frac{1}{12n} \quad (13)$$

Our approach is to use a nonparametric estimator in place of the empirical distribution function. The nonparametric estimator used in this research is Gaussian kernel estimator which has the form:

$$\begin{aligned} \hat{F}(x) &= \int_{-\infty}^x \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2\right) dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^x \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2\right) dx \\ &= \frac{1}{n} \sum_{i=1}^n \Phi\left(\frac{x-X_i}{h}\right) \end{aligned} \tag{14}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. Hence W^2 takes the form:

$$W^2 = \sum_{j=1}^n \left(F_t(x_j) - \frac{1}{n} \sum_{i=1}^n \Phi\left(\frac{X_j - X_i}{h}\right) \right)^2 + \frac{1}{12n} \tag{15}$$

with h as the window width of the kernel. The optimal value of the h parameter is:

$$h_{opt} = \frac{1}{2\sqrt{\pi}} \left[\int f_t''^2(x) dx \right]^{\frac{-1}{5}} n^{\frac{-1}{5}} \tag{16}$$

This form depends on finding $f_t''(x)$ which is unknown. An empirical choice of $h = sn^{-\frac{1}{5}}$ with s as the sample standard deviation proved to work well.

Estimating the parameters when using a nonparametric density estimator for the mixture can be summarized in the following steps:

1. Choose the sample size (sample sizes used here are 30, 50, 70, 90) and choose a model by defining the values for the parameters $(p, \mu, \alpha, \beta, \lambda)$ from the six models considered.
2. Generate 10000 different samples from the true density of the chosen model f_t for the chosen sample size.
3. Optimize W^2 over the space of the unknown five parameters of the model. This gives an estimator f_{CVM} of the true density f_t .
4. Compute the integrated square error measuring the distance between the true density and the estimated one as:

$$ISE_{CVM} = \int (f_t - f_{CVM})^2 dx \tag{17}$$
 then compute the mean integrated square error for the 10000 samples generated and denote it as $MISE_{CVM}$ together with the corresponding standard deviation.
5. Similarly the mean integrated square error associated with maximum likelihood estimator is found.

4. RESULTS AND CONCLUSIONS

The following are the results of estimating the parameters of the distribution of the mixture model with five parameters $(p, \mu, \alpha, \beta, \lambda)$ for samples of different sizes 30, 50, 70, 90 to cover the spectrum of various shapes of the mixture mentioned above. Both the MLE and the proposed MDE are considered and results are given in table 1 to table 6 for $MISE_{MLE}$ and $MISE_{CVM}$ with the corresponding standard deviation between brackets for the 6 models respectively.

$(p, \mu, \alpha, \beta, \lambda)$	Sample size	$MISE_{MLE}$	$MISE_{CVM}$
(0.2, 1.5, 2.4, 3.3, 1.5)	30	0.1371 (0.14717)	0.1180 (0.0659)
	50	0.0893 (0.1134)	0.0793 (0.0832)
	70	0.0946 (0.1595)	0.08455 (0.0424)
	90	0.0409 (0.0303)	0.0308 (0.0237)

Table - 1. Results of model 1 for different sample sizes

$(p, \mu, \alpha, \beta, \lambda)$	Sample size	$MISE_{MLE}$	$MISE_{CVM}$
(0.5, 1.5, 2.4, 3.3, 1.5)	30	0.06948 (0.0714)	0.03080 (0.0151)
	50	0.06309 (0.1144)	0.02705 (0.0790)
	70	0.03598 (0.0528)	0.01320 (0.0049)
	90	0.0360 (0.0446)	0.0135 (0.0116)

Table - 2. Results of model 2 for different sample sizes

$(p, \mu, \alpha, \beta, \lambda)$	Sample size	$MISE_{MLE}$	$MISE_{CVM}$
(0.9,1.5,2.4,3.3,1.5)	30	0.1003 (0.0708)	0.0916 (0.0608)
	50	0.1114 (0.5028)	0.0528 (0.0655)
	70	0.0546 (0.1811)	0.04108 (0.0187)
	90	0.0378 (0.1063)	0.0283 (0.0052)

Table - 3. Results of model 3 for different sample sizes

$(p, \mu, \alpha, \beta, \lambda)$	Sample size	$MISE_{MLE}$	$MISE_{CVM}$
(0.2,4.0,0.5,4.0,0.6)	30	0.1597 (0.6881)	0.0451 (0.0519)
	50	0.2810 (0.0097)	0.0678 (0.0059)
	70	0.1656 (0.4322)	0.0305 (0.0107)
	90	0.1333 (0.6253)	0.0298 (0.0095)

Table - 4. Results of model 4 for different sample sizes

$(p, \mu, \alpha, \beta, \lambda)$	Sample size	$MISE_{MLE}$	$MISE_{CVM}$
(0.5,4.0,0.5,4.0,0.6)	30	0.4781 (0.8512)	0.1648 (0.1445)
	50	0.1966 (0.6085)	0.1472 (0.0211)
	70	0.3802 (0.9718)	0.1176 (0.0120)
	90	0.5996 (0.9889)	0.1177 (0.0122)

Table - 5. Results of model 5 for different sample sizes

$(p, \mu, \alpha, \beta, \lambda)$	Sample size	$MISE_{MLE}$	$MISE_{CVM}$
(0.9,4.0,0.5,4.0,0.6)	30	0.1030 (0.2202)	0.05255 (0.0167)
	50	0.780 (0.7353)	0.206 (0.1610)
	70	0.4121 (0.6826)	0.1717 (0.1401)
	90	0.1018 (0.3564)	0.01311 (0.0145)

Table - 6. Results of model 6 for different sample sizes

The above results show a better performance of the proposed method over the MLE method. The following tables (Table 7 To Table 12) give examples for a sample size of 30 for the six models considered. Each table shows sample points and the estimators for the MLE and the proposed method together with the true parameters. The integrated square error for both ME and the proposed method denoted by ISE_{MLE} and ISE_{CVM} respectively. Figure 2 to Figure 7 show the corresponding densities from tables.

Sample (n = 30)	1.23	0.32	0.59	0.51	0.95	2.69	0.76	0.33	0.30	1.01
	0.43	0.88	0.37	2.43	3.95	5.44	1.42	0.44	0.81	1.34
	2.31	0.57	3.20	1.09	0.23	0.62	0.45	4.71	0.76	1.42
$h = 0.1852$	Parameter	True	MLE	Proposed method	ISE_{MLE}	ISE_{CVM}				
	p	0.2	0.0006	0.180	0.0284	0.0049				
	μ	1.5	1.385	1.232						
	α	2.4	1.501	2.230						
	β	3.3	2.350	3.120						
	λ	1.5	1.381	1.330						

Table - 7. A case for sample size of 30 from model 1

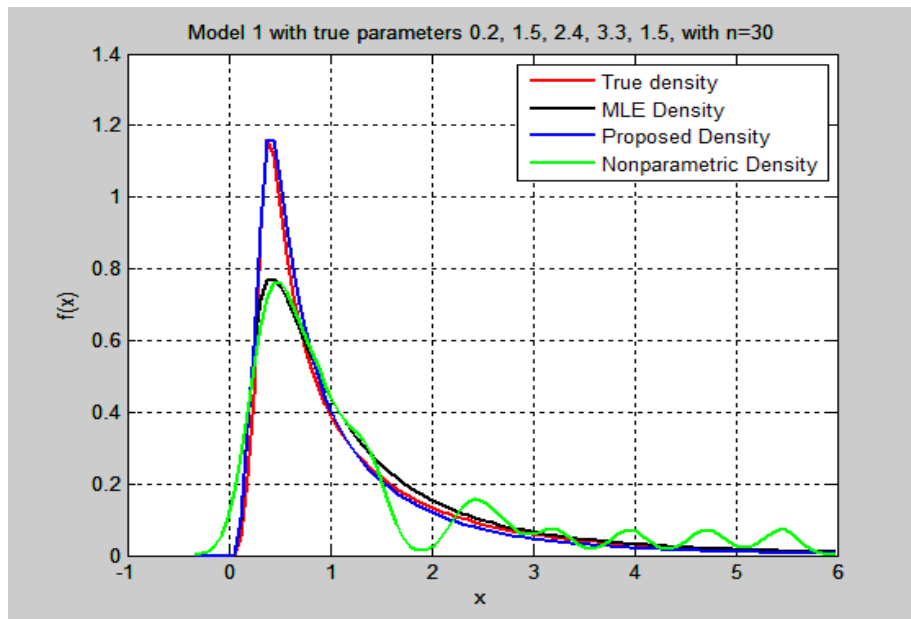


Fig. - 2: Estimated densities for case in Table 7.

Sample (n = 30)	0.41	0.61	0.63	0.36	1.19	1.08	1.33	0.66	1.42	0.84
	0.18	0.40	0.29	0.33	1.88	0.28	1.91	0.58	0.44	0.33
	1.99	0.30	0.34	0.34	0.58	0.47	0.63	0.60	0.54	3.42
$h = 0.0916$	Parameter	True	MLE	Proposed method	ISE_{MLE}	ISE_{CVM}				
	p	0.5	0.617	0.576	0.0539	0.0083				
	μ	1.5	1.206	1.229						
	α	2.4	2.527	2.230						
	β	3.3	2.259	2.942						
	λ	1.5	1.475	1.114						

Table - 8. A case for sample size of 30 from model 2

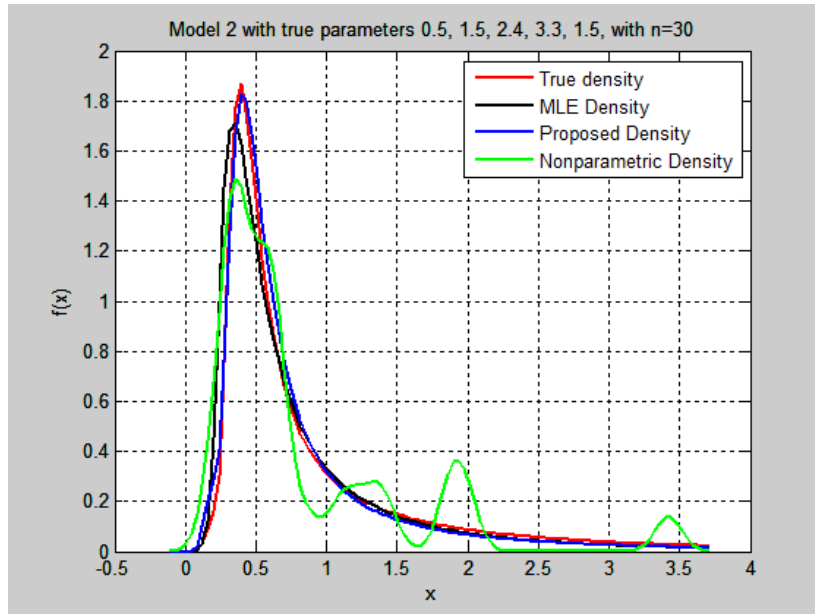


Fig. - 3: Estimated densities for case in Table 8.

Sample (n = 30)	0.33	0.75	0.81	0.72	0.45	0.40	0.48	0.56	0.12	0.58
	0.92	0.35	1.56	0.32	0.41	0.59	0.45	0.38	0.35	0.34
	0.69	0.34	0.28	0.33	1.45	0.31	0.87	0.50	0.43	0.55
$h = 0.48$	Parameter	True	MLE	Proposed method	ISE_{MLE}	ISE_{CVM}				
	p	0.9	0.890	0.933						
	μ	1.5	1.054	1.164						
	α	2.4	2.431	2.398						
	β	3.3	3.588	3.355						
	λ	1.5	0.308	0.516						

Table - 9. A case for sample size of 30 from model 3

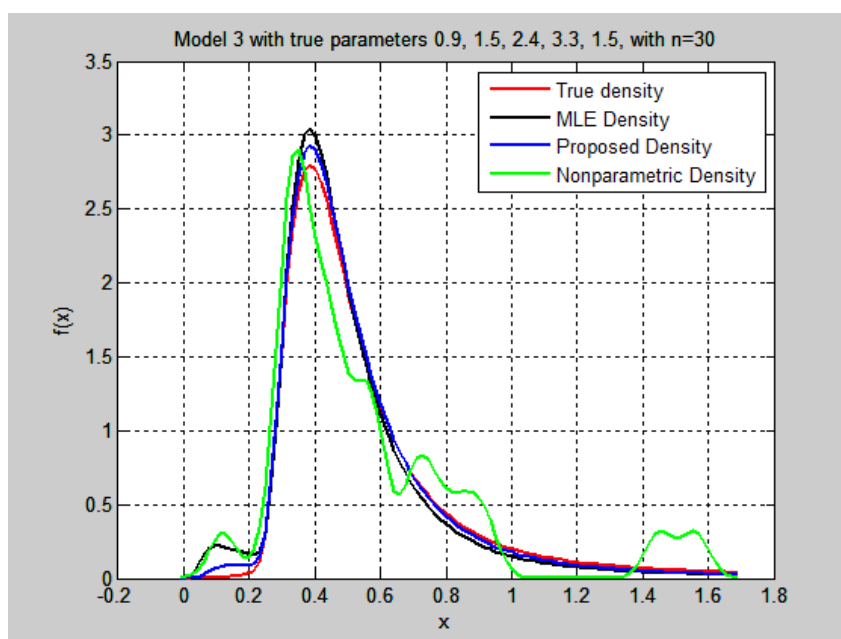


Fig. - 4: Estimated densities for case in Table 9.

Sample (n = 30)	0.12	0.26	0.32	1.17	0.19	4.45	0.39	2.69	0.37	2.25
	2.99	0.24	5.04	0.24	3.11	0.69	11.60	0.15	0.18	0.44
	2.29	0.45	11.66	0.30	1.04	3.26	0.39	1.88	0.67	0.58
$h = 0.3071$	Parameter	True	MLE	Proposed method	ISE_{MLE}	ISE_{CVM}				
	p	0.2	0.120	0.244	0.0173	0.0089				
	μ	4.0	3.065	3.135						
	α	0.5	0.644	0.593						
	β	4.0	1.280	1.525						
	λ	0.6	0.597	0.582						

Table - 10. A case for sample size of 30 from model 4

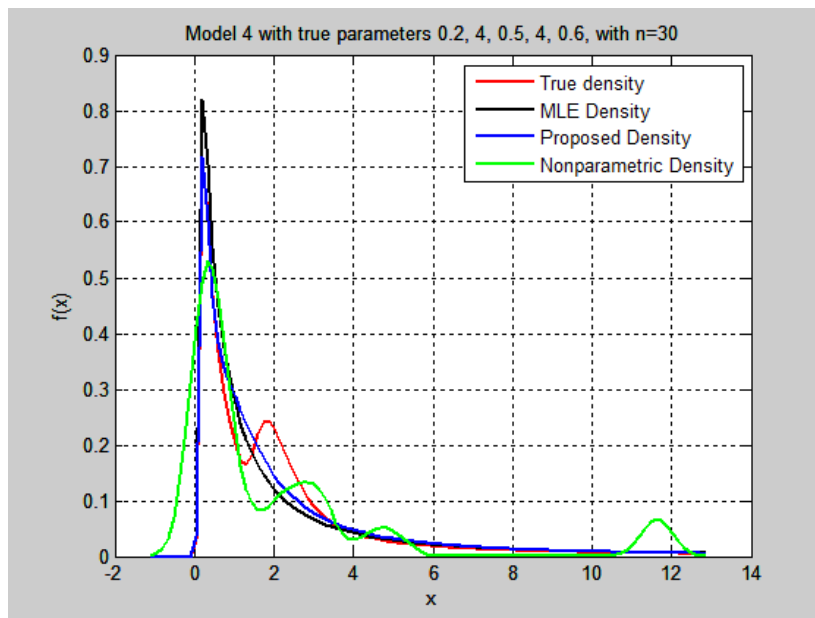


Fig. - 5: Estimated densities for case in Table 10.

Sample (n = 30)	2.46	0.12	1.65	1.17	3.04	0.63	1.85	0.41	0.18	2.30
	2.48	2.03	0.39	2.06	11.58	1.94	3.85	0.23	3.06	0.11
	2.68	2.48	0.63	1.57	1.29	3.35	2.03	1.71	0.30	2.26
$h = 0.1861$	Parameter	True	MLE	Proposed method	ISE_{MLE}	ISE_{CVM}				
	p	0.5	0.604	0.624	0.0298	0.0043				
	μ	4.0	1.565	1.567						
	α	0.5	0.563	0.488						
	β	4.0	2.864	3.203						
	λ	0.6	0.812	0.707						

Table - 11. A case for sample size of 30 from model 5

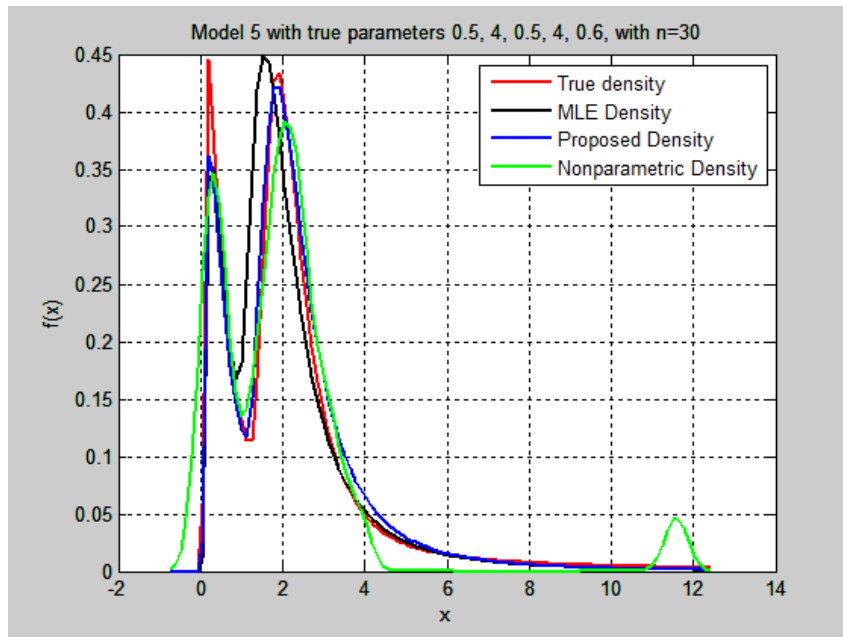


Fig. - 6: Estimated densities for case in Table 11.

Sample (n = 30)	1.93	2.64	2.79	1.71	6.13	1.67	3.69	4.66	2.21	2.48
	3.17	2.17	2.59	4.14	2.17	2.52	2.33	0.13	1.61	3.53
	2.31	2.91	0.28	2.15	1.68	1.62	2.70	1.87	1.77	2.85
$h = 0.3071$	Parameter	True	MLE	Proposed method	ISE_{MLE}	ISE_{CVM}				
	p	0.9	0.896	0.903						
	μ	4.0	3.888	4.025						
	α	0.5	0.473	0.523						
	β	4.0	3.610	3.797						
	λ	0.6	0.474	0.538						

Table - 12. A case for sample size of 30 from model 6

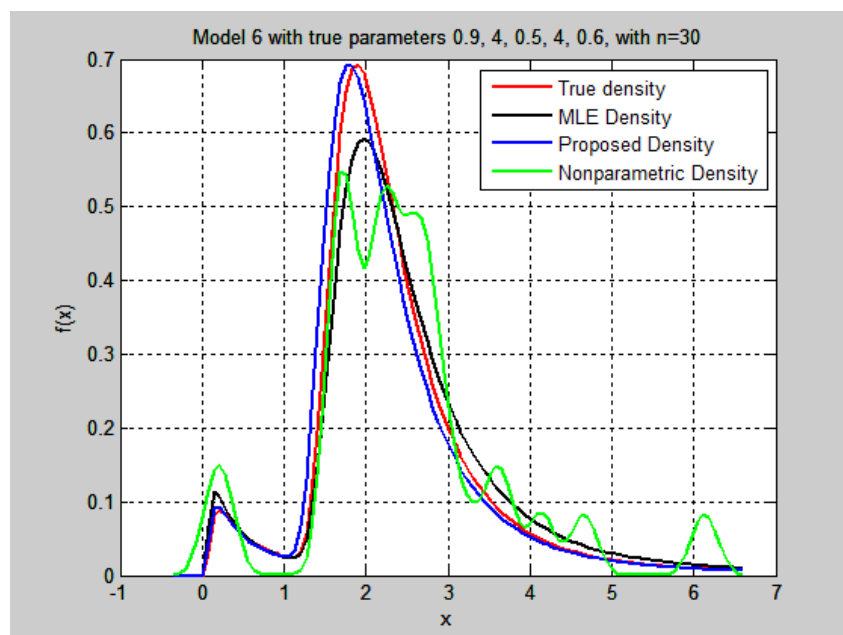


Fig. - 7: Estimated densities for case in Table 12.

5. REFERENCES

1. J. Fonseca, The Application of Mixture Modeling and Information Criteria for Discovering Patterns of Coronary Heart Disease, *Journal of Applied Quantitative Methods*, 3, 292-302, 2008
2. G. Mclachlan and K. Basford, *Mixture Models Inference and Applications to Clustering*, Marcel Dekker, New York, 1988.
3. C. Priebe, Adaptive Mixtures , *American Statistical Association*, 89 ,796-806, 1994.
4. C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press Link, Oxford, 1995.
5. A. Dempster; N. Laird; and D. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society*, 39, 1-38, 1977.
6. G. Mclachlan, *The Em Algorithm and Extensions*, Wiley and Sons, New York, 1997.
7. Z. Keller and R. Kmath, Reliability Analysis of CNC Machine Tools, *Reliability Engineering*, 3, 449-473, 1982.
8. P. Pawlas and D. Szynal Characterizations of the inverse Weibull distribution and generalized extreme value distributions ny moments of K^{th} record values, *Applicationes Mathematicae*, 27,2 , 197–202, 2000.
9. P. Erto, Genesis, properties and identification of the inverse Weibull lifetime model, *Statistica Applicata*, 1, 117-128, 1989.
10. N. Johnson; and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*, Wiley and sons, New York, 1970.
11. V. Hasselblad, Estimation of Finite Mixture of Distributions from the Exponential Family, *Journal of the American Statistical Association*, 64, 1459-1471, 1969.
12. R. Hobbs; A. Moore; and R. Miller, Minimum distance estimation of the three parameter Weibull distribution, *IEEE Transaction in Reliability*, 5, 459-469, 1985
13. Sultan; A, A data driven parameter estimation for the three parameter from censored samples, *Mathematical and Computational Applications*, 13 No. 2, 129-132, 2008.

Source of support: Nil, Conflict of interest: None Declared