A comparison of recent procedures in Weibull mixture testing

Karl Mosler and Lars Haferkamp

Universität zu Köln
D-50923 Köln, Germany
(e-mail: mosler@statistik.uni-koeln.de)

Abstract. The paper considers recent approaches to testing homogeneity in a finite mixture model, the modified likelihood ratio test (MLRT) of Chen *et al.* (2001) and the D-tests of Charnigo and Sun (2004). They are adapted to Weibull mixtures with and without a Weibull-to-exponential transformation of the data. Critical quantiles are calculated by simulation. To cope with the dependency of quantiles on the unknown shape parameter, a corrected D-statistics is implemented and explored. First results are given on the power of these tests in comparison with that of the ADDS test by Mosler and Scheicher (2007).
Keywords: Mixture diagnosis, survival analysis, unobserved heterogeneity, overdispersion, goodness-of-fit.

1 Introduction

A practically important problem is to decide whether for given data a Weibull mixture specification should be preferred over a non-mixed Weibull model, that is, whether the data contain unobserved parameter heterogeneity or not. Various procedures have been proposed in the literature for this specification problem, among them graphical devices (Jiang and Murthy (1995)) and statistical tests (Mosler and Scheicher (2007)). For a comparison of these tests in exponential mixtures, see Mosler and Haferkamp (2007).

In this paper we consider three recent approaches to testing mixture homogeneity, the modified likelihood ratio test (MLRT) of Chen *et al.* (2001), the D-test of Charnigo and Sun (2004), and the ADDS-test. The ADDS test is due to Mosler and Seidel (2001) for exponential and Mosler and Scheicher (2007) for Weibull mixtures.

The subsequent discussion focuses on Weibull scale mixtures that have at most two components and common shape parameter, i.e., on densities

$$\begin{aligned} &f(x;\beta_1,\beta_2,\pi_1,\gamma) \\ &= \pi_1 \frac{\gamma}{\beta_1}\left(\frac{x}{\beta_1}\right)^{\gamma-1} e^{-\left(\frac{x}{\beta_1}\right)^{\gamma}} + (1-\pi_1)\frac{\gamma}{\beta_2}\left(\frac{x}{\beta_2}\right)^{\gamma-1} e^{-\left(\frac{x}{\beta_2}\right)^{\gamma}}, \end{aligned} \tag{1}$$

for $x \geq 0$ and parameters $\gamma, \beta_1, \beta_2 > 0$, $0 < \pi_1 < 1$. If $\gamma = 1$, an exponential mixture arises.

To test for mixture homogeneity, consider a random sample $X_1, \dots, X_n$ from (1). The alternative hypothesis corresponds to $\beta_1 \neq \beta_2$, while the null is signified by $\beta_0 = \beta_1 = \beta_2$ and an arbitrary $\pi_1 \in]0,1[$, say $\pi_1 = \frac{1}{2}$.

2 Three approaches for testing homogeneity

The MLRT of Chen *et al.* (2001) is a penalized likelihood ratio test. It is based on the usual log-likelihood plus a penalty term,

$$l^M(\beta_1, \beta_2, \pi_1, \gamma) = \sum_{i=1}^{n} \log f(\beta_1, \beta_2, \pi_1, \gamma) + C \log[4\pi_1(1-\pi_1)] \,. \qquad (2)$$

Here, $C > 0$ is a constant that weighs the penalty. Following previous authors (Charnigo and Sun (2004)) we use a fixed constant $C = \log 10$. In a first step, $l^M(\beta_1, \beta_2, \pi_1, \gamma)$ is maximized to obtain estimators $\hat{\beta}_1^M, \hat{\beta}_2^M, \hat{\pi}_1^M$, and $\hat{\gamma}^M$. Similarly, under H_0, $l^M(\beta_0, \beta_0, \frac{1}{2}, \gamma_0)$ is maximized to obtain estimators $\hat{\beta}_0^M, \hat{\gamma}_0^M$. In a second step, the test statistic

$$MLRT = 2\left[l^M(\hat{\beta}_1^M, \hat{\beta}_2^M, \hat{\pi}_1^M, \hat{\gamma}^M) - l^M(\hat{\beta}_0^M, \hat{\beta}_0^M, \frac{1}{2}, \hat{\gamma}_0^M)\right] \qquad (3)$$

is evaluated. Asymptotically, under H_0 and some regularity conditions, $MLRT$ is distributed as the fifty-fifty mixture of a χ_1^2 variable and a constant at 0 (Chen *et al.* (2001)). Note that in the special case of exponential mixtures, the MLRT statistic is (3) with both $\hat{\gamma}^M$ and $\hat{\gamma}_0^M$ substituted by the constant 1.

To solve the Weibull homogeneity problem, the MLRT can be used in two different ways, as described above and after a Wei2Exp transformation of the data. Either the four parameters under H_1 and the two under H_0 are estimated via penalized likelihood (2) and the statistic (3) is used as it stands. Or, alternatively, γ_0 is estimated under H_0, and the estimate $\hat{\gamma}_0$ is used to transform the data, $X_i \mapsto X_i^{\hat{\gamma}}$. Then, from the transformed data, the parameters β_1, β_2 and π_1 are estimated via penalized likelihood (2) with $\gamma = 1$ and the MLRT statistic is evaluated.

The D-test, in its original form, measures the area between two densities, one fitted under H_0, and the other fitted under H_1. Firstly, the parameters of the null distribution and the alternative distribution are estimated by some consistent estimators $\hat{\beta}_0, \hat{\gamma}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\pi}_1$, and $\hat{\gamma}$. Then the D-statistic

$$\begin{aligned} D &= \int_0^\infty \left[f(x; \hat{\beta}_1, \hat{\beta}_2, \hat{\pi}_1, \hat{\gamma}) - f(x; \hat{\beta}_0, \hat{\beta}_0, \frac{1}{2}, \hat{\gamma}_0)\right]^2 w(x)dx \qquad (4) \\ &= \sum_{i=0}^{2}\sum_{j=0}^{2} \hat{\pi}_i \; \hat{\pi}_j \frac{\hat{\gamma}_i \hat{\gamma}_j}{\hat{\beta}_i^{\hat{\gamma}_i} \hat{\beta}_j^{\hat{\gamma}_j}} \end{aligned}$$

$$\times \int_0^\infty x^{\hat{\gamma}_i - 1} x^{\hat{\gamma}_j - 1} \exp\left(- \left(\frac{x}{\hat{\beta}_i}\right)^{\hat{\gamma}_i} - \left(\frac{x}{\hat{\beta}_j}\right)^{\hat{\gamma}_j} \right) w(x) dx\,, \qquad (5)$$

where the notation $\hat{\pi}_0 = -1$ and $\hat{\pi}_2 = 1 - \hat{\pi}_1$ is used.

In the special case of an exponential mixture model, D is the same with the constant 1 in place of $\hat{\gamma}_0$ and $\hat{\gamma}_1$. In this case, as Charnigo and Sun (2004) show, D has an asymptotic null distribution which is equivariant to β_0. They provide tables of critical quantiles and report that, in the diagnosis of exponential scale mixtures, the simple D-test is slightly outperformed by the MLRT when n is small ($n \leq 100$). Charnigo and Sun (2004) therefore propose weighted forms of the D-test which put more weight to differences in the tails of the two densities: In place of the differential dx in the integral formula (5) they use $x\,dx$ or $x^2\,dx$. In the sequel, these weighted variants of the D-test will be signified by 'w1D' and 'w2D', respectively.

Like the MLRT, the D-test can be employed either to the original data from a Weibull mixture model or to transformed data that have been subject to a Wei2Exp transformation with an estimated shape parameter $\hat{\gamma}_0$. In Section 3 we will investigate the effect of estimating γ_0 on the critical regions of the D-tests and the MLRT.

The ADDS test combines a dispersion score (DS) test with a classical goodness-of-fit test. The DS statistic,

$$DS = \left(\frac{n(n-1)}{n+1}\right)^{\frac{1}{2}} \frac{1}{(\overline{X})^2} \left[S^2 - \frac{1}{2n} \sum_{i=1}^n T_i^2 \right], \qquad (6)$$

is combined with a goodness-of-fit statistic of Anderson-Darling type,

$$AD = \left(1 + \frac{0.6}{n}\right) \left(n - \frac{1}{n} \sum_{i=1}^n (2i-1) \left(\log\left(1 - e^{-T_{(i)}/\overline{X}}\right) + \frac{T_{(i)}}{\overline{X}} \right) \right), \qquad (7)$$

where $\overline{X}$ and S^2 denote the sample mean and variance and $T_{(i)}$ is the i-th order statistic. Reject H_0 if either DS or AD are too large.

Under H_0, both test statistics do not depend on the scale parameter β. Mosler and Seidel (2001) have demonstrated that the power of the ADDS test for exponential mixtures is always at least comparable to that of a bootstrap LRT, a moment LRT, and a DS test. On large classes of alternatives the tests are outperformed by the ADDS test.

3 Implementing MLRT and D-tests with Weibull alternatives

In order to avoid estimation of all Weibull parameters under H_0 and H_1, we first employ Wei2Exp forms of the MLRT and the D-test. That is, the data $x_1, \ldots, x_n$ are transformed to $x_1^{\hat{\gamma}}, \ldots, x_n^{\hat{\gamma}}$, and MLRT and D-tests for homogeneity in exponential mixtures are done with the transformed data. We simulated the quantiles of each test with estimated γ under H_0.

In our second approach we apply the D-tests and the MLRT directly to the data. Chen *et al.* (2001) and Charnigo and Sun (2004) report implementations of their tests for exponential mixtures[1]. However, when implementing the D-tests in a Weibull mixture model, severe difficulties arise. One difficulty is to choose a good estimation method for the parameters of a Weibull mixture distribution. The other is to find an approximate functional dependency of the relevant quantiles of the D-test statistic and the parameters of the null hypothesis to calculate proper critical values.

The statistic (5) was calculated by numerical integration[2]. Three weightings of the D-test were investigated, $w_1(t) = t$ and $w_2(t) = t^2$, and $w_g(t) = t^{\hat{\gamma}}$. In estimating the parameters under H_0 and H_1, we used the *Nelder-Mead simplex algorithm* (Olsson (1979))[3]. The Nelder-Mead simplex algorithm is included in *R* and works together with some methods of the *MASS* package (Venables and Ripley (2002)). We used a multiple initial value procedure to avoid being trapped at local maxima. The procedure was also applied with the MLRT to estimate the parameters of the penalized likelihood function.

After all we found from our simulations that the critical quantiles of the D-test depend heavily on shape parameter γ_0 under the null (see Figure 1). The same is partially true for the MLRT. The ADDS test, in comparison, shows no relevant dependency on γ_0; with increasing γ_0 it becomes only slightly more conservative. To cope with this observed dependency on γ_0 and with an additional combined dependency on scale β_0 and level α, we introduced corrected statistics as follows,

$$T^* = T \cdot \frac{h(\hat{\beta}_0)}{i(\hat{\gamma}_0, \alpha)} \,.$$

Here, $h(\beta_0)$ is a function that is specific to each test, and $i(\gamma_0, \alpha)$ is an interpolation function obtained from simulation (with $n = 1000$) of the $(1 - \alpha)$-quantiles of the statistic with different shape parameters γ_0. The value of $i(\gamma_0, \alpha)$ has been determined by linear interpolation of the simulated quantiles of the shape parameters $\gamma_0 \in \{1, 1.5, 2, 3, 5\}$. We chose $h(\beta)$ as a linear

[1] We thank these authors for kindly giving us their computer codes.

[2] The integration was done in the interval $[0.5x_{\min}, 2x_{\max}]$ by using the QUADPACK *R*- routines (adaptive quadrature of functions) from Piessens *et al.* (1983). x_min and x_max denote smallest and largest observations.

[3] We tried also the methods of Kaylan and Harris (1981) and Albert and Baxter (1995). The Nelder-Mead algorithm proved to be the fastest one; its likelihood comes near to that of the method of Kaylan and Harris. The *PAEM* of Albert and Baxter yielded a worse likelihood.

function:

$$h(\beta) = \begin{cases} \beta & \text{, D-test,} \\ 1 + 0.3(\beta - 1) & \text{, w1D,} \\ 1 & \text{, w2D,} \\ 1 + 0.3(\beta - 1) & \text{, wgD,} \\ 1 + 0.2(\beta - 1) & \text{, MLRT.} \end{cases}$$

This function has been determined from simulated 95% quantiles of the various tests.

With the corrected statistics, critical quantiles have been determined by simulation for $n = 100, 1000$ and $\alpha = 0.1, 0.05, 0.01$. The number of replications was always 5000.

4 Comparison of power

In a power simulation study we considered both forms of the MLRT and the D-tests, with and without employing a Wei2Exp transformation. In addition we used several weighted versions of the D-test. The power of these tests was evaluated under different alternatives and contrasted with that of the ADDS test.

Figure 2 exhibits the power of the non-weighted D-test, the w2D-test, and the MLRT, each being applied to Wei2Exp transformed data, and the ADDS test. The alternatives are lower contaminations, that is, mixtures of a Weibull distribution with another Weibull distribution having smaller scale. The main result of the simulation study is that under lower contamination the ADDS is the only one that develops reasonable power, while the other three tests break completely down. It has been further shown that under upper contaminations, that is, mixtures with a larger scaled Weibull distributions, the ADDS test is only slighty less powerful than the others.

A similar power comparison has been done for the MLRT and two D-tests that are applied to the non-transformed data. Here, it came out that under lower contaminations the non-weighted D-test outperforms the other three tests when the sample size is small ($n = 100$), while for larger samples ($n = 1000$) the four tests develop similar power. Under upper contaminations, the ADDS test proves to be better than the others, which develop poor power only. This holds true even for fifty-fifty mixtures.

The relatively poor performance of the MLRT and the D-tests may be attributed to parameter estimation on the alternative and, in addition, to the dependency of quantiles on parameters. The power of these tests will possibly improve if correction factors are determined in a less simple way. In particular, instead of employing the same linear function h for all α, one could employ some non-linear interpolations depending on α. Also the number of replications ($R = 5000$) could be increased to obtain more precise results. However, it is rather obvious that the qualitative results of this paper will not change.

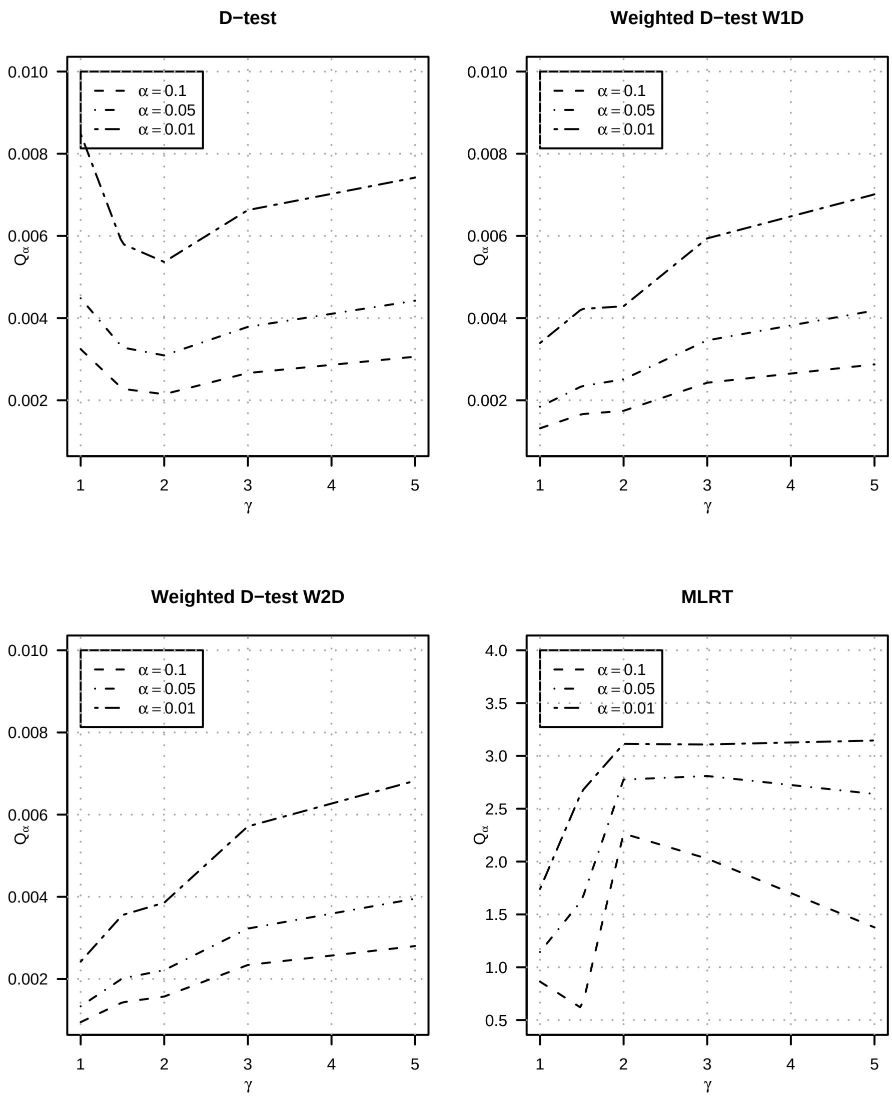

Fig. 1. Dependency of critical quantile Q_α on the true shape parameter γ, for the DTest, the weighted DTests (w1D, w2D) and the MLRT (without Wei2Exp transformation); $n = 1000$, $\alpha = 0.01, 0.05, 0.10$.

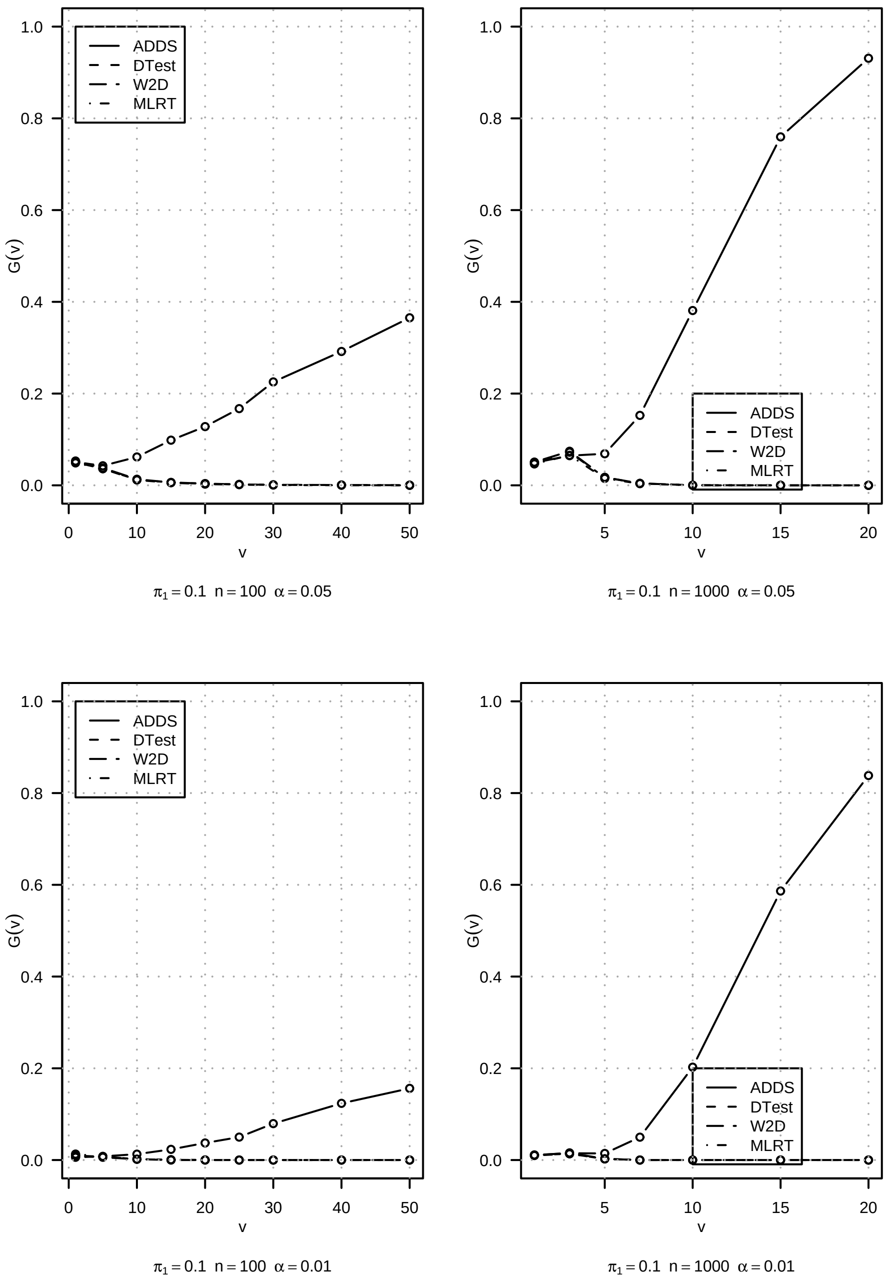

Fig. 2. Power under lower contaminations: D-test, w2D-test (quadratically weighted D-test), and MLRT with Wei2Exp transformation, ADDS test. Comparison of power under the alternative $S(t) = 0.9 \exp(-t^{\gamma}) + 0.1 \exp(-(vt)^{\gamma})$, depending on scale ratio $v \geq 1$.

In a nutshell: While the MLRT and the D-tests perform well in other models (Chen *et al.* (2001), Charnigo and Sun (2004)), their application appears to be not recommendable for homogeneity testing in a Weibull mixture model. Here, the ADDS test provides a reasonable diagnostic alternative.

Bibliography

J. R. G. Albert and L. A. Baxter. Applications of the EM algorithm to the analysis of life length data. *Applied Statistics*, 44:323–341, 1995.

R. Charnigo and J. Sun. Testing homogeneity in a mixture distribution via the l^2 distance between competing models. *Journal of the American Statistical Society*, 99:488–498, 2004.

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

R. Jiang and D.N.P. Murthy. Modeling failure data by mixture of two Weibull distributions: A graphical approach. *IEEE Transactions on Reliability*, 41:241–247, 1995.

A. R. Kaylan and C. M. Harris. Efficient algorithms to derive maximum-likelihood estimates for finite exponential and Weibull mixtures. *Computers and Operations Research*, 8:97–104, 1981.

K. Mosler and L. Haferkamp. Size and power of recent tests for homogeneity in exponential mixtures. *Communications in Statistics - Simulation and Computation*, 36:493–504, 2007.

K. Mosler and C. Scheicher. Homogeneity testing in a Weibull mixture model. *Statistical Papers*, 49:315-332, 2008.

K. Mosler and W. Seidel. Testing for homogeneity in an exponential mixture model. *Australian and New Zealand Journal of Statistics*, 43:231–247, 2001.

D. M. Olsson. Estimation for mixtures of distributions by direct maximization of the likelihood function. *Journal of Quality Technology*, 11:153–159, 1979.

R. Piessens, E. deDoncker Kapenga, C. Uberhuber, and D. Kahaner. *Quadpack: A Subroutine Package for Automatic Integration.* Springer Verlag, 1983.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S-PLUS.* Springer, 4th edition, 2002.