Homogeneity Testing in a Weibull Mixture Model*

Karl Mosler, Christoph Scheicher

Abstract

The mixed Weibull distribution provides a flexible model to analyse random durations in a possibly heterogeneous population. To test for homogeneity against unobserved heterogeneity in a Weibull mixture model, a dispersion score test and a goodness-of-fit test are investigated. The empirical power of these tests is assessed and compared on a broad range of alternatives. It comes out that the dispersion score test, as it is based on a Weibull-to-exponential transformation, often breaks down. A simple new procedure is introduced for Weibull mixtures in scale, which combines the dispersion score test and the goodness-of-fit test. The new test is compared with several known procedures and shown to have a good overall power. To detect mixtures in shape and scale, a goodness-of-fit test is recommended.

MSC 2000 subject classifications. Primary 62N03; secondary 90B25.

Key words and phrases. Mixture diagnosis, survival analysis, hazard models, dispersion score test, goodness-of-fit, Weibull-to-exponential transform.

---

*This research has been partially sponsored by a grant of the Deutsche Forschungsgemeinschaft. We thank Lars Haferkamp for computational assistance and Wilfried Seidel and a referee for their remarks on alternative test procedures.
Corresponding author: Karl Mosler, Statistik und Ökonometrie, Universität Köln, D-50923 Köln, Germany (E-mail: mosler@statistik.uni-koeln.de)

1 Introduction

The *mixed Weibull distribution* is a probability mixture of Weibull distributions which differ in scale and/or shape. It can be seen as a probability model of a heterogeneous population: A given population divides into parts each of which is subject to another reason of failure and described by a pure distribution. When the population consists of k parts and part j has a Weibull distribution $W(\beta_j, \gamma_j)$ with scale parameter $\beta_j > 0$ and shape parameter $\gamma_j > 0$, the survival function of the Weibull k-mixture is

$$S(t) = \sum_{j=1}^{k} p_j \exp\left(-\left(\frac{t}{\beta_j}\right)^{\gamma_j}\right), \tag{1}$$

with $0 < p_j \leq 1, j = 1, \ldots, k$, $\sum_j p_j = 1$. Here p_j corresponds to the relative size of part j. A random duration T that has survival function (1) is named a *Weibull k-mixture*; this is shortly written as $T \sim MW(k, \beta, \gamma, p_1, \ldots, p_{k-1})$. A pure Weibull distribution ($k = 1$) is denoted by $W(\beta, \gamma)$.

Observe that the class of Weibull k-mixtures is closed against exponentiation and multiplication with positive numbers: If $T \sim MW(k, \beta, \gamma, p_1, \ldots, p_{k-1})$, then $bT^c \sim MW(k, b\beta, c^{-1}\gamma, p_1, \ldots, p_{k-1})$, for any $b, c > 0$.

In many applications, mixture models are used in a natural way to model population heterogeneity; see Lindsay (1995), Titterington *et al.* (1985), and others. The assumption that the underlying distribution is a mixture of certain lifetime distributions is widely invoked in the analysis of lifetime or, more general, duration data. This model arises from incomplete observation of an underlying conditional model.

The mixed model can also be seen as a parametric proportional hazards model with unobserved heterogeneity (Lancaster, 1990) when no observed covariates are present.

In the case of two *competing risks*, we get a special case of the 2-mixture model. Let a member of the population fail if the minimum of two continuous lifelengths is attained which have (possibly dependent) survival functions S_1 and S_2. Then its survival function is given by (1) with $k = 2$, $p_1 = p_{cr}$,

$$p_{cr} = \int_0^\infty (S_1(t_2|t_2) - 1)\; dS_2(t_2)\,. \tag{2}$$

Also the popular model of a *competing risk mixture* (e.g. Tarum, 1999) with independent survivals S_1 and S_2, $S = p^* S_1 S_2 + (1-p^*) S_2$, can be written as a 2-mixture with $p_1 = p_{crm}$,

$$p_{crm} = p_1^* \cdot \int_0^\infty (S_1(t_2) - 1)\, dS_2(t_2)\,.$$

The mixture of two Weibull distributions provides a rather flexible model to be fitted to data and is also able to depict non-monotonous hazard rates. It is known under the heading *BiWeibull*. Due to its simplicity and flexibility, the BiWeibull has been widely used in engineering and other application fields (Abernethy, 1996; Tarum, 1999). Figure 1 illustrates the diversity of hazard functions of Weibull 2- and 3-mixtures in scale and/or shape. A 2-mixture in scale has decreasing hazard rate when its common shape parameter does not exceed 1, which is illustrated by Figure 1a. When the shape parameter is greater than 1, non-monotonous hazard rates arise as shown in Figure 1b.

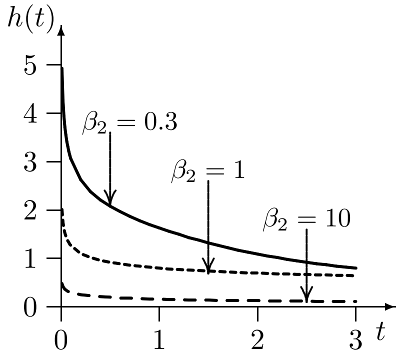
To determine the parameters of a BiWeibull model, Falls (1970) employs moments estimators, and Cheng & Fu (1982) least squares estimators. Kaylan & Harris (1981) determine ML estimators for finite Weibull mixtures; see also Chapter 4 in Sinha (1986). Albert & Baxter (1995) provide a modification of the EM algorithm to calculate ML estimates of finite Weibull mixtures. Marín *et al.* (2003) propose an MCMC approach for the case when the number of mixture components is not known.

An important question is, whether for given data a Weibull mixture model should be preferred over a non-mixed Weibull specification. Kao (1959) and Jiang & Murthy (1995) propose graphical procedures to decide the appropriateness of a two-components Weibull mixture. In this paper we will investigate statistical tests for a non-mixed Weibull specification against a Weibull mixture model. These specification tests can be seen as tests for homogeneity against unobserved heterogeneity in the population.

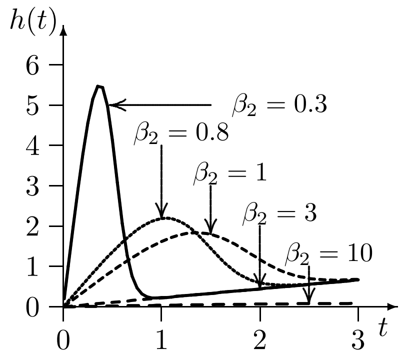
When the shape parameter γ is known, the Weibull model can be put down to the exponential model. If $T \sim W(\beta, \gamma)$, then $T^\gamma \sim Exp(1/\beta^\gamma)$, that is, T^γ has an exponential distribution with hazard rate $1/\beta^\gamma$.

Mosler & Seidel (2001) have compared several diagnostic tests for homogeneity in exponential mixtures, among them a dispersion score

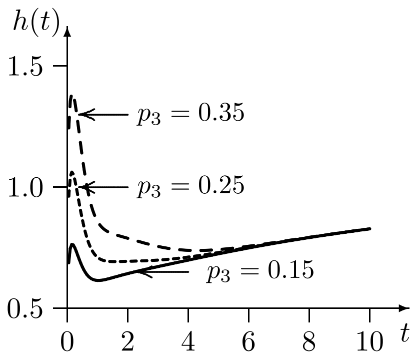
(a) Weibull 2-mixtures in scale, $p = 0.1$, $\gamma_1 = \gamma_2 = 0.8$, $\beta_1 = 1.0$, $\beta_2 \in \{0.3, 1.0, 10.0\}$



(b) Weibull 2-mixtures in scale, $p = 0.1$, $\gamma_1 = \gamma_2 = 2$, $\beta_1 = 3$, $\beta_2 \in \{0.3, 0.8, 1, 3, 10\}$



(c) Weibull 3-mixtures in scale, $p_1 = p_2$, $p_3 \in \{0.15, 0.25, 0.35\}$, $\gamma_1 = \gamma_2 = \gamma_3 = 1.2$, $\beta_1 = 0.3$, $\beta_2 = 1.0$, $\beta_3 = 2$



(d) Weibull 2-mixtures in scale and shape, $p = 0.5$, $\beta_1 = 0.1$, $\gamma_1 = 0.1$, $\beta_2 = 3.0$, $\gamma_2 \in \{2.0, 5.0, 20.0\}$

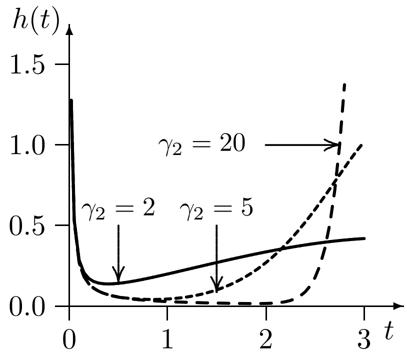


Figure 1: Hazard rates of several Weibull mixtures

test, a goodness-of-fit test and a combination of both tests. Their results hold as well for scale mixtures from a known baseline distribution. In the sequel we will extend their approach to a model where the baseline distribution is not fully known, namely, a Weibull distribution with unknown parameters.

We consider mixtures of Weibull distributions that have a common but unknown shape parameter γ. We employ several tests on exponentiality from Mosler & Seidel (2001) and investigate their behavior when γ is estimated from the data instead of being known. The first test incorporates a Weibull-to-exponential transformation and is designed to detect overdispersion in the transformed data. A simulation study shows that the power of the dispersion score test appears to be reasonably good in certain situations and very bad in others. This appears to be due to the Weibull-to-exponential transformation used in in the test. The reason is that the values of T^γ are very sensitive to variations of the exponent γ. Consequently, if $T_1, \ldots, T_n$ is an i.i.d. sample from $W(\beta, \gamma)$ and $\hat{\gamma}$ is some estimate for γ, the transformed sample $T_1^{\hat{\gamma}}, \ldots, T_n^{\hat{\gamma}}$ may come out far from being exponential, and statistical inference based on exponentiality be possibly misleading. This is in line with Keats *et al.* (2000), who demonstrate that the resulting confidence intervals and tests for β can be very poor.

As an alternative procedure we consider a goodness-of-fit procedure, which is a special Anderson-Darling test, and compare the empirical power of the two tests on various mixture alternatives. In order to combine the strengths of the two tests we introduce a new test procedure 'Reject the null hypothesis if at least one of two tests rejects.' and demonstrate that the combined test has a good overall power.

Further, to test for homogeneity against general Weibull mixtures, that is, mixtures in scale and in shape, we use the same goodness-of-fit test and investigate its power.

Overview: In Section 2 the Weibull model and the estimation of its parameters are presented. The dispersion score test for homogeneity is given and its power investigated; the same is done with the Anderson-Darling test. Then the combined test procedure is introduced and its power compared with the dispersion score test and two recently proposed tests, the penalized likelihood ratio test (Chen *et al.*, 2001) and the D-test (Charnigo & Sun, 2004). In Section 3 general Weibull mixtures are tested. Section 4 concludes.

2 Mixtures in scale

In this section we investigate mixtures of Weibull distributions that have different scale but a common shape parameter. Given a random sample $T_1, \ldots, T_n$, where each T_i has the same survival function

$$S(t) = \sum_{j=1}^{k} p_j \exp\left(-\left(\frac{t}{\beta_j}\right)^{\gamma}\right), \tag{3}$$

$\beta_j, \gamma > 0, p_j \in]0,1], j = 1, \ldots, k, \sum_j p_j = 1$, we want to test for

$$H_0 : k = 1 \quad \text{against} \quad H_1 : k > 1\,. \tag{4}$$

The null hypothesis says that there exists some $\beta_0 > 0$ such that

$$S(t) = \exp\left(-\left(\frac{t}{\beta_0}\right)^{\gamma}\right).$$

The common shape parameter γ is generally not known. The test problem can be seen as a problem of detecting overdispersion in a one-parameter exponential family that has a nuisance parameter γ. Conditional on γ, the variance of a distribution in the family is determined by its mean. Further, given γ, any mixture of distributions from the family is a dilation from the pure distribution that has the same mean as the mixture (Shaked, 1980); consequently, the mixture has larger variance than the pure distribution.

Given an i.i.d. sample $T_1, \ldots, T_n \sim W(\beta, \gamma)$, maximum likelihood estimators (MLE) of β and γ are obtained from the equations

$$\beta = \left(\frac{1}{n}\sum_{i=1}^{n} T_i^{\gamma}\right)^{1/\gamma}, \qquad \frac{1}{\gamma} = \frac{1}{n}\left(\frac{1}{\beta^{\gamma}}\sum_{i=1}^{n}(T_i^{\gamma}\ln T_i) - \sum_{i=1}^{n}\ln T_i\right).$$

Since the MLE of γ is heavily biased in small samples, Yang & Xie (2003) propose a slight bias reducing modification:

$$\frac{1}{\gamma} = \frac{1}{n-2}\left(\frac{\sum_{i=1}^{n} T_i^{\gamma}\ln T_i}{\frac{1}{n}\sum_{i=1}^{n} T_i^{\gamma}} - \sum_{i=1}^{n}\ln T_i\right). \tag{5}$$

Note that the modified MLE is asymptotically equivalent to the usual MLE, since it differs only by a factor $n-2$ instead of n in the determining equation (5). We shall use this modified MLE in the sequel.

2.1 Dispersion score test

To test for (4), we first assume that γ is known. The following dispersion score test is a variant of Neyman's $C(\alpha)$ test (Neyman & Scott, 1966). It is scale invariant, locally most powerful in any direction, and makes an optimal use of the local information on the parameters. Let

$$C_n = \frac{1}{n-1}\sum_{i=1}^{n}(T_i^\gamma - \overline{T^\gamma})^2 - \frac{1}{2n}\sum_{i=1}^{n} T_i^{2\gamma}\,, \quad \text{where } \overline{T^\gamma} = \frac{1}{n}\sum_{i=1}^{n} T_i^\gamma\,. \tag{6}$$

From Mosler & Seidel (2001) follows that, for distributions π with nonnegative support and finite fourth moment and $\int_0^\infty u d\pi(u) > 0$, C_n is an unbiased estimator of the variance $\mathrm{var}(\pi)$ of the mixing distribution π and asymptotically normal. It follows further that under H_0

$$\mathrm{var}(C_n) = \frac{\beta_0^{4\gamma}}{n}\,\frac{n+1}{n-1}$$

holds. However, the null distribution of C_n depends on β_0. As $E(T^\gamma|\beta_0) = \beta_0^\gamma$, the unknown β_0^γ can be estimated by $\overline{T^\gamma}$. Hence, the standardized statistic

$$O_n = \frac{C_n}{\overline{T^\gamma}^2}\left(\frac{n(n-1)}{n+1}\right)^{\frac{1}{2}} \tag{7}$$

is invariant to multiplication of all T_i by a positive number b, that is, to a change of scale in model (3). Especially, under H_0, the distribution of O_n does not depend on β_0.

Now let γ be unknown and estimated by the modified MLE $\hat{\gamma}$. To test for H_0 we use the statistics $\widehat{C}_n$ and $\widehat{O}_n$, which amount to C_n and O_n with γ replaced by $\hat{\gamma}$. Note that, as C_n is a consistent estimate of $\mathrm{var}(\pi)$, also $\widehat{C}_n$ converges in probability to $\mathrm{var}(\pi)$. Thus, the null hypothesis of no mixture should be rejected if $\widehat{C}_n$ appears to be too large.

Quantiles of the standardized test statistic $\widehat{O}_n$ under H_0 have been obtained by Monte-Carlo simulation. We calculated them for different values of the test size α, the sample size n, and the shape parameter γ. As expected, these quantiles do not depend on γ, but only on α and n. For selected α and n, the quantiles of $\widehat{O}_n$ under H_0 are presented in Table 1 of the Appendix. This makes a meaningful dispersion score test for homogeneity in the Weibull mixture model: Reject H_0 if $\widehat{O}_n$ is larger than the proper quantile in Table 1.

An extensive power study was performed, which investigates the behaviour of the dispersion score (DS) test on various Weibull scale mixture alternatives. Depending on the alternative, the empirical power proves to be very different. Consider 2-mixture alternatives and let $v = \frac{\beta_2}{\beta_1} > 1$. The power of the DS test depends on v only; it is reasonably good if p is close to 1. This case is named an *upper contamination*. Figure 2 exhibits, amongs others, the power depending on v in case $p = 0.9$, when sample size is $n = 100$ or $n = 1000$ and test size is $\alpha = 0.01$ or $\alpha = 0.05$. But, for smaller p the DS test develops much less power, which is illustrated in Figure 3 for $p = 0.7$. In case $p = 0.1$, which we call a *lower contamination*, the power is very poor; it becomes zero even for large values of the scale ratio $v = \frac{\beta_2}{\beta_1}$ and large sample sizes n. Figure 4 illustrates this result. Other α and n yield similar pictures.

To understand why the power of the DS test breaks down on lower contamination alternatives, we take a closer look on the procedure. W.l.o.g. assume that $\gamma = 1$. Conditional on the estimated $\hat{\gamma}$, it holds that

$$T^{\hat{\gamma}}|\hat{\gamma} \sim W\left(\beta^{\hat{\gamma}}, \frac{1}{\hat{\gamma}}\right),$$

$$\operatorname{var}\left(T^{\hat{\gamma}}|\hat{\gamma}\right) - \frac{1}{2}\mathrm{E}\left(T^{2\hat{\gamma}}|\hat{\gamma}\right) = \beta^{2\hat{\gamma}}\hat{\gamma}\left(\Gamma(2\hat{\gamma}) - \Gamma(\hat{\gamma})\,\Gamma(1+\hat{\gamma})\right), \tag{8}$$

where Γ denotes the Gamma function. If $\hat{\gamma} < \gamma = 1$ and $\hat{\gamma}$ is not too far from 1, it follows that $\Gamma(2\hat{\gamma}) < \Gamma(1+\hat{\gamma}) < 1$ and therefore $\Gamma(2\hat{\gamma}) < \Gamma(\hat{\gamma})\,\Gamma(1+\hat{\gamma})$, hence

$$\operatorname{var}\left(T^{\hat{\gamma}}|\hat{\gamma}\right) - \frac{1}{2}\mathrm{E}\left(T^{2\hat{\gamma}}|\hat{\gamma}\right) < 0\,.$$

It is seen from our simulations that on lower contamination mixtures the parameter γ is nearly always underestimated, $\hat{\gamma} < \gamma$, when v is greater than and not too close to 1. As, conditional on $\hat{\gamma}$, $\widehat{C}_n$ is an estimate of the left hand side of (8), it follows, particularly often for large n, that $\widehat{C}_n < 0$ and, consequently, $\widehat{O}_n < 0$. Therefore, on a lower contamination the null hypothesis H_0 will hardly ever be rejected if v and n are large enough.

2.2 Goodness-of-fit test

As an alternative procedure for testing homogeneity we use a goodness-of-fit test, in particular, an Anderson-Darling (AD) test for Weibull distributed observations, which is described in D'Agostino & Stephens (1986, pp 149f). The test may be summarized as follows.

Let $T_1, \dots, T_n$ denote a sample of i.i.d. variables. The null hypothesis H_0 is that the T_i follow a (non-mixed) Weibull distribution with unknown scale and shape parameter. Then the transformation $Y = -\ln T$ yields an i.i.d. sample $Y_1, \dots, Y_n$ that has a double-exponential distribution. The two parameters are estimated through maximum likelihood and the estimated distribution function $\widehat{F}$ of Y is obtained. To test for H_0, we calculate the order statistics $Z_{(i)} = (\widehat{F}(Y_{(i)}))$ and use the AD statistic

$$A_n^2 = -n - \frac{1}{n} \sum_i \Big((2i-1) \ln Z_{(i)} + (2n+1-2i) \ln \Big(1 - Z_{(i)}\Big) \Big).$$

H_0 is rejected if A_n^2 is too large. Table 2 presents some simulated critical values.

Now let T have survival function (9) and consider a scale-exponential transformation $T_i \mapsto \tilde{T}_i = (bT)^c$ with $b, c > 0$. Then $\tilde{Y}_i = -\ln \tilde{T}_i = -c \ln b + cY_i$ is a location-scale transformation of Y_i. The distribution of our AD statistic does not depend on the true values of estimated location and scale parameters of Y, that is, neither on $-c \ln b$ nor on c. Therefore, its distribution is invariant against scalar-exponential transformations of form $T_i \mapsto \tilde{T}_i = (bT)^c$.

The power of this test has been evaluated on a large number of Weibull mixtures in scale. Some typical comparative results on 2-mixture alternatives can be seen in Figures 2–4. W.l.o.g., $\gamma_1 = \gamma_2 = 1$. The power functions of the AD test and the DS test are plotted, depending on the scale ratio $v = \frac{\beta_2}{\beta_1}$ and for different values of p, α and n. If p is close to 1, the AD test is clearly outperformed by the DS test. But for smaller p the AD test retains reasonable power, while the DS test breaks down. The larger n and α, the larger is the difference in power. However note that in a small neighbourhood of the null hypothesis (for v near to 1) the DS test always develops slightly more power than the AD test. This seems to be due to the local optimality of the $C(\alpha)$-procedure.

2.3 Combined test

A further look at the simulated samples and the test decisions arising from them shows that there are many samples for which the AD test rejects the null hypothesis while the DS test does not, and there are many other samples for which the reverse happens.

Therefore it seems to be worthwhile combining the two tests as follows:

$$\text{Reject } H_0 \text{ if } \widehat{O}_n \geq t_1 \text{ or } A_n^2 \geq t_2\,,$$

where t_1 and t_2 are critical values which are properly chosen to meet a given test size α. Here we have determined t_1 and t_2 so that the AD test and the DS test each individually obtain the same size $\alpha^* < \alpha$ and the combination of them obtains size α. (This has been done by systematic variation of α^*). This combined test is denoted as the *AD-DS test*. Table 3 presents the critical values t_1 and t_2 of the test for various n and α.

The empirical power of the AD-DS test has been calculated and compared with the power of the AD and DS tests. It comes out that the combined test works reasonably well on all considered alternatives and that its power comes close to the maximum power of the two tests on which it is based. Figures 2–4 illustrate this useful and somewhat surprising result.

We also combined AD and DS tests having unequal sizes α_1^*, α_2^* of the two component tests, e.g. with $\alpha_1^* = 2\alpha_2^*$ or $\alpha_2^* = 2\alpha_1^*$. The resulting power curves are similar, but shifted towards the power curve of the test that is given the larger size.

2.4 Likelihood ratio tests and D-tests

The tests considered so far are tests for homogeneity against a general class of mixture alternatives (Weibull k-mixtures with any k). The DS test is sensitive to k-mixtures in scale and the AD test is a general purpose goodness-of-fit test. To test for homogeneity against specific mixture alternatives, likelihood ratio (LR) tests suggest themselves. The asymptotic distribution of the LR statistic is not chi-square and difficult to evaluate. So, in practice and not only for small samples, these tests have to be based on empirical quantiles, where the maximum likelihood under the alternative is numerically determined by some variant of the EM algorithm. Seidel *et al.* (2000a, 2000b) demonstrate that in an exponential mixture model the EM algorithm depends heavily on initial values (and also on stopping rules); therefore, in each application of the test, a careful multistart strategy is needed to find a 'global' maximum. By this, the LR test for homogeneity in exponential mixtures is awkward and expensive to calculate. Also, the ML estimate under the alternative often yields a component having very small probability weight and/or scale parameter close to zero. A component like this may be regarded as 'spurious'. However, whether it is worth being estimated or not depends on the real problem to which the test is applied. What is true for exponential mixtures is *a fortiori* true in a Weibull mixture model.

To overcome some of the problems connected with LR tests for homogeneity in mixtures of life distributions, a penalized LR test, named MLRT, has been introduced by Chen *et al.* (2001); see also Charnigo & Sun (2004). It has simple chi-square asymptotics. Particularly, when $k = 2$, the test employs the usual loglikelihood function plus a term $C \ln(4p(1-p))$, which penalizes values of p that are close to 0 or 1. Depending on the penalty constant C, more or less 'spurious' solutions are excluded.

An alternative test procedure, called D-test, has been recently proposed by Charnigo & Sun (2004). Test statistic is the L^2-distance between the density under H_0 : 'no mixture' and the density under H_1 : '2-mixture'. Two variants of the D-test employ weighted L^2-distances, with weight functions $w(t) = t$ and $w(t) = t^2$, which correspond to transformations $\ln(T)$ and $\frac{1}{T}$, respectively, of the random duration T. In calculating the D-test statistics, estimates of the Weibull parameters are needed under both H_0 and H_1. In our study we simulated Weibull scale mixtures with $\gamma = 1$ and employed two alternative estimation procedures, that by Kaylan & Harris (1981) and that by Nelder & Mead (see Venables & Ripley, 2002). The D-statistic with weight $w(t) = t^2$ provided relatively best power among the three variants of the D-test. (We always calculated critical values by Monte Carlo and did not rely on their asymptotics.) However, with both estimation procedures, the power of this D-test comes out to be much worse than that of our AD-DS procedure; see Figure 7.

We also applied the D-test after a Weibull-to-exponential transform of the data (with shape parameter being estimated under H_0). The power of this procedure resembles that of the DS test: In particular, it is fine at upper contaminations, but breaks down at lower contaminations; see Figures 8 and 9. We did the same with the penalized LR test of Chen *et al.* (2001) and obtained very similar power results, which are also exhibited in Figures 8 and 9.

3 Mixtures in scale and shape

The Weibull finite mixture distribution in scale and shape has survival function

$$S(t) = \sum_{j=1}^{k} p_j \exp\left(-\left(\frac{t}{\beta_j}\right)^{\gamma_j}\right) \tag{9}$$

with some $\beta_j > 0, \gamma_j > 0, p_j \in [0,1], j = 1,\ldots,k$, $\sum_j p_j = 1$. We test for

$$H_0 : S(t) = \exp\left(-\left(\frac{t}{\beta}\right)^{\gamma}\right) \quad \text{for some} \quad \beta > 0, \gamma > 0$$

against H_1: (9), but not H_0. Here the DS test is not feasible. Instead we employ the above AD test for two reasons. Firstly, this test is a general goodness-of-fit test; secondly, it develops satisfactory power on scale mixture alternatives and, more general, on alternatives which have a 'more decreasing hazard rate' than the null hypothesis (D'Agostino & Stephens, 1986).

The power of this test has been evaluated on various Weibull mixtures in scale and shape. Recall that the distribution of this test statistic does not depend on the true values of the estimated parameters. Moreover, it is invariant against scale transformations $T \mapsto bT, b > 0$ as well as against exponential transformations $T \mapsto T^c, c > 0$. From the invariance property it is clear that the power depends on the ratios $\frac{\beta_2}{\beta_1}$ and $\frac{\gamma_2}{\gamma_1}$ only.

Figures 5 and 6 present some exemplary results on Weibull 2-mixtures in scale alone and on Weibull 2-mixtures in scale and shape. Figure 5 exhibits the power of the AD test on the mixture alternatives

$$S(t) = p \exp\left(-\left(\frac{t}{\beta}\right)^{\gamma_1}\right) + (1-p) \exp\left(-\left(\frac{t}{\beta}\right)^{\gamma_2}\right), \quad \frac{\gamma_2}{\gamma_1} \geq 1,$$

that is, mixtures of a Weibull distribution with a second Weibull distribution having the same scale parameter, but a larger shape parameter, $p \in \{0.1, 0.5, 0.9\}$, $n = 100, 1000$. The mixture is best detected if the components have equal weights. The power is also reasonable if the first component outweighs the second one, a case we call *more increasing hazard contamination*, but it is poor if the second component predominates, which is the case called *less increasing hazard contamination*.

Figure 6 exhibits the power of the AD test on mixtures of two Weibull distributions differing in both scale and shape, with $\frac{\beta_2}{\beta_1} = 30$ and $\frac{\gamma_2}{\gamma_1} \geq$

1, which is illustrated in Figure 1d. Again, the fifty-fifty mixture is best detected, the power on a less increasing hazard contamination is reasonable, and the power on a more increasing hazard contamination is poor.

4 Concluding remarks

The paper has studied diagnostic procedures to analyze whether a given distribution of data stems from some pure Weibull distribution or rather a mixture of such distributions.

1. Firstly, a dispersion score (DS) test for detecting mixtures in scale has been investigated. This test is locally most powerful in any direction, and makes an optimal use of the local information on the parameters. Nevertheless, as we have shown, the practical use of the DS test is rather limited since on many alternatives its power is poor. The power to detect a Weibull 2-mixture appears reasonable if the mixture is an upper contamination or if the scale ratio of the two components comes close to 1, but it is poor if, e.g., the mixture is a lower contamination. One reason for this power behavior is that the test incorporates a Weibull-to-exponential transformation with an estimated shape parameter. But also if the shape parameter is known, which reduces the problem to one of detecting a mixture of exponentials, the DS test has poor power on lower contaminations; see Mosler & Seidel (2001).

2. As an alternative to the DS test we have employed a goodness-of-fit test and shown that for other alternatives this test is preferable. A proper version of the Anderson-Darling (AD) test develops reasonable power also in situations where the DS test breaks down. On the other hand the DS outperforms the AD test on upper contaminations.

3. The AD test can be profitably combined with the DS test. A combination of the two tests has been introduced and it has been demonstrated that the combined test obtains a good overall power, which in all situations considered comes close to the power of the better of the two tests. The combined test also often outperforms the MLRT and the D-test.

4. For general Weibull mixtures regarding also the shape parameter, the DS test is not applicable, but the AD test proves to be a useful procedure to detect mixture alternatives.

5. To keep things simple, we have restricted our analysis to non-censored data. The above procedures may be adjusted for various situations of censored or truncated data; see D'Agostino & Stephens (1986, p. 114f), Shorrack and Wellner (1986) and Andersen *et al.* (1993, ch. 5) for Anderson-Darling tests under random censoring and more general sampling situations, and Lancaster (1990), Commenges & Andersen (1995) and Jaggia (1997) for dispersion score tests under different censoring schemes.

References

Abernethy, R.B. (1996). *The New Weibull Handbook.* 2nd Edition. R.B. Abernethy, North Palm Beach, Fla.

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

Andersen, P.K., Borgan, O., Gill, R.D. & Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

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

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* **B 63** 19–29.

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

Commenges, D. & Andersen, P.K. (1995). Score test of homogeneity for survival data. *Lifetime Data Analysis* **1** 145–160.

D'Agostino, R.B. & Stephens, M.A. (1986). *Goodness-of-fit techniques.* Dekker, New York.

Falls, L.W. (1970). Estimation of parameters in compound Weibull distributions. *Technometrics* **12** **2** 399–407.

Jaggia, S. (1997). Alternative forms of the score test for heterogeneity in a censored exponential model. *Review of Economics and Statistics* **79** 340–343.

Jiang, R. & Murthy, D.N.P. (1995). Modeling failure-data by mixture of 2 Weibull distributions: A graphical approach. *IEEE Transactions on Reliability* **44** **3** 477–488.

Kao, J.H.K. (1959). A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. *Technometrics* **1** 389–407.

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

Keats, J.B., Nahar, P.C., & Korbel, K.M. (2000). A study of the effects of mis-specification of the Weibull shape parameter on confidence bounds based on the Weibull-to-exponential transformation. *Quality and Reliability Engineering International* **16** 27–31.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press, Cambridge.

Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry and Applications.* Institute of Mathematical Statistics, Hayward, Cal.

Marín, J.M., Rodríguez Bernal, M.R. & Wiper, M.P. (2003). Using Weibull mixture distributions to model heterogeneous survival data. *Universidad Carlos III de Madrid, Working Paper* **03-32**, Statistics and Econometrics Series 08.

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

Neyman, J. & Scott, E.L. (1966). On the use of $C(\alpha)$ optimal tests of composite hypotheses. *Bull. Inst. Int. Statist.* **41** **I** 477–497.

Seidel, W., Mosler, K. & Alker, M. (2000a). Likelihood ratio tests based on subglobal optimization: A power comparison in exponential mixture models. *Statistical Papers* **41**, 85–98.

Seidel, W., Mosler, K. & Alker, M. (2000b). A cautionary note on likelihood ratio tests in mixture models. *Annals of the Institute of Statistical Mathematics* **52**, 481–487.

Shaked, M. (1980). On mixtures from exponential families. *Journal of the Royal Statistical Society, Ser. B* **42** 192–198.

Shorack, G.R. & Wellner, J.A. (1986). *Empirical Processes.* Wiley, New York.

Sinha, S.K. (1986). *Reliability and Life Testing.* Wiley, New York.

Tarum, C.D. (1999). Classification and analysis of Weibull mixtures. *SAE Technical Paper Series* **1999-01-0055**.

Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S-PLUS.* 4. Ed., Springer, New York.

Yang, Z. & Xie, M. (2003) Efficient estimation of the Weibull shape parameter based on a modified profile likelihood. *Journal of Statistical Computation and Simulation* **73 2** 115-123.

Appendix

The quantiles in Tables 1 and 2 have been calculated from 10^6 replications. The quantiles in Table 3 have been choosen by calculating (also 10^6 replications) the power of the AD-DS test under H_0 for various $\alpha^* < \alpha$. For n not in the tables the quantile $Q_n(\alpha)$ may be interpolated by fitting a second degree polynomial to three adjacent points $(n_j, Q_{n_j}(\alpha))$.

Table 1: Quantiles of the dispersion score statistic O_n

α	n						
	10	20	50	100	200	500	1000
0.10	0.23	0.39	0.56	0.65	0.71	0.76	0.78
0.05	0.38	0.60	0.82	0.92	0.99	1.03	1.04
0.01	0.68	1.07	1.41	1.55	1.60	1.59	1.57

Table 2: Quantiles of the Anderson-Darling statistic A_n^2

α	n						
	10	20	50	100	200	500	1000
0.10	0.62	0.62	0.63	0.63	0.63	0.63	0.64
0.05	0.73	0.74	0.75	0.75	0.76	0.76	0.76
0.01	0.98	1.01	1.03	1.04	1.04	1.04	1.04

Table 3: Quantiles for the combination AD-DS of O_n and A_n^2 statistics

α	test component	n						
		10	20	50	100	200	500	1000
0.10	A_n^2	0.71	0.72	0.73	0.73	0.73	0.74	0.74
	O_n	0.36	0.57	0.77	0.87	0.94	0.99	1.00
0.05	A_n^2	0.82	0.84	0.85	0.86	0.86	0.87	0.87
	O_n	0.50	0.77	1.04	1.15	1.22	1.25	1.25
0.01	A_n^2	1.07	1.11	1.13	1.15	1.15	1.16	1.15
	O_n	0.77	1.24	1.65	1.81	1.85	1.81	1.76

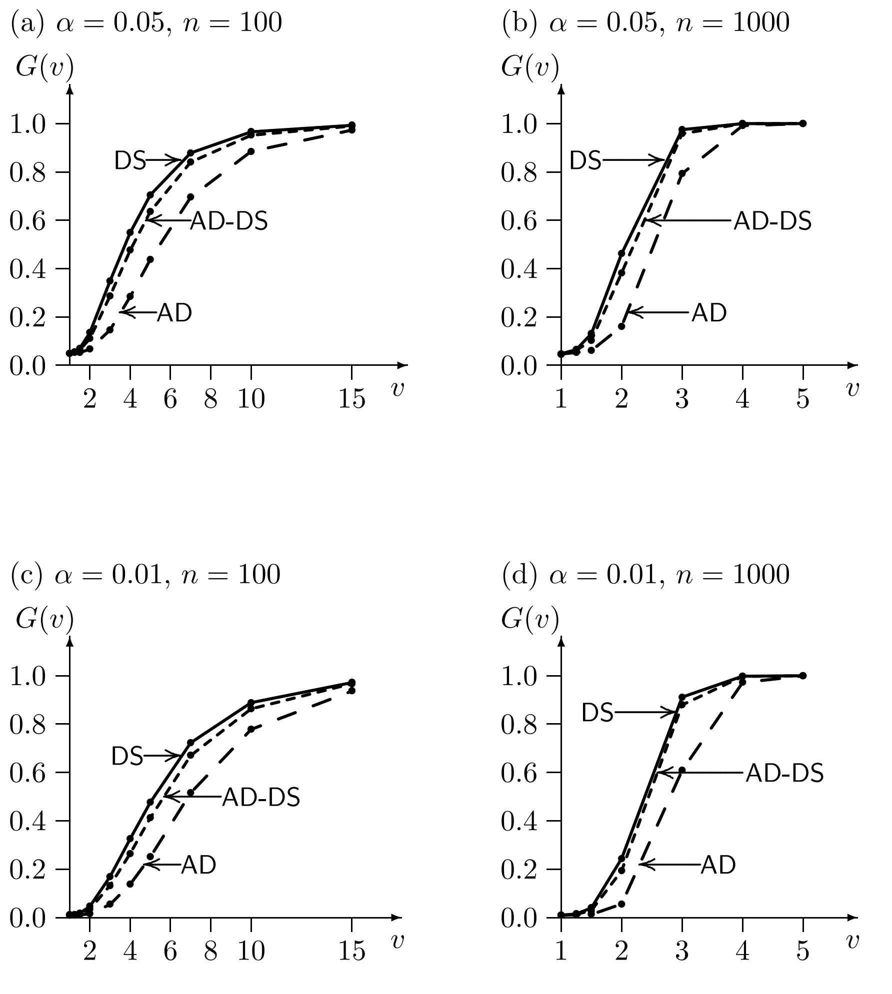


Figure 2: Power $G(v)$ of overdispersion (DS), Anderson-Darling (AD) and their combination (AD-DS) on alternatives $S(t) = 0.9\exp\left(-t^{\gamma}\right) + 0.1\exp\left(-\left(\frac{t}{v}\right)^{\gamma}\right)$ (upper contamination).

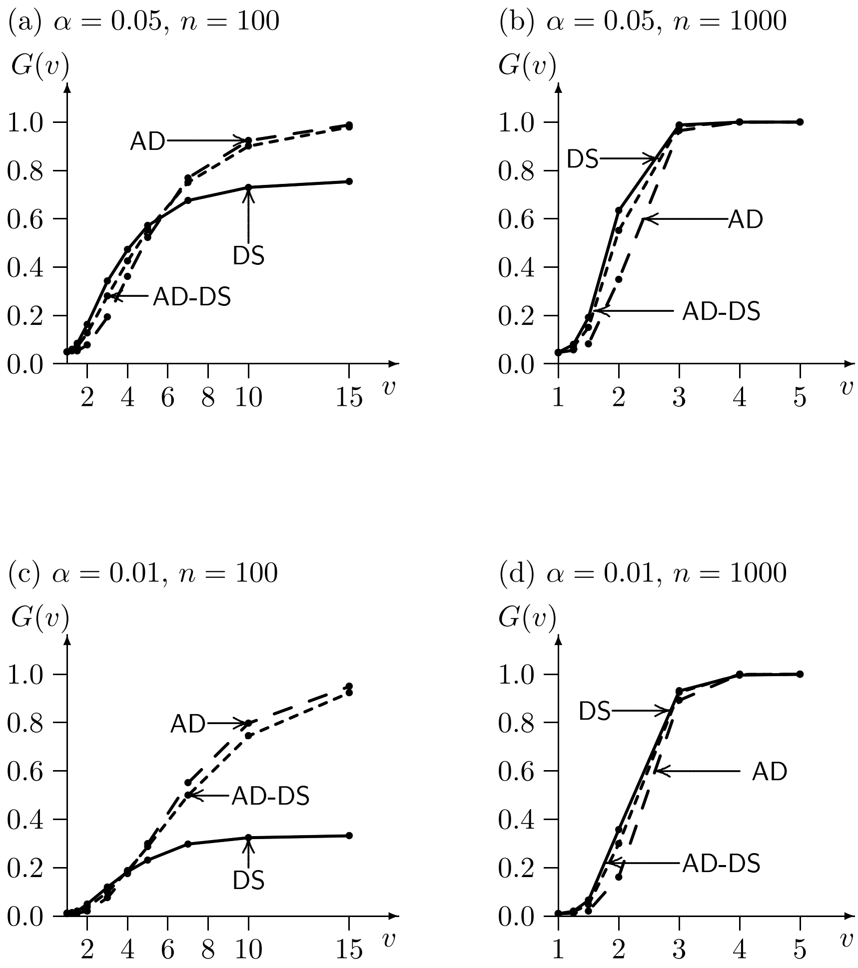


Figure 3: Power $G(v)$ of overdispersion (DS), Anderson-Darling (AD) and their combination (AD-DS) on alternatives $S(t) = 0.7\exp\left(-t^{\gamma}\right) + 0.3\exp\left(-\left(\frac{t}{v}\right)^{\gamma}\right)$. (upper contamination)

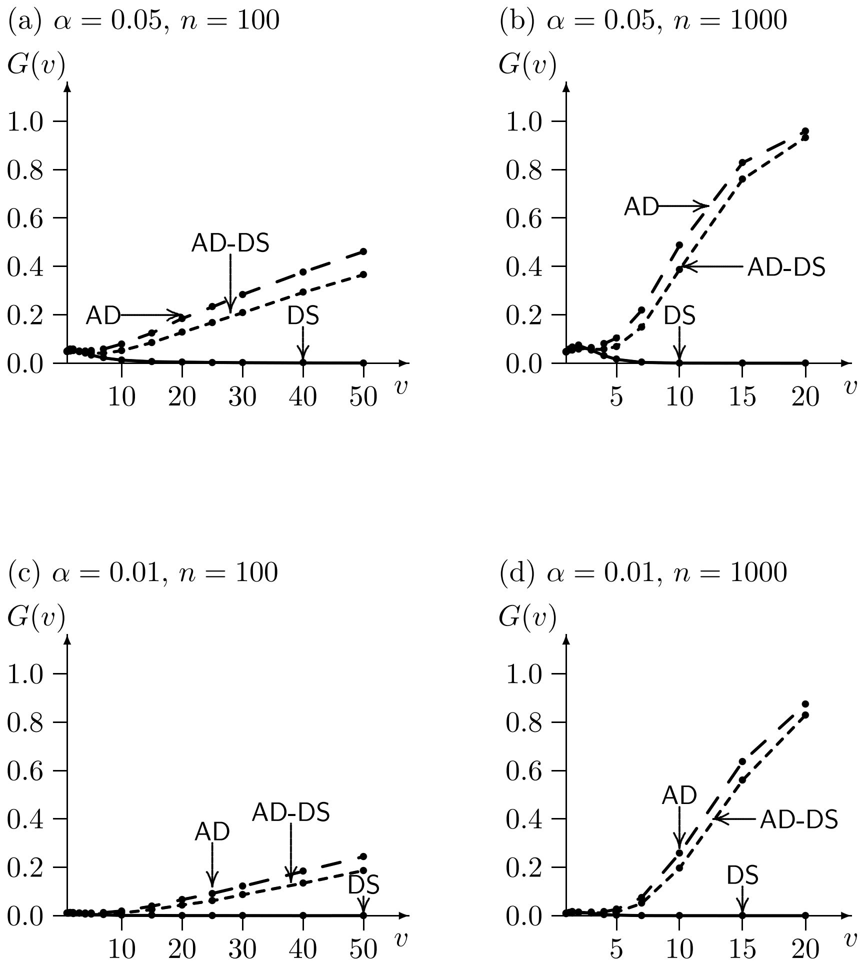


Figure 4: Power $G(v)$ of overdispersion (DS) , Anderson-Darling (AD) and their combination (AD-DS) on alternatives $S(t) = 0.1 \exp\left(-(vt)^{\gamma}\right) + 0.9 \exp\left(-t^{\gamma}\right)$ (lower contamination).

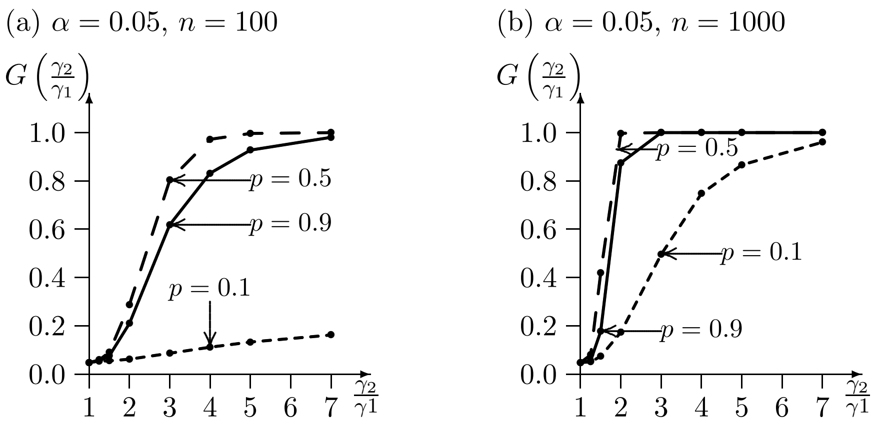


Figure 5: Power $G\left(\frac{\gamma_2}{\gamma_1}\right)$ of Anderson-Darling (AD) test on alternatives $S(t) = p \exp\left(-\left(\frac{t}{\beta}\right)^{\gamma_1}\right) + (1-p)\exp\left(-\left(\frac{t}{\beta}\right)^{\gamma_2}\right)$.

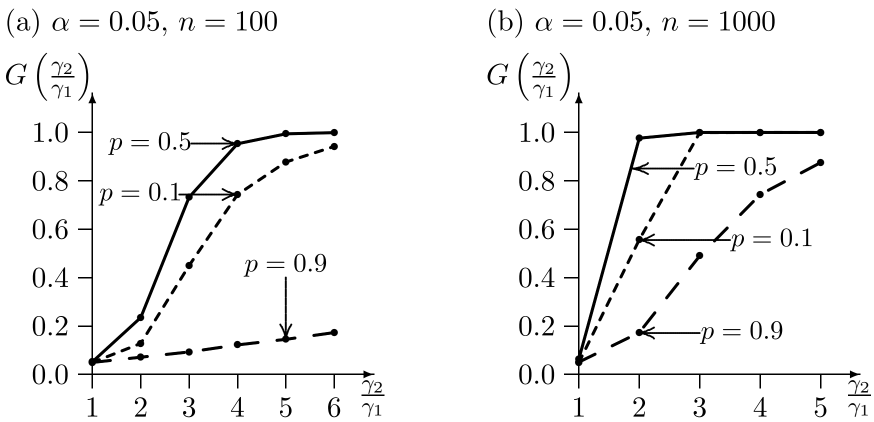


Figure 6: Power $G\left(\frac{\gamma_2}{\gamma_1}\right)$ of Anderson-Darling (AD) test on alternatives $S(t) = p \exp\left(-\left(\frac{t}{\beta}\right)^{\gamma_1}\right) + (1-p)\exp\left(-\left(\frac{t}{30\beta}\right)^{\gamma_2}\right)$.

(a) $\alpha = 0.05$, $n = 1000$

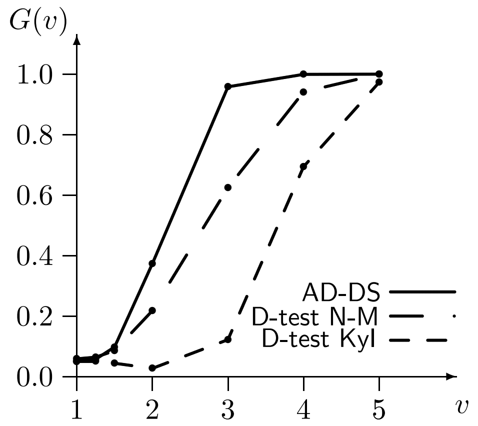


Figure 7: Power $G(v)$ of the AD-DS test and the D-Test for Weibull Mixtures ($\gamma = 1$) using the Nelder-Mead estimator (D-test N-M) resp. the Kaylan estimator (D-test KyI) on alternatives $S(t) = 0.9 \exp\left(-t^{\gamma}\right) + 0.1 \exp\left(-\left(\frac{t}{v}\right)^{\gamma}\right)$ (upper contamination).

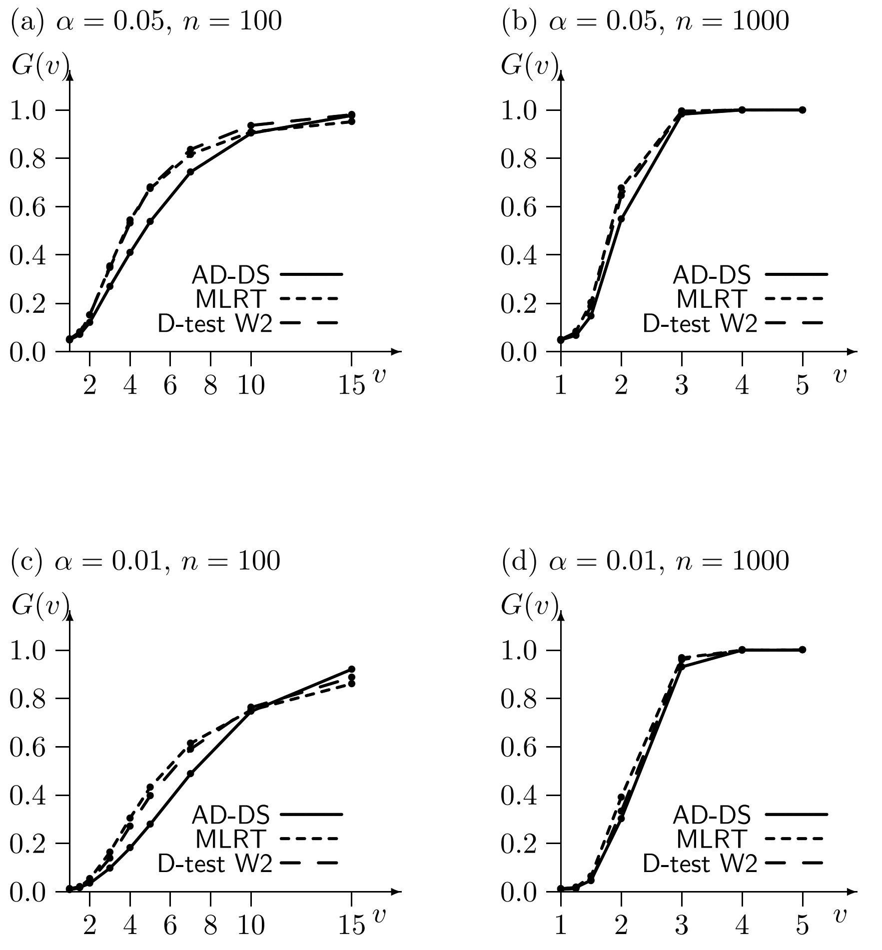


Figure 8: Power $G(v)$ of the AD-DS test, the D-test with weighting function $w_2(t) = t^2$ (D-test W2) and the modified likelihood ratio test (MLRT) on alternatives $S(t) = 0.7 \exp(-t^\gamma) + 0.3 \exp\left(-\left(\frac{t}{v}\right)^\gamma\right)$ (upper contamination), both D-test and MLRT after Weibull-to-exponential transformation, $\gamma = 1$.

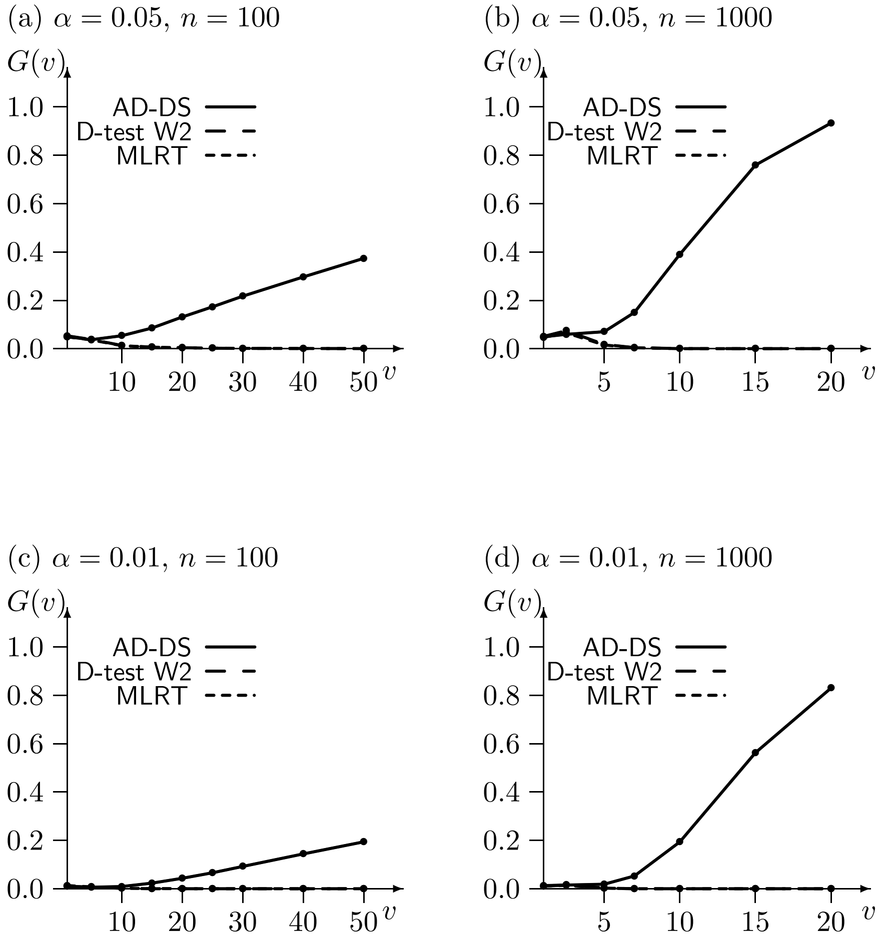


Figure 9: Power $G(v)$ of the AD-DS test, the D-test with weighting function $w_2(t) = t^2$ (D-test W2) and the modified likelihood ratio test (MLRT) on alternatives $S(t) = 0.9 \exp(-t^{\gamma}) + 0.1 \exp(-(vt)^{\gamma})$ (lower contamination), both D-test and MLRT after Weibull-to-exponential transformation. Note that both the power of the MLRT and the D-test approaches zero when v increases, $\gamma = 1$.