Aligarh Journal of Statistics

Vol. 33 (2013), 99-117

A STUDY OF MIXTURE DISTRIBUTIONS (IDENTIFICATION AND ESTIMATION)

G. Lavanya, V. V. HaraGopal and S. N. N. Pandit

ABSTRACT

This paper considers the problem of estimating parameters of a mixture distribution (Rectangular/Exponential) using the method of moments, their functions and ratios mainly by simulation. Estimated the values of parameters, for the case (1) ‘ p ’ unknown, (2) ‘ p ’ is known and are presented in tables and graphs. As sample size ‘m’ increases the proportion of acceptable sample estimates are also increases. However, it is possible that a sample gives estimates which are not in the acceptable range (negative, complex, $p > 1$ or $p < 0$).

1. INTRODUCTION

The present study deals with the problem of estimating parameters in mixture of two distributions from the same family. In particular, the cases of exponential and rectangular populations are considered.

Methodology adapted is to find the estimates of parameters by functions of raw moments as obtained by a sample. These functions are polynomial functions of raw moments: their ratios turn out to be simple functions of the parameters from which estimation will be possible. A major problem in this context is that higher order sample moments have a high variability from sample to sample and their ratios will be even more fluctuating. Hence, one may expect relatively less efficiency of estimates obtained; often the estimates may even be outside acceptable range. Simulation studies are carried out to find how the quality of ‘estimates’ improves with varying sample size. The application of mixture distributions are very much used in Reliability and many of the Biometric, Engineering and Social Sciences.

2. A BRIEF REVIEW

Two essentially different approaches can be seen in attempts for estimating the parameters of the mixture distributions.

1. Individual values of observations 'x_i' are directly used, for example in method of moments or maximum likelihood.
2. Cumulative distributions of mixture models are fitted to the empirical cumulative distribution function from the sample. (Essentially, by least squares, weighted and / or non linear as the case may be).

In this case the graphical methods using special graph papers like the normal, log, double log are used to study the possibility of data from mixture of specific types of distributions. In particular this approach has been attempted by Harding (1948).

Edward and Flowsky (979), consider the case of mixture of two normal (Log normal) distributions: A method, the ¤ $-P$ versus $Q-$ plot, is developed that is, more sensitive to the presence of mixtures than more familiar methods. A series of ad-hoc methods that produce initial guesses for parameter values is also presented. Two methods of global estimation given initial parameter guesses are presented and compared. These methods are the maximum likelihood and a least squares procedure that fits the sample quantiles to the inverse distribution function of the mixture. The sensitivity of the method to poor choices for initial guesses is also considered.

Smiley W. Cheng and James C. Fu (1982) considered cumulative distribution function of mixture of two Weibull distributions. Weighted least squares is proposed for estimating parameters when data are grouped and censored, Dan Ling *et al.* (2009), discusses Nonlinear least square versus graphical methods for the same case.

The other approach is based on functions of individual sample values like moments, maximum likelihood function etc. Brief comments on attempts using this approach are given below:

Rider (1960), considered the problem of mixture of two exponential distributions. He noted that method of moments gives unacceptable estimates; however when $a \neq b$ the proposed estimators are consistent. Also, he found variance of the asymptotic distributions of $a^{\wedge}$ and $b^{\wedge}$. He also, finds out that a chi-square test for single exponential can be misleading; a null hypothesis that data are from a single exponential often gets accepted, when a chi-square test for goodness of it is carried out.

Blischke (1962) considers mixture of two binomials with same 'n' but with different parameters p_1 and p_2 and with mixture parameter p and $q = 1 - p$ and points out some theoretical anomalies about relative efficacy of estimators.

Tallis and Light (1968) use fractional moments for estimates of parameters, and compare method of moments with maximum likelihood method and observes that maximum likelihood method involves much larger amount of calculation without increasing in acceptability of estimates of mixed exponential distributions.

Maximum likelihood method for the estimation of the parameters of the mixed normal was first considered for $\pm_1 = \pm_2$ by Rao (1948). Hasselblad (1966) considered the $k-$component normal mixtures where k is known: here the k means, k variances and $k-1$proportions are to be estimated using maximum likelihood estimates for single truncated normal populations. A method of steepest descent which always converged to maximize the likelihood function of the entire sample is used. Special cases of equal variances and variances proportional to the square of the mean are also considered.

Day (1969) discusses the method of moments, minimum chi-square and Bayes estimators. He also points out that they appear greatly inferior to maximum likelihood in a mixture of two Normal, multivariate, but with unknown covariance matrices.

Craigmile and Titterington (1977) extend the result of Gupta and Miyawaki (1978) and deals with 2 and 3 component mixture of Rectangular distributions. Method of moments and maximum likelihood was discussed.

Hussain and J-Liu derived (2009) $L-$moment estimators of parameters of mixture of two rectangular distributions. They considered possible cases namely, (1) neither over lap or gap, (2) with over lap and with gap, they note that comparison of the two usual maximum likelihood, and method of moments and modified maximum likelihood method. The $L-$ moment estimates have less bias and more efficient in many cases.

It is pointed out that, most of the above methods require large samples.

3. METHODOLOGY OF THE PRESENT APPROACH

The Case of Rectangular Mixture

Let the two rectangular distributions be in the ranges as $0 \to a$ and $0 \to b$ and mixture parameter be given as $p:(1-p)$, where $0 \le p \le 1$, the observation 'x' come from either the 1st or 2nd population with probabilities p and $q = 1-p$.

The $k-$th raw moment is given by

$$m'(k) = \frac{pa^k + qb^k}{k+1} \tag{1}$$

Then, the raw moments are

$$pa + qb = 2m_1' \tag{1a}$$

$$pa^2 + qb^2 = 3m_2' \tag{1b}$$

$$pa^3 + qb^3 = 4m_3' \tag{1c}$$

$$pa^4 + qb^4 = 5m_4' \tag{1d}$$

Using the above functions of moments one can get the following relationship.

$$s = 3m_2' - (2m_1')^2 = pq(a-b)^2 \tag{2a}$$

$$s_1 = (4m_3') - (6m_2'm_1') = pq(a-b)^2(a+b) \tag{2b}$$

$$s_2 = (3m_4') - (8m_3'm_1') = pq(a-b)^2(a^2+b^2+ab) \tag{2c}$$

One can estimate the simple functions $a+b$ and $a-b$ of the parameters by taking the ratios of these functions as given below:

$$(a+b) = \frac{s1}{s} \tag{3a}$$

$$(a-b) = \pm\sqrt{4\left(\frac{s2}{s}\right) - 3\left(\frac{s1}{s}\right)^2} \tag{3b}$$

Solving (3a) and (3b) using the observed (that is sample) moments, one gets

$$\hat{a} = \frac{\left(\frac{s2}{s}\right) \pm \sqrt{4\left(\frac{s2}{s}\right) - 3\left(\frac{s1}{s}\right)^2}}{2}$$

and

$$\hat{b} = \frac{s1}{s} - \hat{a} \tag{4a}$$

From (4a)

$$\hat{p} = \frac{\left(2m_1 - \hat{b}\right)}{\hat{a} - \hat{b}}$$

3.1 The case of Exponential Distribution

Let the two exponential distributions be with the parameters a and b with mixture parameter p and $q = 1 - p$; that is, the observation 'x_i' comes from population with density function $\frac{1}{ae} - \frac{x}{a}$ or $\frac{1}{be} - \frac{x}{a}$ with probabilities p and $q = 1 - p$.

The four corresponding moments are theoretically

$$m'(k) = \frac{pa^k + qb^k}{k!} \tag{5}$$

Then the raw moments are

$$pa + qb = 2m_1' \tag{5a}$$

$$2\left(pa^2 + qb^2\right) = m_2' \tag{5b}$$

$$6\left(pa^3 + qb^3\right) = m_3' \tag{5b}$$

$$24\left(pa^4 + qb^4\right) = m_4' \tag{5d}$$

Using the above functions of moments one can get the following relationship.

$$s = m_2' - \left(2m_1'\right)^2 = 2pq\left(a-b\right)^2 \tag{6a}$$

$$s_1 = m_3' - \left(3m_2'm_1'\right) = 6pq\left(a-b\right)^2\left(a+b\right) \tag{6b}$$

$$s_2 = m_4' - \left(4m_3'm_1'\right) = 24pq\left(a-b\right)^2\left(a^2+b^2+ab\right) \tag{6c}$$

From the above functions it follows that

$$\left(a+b\right) = \frac{s1}{3s} \tag{7a}$$

$$\left(a-b\right) = \pm\sqrt{4\left(\frac{s2}{12s}\right) - 3\left(\frac{s1}{3s}\right)^2} \tag{7b}$$

Solving (7a) and (7b)

$$\hat{a} = \frac{\left(\frac{s1}{3s}\right) \pm \sqrt{4\left(\frac{s2}{12s}\right) - 3\left(\frac{s1}{3s}\right)^2}}{2}$$

and

$$\hat{b} = \frac{s2}{3s} - \hat{a} \tag{8a}$$

From (8a)

$$\hat{p} = \frac{\left(m_1' - \hat{b}\right)}{\hat{a} - \hat{b}}$$

In each of the cases, two sub-cases namely (1) when p is unknown and (2) when p is known (not to be estimated) are considered, with obvious modifications in the estimators. However, the present approach though often gives unacceptable estimates, can be modified by selective sub sampling, to get acceptable and useful estimates.

4. METHODOLOGY OF SIMULATION IN THE PRESENT STUDY

Below is given the procedure to generate mixture of two rectangular distributions; the same with obvious modifications is used to generate the exponential mixtures.

Consider a particular mixture, with parameters a and b and mixture parameter p and $q = 1 - p$; for a given sample size 'm', n samples are generated from each of the rectangular distributions x_1 and x_2 respectively. Let 'r' be a random matrix of size $m \times n$ from uniform $0 \to 1$. The $m \times n$ matrix $x = x(i, j)$, is constructed as follows:

If $r(i,j) \le p, x(i,j) = x_1(i,j)$; else, $r(i,j) \le p, x(i,j) = x_1(i,j)$.

This matrix 'x' has its 'n' columns as independent samples from the desired mixture distribution. Using these samples along with the formulae given above (2a, 2b, 2c) we get estimates of a, b, p. However, of these estimates quite a few turned out to be unacceptable, of being either negative or complex or $p > 1$ or $p < 0$. Excluding these unacceptable values the mean, s.d, min, max etc., of these retained estimates are computed and reported. In this filtering, even if one of the estimates of parameters is unacceptable the other estimates are excluded from consideration. The proportions of acceptable estimates are recorded in tables.

In such cases, samples which give an unacceptable estimates are obtained it is proposed to go for some sort of sub sampling (like bootstrap method), and take a weighted average estimates for the same.

4.1 The Procedure Followed as given above is Illustrated Below

Here we have taken $n = 10$ samples, each of size $m = 5$ of two rectangular distributions x_1 and x_2 with range $0 \to a(=5)$ and $0 \to b(=3)$ of size $m \times n$. From these samples a mixture sample $p = 0.2$ p=0.2 is generated as follows: Another random matrix 'r' from uniform $0 \to 1$ of size $m \times n$ is generated. Now mixture distribution 'x' is constructed as follows:

If

$$r(i,j) \le p, x(i,j) = x_1(i,j);$$

else,

$$r(i,j) \le p, x(i,j) = x_1(i,j);$$

For illustration Let us consider:

x_1 = Simulated Rectangular distribution with parameter ‘ a ’, ‘ p ’ of size 5×10

4.7506	3.8105	3.0772	2.0285	0.2895	1.0138	0.0764	2.0932	4.1906	2.5141
1.1557	2.2823	3.9597	4.6773	1.7643	0.9936	3.7339	4.2311	0.0982	3.5474
3.0342	0.0925	4.6091	4.5845	4.0658	3.0190	2.2255	2.6258	3.4064	2.1445
2.4299	4.1070	3.6910	2.0514	0.0493	1.3609	4.6591	1.0132	1.8974	1.5231
4.4565	2.2235	0.8813	4.4682	0.6945	0.9941	2.3300	3.3607	4.1590	0.9483

x_2 = Simulated Rectangular distribution with parameter ‘ b ’ , ‘ q ’ of size 5×10

0.5803	2.0937	1.487	1.9807	2.1813	2.1082	2.3845	2.9392	0.4096	1.9843
2.0467	1.1351	2.6993	1.0259	0.9279	1.6397	2.8705	0.8143	0.0353	0.8532
0.9083	2.5800	2.4649	0.8692	2.5155	1.3346	1.5678	0.7570	2.6817	1.4077
1.6250	2.5610	1.9347	1.0236	1.7042	2.0837	2.6404	2.6272	0.5974	0.1943
0.4526	1.7807	2.4539	1.6022	1.1112	1.8639	0.5189	2.2119	0.8962	2.9650

r = Random matrix of size 5×10

0.5828	0.2259	0.2091	0.5678	0.4154	0.9708	0.2140	0.4120	0.6833	0.2071
0.4235	0.5798	0.3798	0.7942	0.3050	0.9901	0.6435	0.7446	0.2126	0.6072
0.5155	0.7604	0.7833	0.0592	0.8744	0.7889	0.3200	0.2679	0.8392	0.6299
0.3340	0.5298	0.6808	0.6029	0.0150	0.4387	0.9601	0.4399	0.6288	0.3705
0.4329	0.6405	0.4611	0.0503	0.7680	0.4983	0.7266	0.9334	0.1338	0.5751

x = Simulated Mixture distribution with parameters ‘ a ’ b ’ ,‘ p ’

0.5803	2.0937	1.4897	1.9807	2.1813	2.1082	2.3845	2.939	0.4096	1.9843
2.0467	1.1351	2.6993	1.0259	0.9279	1.6397	2.8705	0.814	0.0353	0.8532
0.9083	2.5800	2.4649	4.5845	2.5155	1.3346	1.5678	0.757	2.681	1.4077
1.6250	2.5610	1.9347	1.0236	0.0493	2.0837	2.6404	2.6272	0.5974	0.1943
0.4526	1.7807	2.4539	4.4682	1.1112	1.8639	0.5189	2.2119	0.8962	2.9650

Now, the matrix has its columns as samples from the desired mixture.

5. NUMERICAL ILLUSTRATION

$r(1,1) = 0.58 > 0.2$; $x(1,1) = x_2(2,2) = 0.5803$;

$x(1,2) = 0.22 > 0.2$, $x(1,2) = x_2(1,2) = 2.09$;

$x(4,5) = 0.01 < 0.2$; $x(4,5) = x_1(4,5) = 0.04$;

$x(5,4) = 0.01 \leq 0.2$; $x(5,4) = x_1(5,4) = 4.4654$;

The first 4 sample moments for each of the 10 samples are computed. Using these functions of raw moments and their ratios as stated earlier, these are used to estimate the parameters of 'a', 'b', 'p' respectively. The first 4 sample raw moments (m_1', m_2', m_3', m_4') functions and their ratios are:

S.No.	1	2	3	4	5	6	7	8	9	10
$m_1' =$	0.1784	0.2227	0.1883	0.1614	0.3027	0.4077	0.0834	0.2567	0.2098	0.2936
$m_2' =$	0.9411	1.8804	1.1157	0.8586	2.3515	4.4968	0.6079	1.8564	1.3619	2.4757
$m_3' =$	1.0630	3.3676	1.5623	1.0503	3.7639	10.6427	1.0352	3.0263	1.9640	4.3480
$m_4' =$	1.2408	6.1380	2.4413	1.3803	6.2178	26.6259	1.7872	5.4479	2.9888	7.7547

Functional values are:

S.No.	1	2	3	4	5	6	7	8	9	10
$s = 3m_2' - (2m_1')^2$	- 0.123	-3.250	-4.302	0.817	0.544	-3.008	-1.767	-0.984	5.066	0.513
$s_1 = 4m_2' - 6m_2' m_1'^2$	0.080	-13.530	-19.197	8.795	1.304	-10.916	-8.907	-4.368	25.879	2.335
$s_2 = 5m_4' - 8m_3' m_1'$	0.386	-44.988	-66.707	39.53	3.134	-30.463	35.472	-18.537	120.234	8.873
$w = s_1 / s$	-0.654	4.162	4.462	10.75	2.397	3.628	5.0383	4.436	5.108	4.550
$w_4 = 4\left(\frac{s_2}{s}\right) - 3\left(\frac{s_1}{s}\right)^2$	-13.840	3.380	2.287	-153.71	5.809	1.010	4.097	16.246	16.649	7.046
$sqrt(w_4)$	1.615	2.446	2.242	1.843	2.311	3.537	2.161	2.596	2.124	2.444

From the above set, the sample numbers 1 and 4 are eliminated, Since w and, $W_4 = (a-b)^2$ should not be negative and the remaining values are substituted in the given formula.

$$\hat{a} = \frac{\left(\frac{11}{s}\right) \pm \sqrt{4\left(\frac{s_2}{s}\right) - 3\left(\frac{s_1}{s}\right)^2}}{2}$$

and

$$\hat{b} = \left(\frac{s_1}{s}\right)^2 - \hat{a} \qquad (4a)$$

From (4a)

$$\hat{p} = \left(\frac{2m_1' - \hat{b}}{\hat{a} - \hat{b}}\right)$$

Then using 8 samples we have the following estimates as:

Table 1: Estimates of parameters of mixture distribution

S.No.	2	3	5	6	7	8	9	10
$\hat{a}$	3.0007	2.9872	2.4036	2.3167	3.5314	4.2338	4.5943	3.6023
$\hat{b}$	1.1620	1.4748	-0.0067	1.3114	1.5072	0.2031	0.5140	0.9479
$\hat{p}$	1.5762	1.9454	1.1288	2.2886	1.2280	0.8775	0.6468	0.7587

However, from the above, samples giving estimates which are $p > 1$ (that is samples $2,3,5,6,7$) are excluded; thus in this case, out of 10 samples by eliminating (2 samples negative and 5 samples $p > 1$) that is 7 samples as giving un acceptable estimates only 3 samples give acceptable estimates . For this set, the parameter estimates and the statistics namely mean, s.d, etc, are given below:

Table 2: Acceptable estimates of parameters

S.No.	8	9	10
$\hat{a}$	4.2338	4.5943	3.6023
$\hat{b}$	0.2031	0.5140	0.9479
$\hat{p}$	0.8775	0.6468	0.7587

Table 3: Acceptable estimates of parameters of their statistics

	Mean	s.d	min	max	rms
$\hat{a}$	4.1435	0.5021	3.6023	4.5943	0.94
$\hat{b}$	0.5550	0.3741	0.2031	0.9479	2.46
$\hat{p}$	0.7610	0.1153	0.6468	0.8775	0.56

In short, though 10 samples were generated from the mixture only 3 gave acceptable estimates, the remaining 7 gives unacceptable estimates.

The above illustration refers to the case when p is to be estimated.

When p is known:

We now illustrate the case when p is known.

Here also, we have taken $n = 10$ samples, each of size $m = 5$ in the same way as in the case where 'p' is also to be estimated. Now mixture distribution 'x' is constructed as follows:

If

$$r(i, j) \le p, x(i, j) = x_1(i, j);$$

else,

$$x(i, j) = x_2(i, j);$$

In this case, only two raw moments and their functions of moments and their ratios are used namely.

Table 4: 1st and 2nd raw moments and their functions and ratios.

S.No.	1	2	3	4	5	6	7	8	9	10
m_1'	1.122	2.030	2.208	2.616	1.357	1.806	1.996	1.869	1.576	1.480
m_2'	1.639	4.411	5.069	9.401	2.636	3.346	4.724	4.334	5.002	3.095
$s_4 = m_2 -(2m_1')^{2'}$	0.123	-3.250	-4.302	0.817	0.544	-3.008	-1.767	-0.984	5.066	0.513
$s_4.(p*q)$	-0.769	-20.315	-26.889	5.110	3.401	-18.805	-11.049	-6.153	31.664	3.207

From the above set samples giving negative values namely $1,2,3,6,7,8$ are omitted and the remaining values are substituted in the given formulae.

$$\hat{a} = 2m_1' + q\sqrt{\frac{s_4}{pq}} \; ;$$

$$\hat{b} = \frac{(2m_1' - p\hat{a})}{q}$$

In this case, out of 10 samples, by eliminating 6 samples as giving un acceptable estimates. For this set the parameters are estimated and statistics namely mean, s.d, etc are given below:

Table 5: Acceptable estimates of parameters

$\hat{a} =$	7.0417	4.1894	7.6548	4.3946
$\hat{b} =$	4.7810	2.3453	2.0278	2.6036

Table 6: Estimates of parameters of their statistics

	mean	std	min	max	r. m. s
$\hat{a}$	5.82	1.78	4.18	7.65	1.74
$\hat{b}$	2.93	1.25	2.02	1.74	1.08

Table-1(a): Estimated parameters of Mixture of Two Rectangular distributions with parameters a, b, p, q, 4 sets of parameter values (a,b); $(5,3)$; $(6,3)$; $(7,3)$; $(10,3)$ with $p = 0.2:0.1:0.5$, sample size $m = 50$, number of samples $n = 500$, proportion of acceptable estimates of a, b, p.

1)a=5	Mean	Std	Min	Max	Rms	Mean	Std	Min	Max	Rms	Mean	Std	Min	Max	Rms	acceptable estimates
0.2	5.10	0.76	3.01	8.16	0.75	1.33	0.87	0.06	3.7	1.88	0.50	0.16	0.03	0.99	0.34	44%
0.3	5.31	0.68	3.53	11.70	0.58	1.55	0.83	0.01	3.73	1.66	0.50	0.13	0	0.81	0.24	45%
0.4	5.25	0.53	2.85	7.41	0.73	1.53	0.76	0.09	3.62	1.65	0.55	0.11	0.04	0.94	0.19	47%
0.5	5.32	0.65	3.92	11.85	0.68	1.59	0.76	0.03	3.26	1.59	0.59	0.11	0.06	0.87	0.15	46%
2)a=6																
0.2	6.06	0.67	3.14	7.51	0.48	1.87	0.81	0.09	3.89	1.38	0.38	0.12	0.04	0.89	0.21	47%
0.3	6.04	0.48	4.07	7.41	0.45	1.95	0.84	0.19	3.92	1.34	0.42	0.11	0.1	0.87	0.17	50%
0.4	6.09	0.44	3.98	7.07	0.47	2.14	0.87	0.01	4.44	1.21	0.48	0.12	0.04	0.78	0.15	55%
0.5	6.12	0.45	4.88	7.75	0.66	2.10	0.89	0.35	4.67	1.26	0.54	0.13	0.02	0.99	0.14	61%
3)a=7																
0.2	6.93	0.66	4.46	8.25	0.48	2.18	0.87	0.06	4.37	0.31	0.12	0.01	0.01	0.6	0.16	56%
0.3	7	0.488	4.58	8.06	0.49	2.49	0.86	0.33	4.68	0.99	0.36	0.12	0.01	0.86	0.14	64%
0.4	6.93	0.49	4.55	8.30	0.44	2.59	0.88	0.60	5.17	0.97	0.42	0.13	0.0006	0.87	0.14	68%
0.5	6.95	0.44	5.35	8.13	0.45	2.56	0.89	0.65	4.86	0.98	0.52	0.14	0.04	0.95	0.14	67%
4)a=10																
0.2	9.7	0.88	6.4	11.42	0.92	2.90	0.88	0.12	5.98	0.89	0.22	0.09	0	0.51	0.09	78%
0.3	9.77	0.71	6.6	11.14	0.75	3.07	0.84	1.50	5.81	0.84	0.30	0.11	0.01	0.58	0.11	89%
0.4	9.77	0.62	6.82	11.30	0.66	3.12	0.93	0.86	6.56	0.94	0.39	0.11	0	0.75	0.11	93%
0.5	9.77	0.57	7.35	11.09	0.61	3.12	0.99	1.12	6.45	1.00	0.49	0.13	0.03	0.99	0.13	94%

Table 2(a): A mixture of two Rectangular distributions with estimated parameters a and b, when p is known, 4 sets of parameter values (a,b); 1) $(5,3)$; 2) $(6,3)$; 3) $(7,3)$; 4) $(10,3)$, sample size $m = 50$, number of samples $n = 500$, proportion of acceptable estimates of a, b, p.

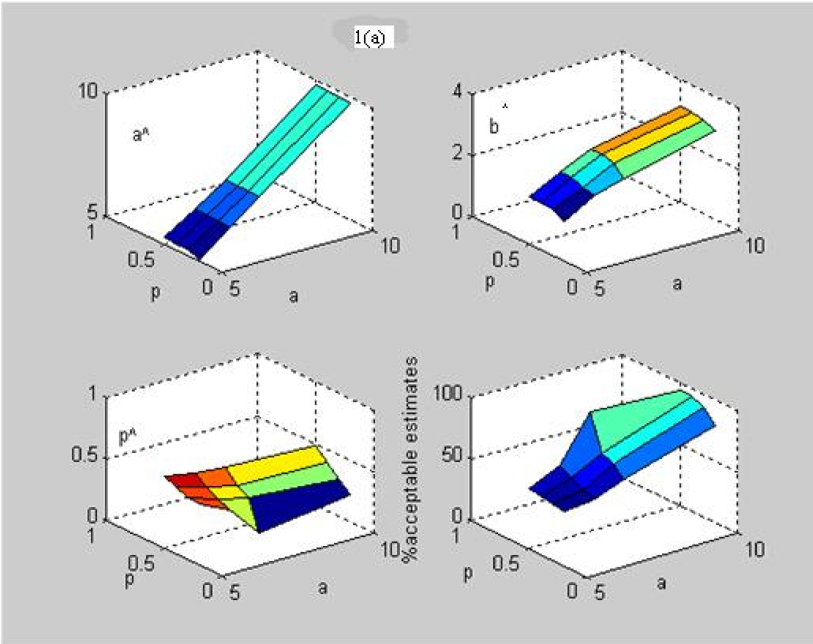
1)a=5	Mean	Std	Min	Max	Rms	Mean	Std	Min	Max	Rms	Mean	Std	Min	Max	Rms	acceptable estimates
0.2	5.10	0.76	3.01	8.16	0.75	1.33	0.87	0.06	3.7	1.88	0.50	0.16	0.03	0.99	0.34	44%
0.3	5.31	0.68	3.53	11.70	0.58	1.55	0.83	0.01	3.73	1.66	0.50	0.13	0	0.81	0.24	45%
0.4	5.25	0.53	2.85	7.41	0.73	1.53	0.76	0.09	3.62	1.65	0.55	0.11	0.04	0.94	0.19	47%
0.5	5.32	0.65	3.92	11.85	0.68	1.59	0.76	0.03	3.26	1.59	0.59	0.11	0.06	0.87	0.15	46%
2)a=6																
0.2	6.06	0.67	3.14	7.51	0.48	1.87	0.81	0.09	3.89	1.38	0.38	0.12	0.04	0.89	0.21	47%
0.3	6.04	0.48	4.07	7.41	0.45	1.95	0.84	0.19	3.92	1.34	0.42	0.11	0.1	0.87	0.17	50%
0.4	6.09	0.44	3.98	7.07	0.47	2.14	0.87	0.01	4.44	1.21	0.48	0.12	0.04	0.78	0.15	55%
0.5	6.12	0.45	4.88	7.75	0.66	2.10	0.89	0.35	4.67	1.26	0.54	0.13	0.02	0.99	0.14	61%
3)a=7																
0.2	6.93	0.66	4.46	8.25	0.48	2.18	0.87	0.06	4.37	0.31	0.12	0.01	0.01	0.6	0.16	56%
0.3	7	0.488	4.58	8.06	0.49	2.49	0.86	0.33	4.68	0.99	0.36	0.12	0.01	0.86	0.14	64%
0.4	6.93	0.49	4.55	8.30	0.44	2.59	0.88	0.60	5.17	0.97	0.42	0.13	0.0006	0.87	0.14	68%
0.5	6.95	0.44	5.35	8.13	0.45	2.56	0.89	0.65	4.86	0.98	0.52	0.14	0.04	0.95	0.14	67%
4)a=10																
0.2	9.7	0.88	6.4	11.42	0.92	2.90	0.88	0.12	5.98	0.89	0.22	0.09	0	0.51	0.09	78%
0.3	9.77	0.71	6.6	11.14	0.75	3.07	0.84	1.50	5.81	0.84	0.30	0.11	0.01	0.58	0.11	89%
0.4	9.77	0.62	6.82	11.30	0.66	3.12	0.93	0.86	6.56	0.94	0.39	0.11	0	0.75	0.11	93%
0.5	9.77	0.57	7.35	11.09	0.61	3.12	0.99	1.12	6.45	1.00	0.49	0.13	0.03	0.99	0.13	94%

Table 3(a): Estimated parameters of Mixture of Two Exponential distributions with parameters a, b, p, q, 4 sets of parameter values (a,b); 1) $(5,3)$; 2) $(6,3)$; 3) $(7,3)$; 4) $(10,3)$, with $p = 0.20:0.1:0.5$, sample size $m = 50$, number of samples $n = 500$, proportion of acceptable estimates of a, b, p.

						b=3										
1)a=5	Mean	Std	Min	Max	Rms	Mean	Std	Min	Max	Rms	Mean	Std	Min	Max	Rms	acceptable estimates
0.2	5.95	2.11	2.37	13.76	2.31	0.88	0.59	0.01	3.1	2.19	0.56	0.20	0.02	0.99	0.41	44%
0.3	6.23	2.27	2.67	16.63	2.58	0.97	0.72	0.01	3.71	2.15	0.54	0.19	0.11	0.98	0.31	44%
0.4	6.48	2.36	2.99	18.59	2.78	1.02	0.71	0.02	3.47	2.09	0.56	0.19	0.08	0.97	0.25	41%
0.5	7.05	2.65	2.88	16.71	3.34	1.14	0.81	0.00	4.61	2.02	0.54	0.21	0.07	0.99	0.21	43%
2)a=6																
0.2	6.62	2.32	2.73	15.36	2.40	0.99	0.80	0.00	3.67	2.15	0.52	0.20	0.05	0.99	0.38	50%
0.3	6.81	2.45	2.58	14.81	2.57	1.17	0.76	0.01	3.79	2.57	0.52	0.76	0.00	3.79	1.97	47%
0.4	7.39	2.49	3.34	18.47	2.85	1.23	0.91	0.01	4.85	1.98	0.53	0.21	0.003	0.99	0.25	39%
0.5	7.93	2.77	3.56	16.65	3.37	1.24	0.95	0.00	5.54	2.00	0.52	0.2	0.02	0.95	0.2	39%
3)a=7																
0.2	7.14	2.51	2.56	15.20	2.51	1.24	0.95	0.00	4.74	2.00	0.49	0.21	0.02	0.99	0.36	44%
0.3	7.56	2.58	3.37	18.16	2.63	1.41	0.94	0.01	4.58	1.84	0.5	0.2	0.04	0.99	1.84	41%
0.4	7.99	2.58	3.41	19.11	2.75	1.42	1.06	0.01	5.21	1.89	0.51	0.19	0.03	0.99	1.89	43%
0.5	8.84	2.98	4.01	23.55	3.50	1.42	1.06	0.01	6.04	1.90	0.53	0.19	0.01	0.99	0.19	40%
4)a=10																
0.2	8.44	2.65	3.48	20.20	3.07	1.74	1.27	0.01	6.15	1.78	0.45	0.2	0.04	0.99	0.32	59%
0.3	9	2.94	3.88	22.92	3.10	1.93	1.31	0.01	6.88	1.69	0.49	0.21	0.01	0.97	0.29	56%
0.4	9.51	2.84	3.93	18.55	2.87	2.09	1.44	0.01	6.82	1.70	0.51	0.2	0.03	0.97	0.23	49%
0.5	10.41	3.46	4.71	29.52	3.48	2.05	1.51	0.01	7.69	1.78	0.55	0.21	0.001	0.98	0.22	47%

Table 4(a): Estimated parameters of Mixture of Two Exponential distributions with parameters a and b when p is known, 4 sets of parameter values (a,b); 1) $(5,3)$; 2) $(6,3)$; 3) $(7,3)$; 4) $(10,3)$, sample size $m = 50$, number of samples $n = 500$, proportion of acceptable estimates of a, b.

1)	Mean	Std	Min	Max	Rms	Mean	Std	Min	Max	Rms	acceptable estimates
0.2	5.64	1.39	2.42	13.73	1.53	2.86	0.51	1.34	4.57	0.52	100%
0.3	5.32	1.2	2.89	9.76	1.24	2.82	0.54	1.05	5.54	0.57	100%
0.4	5.3	1.08	3.06	10.36	1.12	2.81	0.62	0.58	5.17	0.64	100%
0.5	5.24	1.02	3.08	10.77	1.05	2.78	0.74	0.28	5.11	0.77	100%
2)											
0.2	6.23	1.73	2.88	16.40	1.74	2.97	0.55	1.20	5.19	0.55	100%
0.3	6	1.55	2.86	11.63	1.55	2.96	0.60	0.90	5.27	0.60	100%
0.4	6.02	1.33	3.11	12.27	1.33	3.01	0.74	0.12	5.51	0.74	100%
0.5	6.002	1.27	3.44	12.79	1.27	3.03	0.85	0.10	5.47	0.85	100%
3)											
0.2	6.75	2.17	3.01	14.93	2.18	3.07	0.58	1.45	5.03	0.58	100%
0.3	6.87	1.82	3.2	13.29	1.82	3.11	0.75	0.79	5.47	0.76	100%
0.4	6.64	1.69	3.49	12.85	1.73	3.22	0.87	0.33	5.69	0.90	100%
0.5	6.77	1.45	3.74	13.26	1.47	3.16	1.05	0.01	*6.12*	1.06	100%
4)											
0.2	9	3.43	3.26	21.82	3.57	3.24	0.69	0.88	5.54	0.73	100%
0.3	9.04	2.78	3.5	18.79	2.94	3.34	0.88	0.91	7.50	0.94	100%
0.4	9.19	2.5	3.9	17.33	2.63	3.39	1.12	0.78	6.72	1.19	100%
0.5	9.53	2.07	4.49	17.57	2.12	3.51	1.37	0.10	8.73	1.46	100%



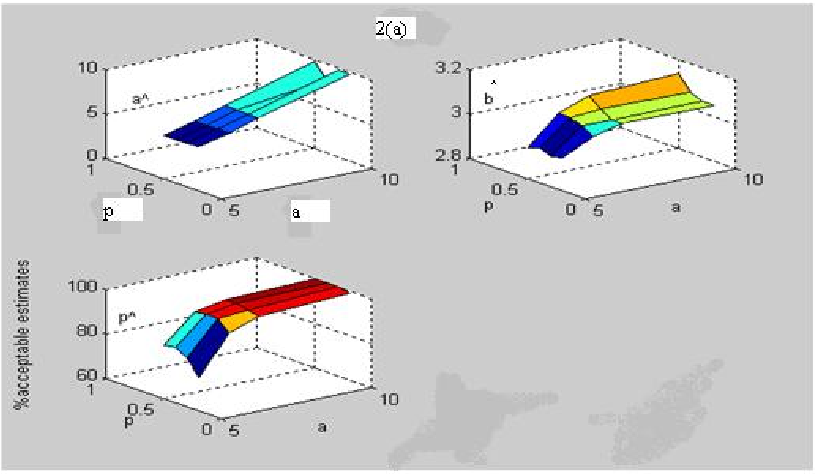
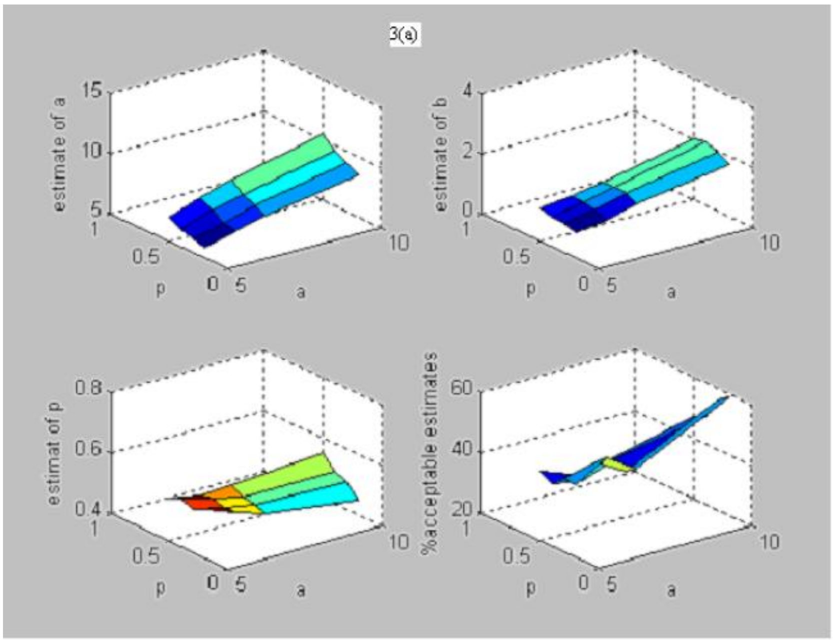


Figure 1(a), 2(a) m=50,n=500 Estimates for parameters (Rectangular Mixture distribution 'p' is not known, 'p' is known, where b=3 fixed)



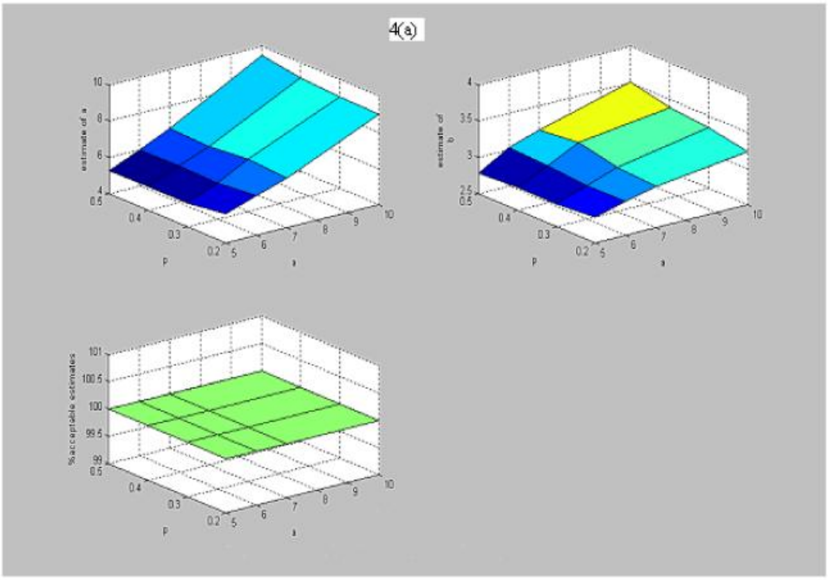


Figure 3(a), 4(a) m=50, n=500 Estimates for parameters (Exponential mixture distribution 'p' is not known, 'p' is known, where b=3 fixed)

Results of Simulation (Tables & Graphs)

Table 1(a): Gives simulational results for the case of rectangular mixtures with different combinations of a and p where b is fixed $(b=3)$. For each of size $m=50$, number of samples $n=500$,were generated and their respective graph are also presented in Figure 1.1(a).

Table 2(a): Presents results of simulation for the case where p is known. In the same way as table 1(a) the only difference being ‘ p ’ is known.

Tables 3(a) & 4(a): presents the case for exponential mixtures with similar combinations of a, b and p where a and p are exponential parameters (with $b = 3$ fixed) a and p taking the values a and $5, 6, 7, 10$ and a and $p = 0.2$ to 0.5. Their respective graphs are also presented in figures.

Observations and Comments from Tables and Graphs

Example: Table -1(a) (Rectangular

It shows when $a = 5, p = 0.2$ and $b = 3$, the mean value of estimates of $a = 5.10, b = 1.33$, $p = 0.50$. As p increases $(0.2 : 0.1 : 0.5)$ actual value of parameters are nearer to its estimated value. And the proportional of acceptable estimates are found to be 44% to 60%. For the same case where $a = 6, p = 0.2, b = 3$, the mean value of estimates of $a = 6.06, p = 0.38, b = 1.87$. As ‘ p ’ increases parameters a, b, p are better estimated. Similarly $a = 7$ and 10 for $p = 0.2, b = 3$ the estimated values of parameters are $a = 6.93, p = 0.12, b = 0.87$; $a = 9.7, p = 0.22, b = 2.90$. Here the estimates of parameters are quite nearer to its true value, proportional of acceptable estimates are found to be 56% to 94%.

These are presented in graph. From this graph one can observe that when the difference of a and b is small as p increases one f the parameter (larger) is estimated well, the other b and p are under estimated. When ‘ a ’ and ‘ b ’ quite far away one another as ‘ p ’ increases parameters are better estimated.

In general estimated values of ‘ p ’ are sufficiently nearer to the true value. The averages being almost equal to the true value expect for the case $a = 5, b = 3, p = 0.2$ where estimate of ‘ p ’ is 0.5. About the proportion of samples gives acceptable estimates we find that as ‘ p ’ increases 0.2 to 0.5 acceptability also increases. Similarly as the difference between a and b increases proportional of acceptable samples also increases giving almost 100 percent acceptability when $a = 10, b = 3$.

As ‘ m ’ increases from $m = 50, 100, 200$, the situation considerably improves; proportion of rejects gets reduced. With acceptable samples the estimates will be sufficiently nearer to the true values. When two populations are sufficiently distinct i.e., $a - b$ is not small, the estimates will be generally better.

Tables 1(a) and 2(a) are the proportion of acceptable in the two cases (1) ‘ p ’ is known (2) ‘ p ’ is un-known. The proportion of acceptable samples are considerably high when ‘ p ’ is known, than when ‘ p ’ is un-known and the estimates (a and b) are quite nearer to the true value.

The above tables are shown in graphs as Fig 1.1(a) shows estimates of 'a' from the acceptable sets of sample is quite satisfactory giving estimates very nearer to the true of 'a' irrespective of the proportion 'p'. However for 'b' which is always true value is 3 the estimates will be highly underestimates when $a=5$. As 'a' increases (becomes more different from 'b'). The corresponding estimate of 'b' become less unsatisfactory and when $a=10$ they are almost good becoming even better as 'p' increases towards 0.5. When 'a' is 5 proportion of acceptable estimates are quite low in the range 44% to 51%. As 'a' increases this proportional of acceptable estimates also increases quite rapidly, and when $a=10$ it starts from 80% to 100% as 'p' increases to 0.5. When $a=5$ and 6 the proportion of acceptable estimates itself is quite low and the estimates of 'p' in this case also are not satisfactory, being high over estimates for $a=5$ and for $a=6$ less unsatisfactory. In case of $a=7$ and 10 these estimates expect for $p=0.2$ in other cases the estimates are relatively satisfactory. However for the case $a=10$ proportional of acceptable estimates is also high (above 80%) and estimates of 'p' are quite nearer to the true values though they have moderately high s.d.

In fig 2(a) rectangular ('p' is known): In this case estimates of 'a' are quite satisfactory for $a=5$, while for other cases they are slightly under estimating. The degree of under estimating being reduces as 'p' increases from 0.2 to 0.5. However regarding estimates of 'b' there is a very slight under estimation when $a=5$, while for other cases that is $a=6,7,10$ estimates are slightly over estimating.

Fig 3(a) in exponential case: When $a=5,6$ and 7 as 'p' increase to 0.2 to 0.5, the estimates of a and p are slightly over estimated. If $a=10$, for $p=0.2$ the estimate of 'a' is being under estimated, approaching the true value as 'p' increases to 0.5. When $b=3$ the estimates of 'b' are under estimated as 'p' increases. From this case $(m=50, n=500)$ the proportional of acceptable samples are found to be 44% to 60%.

Fig 4(a) in exponential (when 'p' is known) $a=5$ and 6 for $p=0.2$ the estimate of 'a' is being over estimated and 'p' increases to 0.5 the estimates of 'a' is nearer to its true value. For the corresponding 'b' is nearer to its true value. If $a=7$ and 10 as 'p' increases the estimates of 'a' is slightly under estimated. From this case estimates of 'b' are over estimated. Here the proportional of acceptable samples found to be 100%.

Similarly, generated estimated parameters for various combinations of parameters (a,b,p), with different sample sizes (m) and different number of samples (n) for a mixture of two rectangular / exponential distributions of size $m\times n$.

As 'm' increases $(50, 100, 200)$ when 'p' is unknown, the proportional of acceptable samples are found to be 44% to 80%. It is interesting to note that when 'p' is known for the case of exponential the samples giving almost 100% acceptable estimates. While in the rectangular it ranges 75% to 84% $(a = 5, b = 3)$. In other cases it is around 75% to 100%.

Hence on the whole this approach does not seem to give useful estimates in the case of rectangular / exponential these are presented in tables and graphs.

This fact shows the importance of developing acceptable estimates as often happens from samples which lead to unacceptable estimates.

6. CONCLUSIONS

1. Two parameters of a mixture distribution are quite nearer, the proportions of acceptable estimates are found to be 40% to70%. And things become better when parameters are relatively far off (difference is large enough).
2. For rectangular even among the acceptable estimates, the bigger parameter is better estimated and the other one is mostly underestimated. About estimate of 'p', not bad in rectangular, except for $a = 5$, $b = 3$; for exponential case it is always bad.
3. When p is known, unacceptable estimates are reduced and found proportional to acceptable estimates of a and b, and estimated parameter values are very close to the true value.

When 'p' is known, proportion of acceptable estimates is in nearer to be 75% and more will be acceptable: and both estimates of a and b will be generally good, in rectangular and exponential mixture distributions.

REFERENCES

Blischke, W. R. (1962): Moment estimators for the parameters of a mixture of two binomial distributions. *Ann. Math. Statist.*, **33**, 2, 444-454.

Craigmile, P. F. and Titterington, D. M. (1997): Parameter estimation for finite mixtures of uniform distributions. *Comm. Statist. Theory Methods*, **26**, 1981-1995.

Cheng, S. W. and James C. Fu. (1982): Estimation of mixed Weibull parameters in life testing. IEEE *Transactions on Reliability*, ***31***, 4, 377-381.

Dan Ling, Hong-Zhong Huang, Yu Liu (2009): *A method for parameter estimation of Mixed Weibull distribution' Annual Reliability and Maintainability*, Symposium – Proceedings (Session-06C)-Fort Worth, Texas, USA, January, 26-29.

Day, N. E. (1969): Estimating the components of a mixture of normal distributions. *Biometrika*, **56**, 463-474.

Flowsky, E. B. (1979): Some methods for studying the mixture of two normal (lognormal) Distributions. *J. Amer. Statist. Assoc*., **74**, 367, 561-575.

Harding, I. P. (1948): The use of probability paper for the graphical analysis of polymodel frequency distributions, *J. Amer. Statist. Assoc.*, **28**, 141-153.

Hasselblad, V. (1966): Estimation of parameters for a mixture of normal distributions. *Techno metrics*, **8**, 431-444.

Hussein, A. and Liu, J. (2009): Parametric estimation of mixtures of two uniform distributions, *J. Stat. Comput. Simul.*, **79**, 4, 395-410.

Rao, C. R. (1952): *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, New York.

Rider, P. R. (1961): The method of moments applied to a mixture of two exponential distributions. *Ann. Math. Statist*., **32**, 1, 143-147.

Tallis, G. M. and Light, R. (1968): The use of fractional moments for estimating the parameters of a mixed exponential distributions. *Technometrics*, **10**, 1, 161-175.

Received: 15.07.2012

Revised: 28.02.2013

G. Lavanya
Department of Statistics
Kakatiya University, Warangal
Email:1naresh_mba206@yahoo.co.in

V.V. HaraGopal and S. N. N. Pandit

Department of Statistics
Osmania University
Hyderabad
Email: haragopalvajjha@gmail.com