Journal of the Korean
Data & Information Science Society
1998, Vol. 9, No. 2 pp. 289 ~ 298

The Null Distribution of the Likelihood Ratio Test for a Mixture of Two Gammas

Dae kee Min [1]

Abstract

We investigate the distribution of likelihood ratio test(LRT) of null hypothesis a sample is from single gamma with unknown shape and scale against the alternative hypothesis a sample is from a mixture of two gammas, each with unknown scale and unknown (but equal) scale. To obtain stable maximum likelihood estimates(MLE) of a mixture of two gamma distributions, the EM(Dempster, Laird, and Rubin(1977))and Modified Newton(Jensen and Johansen(1991)) algorithms were implemented. Based on EM, we made a simple structure likelihood equation for each parameter and could obtain stable solution by Modified Newton Algorithms. Simulation study was conducted to investigate the distribution of LRT for sample size n = 25, 50, 75, 100, 50, 200, 300, 400, 500 with 2500 replications.

To determine the small sample distribution of LRT, I considered the model of a gamma distribution with shape parameter equal to $1 + f(n)$ and scale parameter equal to 2.

The simulation results indicate that the null distribution is essentially invariant to the value of the shape parameter. Modeling of the null distribution indicates that it is well approximated by a gamma distribution with shape parameter equal to the quantity $0.927 + 1.18/\sqrt{n}$ and scale parameter equal to 2.16.

Key Words and Phrases: EM Algorithm, Modified Newton Algorithm, Finite Mixture Models

1. Introduction

We consider the problem of testing the hypothesis that we have a single gamma distribution versus a mixture of two gamma distributions. The problem arise in

[1]Full Time Insrtuctor, Department of Statistical Information, Catholic University of Taegu-Hyosung, 330 Kumnak-Dong, Hayang-Up, Kyungsan-City, Kyungbuk, 713-702, Korea.

practice when failure can occur for more than one reason and the failure distribution for each reason can be adequately approximated by a simple density function. The overall failure distribution is then a mixture of gamma distributions.

There are three main problems that are encountered when developing an inference procedure for mixture distributions. The first is the problem of testing the null hypothesis that a sample is from a single distribution against the alternative hypothesis that a sample is from mixture. A natural candidate for a test statistic is the likelihood ratio test and this has been discussed by Wolfe (1970, 1971) and Hasseblad (1969). The null hypothesis can be tested against the alternative hypothesis by computing the likelihood ratio,λ, given by $\lambda = \mathcal{L}_0/\mathcal{L}_1$ provided we know the sampling distribution of λ under the null hypothesis. This distribution was studied originally by Wilks(1938) who showed that under the regularity conditions $-2\log\lambda$ is asymptotically distributed as χ^2 with degrees of freedom equal to the difference in the number of parameters between the two hypotheses.

The second problem in considering mixture distribution is related to estimation. There are both theoretical and practical difficulties in estimation of the parameters of the mixture distribution. Theoretically, difficulties arise with multiple local maxima of the likelihood surface. This problem is particularly troublesome because the iterative maximization methods necessary to compute the maximum likelihood may converge to a local rather than the global maximum. Practically, convergence of the numerical procedure used to obtain estimates may be another issue. The computational procedures for obtaining maximum likelihood estimates for mixture are usually fairly lengthy iterative processes that may not converge to an acceptable solution. However, these computational difficulties seem to have been resolved somewhat with the application of the EM algorithm.

The EM algorithm is an iterative technique for computing maximum likelihood estimates for incomplete data. The algorithm has been widely used in a variety of settings, with early application to genetics, grouping and censoring, and missing data. Dempster, Laird and Rubin (1977) gave a theoretical basis for the algorithm, and named it EM since the two computational steps involve expectation and maximization. The phrase incomplete data is used quite broadly to represent variety of situations, including mixtures, censored data and missing observations.

Finite mixture models have been widely studied and used in applications (Titterington, Smith, and Markov, 1985). However only a few articles mention the gamma mixture. The mixed gamma distribution does in medical science in the study of ages of onset of autoimmune disease. Kanno (1982) studied maximum likelihood estimation of parameters for a mixture of two gamma distributions using the Newton-Raphson method. Since the Newton-Raphson method was very sensitive to initials values, he obtained the initial values by the method of moments. He simplified the form of the Hessian matrix based on his assumption that the overlap is sufficiently small between component distributions so that all the elements which contain multiple of estimates of mixing proportions would be zero in the Hessian matrix. Without this assumption, he couldn't get a simple form of the Hessian matrix. Eventually he failed to obtain stable estimates. Because a gamma distribution has a long right tail, he couldn't always have sufficiently small overlap.

For the case of a two-component gamma mixture with two unknown and unequal shape parameter and unknown and equal scale parameters, detailed study has been conducted to obtain a stable algorithm for the MLE and to investigate the null distribution of the LRT and its asymptotic properties.

2. Obtaining the MLE of the Parameters of the Two Component Gamma Distribution Using an EM PLUS Modified Newton Algorithm

The definitive reference for the Modified Newton Procedure applied to the derivative of reciprocal likelihood function in a one-dimensional exponential family is the paper by Jensen and Johansen (1991). In this paper they showed global convergence under very general assumptions of iterative maximization procedures with cyclic fixing of groups of parameters, maximizing over the remaining parameter.The EM algorithm for finding the MLE is a powerful numerical technique useful for incomplete data problems. The primary conceptual power of this iterative algorithm lies in converting a maximization problem involving a complicated likelihood, into a sequence of "pseudo-complete" problems, where at each step the updated parameter estimates can be in a closed form. Unlike the Newton-Raphson method, gradient matrices don't need to be derived. The definitive reference for the EM algorithm is the paper by Dempster, Laird, and Rubin (1977). The general idea behind the EM is to represent the observed data vector, y, as the realization of some incompletely or indirectly observed data vector, say x, which we term the complete data. The observed data y has density $g(y|\Psi)$ where Ψ is the group of unknown parameters.

For a k component univariate gamma mixture, the observations $\{x_1, x_2, \cdots, x_n\}$ can be regarded as a sample of incomplete data by considering each x_i to be the known part of complete data $y_i = (x_i, z_i)$. This emphasizes the interpretation of mixture data as incomplete data, with the indicator vectors as missing values.

Based on complete data, the likelihood corresponding to $(y_1, y_2 \ldots, y_n)$ can be written as

$$g(y_1, y_2 \ldots, y_n|\Psi) = \prod_{i=1}^{n}\prod_{j=1}^{k} \pi_j^{z_{ij}} f_j(x_i|\theta_j)^{z_{ij}} \tag{1}$$

where $f_j(x_i|\theta_j) = \frac{1}{\Gamma(\alpha_j)}\beta^{-\alpha_j}x^{\alpha_j-1}exp(-\frac{x}{\beta})$.

With logarithm

$$\mathrm{L}_0(\Psi) = \sum_{i=1}^{n} z_i^T V(\pi) + \sum_{i=1}^{n} z_i^T U_i(\theta) \tag{2}$$

where $V(\pi)$ has the jth component $\log \pi_j$ and $U_i(\Theta)$ has the jth component $\log f_j(x_i| \Theta_j)$.

The form of the mixture likelihood $L_0(\Psi)$ corresponds to the marginal density of $x_1, x_2 \ldots, x_n$ obtained by summing over $z_1, \ldots, z_n$.

The EM algorithm generates, from some initial approximation, $\Psi^{(0)}$, a sequence $\left\{\Psi^{(m)}\right\}$ of estimtes. Each iteration consists of expectation and maximization step. The expectation step, calculation of expectation of log-likelihood procedure is performed as follows.

$$\begin{aligned}
Q(\Psi, \Psi^{(m)}) &= E[\log g(y|\Psi)|x, \Psi^{(m)}] \\
&= \sum_{i=1}^{n} E(z_i|x_i, \Psi^{(m)})V(\pi) + \sum_{i=1}^{n} E(z_i|x_i, \Psi^{(m)})U_i(\Theta) \\
&= \sum_{i=1}^{n} w_i(\Psi^{(m)})^T V(\pi) + \sum_{i=1}^{n} w_i(\Psi^{(m)})^T U_i(\Theta)
\end{aligned}$$

where $w_i(\Psi^{(m)}) = E(z_i|x_i, \Psi^{(m)})$ $\Theta = \{\alpha_1, \alpha_2, \beta\}$.

$w_{ij}(\Psi(m)) = [w_i(\Psi^{(m)})]_j = \frac{\pi_j^{(m)} f_j(x_i|\Theta_j^{(m)})}{p(x_i|\Psi^{(m)})}$ is the probability of category membership for the i-th the observation conditional on x_i and given that the parameter is $\Psi^{(m)}$ for each i, j.

Next maximization $Q(\Psi, \Psi^{(m)})$ step to find $\Psi = \Psi^{(m+1)}$ is done by maximizing the quantities $\sum_{i=1}^{n} w_i(\Psi^{(m)})^T V(\pi)$ and $\sum_{i=1}^{n} w_i(\Psi^{(m)})^T U_i(\Theta)$ respectively. Thus for each parameter the M-step is as follows.

1. M - step for π;

$$\begin{aligned}
\sum_{i=1}^{n} E(z_i|x_i, \Psi^{(m)}) &= \sum_{i=1}^{n} E(z_i|\Psi) \\
\sum_{i=1}^{n} w_{ij}(\Psi^{(m)}) &= n\pi_j \\
\hat{\pi} &= \frac{1}{n} \sum_{i=1}^{n} w_{ij}(\Psi).
\end{aligned}$$

2. M - step for β;

$$\begin{aligned}
\sum_{i=1}^{n} w_i(\Psi^{(m)})^T U_i(\Theta) &= \sum_{i=1}^{n} w_i(\Psi^{(m)}) \log f_j(x_i|\Theta_j^{(m)}) \\
&= \sum_{i=1}^{n} w_i(\Psi^{(m)})[-\log\Gamma(\alpha_j) - \alpha_j \log\beta \\
&\quad +(\alpha_j - 1)\log x_i - \frac{x_i}{\beta}] \\
\frac{\partial}{\partial\beta}\sum_{j=1}^{k}\sum_{i=1}^{n} w_i(\Psi^{(m)})U_i(\Theta) &= \sum_{j=1}^{k}\sum_{i=1}^{n} w_i(\Psi^{(m)})(-\frac{\alpha_j}{\beta} + \frac{x_i}{\beta^2}) \\
\hat{\beta} &= \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} w_i(\Psi^{)} x_i}{\sum_{j=1}^{k}\sum_{i=1}^{n} w_i(\Psi)\hat{\alpha}_j} \\
&= \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} w_i(\Psi) x_i}{\sum_{j=1}^{k} \hat{\pi}\hat{\alpha}_j}.
\end{aligned}$$

3. M - step for α_j;

$$\begin{aligned}
\sum_{i=1}^{n} w_i(\Psi^{(m)})^T U_i(\Theta) &= \sum_{i=1}^{n} w_i(\Psi^{(m)}) \log f_j(x_i|\Theta_j^{(m)}) \\
&= \sum_{i=1}^{n} w_i(\Psi^{(m)})[-\log\Gamma(\alpha_j) - \alpha_j \log\beta \\
&\quad +(\alpha_j - 1)\log x_i - \frac{x_i}{\beta}]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\alpha_j}\sum_{i=1}^{n} w_i(\Psi^{(m)})U_i(\Theta) &= \sum_{i=1}^{n} w_i(\Psi^{(m)})[-\varphi(\alpha_j) - \log\beta + \log x_i] \\
\text{Let} \quad f(\alpha_j) &= -\varphi(\alpha_j) - \log\beta + \frac{\sum_{i=1}^{n} w_i(\Psi)\log x_i}{\sum_{i=1}^{n} w_i(\Psi)}
\end{aligned}$$

By The Newton Method

$$\begin{aligned}
\hat{\alpha}_{j,m+1} &= \alpha_{j,m} - \frac{f(\alpha_{j,m})}{\acute{f}(\alpha_{j,m})} \\
&= \alpha_{j,m} - \frac{f(\alpha_{j,m})}{\acute{\varphi}(\alpha_{j,m})}
\end{aligned}$$

and by The Modified Newton

$$\hat{\alpha}_{j,m+1} = \alpha_{j,m} - \frac{f(\alpha_{j,m})}{\acute{\varphi}(\alpha_{j,m}) + f(\alpha_{j,m})}.$$

3. The Null Distribution of the LRT Under the Null Hypothesis of a Single Gamma Distribution

3.1 Invariance of the LRT to the Shape Parameter α

As we know, the standard gamma density

$$f_X(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}\exp(-x), \quad \text{with} \quad \alpha > 0 : x > 0 \tag{3}$$

is a special case of the two parameter gamma density with $\beta = 1$. Thus we can tell clearly that the likelihood value depends on α from 3. So the distribution of the likelihood ratio test statistic may depend on α.

To study whether the LRT depends on α, 2500 samples of size $n = 100$ from a standard gamma were generated for $\alpha = 4, 8, 12, 16, 20, 30, 40$. The SAS function UNIVARIATE was used to obtain the simulated mean, standard deviation, and percentage points of the observed value of the LRT $G^2_{1:2}$. We also considered the possibility that the distribution of the LRT statistic, $G^2_{1:2}$ is pseudo chisquare and hence a gamma variable with shape parameter, $\alpha_{G^2_{1:2}}$ and scale parameter, $\beta = 2.0$. Thus $\alpha_{G^2_{1:2}}$ was estimated for each value of α

As we can see in Table 1, there is no clear trend in mean and variance of the LRT statistic with respect to the α value. The Kruskal-Wallis test was done to test whether the distribution has the same median across the α values. The test results were significant at .05 level when we considered all α values. But the results were not significant on comparing the distributions from $\alpha = 8, 12, 30, 40$. This supports our conclusion that any dependency of $G^2_{1:2}$ on α is very small and the distribution can be considered as being essentially alpha invariant.

Table 1. Comparison of the Null Distribution of the LRT Statistic for Different Values of α

%	VALUE of SHAPE PARAMETER α							
	4	8	12	16	20	30	40	χ^2_2
99	9.74	10.37	9.53	10.14	10.25	10.32	9.82	9.21
95	6.49	6.57	6.55	6.56	7.15	6.52	6.62	5.99
90	5.09	4.99	5.07	5.19	5.53	5.06	5.14	4.60
75	3.2	3.02	3.08	3.23	3.21	3.06	3.06	2.41
50	1.63	1.48	1.51	1.56	1.67	1.45	1.58	1.39
25	0.68	0.63	0.60	0.67	0.69	0.61	0.65	0.56
10	0.25	0.25	0.2	0.23	0.23	0.2	0.24	0.21
5	0.12	0.13	0.09	0.09	0.1	0.1	0.12	0.1
1	0.01	0.03	0.01	0.02	0.01	0.01	0.02	0.02
mean	2.25	2.18	2.17	2.26	2.34	2.18	2.22	2.0
s.e	2.11	2.31	2.13	2.2	2.26	2.34	2.18	2.0

3.2 Modeling the Null Distribution of the LRT

To determine the small sample distribution of $G^2_{1:2}$, the 2500 samples of $G^2_{1:2}$ calculated for each sample size of $n = 25, 50, 75, 100, 150, 200, 300, 400, 500$ were subdivided into 5 subgroups, stratified by their order of occurrence during the generation of the samples. Then the MEL's of $\alpha_{G^2_{1:2}}$ and $\beta_{G^2_{1:2}}$ were computed for each subgroup and sample size.

We fixed $\hat{\beta} = 2.16$, the observed median of the $\hat{\beta}$, MEL of $\hat{\alpha}$ were computed for each subsample, because the estimates $\hat{\alpha}$ are function of sample size and the estimates of $\hat{\beta}$ are independent of sample size.

Since the $\hat{\alpha}_n$ were decreasing as sample size increase, the 45 obtained values of $\hat{\alpha}$ were regressed on the quantity $(\frac{1}{n})^t$ for $t = 0.5, 1, 2$. For $t = 0.5$, the observed R^2 was 0.75 and $F = 122.62, P$-value$= 0.0001$. The lack of fit statistics is not significant at the $P = 0.05$ level. The results $R^2 = .75$ from the regression show that the model for $\hat{\alpha}$ fits very well to data. For $t = 0.5$, the estimated regression equation is

$$\hat{\alpha}_n = 0.927 + \frac{1.18}{\sqrt{n}} \quad (4)$$

$$\hat{\beta} = 2.16 \quad (5)$$

The fitted percentage points of $\Gamma(\alpha_n, 2.16)$ for $n = 25, 50, 75, 100, 150, 200, 300, 400, 500$ are given in Table 2.

Table 2. Fitted Percentage of Null Distribution of LRT

%	SAMPLE SIZE									
	25	50	75	100	150	200	300	400	500	χ^2_2
0.99	10.72	10.39	10.24	10.17	10.05	9.99	9.92	9.88	9.85	9.12
0.95	7.13	6.84	6.72	6.65	6.55	6.51	6.45	6.41	6.38	6.03
0.90	5.56	5.30	5.19	5.14	5.04	5.01	4.95	4.92	4.89	4.71
0.75	3.46	3.26	3.17	3.12	3.05	3.02	2.97	2.95	2.93	2.85
0.50	1.83	1.68	1.62	1.59	1.53	1.51	1.48	1.46	1.45	1.48
0.25	0.82	0.73	0.69	0.67	0.64	0.63	0.61	0.60	0.59	0.6
0.10	0.34	0.28	0.26	0.25	0.24	0.23	0.22	0.21	0.20	0.16
0.05	0.18	0.14	0.13	0.12	0.11	0.11	0.10	0.10	0.10	0.07
0.01	0.044	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.019	0.02

3.3 Goodness of Fit of Null Distribution of LRT

The Kolmogorov-Smirnov goodness-of-fit test·(one-sample test) is applicable to 2500 ungrouped measurements and the null hypothesis simulated samples from a gamma distribution with $\hat{\alpha}$ and $\hat{\beta}$ from the estimated equation for each sample size. A central feature of the K-S test is that it is invariant under the reparameterization of x.

As you see in Table 3, the K-S test results were none significant at the 0.1 level for the all sample sizes. Thus based on this test we conclude that the GAM(0.927 + $1/\sqrt{n}$,2.16) is appropriate for approximating the null distribution of the LRT.

Table 3. Goodness of Fit of GAM(0.927 + 1.18/n)

Sample size	α	β	D_n	P-Value
25	1.16	2.16	0.0127	0.8
50	1.09	2.16	0.023	0.12
75	1.06	2.16	0.0237	0.117
100	1.045	2.16	0.024	0.103
150	1.02	2.16	0.0168	0.47
200	1.01	2.16	0.0217	0.207
300	0.995	2.16	0.0227	0.18
400	0.986	2.16	0.0185	0.35
500	0.98	2.16	0.0233	0.118

4. Conclusions

We successfully found a stable maximization algorithm by applying the EM algorithm. Because the EM algorithm does not require us to calculate gradients in each iterative step, the derivatives of log-likelihood function of gamma mixture is pretty simple for each parameter.

Before investigating the distribution of LRT, we checked the invariance of the LRT to α. We conclude that the dependency of the LRT on α is very small.

Based on a regression analysis of these simulated results, the asymptotic null distribution of LRT is a gamma distribution with shape parameter equal to 0.927 and scale parameter equal to 2.16. We give tables of the percentage points of the LRT based on the model. The relative error between percentage points of simulation and expectation were extremely small. In addition to this, we show the results of K-S goodness of fit (one sample) test for expected parameter value. The test results were not significant for the all sample sizes. This means estimated parameter value is appropriate for the all samples. Distribution of LRT is $\mathrm{GAM}(0.927 + 1/\sqrt{n}, 2.16)$.

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm, *J. R. Statist. Soc.*, 39, 1-38.

2. Hasseblad, V. (1969). Estimation of Parameters for a Mixture of Normal Distribution, *Technometrics*, 8(3), 431-444.

3. Jamshidian, M. and Jennrich, I. R. (1997). Acceleration of the EM Algorithm by using Quasi-Newton Methods, *J. R. Statist. Soc.*, 59, 569-587.

4. Jensen, S. T., Johansen, S., Lauritzen, S. L. (1991). Globally convergent algorithm for maximizing a likelihood function, *Biometrika*, 78, 867-877.

5. Kanno, R., (1982). Maximum likelihood estimation of parameters for a mixture of two gamma distribution, *Rep. Stat. Appl. Res., JUSE*, 29, 14-24

6. McLachlan, G. J. (1997). *Recent Advances in Finite Mixture Models*: Wiley

7. Tipping, E. M., and Bishop, M. C. (1997). Mixtures of Probabilistic Principal Compoent Analysis, *Technical Report NCRG.*

8. Titterington, D. M., Smith, A. F. M., Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Chichester: Wiley.

9. Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses, *Ann. Math. Stat.*, 9, 60-62.

10. Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis, *Multivariate Behavioral Research*, 329-350.

11. Meng, X. L. and Dyk, V. D. (1997). The EM Algorithm-an Old Falk-song Sung toa Fast New Tune, *J. R. Statist. Soc.*, 59, 511-567.