University of Wollongong

## Research Online

2015

# Smooth tests of fit for finite mixture distributions

T Suesse
*University of Wollongong*

J. C. W Rayner
*University of Wollongong*

O Thas
*University of Wollongong*

Follow this and additional works at: https://ro.uow.edu.au/niasrawp

# Smooth tests of fit for finite mixture distributions

## Abstract

Mixture distributions have become a very flexible and common class of distributions, used in many different applications, but hardly any literure can be found on tests for assessing their goodnes of fit. We propose two types of smooth tests of goodness of fit for mixture distributions. The first test is a genuine smooth test, and the second test makes explicitly use of the mixture structure. In a simulation study the tests are compared to some traditional goodness of fit tests that, however, are not customised for mixture distributions. The first smooth test has overall good power and generally outperforms the other tests. The second smooth test is particularly suitable for assessing the fit of each component distribution separately. The tests are applicable to both continuous and discrete distributions and they are illustrated on three example data sets.

# National Institute for Applied Statistics Research Australia

## University of Wollongong

## Working Paper

04-15

Smooth Tests of Fit for Finite Mixture Distributions

T. Suesse, J.C.W Rayner and O. Thas

# Smooth Tests of Fit for Finite Mixture Distributions

T. Suesse[1], J.C.W. Rayner[1,2] and O. Thas[1,3]

[1] National Institute for Applied Statistics Research Australia, University of Wollongong,
NSW 2522, Australia

[2] School of Mathematical and Physical Sciences, University of Newcastle,
NSW 2308, Australia

[3] Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University,
Coupure Links 653, B-9000 Gent, Belgium
*email:*Olivier.Thas@UGent.be

March 11, 2015

## Abstract

Mixture distributions have become a very flexible and common class of distributions, used in many different applications, but hardly any literure can be found on tests for assessing their goodnes of fit. We propose two types of smooth tests of goodness of fit for mixture distributions. The first test is a genuine smooth test, and the second test makes explicitly use of the mixture structure. In a simulation study the tests are compared to some traditional goodness of fit tests that, however, are not customised for mixture distributions. The first smooth test has overall good power and generally outperforms the other tests. The second smooth test is particularly suitable for assessing the fit of each component distribution separately. The tests are applicable to both continuous and discrete distributions and they are illustrated on three example data sets.

Keywords: goodness of fit, hypothesis testing, score test

## 1 Introduction

Although mixture distributions were studied by Karl Pearson more than 100 years ago, it is only in the last one or two decades that they have become popular in daily statistical practice. The book of McLachlan and Peel (2000) gives an excellent overview of the estimation theory and the scope of finite mixture distributions. A mixture distribution of $m$ component distributions is defined as follows. Let $f_i(.; \eta_i)$, $i = 1, \ldots, m$, denote $m$ density functions that make up the mixture, in which the $\eta_i$ are the corresponding nuisance parameter vectors. The $m$ distributions are referred to as the components of the

mixture distribution. The density function of the mixture distribution is given by

$$f(x; p, \eta) = \sum_{i=1}^{m} p_i f_i(x; \eta_i),\tag{1}$$

where the $p_i$ are the mixing proportions, subject to $0 < p_i < 1$ and $\sum_{i=1}^{m} p_i = 1$. The parameters can be estimated by means of the maximum likelihood method, for which the EM-algorithm (Dempster, Laird, & Rubin, 1977) is generally accepted as the most appropriate implementation for mixture models. For notational convenience we will often use $\beta^t = (p^t, \eta^t)$, where $p$ and $\eta$ are vectors containing the $\eta_i$ ($i = 1, \ldots, m$) and $p_i$ ($i = 1, \ldots, m-1$) parameters. Note that $p_m$ is not part of the parameter vector $\beta$, because $p_m = 1 - \sum_{i=1}^{m-1} p_i$.

Mixture distributions are also popular in classification problems. In this paper we consider an example from genomics in which the methylation status (methylated or not) of the MGMT gene is to be predicted. See Vlassenbroeck et al. (2008) for more details. There appears to be clinical evidence that in glioblastoma cancer patients alkylating agent therapy, such as with temozolomide, is more likely to be successful in patients with methylated MGMT genes. Although molecular techniques for the exact determination of the methylation status exist, they are not feasible in daily clinical practice. Instead a high-throughput molecular method (rt-qPCR) is used, but this technique does not give a binary outcome for the methylation status. Figure 1 shows histograms of the continuous outcomes from two data sets, each with 150 patients, some of whom are believed to have a methylated MGMT gene. The histograms clearly show bimodal distributions. Vlassenbroeck et al. (2008) fitted a two-component normal mixture to the data and used this to find the optimal threshold to discriminate between methylated or not-methylated. This approach implicitly assumes that the bio-assay outcomes of methylated MGMT genes can be described by a normal distribution and that the outcomes of non-methylated MGMT genes can be described by another normal distribution. A mixture distribution may also be used to compute posterior probabilities that an observed outcome, say $x$, has been generated by a particular component, say $f_i$. To be more specific, if $p_i$ is given the interpretation of the prior probability P {an observation has been generated by $f_i$}, then, upon using Bayes rule, $p_i f_i(x)/f(x) = $ P {an observation has been generated by $f_i \mid x$}, i.e. given the observed $x$, $p_i f_i(x)/f(x)$ gives the probability that it belongs to the $i$-th component. The result obviously depends on the appropriateness of the component densities $f_i$ and the mixing proportions for describing the data distribution.

Despite the increasing usage of mixture distributions, the authors have no knowledge of published goodness of fit tests for assessing the mixture distributional assumption, except for an ad-hoc method proposed by Hawkins, Muller, and ten Krooden (1982). They suggested first fitting the mixture distribution to the data and using the fitted distribution to classify each sample observation into one of the $m$ clusters by means of the posterior probabilities. Subsequently, upon using a classical test, goodness of fit of each individual component distribution can be tested using the observations classified to cluster $i$.

Part of the goodness of fit is the correctness of the number of components ($m$). Several methods for this problem have been described in the literature; see McLachlan and Peel (2000, chapter 6) for an overview of these methods. The methods that we introduce in this paper, however, assume $m$ is known and focus on the assessment of the fit of the mixture distribution.

In this paper we propose two smooth tests of goodness of fit for mixture distributions. In the first approach the mixture distribution is considered as any other distribution, and the test follows from methods for the construction of smooth tests (Rayner, Thas, & Best, 2009, chapter 6). The second approach explicitly makes use of the structure of a mixture distribution and allows for the assessment of the quality of the fit of each component distribution separately. For example, for the MGMT data sets shown in Figure 1, we will conclude that neither is consistent with a mixture of two normal distributions, and we will be able to identify the second component of MGMT1 and the first component of MGMT2 as the cause of the rejection of the null hypotheses. This is important information for the data-analyst when seeking an improved mixture model.

The structure of this paper is as follows. Section 2 describes the construction of the tests in detail. Their performance is evaluated in a simulation study and reported in Section 3. In Section 4 the methods are applied to some data sets. Finally, conclusions are formulated in Section 5.

## 2   Two Smooth Tests

The construction of a smooth test starts by embedding the hypothesised density into a family of alternative densities, indexed by $k$ parameters, the elements of $\theta$ say, so that the hypothesised density results from setting $\theta = 0$. The smooth test is basically the efficient score test for testing $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ within this alternative density, which is referred to as the smooth alternative, because it varies smoothly from the hypothesised density when $\theta$ moves away from zero. The two tests that we propose, differ in the set up of the smooth alternatives.

### 2.1   The Smooth Mixture Test

We now apply the theory in Rayner et al. (2009, chapter 6) to derive smooth tests of goodness of fit. Given a random sample of i.i.d. observations $X_1, \ldots, X_n$, the construction method for the smooth test statistic is given by Rayner et al. (2009, Theorem 6.1.3, p. 100). The construction requires a smooth alternative density function that involves a set of orthonormal functions. Details follow in the next paragraphs.

Let $\theta^t = (\theta_1, \ldots, \theta_k)$ and let $\{h_r(x; \beta)\}$ be a set of polynomials orthogonal on the mixture density $f(x; \beta)$, i.e. the polynomials satisfy

$$\int_{-\infty}^{+\infty} h_r(x; \beta) h_s(x; \beta) f(x; \beta) dx = \delta_{rs} \quad (r, s = 1, 2, \ldots),$$

where $\delta_{rs} = 1$ if $r = s$ and 0 otherwise. For common distributions (e.g. normal, exponential, ...) the polynomials are often in the literature, but for mixture distributions a more generic method is required. Rayner, Thas, and De Boeck (2008) proposed recurrence relations that may be used for almost any density function for which the moments are known. Here we can take advantage of the linear structure of a mixture distribution that allows the expression of the moments of $f$ as a function of the moments of the component distributions. More details are given in Appendix A.

Let $h^t(x;\beta) = (h_1(x;\beta),\ldots,h_k(x;\beta))$. Consider the smooth alternative given by

$$g(x;\theta,\beta) = C(\theta,\beta)\exp\{\theta^t h(x;\beta)\}f(x;\beta), \tag{2}$$

where $C(\theta,\beta)$ is a normalisation constant and $f(x;\beta)$ is the mixture density function (1). The smooth test requires estimates of the nuisance parameters $p$ and $\eta$ under the null hypothesis. Let $\hat{p}$ and $\hat{\eta}$ denote the maximum likelihood estimators (MLE) of $p$ and $\eta$ in the mixture distribution. Let $\hat{\beta}^t = (\hat{p}^t, \hat{\eta}^t)$. Following the theory of Rayner et al. (2009, Theorem 6.1.3), the smooth test statistic is given by

$$S_k = V^t(\hat{\beta})M^{-1}(\hat{\beta})V(\hat{\beta}),$$

where $V(\hat{\beta}) = (V_1(\hat{\beta}),\ldots,V_k(\hat{\beta}))^t$ with

$$V_r(\hat{\beta}) = \frac{1}{\sqrt{n}}\sum_{j=1}^n h_r(X_j;\hat{\beta}).$$

The matrix $M(\hat{\beta})$ equals the asymptotic covariance matrix of $V(\hat{\beta})$, evaluated under the null hypothesis, and with $\beta$ replaced by its MLE. In particular,

$$M(\beta) = I_k - \mathrm{Cov}_0\left\{h,\frac{\partial\log f}{\partial\beta}\right\}\mathrm{Var}_0\left\{\frac{\partial\log f}{\partial\beta}\right\}^{-1}\mathrm{Cov}_0\left\{\frac{\partial\log f}{\partial\beta},h\right\}. \tag{3}$$

In Appendix B we show how $M(\beta)$ can be calculated. When $\sigma_r^2(\hat{\beta})$ denotes the $r$th diagonal element of $M(\hat{\beta})$, the statistic $V_r(\hat{\beta})/\sigma_r(\hat{\beta})$ is known as the $r$th component of $S_k$ and component can be used for testing $H_{0r} : \theta_r = 0$ against $H_{1r} : \theta_r \neq 0$. As $M(\hat{\beta})$ is not diagonal these components are not uncorrelated.

The statistic $S_k$ is only well defined if $M(\hat{\beta})$ is of full rank. Often estimating functions of the $\eta$ parameters imply that $V_r(\hat{\beta}) = 0$ for one or more $r \leq k$. This results in a singular $M(\hat{\beta})$. One solution is to modify the smooth alternative (2) by removing the corresponding $\theta_r h_r(x;\beta)$ terms; the resulting $S_k$ test statistic is now well defined.

Given that the regularity conditions of Theorem 6.1.3 in Rayner et al. (2009, p.100) hold for the mixture distribution and that the first $q$ $\theta_r h_r(x;\beta)$ terms have been removed, the test statistic $S_k$ asymptotically has a $\chi^2_{k-q}$ null distribution. The test is referred to as the smooth mixture (SM) test. The components $V_r(\hat{\beta})/\sigma_r(\hat{\beta})$ are asymptotically standard normal under the null hypothesis.

## 2.2 The Component Focused Smooth Test

For the second type of smooth test we require $m$ sets of polynomials so that the $i$th set, say $\{h_{ir}(x;\eta_i)\}$, contains polynomials that are orthonormal with respect to the component density $f_i$, i.e. the polynomials satisfy

$$\int_{-\infty}^{+\infty} h_{ir}(x;\eta_i)h_{is}(x;\eta_i)f_i(x;\eta_i)dx = \delta_{rs} \quad (r,s = 1,2\ldots).$$

4

For many common densities $f_i$ the orthonormal polynomials are listed in Rayner et al. (2009) and thus no recurrence relations are required.

Let $h_i(x; \eta_i) = (h_{i1}(x; \eta_i), \ldots, h_{ik}(x; \eta_i))^t$. Consider now smooth alternatives to the hypothesised mixture distribution constructed as a mixture distribution of smooth alternatives to the component densities of $f$,

$$g(x; \theta, \beta) = \sum_{i=1}^{m} p_i C_i(\theta_i, \eta_i) \exp\{\theta_i^t h_i(x; \eta_i)\} f_i(x; \eta_i), \tag{4}$$

where $\theta_i = (\theta_{i1}, \ldots, \theta_{ik})^t$ and $C_i(\theta_i, \eta_i)$ is the normalisation constant of the $i$th component density in (4). In contrast to the $\theta$ parameters in (2), the embedding $\theta_{ir}$ parameters in (4) refer to a particular component density. We anticipate that this will allow for more focussed tests that may be used for assessing the fit of each component distribution separately.

Whereas the SM test is a genuine smooth test in the sense that it is an efficient score test related to smooth alternative (2), the test constructere here is a smooth test of a different kind because its smooth alternative (4) has a different structure. However, the test is still constructed as an efficient score test.

The test again needs the MLEs of $p$ and $\eta$ under the null hypothesis, and thus $\hat{p}$ and $\hat{\eta}$ are found exactly as before. In Appendix C we demonstrate that the score statistic for testing $H_0 : \theta_{ir} = 0$ against $H_1 : \theta_{ir} \neq 0$ is

$$V_{ir}(\hat{\beta}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tau_i(X_j; \hat{\beta}) h_{ir}(X_j; \hat{\eta}_i), \tag{5}$$

where $\tau_i(x; \hat{\beta}) = \hat{p}_i f_i(x; \hat{\eta}_i) / f(x; \hat{\beta})$ is the estimated posterior probability that $x$ has been generated by the $i$th component distribution. The form of the $V_{ir}(\hat{\beta})$ statistic resembles $\sum_{j=1}^{n} h_{ir}(X_j; \hat{\eta}_i)/\sqrt{n}$, which form the basis of the smooth test for assessing the fit of the sample observations to the single density $f_i$. The factor $\tau_i(X_j; \hat{\beta})$, which appears in (5), weights the observations with their estimated a posteriori probability of being generated by $f_i$.

Let

$$V^t(\hat{\beta}) = (V_{11}(\hat{\beta}), V_{12}(\hat{\beta}), \ldots, V_{1k}(\hat{\beta}), V_{21}(\hat{\beta}), \ldots, V_{mk}(\hat{\beta})).$$

The asymptotic covariance matrix of $V^t(\hat{\beta})$, say $M(\beta)$, is of the same form as (3), but in general no simple analytical expressions are available so that numerical integration is required. Details are provided in Appendix B.

From efficient score test theory it follows that $V^t(\hat{\beta}) M^{-1}(\hat{\beta}) V(\hat{\beta})$ asymptotically has a $\chi^2_{mk}$ null distribution, and the individual component test statistics, $V_{ir}(\hat{\beta})/\sigma_{ir}(\hat{\beta})$, with $\sigma^2_{ir}(\hat{\beta})$ the appropriate diagonal element of $M(\hat{\beta})$, have asymptotically standard normal null distributions. We also define the component-specific test statistics ($i = 1, \ldots, m$)

$$S_{ik} = V_i^t(\hat{\beta}) M_i^{-1}(\hat{\beta}) V_i(\hat{\beta})$$

with $V_i^t(\hat{\beta}) = (V_{i1}(\hat{\beta}), V_{i2}(\hat{\beta}), \ldots, V_{ik}(\hat{\beta}))$ and $M_i(\hat{\beta})$ the estimated asymptotic covariance of $\hat{V}_i^t$ (elements taken from $M(\hat{\beta})$). Under the null hypothesis $S_{ik}$ is asymptotically $\chi^2_k$ distributed. The test can be used for testing the goodness of fit for each component distribution separately. The test is referred to as the component focused smooth (CFS) test.

## 2.3 Smooth Tests for Mixtures of Discrete Distributions

Smooth tests 1 and 2 may also be developed for mixtures of discrete distributions of ordinal random variables. The theory essentially remains unaltered, except that the orthonormal polynomials are now defined on the sample space of the ordinal observations. The integrals in the orthonormality conditions have to be replaced with sums over all elements of the sample space. We refer again to Rayner et al. (2009) for details on smooth tests for discrete distributions.

# 3 Simulation Study

## 3.1 Settings

The performance of the proposed smooth tests is assessed in a simulation study. To reduce complexity, but still covering continuous and discrete distributions, only testing for a mixture of normal distributions (with $m = 2, 3$ and 4) and for a mixture of $m = 2$ Poisson distributions is considered.

All computations are performed using R (R Core Team, 2014) and the R packages `mixtools` and `gamlss.mx`. For each setting, 2000 data sets of size $n = 200$ are generated. All tests were performed at the 5% level of significance. The asymptotic null distributions do not give good approximations; see Web Tables 1, 2, 5, 6 and 7. Therefore, based on 500 bootstrap samples, the parametric bootstrap method has been used for the calculation of $p$-values.

Alternatives to the null hypotheses of a normal mixture and a Poisson mixture were constructed by replacing one or more hypothesised component distributions by another distribution. For the mixtures of normal, $t_5$, gamma and lognormal distributions were considered. The alternative component distributions were shifted and scaled to have to the same mean and variance as the normal component distributions. In particular the means were set to $\mu_1 = 1$, $\mu_2 = 6$ (for $m = 2, 3$ and $m = 4$), $\mu_3 = 12$ (for $m = 3$ and $m = 4$) and $\mu_4 = 18$ (for $m = 4$). The mixing probabilities were set to $p_1 = 0.4$ and $p_2 = 0.6$ for $m = 2$, $p_1 = 0.2$, $p_2 = 0.4$ and $p_3 = 0.4$ for $m = 3$, and $p_1 = 0.2$, $p_2 = 0.2$, $p_3 = 0.2$ and $p_3 = 0.4$ for $m = 4$. Tables 1 and 2 show the powers of the smooth tests for $m = 2$ and $m = 3$, respectively, for testing the null hypothesis of a mixture of normals with common variances ($\sigma_i^2 = 2; i = 2, 3$). Results for $m = 4$ and for unequal variances are presented in Web Tables 8 up to 13.

For the mixture of two Poisson distributions, binomial and negative binomial (NB) distributions were considered as alternative component distributions. The means of the Poisson components are set to 4 and 10. The size parameter of the binomial distribution is fixed at 20 and the probability parameter is set to 0.2 or 0.5 so as to obtain the same mean as the Poisson component. The mixing probabilities were set to $p_1 = 0.4$ and $p_2 = 0.6$ The results are presented in Table 4; all powers are based on bootstrap $p$-values (500 bootstrap samples). Results based on the asymptotic null distributions are provided in Web Table 4.

Table 5 shows the powers of three competitor goodness of fit tests: the Chernoff-Lehmann (CL) test (dividing the observations in 10 intervals with equal estimated frequencies across intervals under the null

hypothesis), the Anderson-Darling (AD) test and the Kolmogorov-Smirnov (KS) test. The parametric bootstrap procedure has been used for $p$-value calculations to place all tests on a similar footing in respect of distributional issues.

## 3.2 Results

Under the null hypotheses the type I error rates of the bootstrap tests are quite well controled at the nominal significance level (see Tables 1 to 4). Some tests are slightly conservative. Use of the asymptotic null distribution is less succesful at controlling the type I error rates (see Web Tables 1, 2, 5, 7 and 8), and we recommend use of the parametric bootstrap.

To keep the exposition brief and focussed, we will give a general discussion rather than discussing all tests and scenarios separately. Therefore, the competitor tests will be compared with the SM test. The results of the CFS test will be discussed in terms of its diagnostic ability to identify what component distribution deviates from the null hypothesis

Among the competitor tests, the AD test is always most powerful. See Table 5. Apart from a few alternatives, the SM test does always better than the AD test. The power of the SM test depends on the order $r$. The simulation results suggest that $r = 4$ or $r = 5$ gives overall good results for $S_r$ and $S_{ir}$.

The $V_{ir}$ components of the CFS test are for many scenarios diagnostic in the sense that $V_{ir}$ has power when component distribution $i$ deviates from the component distribution as specified under $H_0$. This diagnostic property seems to diminish when the deviating component distribution has tails that overlap with other component distributions. This can, for example, be clearly seen from the gamma/gamma alternative in Table 1: the $V_{2r}$ tests are powerful, whereas the $V_{1r}$ tests are not (the long right tail of the first component distribution is masked by the second).

The $V_{ir}$ component tests do not succeed well in discovering the moments that deviate from the hypothesised component distribution. For example, for the $t/t$ alternative (Table 1) the $V_{i3}$ and $V_{i5}$ component tests show substantial powers, whereas $t$ distributions are symmetric, as are the hypothesised normal distributions. On other occasions, however, asymmetry can be detected (e.g. $V_{13}$ for gamma/normal, $V_{13}$ and $V_{15}$ for lognormal/normal) as can a deviation in terms of kurtosis (e.g. $V_{14}$ for the $t$/normal and $t$/normal/normal).

From all the results presented here and in Web-based Supplementary Material, we conclude that the new smooth tests are overall more powerful than the competitor tests considered. The SM test seems a good choice if no assessment of individual components is wanted; we recommend $r = 4$ or $r = 5$. The components of the CFS test often succeed in identifying the component distributions that deviate from those hypothesised, but when the tails of the deviating component distributions are masked by the other component distributions, this propery diminishes. In some circumstances the components of the CFS test also succeed in detecting the moments in which the component distributions differ from the hypothesised, but often this property fails.

All conclusions formulated here do equally well apply to $m = 4$ and the unequal-variance situation (Web Tables 8 up to 13).

# 4 Example

## 4.1 MGMT

We analyse the MGMT data sets with the two types of smooth tests. First a mixture of two normal distributions is fitted to the data. The data and the fits are shown in Figure 1. Next the two smooth tests, and their component tests are applied for several orders $r = 3, 4, 5$ and $r = 6$. The $p$-values are computed using the bootstrap with $5000$ bootstrap samples for $p$-value calculation. The results are presented in Table 3. The three competitor tests (CL, AD and KS) have also been applied with parametric bootstrap; the results are also shown in Table 3. All tests are performed at the $5\%$ level of significance.

For both MGMT data sets, the SM test and the competitor tests agree to reject the null hypothesis. From the component tests of the CFS test applied to MGMT1, we conclude that the first component distribution fits the data well, but the second does not. For MGMT2 the component tests suggest that the first component distribution does not fit the data well, but the second does.

## 4.2 Snapper Length

Cassie (1954) reported lengths of 256 red snappers taken by a trawl with a mesh of about 1.5 in. The data are shown in Web Figure 1 and clearly show a multimodal distribution. Since snappers typically spawn in a particular season, and since young snappers show a fast and a slow growing season, growing about 10cm per year, Cassie hypothesised that data can be described by a mixture distribution, with each component distribution describing the length distribution of fish of a particlar age (0, 1, 2, 3, ... years). The question is whether he observed data from four of five age classes and whether the age-specific lengths can be modelled with normal distributions. We fitted mixtures with $m = 3$ and $m = 4$ normal components. The fits are depicted in Web Figure 1 and the test results are presented in Table 3 (tests are performed as for the MGMT data).

The AD and KS tests reject the null hypotheses for $m = 3$ and $m = 4$, whereas the smooth tests do not reject when $m = 4$. When $m = 3$ the CFS test suggests that the first component distribution does not fit well. We have not included the results for $m = 5$, because with $m = 4$ the test results already indicate a good fit.

We should remark that that the data contain many ties as a consequence of severe rounding. The observed AD and KS test statistics, which are functionals of the difference between the empirical distribution function and the fitted mixture distribution function, are therefore large, whereas the test statistics calculated from the parametric bootstrap samples do not suffer from this issue as the bootstrap samples are unlikely to contain ties. This may explain the extremely small $p$-values for KS and AD. The CL test shows decreasing $p$-values with increasing number of components. A possible explanation is that for large $m$ the CL test should perhaps better be calculated from a larger number of intervals, but the optimisation of the CL test is outside the scope of this paper.

8

### 4.3 London Deaths

A classic data set considered by a number of authors starting with Whitaker (1914) considers deaths per day of women over 80 in London during the years 1910, 1911 and 1912 as recorded in the Times newspaper. A mixture of two Poisson distributions is anticipated, caused by different death rates in summer and winter. The data and the fit are depicted in Web Figure 2 and the test results are presented in Table 3 (tests are performed as for the MGMT data). All tests accept the null hypotheses, including the components of the CFS test.

## 5    Conclusion

Despite the enormous literature on goodness of fit tests, hardly any test for assessing the fit of mixtures distributions has been published.

In this paper we have proposed two smooth tests of goodness of fit for finite mixture distributions. A first type of smooth test considers the mixture density function as any other density function, and uses the methodology of Rayner et al. (2009) for the construction of a smooth test. A second type of smooth tests makes use of the mixture structure and constructs smooth test statistics for each component distribution separately.

From an extensive simulation study that includes mixtures of continuous and mixtures of discrete distributions, with two up to four component distributions, we have concluded that the parametric bootstrap works well for controlling the type I error rates. The first smooth test has better overall powers than any competitor test, and the second smooth test succeeds quite well in indentifying the component distributions that deviate from the hypothesised. However, when the tails of the deviating component distribution are masked by other component distributions, this diagnostic proporty diminishes.

The second smooth test may be considered as an improvement over the method of Hawkins et al. (1982) in the sense that their approach applies a crisp classification of observations to component distributions (before applying whatever test for goodness of fit), whereas we use all observations, but weighed with their posterior probabilities.

## Acknowledgements

# References

Cassie, R. (1954). Some uses of probability paper in the analysis of size frequency distributions. *Australian Journal of Marine Freshwater Research*, *5*, 513-522.

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

Hawkins, D., Muller, M., & ten Krooden, J. (1982). Topics in applied multivariate analysis. In D. Hawkins (Ed.), (p. 303-356). Cambridge University Press.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, USA: Wiley.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rayner, J., Thas, O., & Best, D. (2009). *Smooth tests of goodness of fit: Using R*. Singapore: Wiley.

Rayner, J., Thas, O., & De Boeck, B. (2008). A generalised Emerson recurrence relation. *Australian and New Zealand Journal of Statistics*, *50*, 235240.

Vlassenbroeck, I., Califice, S., Diserens, A., Migliavacca, E., Straub, J., Di Stefano, I., et al. (2008). Validation of real-time methylation-specific PCR to determine $O^6$-methylguanine-DNA methyl-transferase gene promoter methylation in glioma. *Journal of Molecular Diagnostics*, *10*, 332-337.

Whitaker, L. (1914). On the Poisson law of small numbers. *Biometrika*, *10*, 36-71.

# Appendices

# A  Moments of the mixture distribution

Let $\mu'_{ik} = \mathrm{E}_{f_i}\{X\}$ denote the $k$th noncentral moment of the $i$th component density $f_i$. The $k$th noncentral moment of the mixture distribution $f$ then follows immediately from

$$\mu'_k = \mathrm{E}_f\{X^k\} = \int_{-\infty}^{+\infty} x^k \sum_{i=1}^m p_i f_i(x) = \sum_{i=1}^m p_i \mathrm{E}_{f_i}\{X^k\} = \sum_{i=1}^m p_i \mu'_{ik}.$$

Rayner et al. (2008) and Rayner et al. (2009) give details on how orthonornal polynomials may be constructed starting from a set of moments. Starting from the moments, the recurrence relations of Rayner et al. (2008) may be used for constructing the orthonormal polynomials.

# B  Covariance matrices for smooth tests

The elements of the information matrices for tests 1 and 2 are listed below. For notational comfort we write $h$ instead of $h(X;\beta)$ (similarly for other functions) and we have suppressed the dependence on the parameters.

## The SM test

For the matrix $M(\beta)$ of the SM test we need

$$\mathrm{Cov}_f\left\{h, \frac{\partial \log f}{\partial \beta}\right\} = \left(\mathrm{Cov}_f\left\{h, \frac{\partial \log f}{\partial p}\right\} \quad \mathrm{Cov}_f\left\{h, \frac{\partial \log f}{\partial \eta}\right\}\right)^t,$$

where

$$\mathrm{Cov}_f\left\{h, \frac{\partial \log f}{\partial p}\right\} = \mathrm{Cov}_f\left\{h, \frac{f_i - f_m}{f}\right\} = \mathrm{E}_{f_i}\{h\} - \mathrm{E}_{f_m}\{h\}.$$

Because of the polynomial nature of $h$, we may write $h = Az$ with $z^t = (1, x, x^2, \ldots x^k)$ for $h$ containing polynomials up to order $k$. Hence, $\mathrm{E}_{f_i}\{h\} = A\mathrm{E}_{f_i}\{z\} = A\mu_i'$ with $\mu_i' = (1, \mu_{i1}', \ldots, \mu_{ik}')^t$ the vector with the noncentral moments of $f_i$ up to order $k$. Hence,

$$\mathrm{Cov}_f\left\{h, \frac{\partial \log f}{\partial p}\right\} = \mathrm{Cov}_f\left\{h, \frac{f_i - f_m}{f}\right\} = A\left(\mu_i' - \mu_m'\right).$$

Since each component distribution density function $f_i$ only depends on $\eta_i$ (and not on the other nuisance parameters in $\eta$), we find

$$\frac{\partial f}{\partial \eta} = \left(\tau_1 \frac{\partial f_1}{\partial \eta_1}, \ldots, \tau_m \frac{\partial f_m}{\partial \eta_m}\right)^t.$$

The matrix $\mathrm{Cov}_f\left\{h, \frac{\partial \log f}{\partial \eta}\right\}$ thus requires the elements $\mathrm{Cov}_f\left\{h, \tau_i \frac{\partial f_i}{\partial \eta_i}\right\}$. For many distributions the score function $\frac{\partial f_i}{\partial \eta_i}$ is polynomial in $x$ so that it can be written as $\frac{\partial f_i}{\partial \eta_i} = C_i h$, with $h = AZ$ as before. Under this condition,

$$\begin{aligned}
\mathrm{Cov}_f\left\{h, \tau_i \frac{\partial f_i}{\partial \eta_i}\right\} &= \mathrm{Cov}_f\left\{h, \tau_i h^t\right\} C_i^t \\
&= \int h(x) h^t(x) p_i f_i(x) dx \\
&= p_i \mathrm{E}_{f_i}\left\{hh^t\right\} \\
&= p_i A \mathrm{E}_{f_i}\left\{zz^t\right\} A^t,
\end{aligned}$$

which requires the noncentral moments of the component distribution $f_i$.

The SM test also requires the matrix

$$\mathrm{Var}_f\left\{\frac{\partial \log f}{\partial \beta}\right\} = -\begin{pmatrix} \mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial p \partial p}\right\} & \mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial \eta \partial p}\right\} \\ \mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial \eta \partial p}\right\} & \mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial \eta \partial \eta}\right\} \end{pmatrix}, \tag{6}$$

with elements

$$
\begin{aligned}
\mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial \eta_i \partial \eta_j^t}\right\} &= -\mathrm{E}_f\left\{\tau_i \tau_j \frac{\partial \log f_i}{\partial \eta_i}\frac{\partial \log f_j}{\partial \eta_j^t}\right\} \quad \text{if } i \neq j \\
&= \mathrm{E}_f\left\{\tau_i(1-\tau_i)\frac{\partial \log f_i}{\partial \eta_i}\frac{\partial \log f_i}{\partial \eta_i^t} + \tau_i\frac{\partial^2 \log f_i}{\partial \eta_i \partial \eta_i^t}\right\} \quad \text{if } i = j \\
\mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial p_i \partial p_j}\right\} &= -\mathrm{E}_f\left\{\left(\frac{\tau_i}{p_i} - \frac{\tau_m}{p_m}\right)\left(\frac{\tau_j}{p_j} - \frac{\tau_m}{p_m}\right)\right\} \\
\mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial \eta_i \partial p_j}\right\} &= -\mathrm{E}_f\left\{\tau_i\left(\frac{\tau_j}{p_j} - \frac{\tau_m}{p_m}\right)\frac{\partial \log f_i}{\partial \eta_i}\right\} \quad \text{if } i \neq j \\
&= \mathrm{E}_f\left\{\tau_i\left(\frac{1-\tau_i}{p_i} + \frac{\tau_m}{p_m}\right)\frac{\partial \log f_i}{\partial \eta_i}\right\} \quad \text{if } i = j.
\end{aligned}
$$

The estimator of $M(\beta)$ is obtained by calculating all matrices involved and subsequently substituting $\beta$ with $\hat{\beta}$. Sometimes the expression involves $\mathrm{E}_{f_i}\{zz^t\}$, which can be estimated by expressing the noncentral moments of $f_i$ in terms of $\eta_i$ and replacing the latter with its MLE. Some the integrals have no direct analytical expressions, but they may be numerically approximated for a given $\hat{\beta}$.

### The CFS test

The matrix $M(\beta)$ of the second smooth test (efficient score test) is given by

$$
M(\beta) = C(\beta) - A(\beta)B^{-1}(\beta)A(\beta),
$$

in which, with $g$ given by (4),

$$
\begin{aligned}
A(\beta) &= \mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial \theta \partial \theta^t}\right|_{\theta=0}\right\} \\
B(\beta) &= \mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial \theta \partial \beta^t}\right|_{\theta=0}\right\} \\
C(\beta) &= \mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial \beta \partial \beta^t}\right|_{\theta=0}\right\} = \mathrm{E}_f\left\{\frac{\partial^2 \log f}{\partial \beta \partial \beta^t}\right\}.
\end{aligned}
$$

Note that $C(\beta)$ equals matrix (6) of the SM test. The elements of the other two matrices are given by the following expectations:

$$\mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial\theta_{ir}\partial\theta_{js}}\right|_{\theta=0}\right\} = \mathrm{E}_f\left\{\tau_i\tau_j h_{ir} h_{js}\right\}$$

$$\mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial\theta_{ir}\partial\eta_{js}}\right|_{\theta=0}\right\} = \mathrm{E}_f\left\{\tau_i\tau_j h_{ir}\left(\frac{\partial \log f_j}{\partial\eta_{js}} - E_{f_j}\left[\frac{\partial \log f_j}{\partial\eta_{js}}\right]\right)\right\}$$

$$\mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial\theta_{ir}\partial p_i}\right|_{\theta=0}\right\} = \mathrm{E}_f\left\{\tau_i h_{ir}\left[\frac{\tau_j}{p_j} - \frac{\tau_m}{p_m}\right]\right\}$$

$$\mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial\eta_{ir}\partial\eta_{js}}\right|_{\theta=0}\right\} = \mathrm{E}_f\left\{\tau_i\tau_j\left(\frac{\partial \log f_i}{\partial\eta_{is}} - E_{f_i}\left[\frac{\partial \log f_i}{\partial\eta_{is}}\right]\right)\left(\frac{\partial \log f_j}{\partial\eta_{js}} - E_{f_j}\left[\frac{\partial \log f_j}{\partial\eta_{js}}\right]\right)\right\}$$

$$\mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial\eta_{ir}\partial p_j}\right|_{\theta=0}\right\} = \mathrm{E}_f\left\{\tau_i\left(\frac{\partial \log f_i}{\partial\eta_{is}} - E_{f_i}\left[\frac{\partial \log f_i}{\partial\eta_{is}}\right]\right)\left(\frac{\tau_j}{p_j} - \frac{\tau_m}{p_m}\right)\right\}$$

$$\mathrm{E}_f\left\{\left.\frac{\partial^2 \log g}{\partial p_i\partial p_j}\right|_{\theta=0}\right\} = \mathrm{E}_f\left\{\left[\frac{\tau_i}{p_i} - \frac{\tau_m}{p_m}\right]\left[\frac{\tau_j}{p_j} - \frac{\tau_m}{p_m}\right]\right\}.$$

For many distributions it will be difficult to find an analytical expression in terms of the parameters. In these situations the expectations may be approximated through numerical integration. The estimator of $M(\beta)$ is obtained by replacing $\beta$ by $\hat{\beta}$.

# C   Score statistic for the CFS test

The $r$th component of the score statistic to be used in the smooth test statistic for testing $H_0 : \theta_i = 0$ is given by

$$
\begin{aligned}
\left.\frac{\partial \log L(\theta,\beta)}{\partial\theta_{ir}}\right|_{\theta_i=0} &= \sum_{j=1}^{n}\left.\frac{\partial \log g(x_j;\theta,\beta)}{\partial\theta_{ir}}\right|_{\theta_i=0} \\
&= \sum_{j=1}^{n}\frac{1}{g(x_j;\theta,\beta)}\sum_{l=1}^{m}p_l\left.\frac{\partial g_l(x_j;\theta_l,\eta_l)}{\partial\theta_{ir}}\right|_{\theta_i=0} \\
&= \sum_{j=1}^{n}\frac{1}{g(x_j;\theta,\beta)}p_i g_i(x_j;\eta_i)\left.\frac{\partial \log g_i(x_j;\theta_i,\eta_i)}{\partial\theta_{ir}}\right|_{\theta_i=0} \\
&= \sum_{j=1}^{n}\frac{p_i g_i(x_j;\eta_i)}{g(x_j;\theta,\beta)}\left(\frac{\partial \log C_i(x_j;\theta_i,\eta_i)}{\partial\theta_{ir}} + h_{ir}(x_j;\eta_i)\right)\bigg|_{\theta_i=0} \\
&= \sum_{j=1}^{n}\tau_i(x_j;\beta)h_{ir}(x_j;\eta_i).
\end{aligned}
$$

The final step follows from

$$\frac{\partial \log C_i(\theta_i,\eta_i)}{\partial\theta_i} = -\mathrm{E}_{g_i}\left\{h_{ir}(X;\eta_i)\right\}, \tag{7}$$

13
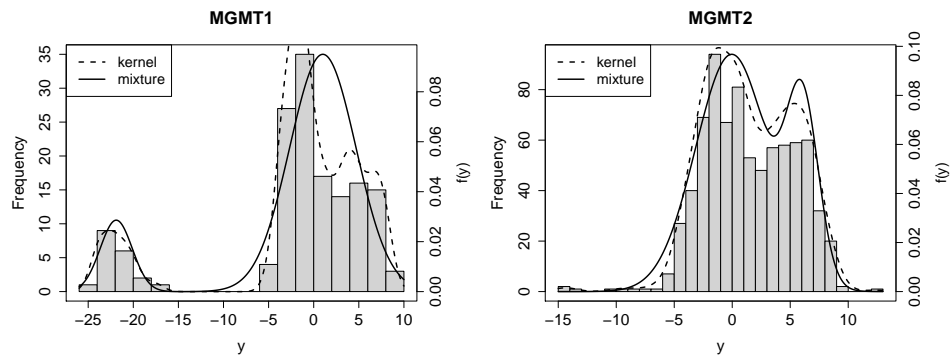
Figure 1: Histograms, nonparametric kernel density estimates and the fitted two-component normal mixtures of the continuous outcome of the MGMT-methylation assay for two data sets (MGMT1 and MGMT2).

which is zero when $\theta_i = 0$. Identiy (7) follows from differentiating w.r.t. $\theta_i$ of both sides of

$$\int g(x; \theta, \beta) dx = 1.$$

Table 1: Powers (%) of smooth tests for normal mixtures using parametric bootstrap. The mixtures have $m = 2$ components with means $\mu_1 = 1$ and $\mu_2 = 7$, and variances $\sigma_1^2 = \sigma_2^2 = 2$. Normal (N), $T_5$ (T), gamma (G) and lognormal (L) component distributions were considered. Powers are based on 2000 Monte Carlo runs of samples of size $n = 200$

| | order | SM | | CFS | | | |
| | | $V_r$ | $S_r$ | $V_{1r}$ | $V_{2r}$ | $S_{1r}$ | $S_{2r}$ |
|---|---|---|---|---|---|---|---|
| N/N | 3 | 6.4 | 6.4 | 5.9 | 5.6 | 5.9 | 5.6 |
| | 4 | 5.8 | 6.8 | 5.6 | 5.6 | 5.6 | 6.1 |
| | 5 | 6.4 | 6.5 | 5.9 | 4.3 | 5.6 | 5.5 |
| | 6 | 6.2 | 6.6 | 5.6 | 4.7 | 5.2 | 5.2 |
| T/N | 3 | 19.9 | 19.9 | 28.9 | 9.5 | 28.9 | 9.5 |
| | 4 | 27.2 | 38 | 41.8 | 8.0 | 41 | 9.9 |
| | 5 | 29.9 | 37.6 | 27.5 | 6.9 | 39.9 | 12 |
| | 6 | 33.7 | 37.0 | 31.9 | 6.1 | 38.5 | 11.0 |
| T/T | 3 | 29.1 | 29.1 | 36.7 | 41.0 | 36.7 | 41.0 |
| | 4 | 61.5 | 60.4 | 44.6 | 54.9 | 42.8 | 51.4 |
| | 5 | 52.8 | 63.4 | 32.4 | 38.3 | 42.4 | 51.2 |
| | 6 | 71.9 | 67.3 | 34.2 | 42.9 | 42.4 | 52.0 |
| G/N | 3 | 13.4 | 13.4 | 17.1 | 6.2 | 17.1 | 6.2 |
| | 4 | 12.6 | 13.2 | 1.0 | 7.2 | 11.5 | 6.3 |
| | 5 | 6.4 | 8.1 | 10.6 | 5.6 | 9.3 | 7.5 |
| | 6 | 3.1 | 8.4 | 3.5 | 5.8 | 7.5 | 8.2 |
| G/G | 3 | 55.2 | 55.2 | 6.8 | 58.3 | 6.8 | 58.3 |
| | 4 | 23.8 | 50.9 | 1.2 | 33.6 | 11.2 | 54.6 |
| | 5 | 50.7 | 50.0 | 4.3 | 29.8 | 8.4 | 51.6 |
| | 6 | 26.7 | 45.3 | 3.3 | 23.5 | 8.1 | 50.1 |
| L/N | 3 | 28.2 | 28.2 | 38.5 | 8.0 | 38.5 | 8.0 |
| | 4 | 17.6 | 29.6 | 8 | 8.8 | 29.8 | 8.5 |
| | 5 | 10.6 | 17.9 | 24.2 | 7.0 | 23.0 | 8.7 |
| | 6 | 3.6 | 17.6 | 3.0 | 6.2 | 18.0 | 11.0 |
| L/L | 3 | 79.9 | 79.9 | 10.7 | 84.1 | 10.7 | 84.1 |
| | 4 | 52.5 | 78.2 | 2.6 | 58.0 | 21.3 | 80.7 |
| | 5 | 82.0 | 78.5 | 9.4 | 50.0 | 16.5 | 78.6 |
| | 6 | 53.4 | 74.9 | 2.9 | 42.3 | 15.1 | 78.1 |

Table 2: Powers (%) of smooth tests for normal mixtures using parametric bootstrap. The mixtures have $m = 3$ components with means $\mu_1 = 1$, $\mu_2 = 6$ and $\mu_3 = 12$ and variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$. Normal (N), $T_5$ (T), gamma (G) and lognormal (L) component distributions were considered. Powers are based on 2000 Monte Carlo runs of samples of size $n = 200$.

| | order | SM | | CFS | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | $V_r$ | $S_r$ | $V_{1r}$ | $V_{2r}$ | $V_{3r}$ | $S_{1r}$ | $S_{2r}$ | $S_{3r}$ |
| N/N/N | 3 | 5.7 | 5.7 | 5.2 | 3.9 | 4.5 | 5.2 | 3.9 | 4.5 |
| | 4 | 5.1 | 5.1 | 4.4 | 4.5 | 4.4 | 4.8 | 3.9 | 4.6 |
| | 5 | 5.6 | 5.3 | 4.9 | 4.6 | 5.6 | 5.1 | 2.8 | 4.1 |
| | 6 | 5.1 | 4.9 | 4.8 | 5.3 | 5.0 | 4.2 | 2.1 | 4.5 |
| T/N/N | 3 | 12.0 | 12.0 | 19.3 | 4.4 | 6.6 | 19.3 | 4.4 | 6.6 |
| | 4 | 14.0 | 15.8 | 24.2 | 7.6 | 5.3 | 23.6 | 5.0 | 6.3 |
| | 5 | 16.3 | 19.9 | 16.0 | 4.2 | 6.0 | 24.0 | 4.1 | 6.4 |
| | 6 | 20.3 | 22.0 | 21.8 | 6.4 | 5.4 | 22.0 | 4.0 | 6.5 |
| T/T/T | 3 | 25.1 | 25.1 | 27.7 | 5.2 | 36.1 | 27.7 | 5.2 | 36.1 |
| | 4 | 37.4 | 39.2 | 27.5 | 14.2 | 41.8 | 27.7 | 12.0 | 40.0 |
| | 5 | 40.7 | 46.8 | 23.6 | 6.1 | 32.3 | 27.7 | 8.7 | 39.9 |
| | 6 | 54.2 | 52.1 | 23.4 | 12.2 | 33.5 | 27.9 | 9.1 | 41.2 |
| G/N/N | 3 | 5.9 | 5.9 | 5.1 | 4.6 | 5.8 | 5.1 | 4.6 | 5.8 |
| | 4 | 7.1 | 5.8 | 2.0 | 4.8 | 5.3 | 4.8 | 4.6 | 5.2 |
| | 5 | 6.0 | 6.1 | 6.6 | 4.9 | 5.0 | 5.0 | 3.0 | 5.1 |
| | 6 | 3.6 | 5.1 | 2.8 | 4.6 | 5.0 | 3.7 | 3.0 | 4.9 |
| G/G/G | 3 | 40.9 | 40.9 | 5 | 8.0 | 46.3 | 5 | 8.0 | 46.3 |
| | 4 | 16.6 | 34.5 | 1.6 | 6.5 | 29.8 | 2.9 | 6.9 | 43.8 |
| | 5 | 49.1 | 37.6 | 2.9 | 8.6 | 26.5 | 2.7 | 4.8 | 40.8 |
| | 6 | 24.3 | 35.5 | 2.8 | 4.8 | 22.1 | 2.1 | 6.0 | 40.5 |
| L/N/N | 3 | 11.1 | 11.1 | 7.7 | 4.6 | 6.3 | 7.7 | 4.6 | 6.3 |
| | 4 | 11.0 | 12.1 | 1.6 | 6.3 | 5.3 | 10.3 | 5.6 | 5.3 |
| | 5 | 9.2 | 8.7 | 11.4 | 5.3 | 6.1 | 7.5 | 3.8 | 5.7 |
| | 6 | 4.0 | 7.0 | 2.1 | 5.3 | 6.3 | 5.6 | 4.2 | 6.1 |
| L/L/L | 3 | 63.6 | 63.6 | 1.2 | 11.6 | 72.3 | 1.2 | 11.6 | 72.3 |
| | 4 | 35.0 | 59.9 | 2.8 | 10.4 | 53.3 | 5.2 | 13.2 | 69.0 |
| | 5 | 74.6 | 65.1 | 4.1 | 15.3 | 45.6 | 4.2 | 9.2 | 66.8 |
| | 6 | 47.8 | 63.2 | 3.6 | 5.8 | 39.9 | 2.9 | 11.0 | 66.7 |

Table 3: Bootstrap $p$-values for the example data sets. The order only refers to the order of the SM and CFS tests.

| data | order | SM | | CFS | | | | other tests | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $V_r$ | $S_r$ | $V_{1r}$ | $V_{2r}$ | $V_{3r}$ | $V_{4r}$ | CL | AD | KS |
| MGMT1 | 3 | 0.023 | 0.023 | 0.173 | 0.027 | | | 0.000 | 0.000 | 0.000 |
| | 4 | 0.157 | 0.006 | 0.879 | 0.014 | | | | | |
| | 5 | 0.820 | 0.011 | 0.292 | 0.009 | | | | | |
| | 6 | 0.002 | 0.012 | 0.906 | 0.006 | | | | | |
| MGMT2 | 3 | 0.006 | 0.006 | 0.006 | 0.708 | | | 0.000 | 0.000 | 0.001 |
| | 4 | 0.000 | 0.000 | 0.000 | 0.353 | | | | | |
| | 5 | 0.000 | 0.000 | 0.000 | 0.484 | | | | | |
| | 6 | 0.000 | 0.000 | 0.000 | 0.055 | | | | | |
| Snapper length $(m=3)$ | 3 | 0.068 | 0.068 | 0.017 | 0.021 | 0.312 | | 0.066 | 0.000 | 0.000 |
| | 4 | 0.021 | 0.054 | 0.020 | 0.920 | 0.695 | | | | |
| | 5 | 0.169 | 0.061 | 0.494 | 0.147 | 0.842 | | | | |
| | 6 | 0.058 | 0.093 | 0.043 | 0.874 | 0.437 | | | | |
| Snapper length $(m=4)$ | 3 | 0.938 | 0.938 | 0.904 | 0.518 | 0.471 | 0.942 | 0.041 | 0.000 | 0.000 |
| | 4 | 0.815 | 0.811 | 0.917 | 0.443 | 0.888 | 0.919 | | | |
| | 5 | 0.956 | 0.929 | 0.875 | 0.890 | 0.264 | 0.813 | | | |
| | 6 | 0.814 | 0.960 | 0.755 | 0.349 | 0.883 | 0.966 | | | |
| London deaths | 3 | 0.504 | 0.504 | 0.740 | 0.510 | | | 0.975 | 0.899 | 0.713 |
| | 4 | 0.468 | 0.735 | 0.870 | 0.500 | | | | | |
| | 5 | 0.699 | 0.564 | 0.766 | 0.631 | | | | | |
| | 6 | 0.937 | 0.433 | 0.440 | 0.757 | | | | | |

Table 4: Powers (%) of smooth tests for Poisson mixtures using parametric bootstrap. The mixtures have $m = 2$ components with means $\mu_1 = 4$ and $\mu_2 = 10$. Poisson (P), Binomial (B), and negative binomial (NB) component distributions were considered. Powers are based on 2000 Monte Carlo runs of samples of size $n = 200$

|  | order | SM | | CFS | | | |
|---|---|---|---|---|---|---|---|
|  |  | $V_r$ | $S_r$ | $V_{1r}$ | $V_{2r}$ | $S_{1r}$ | $S_{2r}$ |
| P/P | 2 | 4.8 | 5 | 3.5 | 4.2 | 3.5 | 4.2 |
|  | 3 | 5.4 | 5 | 3.7 | 5.1 | 3.6 | 3.8 |
|  | 4 | 5.4 | 1.4 | 4.3 | 5.7 | 2.4 | 3.7 |
|  | 5 | 5.7 | 3.9 | 4.4 | 5.0 | 1.5 | 2.5 |
|  | 6 | 5.1 | 5.6 | 4.0 | 4.1 | 5 | 2.1 |
| P/B | 2 | 87.2 | 87.2 | 39.5 | 88.2 | 39.5 | 88.2 |
|  | 3 | 67.8 | 82.1 | 14.3 | 44.8 | 54.8 | 67.3 |
|  | 4 | 0 | 67.8 | 2.8 | 38.9 | 53.0 | 53.9 |
|  | 5 | 72.3 | 63.4 | 7.2 | 8.2 | 31.4 | 36.2 |
|  | 6 | 15.8 | 57.0 | 4.3 | 46.4 | 10.4 | 17.1 |
| B/P | 2 | 8.0 | 8.0 | 8.0 | 6.3 | 8.0 | 6.3 |
|  | 3 | 4.4 | 5.3 | 8.2 | 4.1 | 5.7 | 5.2 |
|  | 4 | 3.8 | 4.6 | 4.6 | 4.4 | 3.8 | 4.0 |
|  | 5 | 6.5 | 4.9 | 4.2 | 4.2 | 1.7 | 2.5 |
|  | 6 | 5.0 | 4.3 | 4.0 | 4.9 | 5.0 | 1.7 |
| B/B | 2 | 94.2 | 94.2 | 61.3 | 93.8 | 61.3 | 93.8 |
|  | 3 | 64.3 | 84.7 | 30.5 | 54.9 | 70.2 | 77.7 |
|  | 4 | 5 | 69.8 | 1.2 | 49.4 | 61.1 | 61.8 |
|  | 5 | 82.8 | 71.0 | 8.4 | 8.8 | 35.4 | 40.8 |
|  | 6 | 22.9 | 67.8 | 4.7 | 57.7 | 10.3 | 17.0 |
| P/NB | 2 | 99.9 | 99.9 | 71.3 | 99.9 | 71.3 | 99.9 |
|  | 3 | 100.0 | 100.0 | 9.9 | 24.7 | 88.1 | 99.9 |
|  | 4 | 85.6 | 100.0 | 16.0 | 100.0 | 89.1 | 99.9 |
|  | 5 | 99.9 | 100.0 | 7.1 | 78.9 | 89.5 | 99.9 |
|  | 6 | 98.7 | 100.0 | 6.3 | 100.0 | 89.6 | 99.9 |
| NB/P | 2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|  | 3 | 82.9 | 99.9 | 100.0 | 31.0 | 100.0 | 100.0 |
|  | 4 | 304 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|  | 5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|  | 6 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| NB/NB | 2 | 98.0 | 98.0 | 98.0 | 98.0 | 98.0 | 98.0 |
|  | 3 | 88.7 | 100.0 | 100.0 | 61.8 | 98.0 | 98.0 |
|  | 4 | 33.7 | 100.0 | 34.5 | 98.1 | 97.9 | 98.0 |
|  | 5 | 100.0 | 100.0 | 100.0 | 100.0 | 96.5 | 90.6 |
|  | 6 | 99.8 | 100.0 | 100.0 | 96.5 | 89.0 | 89.0 |

Table 5: Powers (%) of the CL, AD and KS tests for normal and Poisson mixtures using parametric bootstrap. The settings and abbreviations are as for Tables 1, 2 and 4. Powers are based on 2000 Monte Carlo runs of samples of size $n = 200$

|  | CL | AD | KS |
|---|---|---|---|
| N/N | 4.8 | 4.3 | 3.9 |
| T/N | 13.8 | 21.8 | 10.2 |
| T/T | 22.8 | 44.9 | 28.4 |
| G/N | 6.0 | 11.4 | 7.2 |
| G/G | 11.2 | 42.6 | 24.0 |
| L/N | 10.0 | 24.1 | 9.4 |
| L/L | 266 | 738 | 453 |
| N/N/N | 3.7 | 4.8 | 4.4 |
| T/N/N | 6.8 | 10.3 | 5.0 |
| T/T/T | 15.5 | 29.3 | 16.4 |
| G/N/N | 5.4 | 6.0 | 4.5 |
| G/G/G | 9.6 | 28.5 | 14.5 |
| L/N/N | 5.4 | 9.4 | 5.6 |
| L/L/L | 15.4 | 57.2 | 29.0 |
| P/P | 4.8 | 5.4 | 1.8 |
| P/B | 16.0 | 79.4 | 61.0 |
| B/P | 5.5 | 7.5 | 2.1 |
| B/B | 24.6 | 86.2 | 64.5 |
| B/NB | 62.5 | 100.0 | 100.0 |
| NB/B | 99.0 | 100.0 | 100.0 |