Title	Test for homogeneity in gamma mixture models using likelihood ratio
Author(s)	Wong, ST; Li, WK
Citation	Computational Statistics & Data Analysis, 2014, v. 70, p. 127-137
Issued Date	2014
URL	http://hdl.handle.net/10722/194985
Rights	NOTICE: this is the author's version of a work that was accepted for publication in Computational Statistics & Data Analysis. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Computational Statistics & Data Analysis, 2014, v. 70, p. 127-137. DOI: 10.1016/j.csda.2013.09.001

Test for homogeneity in gamma mixture models using likelihood ratio

Tony Siu Tung Wong*, Wai Keung Li

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong

Abstract

A testing problem of homogeneity in gamma mixture models is studied. It is found that there is a proportion of the penalized likelihood ratio test statistic that degenerates to zero. The limiting distribution of this statistic is found to be the chi-bar-square distributions. The degeneration is due to the negative-definiteness of a complicated random matrix, depending on the shape parameter under the null hypothesis. In light of this dependency, bounds on the distribution are introduced and a weighted average procedure is proposed. Simulation suggests that the results are accurate and consistent, and that the asymptotic result applies to the maximum likelihood estimator, obtained via an Expectation-Maximization algorithm.

Keywords: Chi-bar-square distributions, gamma mixture, likelihood ratio, maximum likelihood, negative definite

1. Introduction

In recent years, gamma mixture models have seen a surge of applications in many fields. Craig and Strassels (2010) examined the out-of-pocket prices of analgesic medications using a two-component gamma mixture model. See also Mayrose et al. (2005) for applications in bioinformatics and the references in Liu et al. (2003). Due to their importance, developing effective and handy statistical procedures for gamma mixture models is an imperative

*Corresponding author. Tel.: +852 28592469; fax: +852 28589041.

Email addresses: `wongtonyst@hku.hk` (Tony Siu Tung Wong), `hrntlwk@hku.hk` (Wai Keung Li)

task, in particular for the test of homogeneity. An obvious way of approaching the problem is to use the ordinary likelihood ratio test (LRT). One of the few results available is Liu et al. (2003). The authors showed that when the range of some parameters is unbounded, the LRT statistic diverges to infinity at a rate of $\log \log n$ and that its asymptotic behaviour is of extreme-value type through a highly complex piece of stochastic analysis. However, their simulation results suggested that the limiting distribution is far from converging to the extreme value distribution and that a possible solution is to simulate the finite-sample null distribution. The peculiar behaviour of the statistic arises because the maximum likelihood estimator (MLE) of some parameters may not be consistent. See, for example, the asymptotic result for $R_n(\varepsilon; I)$ in Chen and Chen (2001). Related problems in general mixture models were also addressed by Ghosh and Sen (1985), Dacunha-Castelle and Gassiat (1999), Chen and Chen (2001) and Liu and Shao (2003). In particular, Ghosh and Sen (1985) and Chen and Chen (2001) showed that the asymptotic distribution involves the supremum of a Gaussian process. See also Liu and Shao (2004) in normal mixture models. However, there are several shortfalls of the above results. Firstly, the results lose their appeal because the supremum of a Gaussian process is difficult to compute (Chen et al., 2001). Secondly, the divergence to infinity is so slow that it is not detected in simulation (Liu and Shao, 2004). The convergence of the test statistic, normalized by $\log \log n$, to the extreme value distribution is hardly detectable (Liu et al., 2003). Lastly, Hall and Stewart (2005) provided a theoretical analysis on the reduction of power against alternative hypotheses regarding the above results.

In light of the peculiar behaviour of LRT, a resampling approach (McLachlan, 1987; McLachlan and Peel, 2000; McLachlan and Khan, 2004) can be carried out. However, when some of the regularity conditions are restored, especially consistency of the estimator, it is of great theoretical significance to further investigate the likelihood ratio.

The consistency of the MLE in the test for homogeneity has not been solved until the introduction of a clever penalized procedure proposed by Chen et al. (2001). The authors innovated the modified likelihood ratio test (MLRT) by incorporating a penalty function. The MLRT was also developed by Chen and Kalbfleisch (2005) in normal mixture models and further extended to an EM-test by Li et al. (2009) and Chen and Li (2009). Exact theoretical results on the asymptotic null distribution have been obtained in some special cases. For densities with a single parameter of interest, the

MLRT statistic has the limiting distribution $0.5\chi_0^2+0.5\chi_1^2$ (Chen et al., 2001; Li et al., 2009). For the normal mixture model, the statistic has χ_2^2 when the means and the variances are unequal and unknown (Chen and Li, 2009). Conceivably, the MLRT falls into the type II likelihood ratio problem (Lindsay, 1995, Section 4.4) which generates the chi-bar-square distributions of which some are parameter-dependent limiting null distributions. The above result in the normal mixture models returns to the χ_2^2 distribution due to loss of strong identifiability (Chen and Li, 2009, Example 1). Qin and Smith (2006) investigated an extension of the MLRT in multivariate normal mixture models. The authors showed the asymptotic null distribution being a mixture of distributions and suggested it must be found using numerical methods. For models with multidimensional parameters, Zhu and Zhang (2004) analysed the asymptotic properties of LRT and Niu et al. (2011) considered an EM test. Although the problem of estimator consistency has been solved in MLRT and the EM-test, in many other mixture models, such as the gamma mixture models, the results $0.5\chi_0^2 + 0.5\chi_1^2$ or χ_2^2 cannot be applied directly without theoretical justifications. The general testing problem has not been completely solved and remains as a long-standing open problem. Charnigo and Sun (2004) acknowledged the generalization of the MLRT to higher dimensional problems and suggested that the null distribution can be obtained by simulation. However, the extension is not at all straightforward as presented in this paper and simulation of the null distribution in the absence of a closed-form expression should no longer be tolerated. A clear guideline has been long overdue for practitioners in the rejection or retention of the homogeneity assumption. The purposes of the paper are to fill this research gap in gamma mixture models and to explore how the limiting null distribution depends on the parameters.

Motivated by the above needs and the importance of the gamma mixture models, this paper aims at investigating the limiting distribution of the MLRT statistic. We obtain the condition under which the MLRT statistic degenerates to zero and determine the proportion of degeneration. Then, we can show that the asymptotic null distribution has parameter-dependent chi-bar-square distributions. This subsequently establishes a foundation for quick model selection using the χ_2^2 distribution in practice. Moreover, in light of the popular Expectation-Maximization (EM) algorithm for parameter estimation in finite mixture models, we demonstrate through intensive simulation studies that our results can be applied to the likelihood ratio statistic evaluated at the MLE obtained via the EM algorithm.

The article is organized as follows. In Section 2, we present the asymptotic results. Section 3 lists a number of considerations in the applications of the results. The asymptotic analysis is supplemented by simulation in Section 4. Section 5 presents two data examples and Section 6 gives a conclusion.

2. Asymptotic Results

We consider a two-parameter gamma density function

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

where $\alpha > 1$ and $\beta > 0$ are shape and scale parameters, respectively. Given a set of independent and identically distributed data, we are interested in testing the homogeneity hypothesis H_0 against the alternative hypothesis of a two-component gamma mixture model H_1 where

$$\begin{aligned} H_0 : f(x) &= f(x;\alpha,\beta); \\ H_1 : f(x) &= \pi f(x;\alpha_1,\beta_1) + (1-\pi) f(x;\alpha_2,\beta_2), \end{aligned}$$

and $0 < \pi < 1$ is a mixing proportion. In this paper, we study a very general testing problem that the parameters under the hypotheses are all unknown and unequal. This is completely different from the setting in Liu et al. (2003). For parametric hypothesis testing problems it is customary to use the ordinary LRT based on the statistic which is defined as

$$LR_n = 2\left\{L\left(\hat{\pi},\hat{\alpha}_1,\hat{\beta}_1,\hat{\alpha}_2,\hat{\beta}_2\right) - L\left(0.5,\hat{\alpha},\hat{\beta},\hat{\alpha},\hat{\beta}\right)\right\},$$

where

$$L(\pi,\alpha_1,\beta_1,\alpha_2,\beta_2) = \sum_{i=1}^{n}\log\left\{\pi f(x_i;\alpha_1,\beta_1) + (1-\pi) f(x_i;\alpha_2,\beta_2)\right\} \quad (1)$$

is the log-likelihood function and $\left(\hat{\pi},\hat{\alpha},\hat{\alpha}_1,\hat{\alpha}_2,\hat{\beta},\hat{\beta}_1,\hat{\beta}_2\right)$ is the MLE of parameter $(\pi,\alpha,\alpha_1,\alpha_2,\beta,\beta_1,\beta_2)$. It is well known that the consistency of the MLE, obtained by maximizing (1) directly, is not guaranteed. See for example Ghosh and Sen (1985); Hathaway (1985); Chen and Chen (2001). This motivates a penalized procedure coined by Chen et al. (2001) based on a modified log-likelihood function

$$L^{p}(\pi,\alpha_1,\beta_1,\alpha_2,\beta_2) = L(\pi,\alpha_1,\beta_1,\alpha_2,\beta_2) + c\log\left\{4\pi(1-\pi)\right\}, \quad (2)$$

where c is a positive constant corresponding to the level of modification. Denote by $\left(\hat{\pi}^p, \hat{\alpha}_1^p, \hat{\alpha}_2^p, \hat{\beta}_1^p, \hat{\beta}_2^p\right)$ the penalized MLE of $(\pi, \alpha_1, \alpha_2, \beta_1, \beta_2)$ obtained by maximizing (2) given a suitable value of c. Adding a penalty function to the log-likelihood regains the consistency of the penalized MLE (Chen et al., 2008; Chen and Li, 2009). The MLRT statistic is

$$LR_n^p = 2\left\{L^p\left(\hat{\pi}^p, \hat{\alpha}_1^p, \hat{\beta}_1^p, \hat{\alpha}_2^p, \hat{\beta}_2^p\right) - L\left(0.5, \hat{\alpha}, \hat{\beta}, \hat{\alpha}, \hat{\beta}\right)\right\}. \tag{3}$$

We study the asymptotic distribution of LR_n^p which can be expressed as $LR_n^p = LR_{1n}^p - LR_{0n}$ in terms of the true parameter (α_0, β_0) under H_0, where

$$\begin{aligned} LR_{0n} &= 2\left\{L\left(0.5, \hat{\alpha}, \hat{\beta}, \hat{\alpha}, \hat{\beta}\right) - L\left(0.5, \alpha_0, \beta_0, \alpha_0, \beta_0\right)\right\}; \\ LR_{1n}^p &= 2\left\{L^p\left(\hat{\pi}^p, \hat{\alpha}_1^p, \hat{\beta}_1^p, \hat{\alpha}_2^p, \hat{\beta}_2^p\right) - L\left(0.5, \alpha_0, \beta_0, \alpha_0, \beta_0\right)\right\}. \end{aligned}$$

An immediate asymptotic approximation for LR_{0n} is

$$\left(n^{-1/2}\sum_{i=1}^n Y_i^T\right)\left(n^{-1}\sum_{i=1}^n Y_i Y_i^T\right)^{-1}\left(n^{-1/2}\sum_{i=1}^n Y_i\right) + o_p(1),$$

where Y_i is a random vector given by

$$Y_i = \left\{\begin{array}{c} -\Gamma^{(1)}(\alpha_0) + \log \beta_0 + \log X_i \\ \alpha_0 \beta_0^{-1} - X_i \end{array}\right\} \tag{4}$$

and $\Gamma^{(k)}(\alpha) = d^k \ln \Gamma(\alpha) / d\alpha^k$. In Appendix A, we derive the following asymptotic approximation for LR_{1n}^p

$$\begin{aligned} &\left(n^{-1/2}\sum_{i=1}^n Y_i^T\right)\left(n^{-1}\sum_{i=1}^n Y_i Y_i^T\right)^{-1}\left(n^{-1/2}\sum_{i=1}^n Y_i\right) \\ &\quad + \left(n^{-1/2}\sum_{i=1}^n W_i^T\right)\left(n^{-1}\sum_{i=1}^n W_i W_i^T\right)^{-1}\left(n^{-1/2}\sum_{i=1}^n W_i\right) + o_p(1) \end{aligned} \tag{5}$$

if $n^{-1/2}\sum_{i=1}^n U_i$ is non-negative-definite, where

$$W_i = \left(\sum_{i=1}^n Z_i \tilde{\gamma}_2 Y_i^T\right)\left(\sum_{i=1}^n Y_i Y_i^T\right)^{-1} Y_i - Z_i \tilde{\gamma}_2,$$

$\tilde{\gamma}_2 \neq 0$ is the solution to $\left(\sum_{i=1}^{n} W_i W_i^T\right)^{-1} \sum_{i=1}^{n} U_i = I_2$, I_2 is the two-dimensional identity matrix,

$$U_i = Z_i - V_i Z_i, \quad V_i = \left(\sum_{j=1}^{n} Y_j^T\right) \left(\sum_{j=1}^{n} Y_j^T Y_j\right)^{-1} Y_i, \tag{6}$$

and Z_i is a symmetric random matrix whose elements on the top left, top right and bottom right are, respectively

$$\begin{aligned} Z_{i[11]} &= -\Gamma^{(2)}(\alpha_0) + \left\{-\Gamma^{(1)}(\alpha_0) + \log\beta_0 + \log X_i\right\}^2; \\ Z_{i[12]} &= \beta_0^{-1} + \left\{-\Gamma^{(1)}(\alpha_0) + \beta_0 + \log X_i\right\}\left(\alpha_0\beta_0^{-1} - X_i\right); \\ Z_{i[22]} &= -\alpha_0\beta_0^{-2} + \left(\alpha_0\beta_0^{-1} - X_i\right)^2. \end{aligned} \tag{7}$$

The random quantity V_i is scalar. If $n^{-1/2}\sum_{i=1}^{n} U_i$ is negative-definite, W_i are taken to be zero resulting in $LR_n^p = o_p(1)$. Under H_0, Y_i and Z_i are random with mean zero. Then, by the central limit theorem , $n^{-1/2}\sum_{i=1}^{n} W_i$ converges to a bivariate normal random vector with mean zero. We summarize the result in Theorem 1.

Theorem 1. *Under H_0, the asymptotic distribution of LR_n^p degenerates to zero with a weight $0 < p < 1$ and has a χ_2^2 distribution with a weight $1 - p$ where p is the probability that the random matrix to which the matrix $n^{-1/2}\sum_{i=1}^{n} U_i$ converges as $n \to \infty$ is negative-definite and U_i is defined in (6). That is,*

$$LR_n^p \sim p\chi_0^2 + (1-p)\chi_2^2 \tag{8}$$

for large n, where χ_0^2 is a degenerate distribution with all its mass at zero and the notation $\sim$ means 'is distributed like'.

The limiting distribution in (8) is known as the chi-bar-square distributions (Johnson et al., 1994, pg. 454). A more precise expression for p will be derived in Section 3.2. Hence, the above result will be restated by (10) indicating clearly the dependency on the shape parameter.

3. Practical Considerations

3.1. Estimating p

From the definition of U_i in (6), we observe its dependence on the random vector Y_i and the random matrix Z_i given by (4) and (7), respectively, which

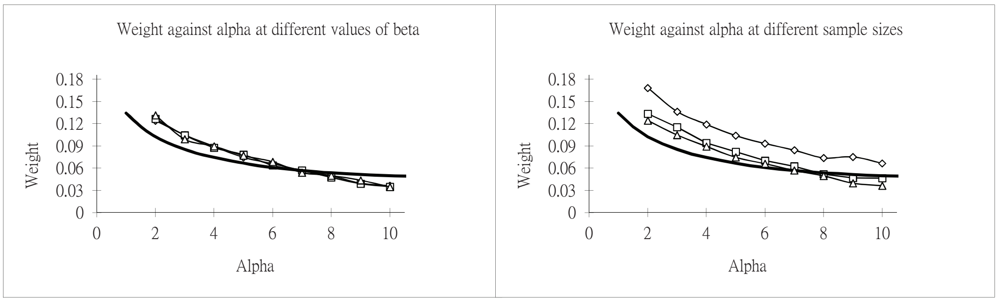

Figure 1: Plots of weight $p_s(\alpha_0, \beta_0, n)$ against α_0. The left panel shows three series of $\beta_0 = 2$ ($\diamond$), $\beta_0 = 6$ ($\square$) and $\beta_0 = 10$ ($\triangle$). The right panel shows three series of $n = 100$ ($\diamond$), $n = 500$ ($\square$) and $n = 1000$ ($\triangle$). The solid line depicts the asymptotic weight $p(\alpha_0)$.

are related to the parameter (α_0, β_0) under H_0. In addition, the estimate of p may also depend on n as the random matrix concerned involves a summation over n random matrices. As a rough visualization of the relations between these variables, we simulate n random variables from $f(x; \alpha_0, \beta_0)$, compute Y_i, Z_i and U_i, and evaluate the proportions in 10000 replications that $n^{-1/2} \sum_{i=1}^{n} U_i$ is negative-definite. Denote by $p_s(\alpha_0, \beta_0, n)$ such a proportion. Fig. 1 displays two plots of $p_s(\alpha_0, \beta_0, n)$ at some selected values of α_0, β_0 and n. The left panel shows three series of $p_s(\alpha_0, \beta_0, n)$ against α_0 at $n = 1000$, each series corresponding to different values of β_0. There is a decreasing trend of $p_s(\alpha_0, \beta_0, n)$ as α_0 increases, this trend being invariant in β_0. The right panel shows another three series of $p_s(\alpha_0, \beta_0, n)$ against α_0 at $\beta_0 = 2$, each series corresponding to different values of n. A similar decreasing trend of $p_s(\alpha_0, \beta_0, n)$ against α_0 is observed. In addition, the values of $p_s(\alpha_0, \beta_0, n)$ get lower at larger sample sizes and seem to converge to some certain level as n grows. Overall, $p_s(\alpha_0, \beta_0, n)$ seems to decreases as α_0 increases, but remains constant as β_0 varies. Its possible convergence as n tends to infinity motivates further investigation. Last, it is worth pointing out some merits of the simulation technique. Apart from quick and easy construction of the weight estimate, its use in the construction of a lower bound in a finite samples will be outlined in Section 4.

3.2. Asymptotic p

We require some general conditions on Y_i, Z_i and the products of their elements. In particular,

$$n^{-1}\sum_{i=1}^{n} Y_i Y_i^T \to M, \quad n^{-1}\sum_{i=1}^{n} Y_{i[j]} Z_i \to \upsilon_j$$

in probability for $j = 1, 2$, where $Y_{i[j]}$ denotes the jth element of vector Y_i. The expression of each of the elements in matrices M and υ are given in Appendix B. Denote by $U_{i[11]}$, $U_{i[12]}$ and $U_{i[22]}$, respectively the elements on the top left, top right and bottom right of U_i. By the central limit theorem, the vector on the left-hand side below

$$n^{-1/2}\begin{pmatrix} \sum_{i=1}^{n} U_{i[11]} \\ \sum_{i=1}^{n} U_{i[12]} \\ \sum_{i=1}^{n} U_{i[22]} \end{pmatrix} \to N_3\left\{0, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}\right\}$$

converges, as $n \to \infty$, to a random vector denoted by $S^T = (S_1, S_2, S_3)$ having a trivariate normal distribution N_3 with zero mean vector and covariance matrix whose elements are

$$\begin{aligned}
\sigma_{11} &= 2\left\{\Gamma^{(2)}(\alpha_0)\right\}^2 + \Gamma^{(4)}(\alpha_0) + \alpha_0\left\{\Gamma^{(3)}(\alpha_0)\right\}^2\left\{1 - \alpha_0\Gamma^{(2)}(\alpha_0)\right\}^{-1}; \\
\sigma_{12} &= \left[-2\Gamma^{(2)}(\alpha_0) + \Gamma^{(3)}(\alpha_0)\left\{1 - \alpha_0\Gamma^{(2)}(\alpha_0)\right\}^{-1}\right]\beta_0^{-1}; \\
\sigma_{13} &= \left[2 + \alpha_0\Gamma^{(3)}(\alpha_0)\left\{-1 + \alpha_0\Gamma^{(2)}(\alpha_0)\right\}^{-1}\right]\beta_0^{-2}; \\
\sigma_{22} &= \left[-1 + \Gamma^{(2)}(\alpha_0)\left\{-1 + \alpha_0^2\Gamma^{(2)}(\alpha_0)\right\}\right]\left\{-1 + \alpha_0\Gamma^{(2)}(\alpha_0)\right\}^{-1}\beta_0^{-2}; \\
\sigma_{23} &= \left[-2\alpha_0 + \left\{-1 + \alpha_0\Gamma^{(2)}(\alpha_0)\right\}^{-1}\right]\beta_0^{-3}; \\
\sigma_{33} &= \alpha_0\left[2 + 2\alpha_0 + \left\{1 - \alpha_0\Gamma^{(2)}(\alpha_0)\right\}^{-1}\right]\beta_0^{-4}.
\end{aligned}$$

Hence, the negative-definiteness condition implies that p can be obtained by the following probability

$$P\left(\{S_1 < 0\} \cap \left\{S_1 S_3 - S_2^2 > 0\right\}\right) = \int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{-\infty}^{s_2^2 s_1^{-1}} g(s_1, s_2, s_3)\, ds_3 ds_2 ds_1, \tag{9}$$

where $g(s_1, s_2, s_3)$ is the density function of the above trivariate normal distribution. The probability can be easily evaluated by numerical integration using, for example, Wolfram Mathematica®. It is important to observe that the integral is independent of β_0. This can be easily verified by simple transformation in the integration. Hence, the probability may precisely be denoted by $p(\alpha_0)$ and the result in (8) is more appropriately written as

$$LR_n^p \sim p(\alpha_0)\,\chi_0^2 + \{1 - p(\alpha_0)\}\,\chi_2^2. \tag{10}$$

A more precise description of the relation of $p(\alpha_0)$ as α_0 varies can be obtained by (9). Fig. 1 overlays a curve of $p(\alpha_0)$ against α_0 for $1 \le \alpha_0 \le 10$. Its match with the simulated weight $p_s(\alpha_0, \beta_0, n)$ suggests that the mysterious dependency of the proportion of degeneration on the shape parameter is further illustrated. An astounding observation is that $p(\alpha_0)$ is small at some large values of α_0. An example $p(12)$=0.0497 suggests that the χ_2^2 distribution is quite accurate to approximate the asymptotic null distribution. In addition, as an empirical rule of thumb, we may use that $p(\alpha_0) \le p(1)$ =0.1345 and $p(\alpha_0) > 0$ to develop a lower bound D_L and an upper bound D_U for the statistic LR_n^p

$$D_L \le LR_n^p \le D_U, \quad D_L \sim 0.1345\chi_0^2 + 0.8655\chi_2^2, \quad D_U \sim \chi_2^2 \tag{11}$$

as a quick guideline. Given a significance level, H_0 is retained if the observed MLRT statistic falls below the critical value evaluated by the above lower bound, and is rejected if it is above the critical value based on the upper bound.

3.3. Weighted Average Procedure

The previous subsections demonstrate the theoretical analysis to the testing problem. However, practical implementation of (10) encounters a drawback in that the value α_0 is unknown. A possible solution is to substitute this value by the parameter estimate, for instance, the maximum likelihood estimate $\hat{\alpha}$ under H_0. Then, the weight $p(\alpha_0)$ is estimated by $\hat{\alpha}$ through (9) and the asymptotic null distribution is established as in (10). However, the substitution may suffer a certain degree of bias because all prior beliefs are placed on $\hat{\alpha}$. Lindsay (1995) suggested the use of the least favourable null distribution by employing the least favourable critical value within a confidence interval for α_0. However, the problem remains unsolved if the observed test statistic falls below this least favourable critical value.

In light of the above difficulties, we propose a weighted average procedure to accommodate the estimation error. It is well-known that $\hat{\alpha} - \alpha_0$ is asymptotically normal with mean zero and variance $v(\alpha_0) = n^{-1}\alpha_0\left\{-1 + \alpha_0\Gamma^{(2)}(\alpha_0)\right\}^{-1}$.

Table 1: Weight p against α using weighted average procedure $p_w(\alpha, r)$ and direct substitution $p(\alpha)$. Different numbers of candidates r and the effect of sample size n for the weighted average procedure are shown.

α	$p_w(\alpha, 10)$		$p_w(\alpha, 20)$		$p(\alpha)$
	$n = 100$	$n = 1000$	$n = 100$	$n = 1000$	
2	0.1023	0.1023	0.1023	0.1023	0.1023
4	0.0744	0.0743	0.0744	0.0743	0.0743
6	0.0612	0.0611	0.0613	0.0611	0.0608
8	0.0540	0.0536	0.0541	0.0536	0.0536
10	0.0509	0.0497	0.0512	0.0497	0.0496

Then, r candidates of $\alpha_{0,k}$ can be obtained from the normal distribution through

$$\frac{k}{r+1} = \int_{-\infty}^{\alpha_{0,k}} \{2\pi v(\hat{\alpha})\}^{-\frac{1}{2}} \exp\left\{-\frac{(x-\hat{\alpha})^2}{2v(\hat{\alpha})}\right\} dx, \quad k = 1, \ldots, r$$

provided that $\alpha_{0,k} \geq 1$. Each of these $\alpha_{0,k}$ forms an asymptotic null distribution given by (8). The assignment of an equal weight to each $\alpha_{0,k}$ leads to the asymptotic null distribution

$$p_w(\hat{\alpha}, r)\chi_0^2 + \{1 - p_w(\hat{\alpha}, r)\}\chi_2^2, \quad p_w(\hat{\alpha}, r) = \frac{1}{r}\sum_{k=1}^{r} p(\alpha_{0,k}).$$

As illustrated in Fig. 1 the convexity of the weight in the shape parameter, the weighted average procedure will give a weight slightly larger than the direct substitution does. The effect of this finite-sample refinement is illustrated in Table 1. The weights $p_w(\hat{\alpha}, r)$ using $n = 100$ are slightly larger than those using $n = 1000$ which are very close to the value obtained by direct substitution $p(\hat{\alpha})$. Hence, this procedure tends to yield a smaller p-value than the method of direct substitution leading to a conclusion which is less favourable to the null hypothesis when information from the sample is scarce. Moreover, the input r seems less important compared to the sample size. We shall fix $r = 10$ in data analysis in Section 5.2.

3.4. MLE *Obtained via* EM *Algorithm*

Mixture models are getting popular in the statistics literature because of its wide range of applications, including examination of homogeneity of populations, assessment of unimodality and identifications of clusters or outliers.

The introduction of the EM algorithm has further pushed up its popularity. Frühwirth-Schnatter (2006) commented that the EM algorithm is the most common method for parameter estimation in finite mixture models nowadays. However, the penalized procedure will not be considered for parameter estimation of a mixture model, except only when a test of homogeneity is conducted. Given the homogeneous and mixture models, the latter one is fitted to the data by the EM algorithm. The goodness-of-fit may be justified by comparing the values of the log-likelihood. Hence, the EM algorithm has retained the convenience of the ordinary LRT. We will investigate through simulation whether the goodness-of-fit justification is appropriate and under what circumstances it can be applied. Another problem inherited in the MLRT is the possible reduction of power under H_1. In the twelve cases under study in the simulation, the power is not seriously affected but the reduction in log-likelihood value due to the penalty function should not be overlooked. In light of this, we may use the following conventional likelihood ratio as an alternative statistic

$$LR_{itr,n}^{EM} = 2\left\{L\left(\hat{\pi}^{EM}, \hat{\alpha}_1^{EM}, \hat{\beta}_1^{EM}, \hat{\alpha}_2^{EM}, \hat{\beta}_2^{EM}\right) - L\left(0.5, \hat{\alpha}, \hat{\beta}, \hat{\alpha}, \hat{\beta}\right)\right\}, \quad (12)$$

where $\left(\hat{\pi}^{EM}, \hat{\alpha}_1^{EM}, \hat{\beta}_1^{EM}, \hat{\alpha}_2^{EM}, \hat{\beta}_2^{EM}\right)$ is the MLE of $(\pi, \alpha_1, \beta_1, \alpha_2, \beta_2)$ obtained via the EM algorithm and itr is the number of EM algorithm iterations given a suitable initial guess. This statistic not only preserves the convenience as the ordinary LRT does, but part of it is also very common in the formation of AIC and BIC in mixture model selection.

The limiting distribution of $LR_{itr,n}^{EM}$ will be given after a brief discussion of the asymptotic characteristics of the EM estimators. In the rest of this subsection, we assume without loss of generality that $\pi \geq 0.5$. The argument in Chen and Chen (2001) points to the problem that in the ordinary LRT under H_0, the products $\left(1 - \hat{\pi}^{EM}\right)\hat{\alpha}_2^{EM}$ and $\left(1 - \hat{\pi}^{EM}\right)\hat{\beta}_2^{EM}$ are consistent but not $\hat{\alpha}_2^{EM}$ and $\hat{\beta}_2^{EM}$. The EM algorithm suffers a similar problem except that it can never reach the boundary point of π and that the iterations will either slowly merge $\hat{\alpha}_1^{EM}$ with $\hat{\alpha}_2^{EM}$ and $\hat{\beta}_1^{EM}$ with $\hat{\beta}_2^{EM}$, or force $\hat{\pi}^{EM}$ towards one (Lindsay, 1995, Section 3.4). Denote by E_I the former event that individual estimators are consistent and by E_{II} the latter event that individual estimators are not consistent. The advantage of $LR_{itr,n}^{EM}$ is on the extremely slow convergence of the EM algorithm under H_0. The occurrence of E_I or E_{II} can be easily observed as the iterations proceed. If E_{II} is observed, we may retain H_0 in the absence of a tolerable significance level;

otherwise, large values of $LR_{itr,n}^{EM}$ may suggest rejection of H_0 according to (10) conditional on E_I. Precisely,

$$LR_{itr,n}^{EM} \mid E_I \sim p(\alpha_0)\,\chi_0^2 + \{1 - p(\alpha_0)\}\,\chi_2^2 \tag{13}$$

as $n \to \infty$. The number of iterations itr may be determined based on some stopping rules as outlined in Lindsay (1995). Our simulation results suggest that when a suitable initial guess, such as the penalized maximum likelihood estimate, is adopted, the increase in the likelihood function is not significant as the iterative run proceeds. Hence, practitioners may pick a number of $LR_{itr,n}^{EM}$ values after the EM algorithm has changed insignificantly and conclude whether to retain or to reject H_0 if these values yield consistent results. The above suggestions essentially preserve the convenience in the use of likelihood-ratio-type tests and avoid power deterioration in applications. The arguments and suggestions in this subsection will be supplemented by the material lifetimes example in Section 5.2.

4. Simulation

We have conducted an extensive simulation study to evaluate the accuracy of the results. Due to the dependency of $p(\alpha_0)$ on α_0, it is interesting to conduct simulations using different values of α_0 and holding $\beta_0 = 1$ with a number of sample sizes. The first statistic under study is LR_n^p given by (3). It is the MLRT statistic with $c = \log 50$ in the penalty function in accordance with the recommendations in Chen et al. (2001). The second statistic $LR_{itr,n}^{EM}$ given by (12) uses the likelihood ratio evaluated at the MLE obtained via the EM algorithm. The extremely slow convergence in the EM algorithm makes simulation studies tedious. Lindsay (1995) pointed out that the solution of the likelihood equations can depend greatly on the initial values. Therefore, we use the penalized MLEs as initial guesses and carry out ten iterations.

We report the empirical sizes obtained from 10000 replications. Two sets of simulations are illustrated in Table 2. Other sets using different values of α_0 share similar results and hence are not reported. The agreement between the theoretical results and the simulation results is obvious. Improvements are generally obtained when we increase the sample size. The simulation also shows the dependency of $p(\alpha_0)$ on α_0. In Table 3, we report the proportions of zero statistics obtained from the simulation and the weight $p_s(\alpha_0, \beta_0, n)$ obtained from simulation in Section 3.1, and the asymptotic weight $p(\alpha_0)$ is

Table 2: Simulation results at selected nominal levels of 0.1, 0.05 and 0.01 using two homogeneous gamma models.

	Empirical significance levels for $H_0 : f\,(x; 2, 1)$					
	0.10		0.05		0.01	
n	LR_n^p	$LR_{10,n}^{EM}$	LR_n^p	$LR_{10,n}^{EM}$	LR_n^p	$LR_{10,n}^{EM}$
100	0.1144	0.1285	0.0598	0.0701	0.0138	0.0176
200	0.1057	0.1117	0.0524	0.0562	0.0112	0.0135
500	0.1025	0.1034	0.0559	0.0569	0.0111	0.0114
1000	0.0972	0.0974	0.0471	0.0476	0.0103	0.0104
	Empirical significance levels for $H_0 : f\,(x; 8, 1)$					
	0.10		0.05		0.01	
n	LR_n^p	$LR_{10,n}^{EM}$	LR_n^p	$LR_{10,n}^{EM}$	LR_n^p	$LR_{10,n}^{EM}$
100	0.1073	0.1207	0.0563	0.0662	0.0121	0.0158
200	0.1044	0.1114	0.0560	0.0600	0.0143	0.0161
500	0.1053	0.1064	0.0532	0.0540	0.0093	0.0097
1000	0.1047	0.1052	0.0535	0.0541	0.0106	0.0107

in the caption. First, it is interesting that both statistics LR_n^p and $LR_{10,n}^{EM}$ result in the same figures. This implies that the EM algorithm no longer increases the likelihood value under the occurrence of degeneration. Second, it is obvious that the asymptotic analysis leading to $p\,(\alpha_0)$ agrees quite well with the simulation results of LR_n^p and $LR_{10,n}^{EM}$. This consistently justifies one of the main results of this paper that the degeneration arises from the negative-definiteness of the random matrix. The relatively weak approximation in the sample size of 100 can be explained by the relatively weak second-order approximation given by (5). Lastly, the value $p_s\,(\alpha_0, \beta_0, n)$ is the largest when the sample size is less than 1000. We may replace the lower bound given by (11) by $p_s\,(\alpha_0, \beta_0, n)$ if being smaller as a more prudent benchmark in a finite-sample situation.

Some insights on the power of the tests can be gained. We consider a number of gamma mixture models which are either entirely different in mixing proportion, shape and scale parameters or with some of these parameters being equal. Each of the following alternative hypotheses is formulated to

Table 3: Proportions of zero statistics and simulated weights $p_s(\alpha_0, \beta_0, n)$ using four homogeneous gamma models. The asymptotic weights for the four cases are $p(2)$ =0.1023, $p(6)$ =0.0576, $p(8)$=0.0536 and $p(12)$=0.0497.

	$H_0 : f(x; 2, 1)$			$H_0 : f(x; 6, 1)$		
n	LR_n^p	$LR_{10,n}^{EM}$	$p_s(2, 1, n)$	LR_n^p	$LR_{10,n}^{EM}$	$p_s(6, 1, n)$
100	0.1195	0.1195	0.1636	0.0516	0.0516	0.0925
200	0.1258	0.1258	0.1596	0.0484	0.0484	0.0839
500	0.1206	0.1206	0.1382	0.0542	0.0542	0.0763
1000	0.1225	0.1225	0.1301	0.0545	0.0545	0.0649
	$H_0 : f(x; 8, 1)$			$H_0 : f(x; 12, 1)$		
n	LR_n^p	$LR_{10,n}^{EM}$	$p_s(8, 1, n)$	LR_n^p	$LR_{10,n}^{EM}$	$p_s(12, 1, n)$
100	0.0465	0.0465	0.0731	0.0349	0.0349	0.0578
200	0.0470	0.0470	0.0674	0.0324	0.0324	0.0439
500	0.0384	0.0384	0.0592	0.0336	0.0336	0.0357
1000	0.0444	0.0444	0.0473	0.0297	0.0297	0.0259

test against H_0

$$
\begin{array}{ll}
H_{101} : 0.2f(x; 8, 1) + 0.8f(x; 2, 4); & H_{102} : 0.8f(x; 8, 1) + 0.2f(x; 2, 4); \\
H_{103} : 0.5f(x; 8, 1) + 0.5f(x; 2, 4); & H_{104} : 0.5f(x; 8, 4) + 0.5f(x; 2, 1); \\
H_{105} : 0.2f(x; 8, 1) + 0.8f(x; 2, 1); & H_{106} : 0.2f(x; 8, 4) + 0.8f(x; 2, 4); \\
H_{107} : 0.2f(x; 8, 1) + 0.8f(x; 8, 4); & H_{108} : 0.2f(x; 2, 1) + 0.8f(x; 2, 4); \\
H_{109} : 0.5f(x; 8, 1) + 0.5f(x; 8, 4); & H_{110} : 0.5f(x; 2, 1) + 0.5f(x; 2, 4); \\
H_{111} : 0.5f(x; 8, 1) + 0.5f(x; 2, 1); & H_{112} : 0.5f(x; 8, 4) + 0.5f(x; 2, 4).
\end{array}
$$

Every simulation experiment consists of 10000 replications, each of sample size 1000. We find that the upper bound χ_2^2 given by (11) is extremely efficient in the testing process. Almost all simulated test statistics of LR_n^p and $LR_{10,n}^{EM}$ are greater than the critical values of the χ_2^2 distribution. The powers are all equal to one at significance levels 0.1, 0.05 and 0.01 except for the test of H_{104}. In this particular case, the statistic LR_n^p gives powers of 0.9999, 0.9997 and 0.9985 at the corresponding significance levels, whereas $LR_{itr,n}^{EM}$ yields powers of 0.9999, 0.9997 and 0.9987. Both methods seem to be equally powerful. However, we should point out that the EM algorithm increases the likelihood value at each cycle in the iterative sequence (Dempster et al., 1977). Meanwhile, the penalty function in the MLRT may reduce the log-likelihood. In Table 4, we report the average test statistics in 10000 replications for each of the alternative hypotheses. That the averages of $LR_{10,n}^{EM}$ are always higher

Table 4: Average values of test statistics under alternative hypotheses.

	H_{101}	H_{102}	H_{103}	H_{104}	H_{105}	H_{106}
LR_n^p	576.1	487.5	780.4	43.54	61.82	61.82
$LR_{10,n}^{EM}$	884.3	783.7	781.2	43.79	76.23	76.21
	H_{107}	H_{108}	H_{109}	H_{110}	H_{111}	H_{112}
LR_n^p	470.8	129.0	341.4	62.63	124.9	61.82
$LR_{10,n}^{EM}$	558.5	140.8	342.2	62.72	125.6	74.64

implies that the use of $LR_{10,n}^{EM}$ may achieve higher power given some extremely small significance level or in some special cases that have not been considered in the simulation.

5. Data Example

5.1. Danish Fire Loss

This example is based on the Danish fire loss data set which consists of 2157 losses exceeding one million Danish Krone from the years 1980 to 1990 inclusive. It is well known that the data set has a heavy right tail in the extreme value literature (Embrechts et al., 1997). The adequacy of the homogeneous gamma model which has a moderate tail is suspected. We apply the proposed results and methods to see if a two-component gamma mixture model will improve the fitting with further verifications, justified by some goodness-of-fit measures. McNeil (1997) provided a time series plot to check for clustering of large losses and a sample mean excess function to determine heavy-tailed behaviour. The results suggest the validity of the independence assumption and the possibility in modelling excesses over high thresholds using the generalized Pareto distribution. Recently, Wong and Li (2010) proposed a threshold model incorporating the generalized Pareto distribution for excesses and a Weibull distribution for the rest of the observations. This threshold model flexibly gives a global fit and an appropriate tail modelling. These two findings suggest that the loss data are likely to be independent but from a heterogeneous population.

The maximum likelihood estimate of a gamma model is $\left(\hat{\alpha}, \hat{\beta}\right) = (1.299, 0.382)$ with a corresponding maximized log-likelihood of -4752. In the gamma mix-

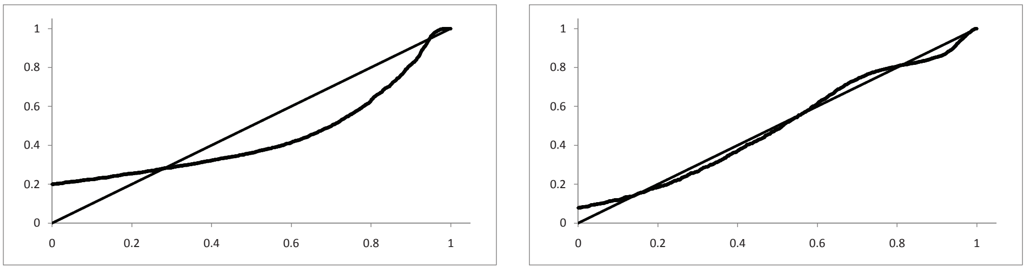

Figure 2: The probability-probability plots. The left panel shows the plot of the fitted gamma model and the right panel shows that of the fitted gamma mixture model using the EM algorithm. A 45-degree straight line is given for reference.

ture model, the penalized procedure and the EM algorithm give

$$\begin{aligned}\left(\hat{\pi}^p, \hat{\alpha}_1^p, \hat{\beta}_1^p, \hat{\alpha}_2^p, \hat{\beta}_2^p\right) &= (0.5098,15.68,10.09,1.226,0.2040)\,; \\ \left(\hat{\pi}^{EM}, \hat{\alpha}_1^{EM}, \hat{\beta}_1^{EM}, \hat{\alpha}_2^{EM}, \hat{\beta}_2^{EM}\right) &= (0.2816,1.256,0.1619,10.19,6.036)\end{aligned}$$

which yield the values of the test statistic of 1829 and 1978, respectively. The evidence for the mixture model is overwhelming as both statistics are far greater than the critical values of the χ_2^2 distribution at any reasonable significance level. Further support for this is a goodness-of-fit assessment based on probability-probability plots as shown in Fig. 2. The gamma mixture model provides a much better fit as the plot exhibits obviously a straight line pattern. This example lends further support to the asymptotic distribution in (10), improvement in power through the use of $LR_{itr,n}^{EM}$, and demonstrates its simplicity in implementation.

5.2. Material Lifetimes

Gamma distributions give useful representations of a number of physical situations such as random processes in time. We consider a set of 101 observations for the lifetime of an aluminium sheet under maximum stress of 21,000 psi. A brief description and the data listed in increasing order are available in Birnbaum and Saunders (1958). The authors demonstrated a realistic adjustment to exponential models in representing lifetimes in a life-testing situation. Therefore, it is interesting to check the redundancy of a mixture structure in the representation. More insight may be gained by applying our results in studying the data set.

Table 5: Estimation results of the EM algorithm for the material lifetimes data. The p-values are calculated using (13) based on (a) direct substitution and (b) weighted average procedure.

itr	Parameter estimate $\left(\hat{\pi}^{EM},\hat{\alpha}_1^{EM},\hat{\beta}_1^{EM},\hat{\alpha}_2^{EM},\hat{\beta}_2^{EM}\right)$	$LR_{itr,n}^{EM}$	p-value (a)	p-value (b)
10	(0.5048,22.73,0.01452,10.25,0.008307)	2.389	0.2879	0.2872
50	(0.5222,22.20,0.01425,10.15,0.008262)	2.397	0.2867	0.2860
100	(0.5446,21.53,0.01389,9.999,0.008175)	2.408	0.2851	0.2845

The parameter estimate of a gamma model is $\left(\hat{\alpha},\hat{\beta}\right)=(11.86,0.008462)$. In the gamma mixture model, the penalized maximum likelihood estimate of

$$\left(\hat{\pi}^p,\hat{\alpha}_1^p,\hat{\beta}_1^p,\hat{\alpha}_2^p,\hat{\beta}_2^p\right) = (0.5001,\ 22.82,\ 0.01457,\ 10.26,\ 0.008307)$$

leads to a value of the MLRT statistic of 2.387. A lower bound of the p-value evaluated through (11) is 0.2624, larger than any reasonable size of a statistical test. In the absence of α_0 under H_0, direct substitution and the weighted average procedure yields $p\left(\hat{\alpha}\right)$=0.0495 and $p_w\left(\hat{\alpha},10\right)$=0.0518, respectively. The corresponding p-values are 0.2882 and 0.2875. The use of the penalized estimates initiates the EM iterative sequence. A series of 1000 iteration steps seems to indicate the occurrence of event E_{II} that individual estimators are not consistent as shown in the left panel of Fig. 3 in which $\hat{\pi}^{EM}$ increases slowly to one as the iteration moves on. Evidence in favour of H_0 is obvious. On the other hand, the behaviour of $LR_{itr,n}^{EM}$ is agonizing, in particular as $\hat{\pi}^{EM}$ is closer to one that a jump in the test statistic is observed in the right panel of Fig. 3. The AIC criterion starts to reject H_0 in the 833th iteration whereas the BIC criterion and the statistic $LR_{itr,n}^{EM}$ at the 5% significance level consistently suggest retention of H_0 in all 1000 iterations. However, the insignificant increase of the statistic in the first 600 iterations suggests early termination of the EM algorithm. Therefore, the suggestions in Section 3.4 are useful. We can consider a number of the test statistics in different iterative steps and apply (13). The results reported in Table 5 consistently suggest the retention of H_0 in agreement with the method of MLRT.

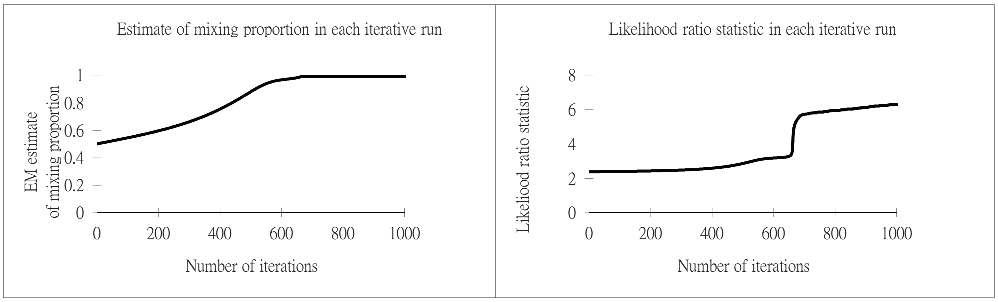

Figure 3: The EM estimate of the mixing proportion (left) and the statistic $LR^{EM}_{itr,n}$ (right) in each iterative run.

6. Conclusion

We investigate the modified likelihood ratio test for homogeneity in two-component gamma mixture models. We have found that the limiting distribution of the test statistic is the parameter-dependent chi-bar-square distributions given by a degeneration to zero with weight $p(\alpha_0)$ and a chi-square distribution with two degrees of freedom with weight $1 - p(\alpha_0)$. This weight is related to the negative-definiteness of a complicated random matrix dependent on the shape parameter of the homogeneous gamma model. An asymptotic approximation using a trivariate normal probability has been developed for $p(\alpha_0)$. All these theoretical results have been revealed through an extensive simulation to be very accurate and reliable.

In applications, the shape parameter is unknown. Based on the behaviour of $p(\alpha_0)$, we have developed a lower bound for the retention of the homogeneous hypothesis and an upper bound for the rejection. The bounds have been proved to be extremely useful in simulation and in two real examples. On the rare occasion that the observed test statistic falls between the bounds or if practitioners require an evaluation of the p-value, we recommend the weighted average procedure which takes into account the estimation error of the shape parameter. This procedure has yielded consistent results in a study of the material lifetimes data.

Due to the popularity of the EM algorithm in the analysis of mixture models, we recommend that the likelihood ratio test statistic be evaluated at the maximum likelihood estimates obtained via the EM algorithm. There are some appealing advantages including the preservation of the convenience of the conventional likelihood ratio test procedure and in the prevention of power reduction. The fact that the EM iterative sequence converges slowly allows the selection of a number of observed test statistics. A decision may therefore be based on these statistics by comparing the derived asymptotic null distribution conditional on the consistency of individual estimators. Its simplicity and convenience have been illustrated in the real examples.

A number of interesting insights have been obtained on the form of the asymptotic null distribution and on its practical implementation. We believe many other mixture models share similar characteristics and this deserves further research and discussion. In particular, the parameter-dependent structure of the limiting distribution may not be as simple as our situation in which only the shape parameter is involved. Developing simple decision criteria such as bounds appears to be very challenging.

Acknowledgement

Partial support by the Area of Excellence Scheme under the University Grants Committee of the Hong Kong Special Administration Region, China (Project AoE/P-04/2004 and GRF grant HKU7036/06P) is acknowledged. We thank Dr. G. Li of The University of Hong Kong, the Co-Editor, an Associate Editor and two reviewers for their constructive comments and suggestions that greatly led to the improvement of the paper.

Appendix A. Approximation for LR_{1n}^{p}

Define

$$Y_i = \frac{1}{f\left(X_i;\alpha_0,\beta_0\right)} \left\{ \begin{array}{c} \frac{\partial f(X_i;\alpha_0,\beta_0)}{\partial \alpha} \\ \frac{\partial f(X_i;\alpha_0,\beta_0)}{\partial \beta} \end{array} \right\}; \quad Z_i = \frac{1}{f\left(X_i;\alpha_0,\beta_0\right)} \left\{ \begin{array}{cc} \frac{\partial^2 f(X_i;\alpha_0,\beta_0)}{\partial \alpha^2} & \frac{\partial^2 f(X;\alpha_0,\beta_0)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 f(X_i;\alpha_0,\beta_0)}{\partial \alpha \partial \beta} & \frac{\partial^2 f(X;\alpha_0,\beta_0)}{\partial \beta^2} \end{array} \right\}$$

whose expressions for the gamma model are given by (4) and (7), respectively. The penalty function regains the consistency and efficiency of the penalized maximum likelihood estimators (Chen et al., 2008; Chen and Li, 2009). Following Section 2.3 of Chen et al. (2000), but in a bivariate context below, the resulting characterization of LR_{1n}^{p} involves the maximum of the following function

$$2\sum_{i=1}^{n}\delta_i - \sum_{i=1}^{n}\delta_i^2$$

plus $o_p(1)$, where

$$\begin{aligned}
\delta_i &= \pi \begin{pmatrix} \alpha_1 - \alpha_0 \\ \beta_1 - \beta_0 \end{pmatrix}^T Y_i + \pi \begin{pmatrix} \alpha_1 - \alpha_0 \\ \beta_1 - \beta_0 \end{pmatrix}^T Z_i \begin{pmatrix} \alpha_1 - \alpha_0 \\ \beta_1 - \beta_0 \end{pmatrix} \\
&\quad + (1-\pi) \begin{pmatrix} \alpha_2 - \alpha_0 \\ \beta_2 - \beta_0 \end{pmatrix}^T Y_i + (1-\pi) \begin{pmatrix} \alpha_2 - \alpha_0 \\ \beta_2 - \beta_0 \end{pmatrix}^T Z_i \begin{pmatrix} \alpha_2 - \alpha_0 \\ \beta_2 - \beta_0 \end{pmatrix}.
\end{aligned}$$

A re-parameterization using vector parameters γ_1 and γ_2, where

$$\gamma_1 = \left\{ \begin{array}{c} \pi(\alpha_1 - \alpha_0) + (1-\pi)(\alpha_2 - \alpha_0) \\ \pi(\beta_1 - \beta_0) + (1-\pi)(\beta_2 - \beta_0) \end{array} \right\}; \quad \gamma_2 = \{0.5\pi(1-\pi)\}^{\frac{1}{2}} \begin{pmatrix} \alpha_1 - \alpha_2 \\ \beta_1 - \beta_2 \end{pmatrix}$$

leads to

$$LR^p_{1n} = \max_{\gamma_1, \gamma_2} q(\gamma_1, \gamma_2) + o_p(1),$$

where

$$q(\gamma_1, \gamma_2) = 2\sum_{i=1}^n \left(\gamma_1^T Y_i + \frac{1}{2}\gamma_1^T Z_i \gamma_1 + \gamma_2^T Z_i \gamma_2 \right) - \sum_{i=1}^n \left(\gamma_1^T Y_i + \frac{1}{2}\gamma_1^T Z_i \gamma_1 + \gamma_2^T Z_i \gamma_2 \right)^2.$$

From Lemma 1 in Charnigo and Sun (2004) and Lemma A2 in Li et al. (2009) that $\hat{\gamma}_1^p = O_p\left(n^{-1/2}\right)$ and by the strong law of large numbers that $n^{-1}\sum_{i=1}^n Z_i = o_p(1)$. It follows that

$$LR^p_{1n} = \max_{\gamma_1, \gamma_2} q^*(\gamma_1, \gamma_2) + o_p(1),$$

where

$$q^*(\gamma_1, \gamma_2) = 2\sum_{i=1}^n \left(\gamma_1^T Y_i + \gamma_2^T Z_i \gamma_2 \right) - \sum_{i=1}^n \left(\gamma_1^T Y_i + \gamma_2^T Z_i \gamma_2 \right)^2.$$

The maximum value of $q^*(\gamma_1, \gamma_2)$ is (5) excluding the term $o_p(1)$.

Appendix B. Asymptotic weight $p(\alpha_0)$

Denote the matrices M and υ_j by

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}^{-1}; \quad \upsilon_j = \begin{pmatrix} \upsilon_{j11} & \upsilon_{j12} \\ \upsilon_{j12} & \upsilon_{j22} \end{pmatrix},$$

where

$$\begin{array}{lll}
m_{11} = \frac{\alpha_0}{-1+\alpha_0\Gamma^{(2)}(\alpha_0)}; & m_{12} = \frac{\beta_0}{-1+\alpha_0\Gamma^{(2)}(\alpha_0)}; & m_{22} = \frac{\beta_0^2\Gamma^{(2)}(\alpha_0)}{-1+\alpha_0\Gamma^{(2)}(\alpha_0)}; \\
\upsilon_{111} = \Gamma^{(3)}\left(\alpha_0\right); & \upsilon_{112} = 0; & \upsilon_{122} = \frac{1}{\beta_0^2}; \\
\upsilon_{211} = 0; & \upsilon_{212} = \frac{1}{\beta_0^2}; & \upsilon_{222} = -\frac{2\alpha_0}{\beta_0^3}.
\end{array}$$

Then, we express $U_{i[11]}$, $U_{i[12]}$ and $U_{i[22]}$ by

$$\begin{array}{l}
U_{i[11]} = Z_{i[11]} - Y_{i[1]}\left(m_{11}\upsilon_{111} + m_{12}\upsilon_{211}\right) - Y_{i[2]}\left(m_{12}\upsilon_{111} + m_{22}\upsilon_{211}\right); \\
U_{i[12]} = Z_{i[12]} - Y_{i[1]}\left(m_{11}\upsilon_{112} + m_{12}\upsilon_{212}\right) - Y_{i[2]}\left(m_{12}\upsilon_{112} + m_{22}\upsilon_{212}\right); \\
U_{i[22]} = Z_{i[22]} - Y_{i[1]}\left(m_{11}\upsilon_{122} + m_{12}\upsilon_{222}\right) - Y_{i[2]}\left(m_{12}\upsilon_{122} + m_{22}\upsilon_{222}\right).
\end{array}$$

The result in Section 3.2 follows from the central limit theorem and the covariance matrix is obtained by

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} = \left\{ \begin{array}{ccc} E\left(U_{1[11]}U_{1[11]}\right) & E\left(U_{1[11]}U_{1[12]}\right) & E\left(U_{1[11]}U_{1[22]}\right) \\ E\left(U_{1[11]}U_{1[12]}\right) & E\left(U_{1[12]}U_{1[12]}\right) & E\left(U_{1[12]}U_{1[22]}\right) \\ E\left(U_{1[11]}U_{1[22]}\right) & E\left(U_{1[12]}U_{1[22]}\right) & E\left(U_{1[22]}U_{1[22]}\right) \end{array} \right\}.$$

References

Birnbaum, Z. W., Saunders, S. C., 1958. A statistical model for life-length of materials. *Journal of the American Statistical Association* **53**, 151–160.

Charnigo, R., Sun, J., 2004. Testing homogeneity in a mxture distribution via L^2 distance between competing models. *Journal of the American Statistical Association* **466**, 488–498.

Chen, H., Chen, J., Kalbfleisch, J. D., 2000. A modified likelihood ratio test for homogeneity in the finite mixture models. Technical Report STAT 2000–01. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo.

Chen, H., Chen, J., 2001. The likelihood ratio test for homogeneity in finite mixture models. *The Canadian Journal of Statistics* **29**, 201–215.

Chen, H., Chen, J., Kalbfleisch, J. D., 2001. A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society: Series B* **63**, 19–29.

Chen, J., Tan, X., Zhang, R., 2008. Inference for normal mixtures in mean and variance. *Statistica Sinica* **18**, 443–465.

Chen, J., Li, P., 2009. Hypothesis test for normal mixture models: the EM Approach. *The Annals of Statistics* **37**, 2523–2542.

Chen, J., Kalbfleisch, J. D., 2005. Modified likelihood ratio test in finite mixture mdoels with a structural parameter. *Journal of Statistical Planning and Inference* **129**, 93–107.

Craig, B. M, Strassels, S. A., 2010. Out-of-pocket prices of opioid analgesics in the United States, 1999-2004. *Pain Medicine* **11**, 240–247.

Dacunha-Castelle D., Gassiat E., 1999. Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *The Annals of Statistics* **27**, 1178–1209.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society: Series B* **39**, 1–38.

Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. *Modelling Extreme Events.* Berlin: Springer-Verlag.

Frühwirth-Schnatter, S., 2006. *Finite Mixture and Markov Switching Models.* New York: Springer.

Ghosh, J. K., Sen, P. K., 1985. On the asymptotic performance of the log likelihood ratio statitsic for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (edited by L. M. Le Cam and R. A. Olshen), Volume 2, 789–806. Belmont: Wadsworth.

Hall, P., Stewart, M., 2005. Theoretical analysis of power in a two-component normal mixture model. *Journal of Statistical Planning and Inference* **134**, 158–179.

Hathaway, R. J., 1985. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics* **13**, 795–800.

JOHNSON, N. L., KOTZ, S., BALAKRISHNAN, N., 1994. *Continuous Univariate Distributions Volume 1.* New York: Wiley & Sons.

LI, P., CHEN, J., MARRIOTT, P., 2009. Non-finite Fisher information and homogeneity: an EM approach. *Biometrika* **96**, 411–426.

LINDSAY, B. G., 1995. *Mixture Models: Theory, Geometry and Applications.* Hayward, California: Institute of Mathematical Statistics.

LIU, X., PASARICA, C., SHAO, Y., 2003. Testing homogeneity in gamma mixture models. *Scandinavian Journal of Statistics* **30**, 227–239.

LIU, X., SHAO, Y., 2003. Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics* **31**, 807–832.

LIU, X., SHAO, Y., 2004. Asymptotics for the likelihood ratio test in a two-component normal mixture model. *Journal of Statistical Planning and Inference* **123**, 61–81.

MAYROSE, I., FRIEDMAN, N., PUPKO, T., 2005. A gamma mixture model better accounts for among site rate heterogeneity. *Bioinformatics* **21**, 151–158.

MCNEIL, A. J., 1997. Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin* **27**, 117–137.

MCLACHLAN, G. J., 1987. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics* **36**, 318–324.

MCLACHLAN, G. J. AND PEEL, D., 2000. *Finite Mixture Models.* New York: Wiley.

MCLACHLAN, G. J. AND KHAN, N., 2004. On a resampling approach for tests on the number of clusters with mixture model-based clustering of tissue samples. *Journal of Multivariate Analysis* **90**, 90–105.

NIU, X., LI, P., ZHANG P., 2011. Testing homogeneity in a multivariate mixture model. *The Canadian Journal of Statistics* **39**, 218–238.

QIN, Y. S., SMITH, B., 2006. The likelihood ratio test for homogeneity in bivariate normal mixtures. *Journal of Multivariate Analysis* **97**, 474-491.

WONG, T. S. T., LI, W. K., 2010. A threshold approach for peaks-over-threshold modeling using maximum product of spacings. *Statistica Sinica* **20**, 1257-1272.

ZHU, H. T., ZHANG, H., 2000. Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society: Series B* **66**, 3–16.