



HAL
open science

Testing the number and the nature of the components in a mixture distribution

Michel Broniatowski, Emilie Miranda, Wolfgang Stummer

► **To cite this version:**

Michel Broniatowski, Emilie Miranda, Wolfgang Stummer. Testing the number and the nature of the components in a mixture distribution. Lecture Notes in Computer Science, 2018, Geometric Science of Information 4th International Conference, GSI 2019, Toulouse, France, August 27–29, 2019, Proceedings, 11712, pp.309-318. 10.1007/978-3-030-26980-7_32 . hal-03897534

HAL Id: hal-03897534

<https://hal.science/hal-03897534v1>

Submitted on 13 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Testing the number and the nature of the components in a mixture distribution

Michel Broniatowski¹, Emilie Miranda², and Wolfgang Stummer³

¹ Corresponding Author. Sorbonne Universite Pierre et Marie Curie, LPSM, 4 place Jussieu, 75252 Paris, France. michel.broniatowski@sorbonne-universite.fr

² Sorbonne Universite Pierre et Marie Curie, LPSM, 4 place Jussieu, 75252 Paris, France.

³ Department of Mathematics, University of Erlangen–Nürnberg, Cauerstrasse 11, 91058 Erlangen, Germany.

Abstract. In this paper, we propose a test procedure for the number of components of mixture distributions in a parametric setting. The test statistic is based on divergence estimators derived through the dual form of the divergence in parametric models. We provide a standard limit distribution for the test statistic under the null hypothesis that holds for mixtures of any number of components $k \geq 2$.

Keywords: Test procedure · Mixture model · Number of component · Divergence estimation

1 Introduction

Consider a k -component parametric mixture model P_θ ($k \geq 2$) defined as follows:

$$P_\theta := \sum_{i=1}^k w_i P_{a_i}^{(i)} \quad (1)$$

where $\{P_{a_1}^{(1)}; a_1 \in A_1\}, \dots, \{P_{a_k}^{(k)}; a_k \in A_k\}$ are k -parametric models and A_1, \dots, A_k are k sets in $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_k}$ with $d_1, \dots, d_k \in \mathbb{N}^*$ and $0 \leq w_i \leq 1$, $\sum w_i = 1$. Note that we consider an unstandard framework in which the weights are allowed to be equal to 0. Note Θ the parameter space:

$$\theta \in \Theta := \left\{ (w_1, \dots, w_k, a_1, \dots, a_k)^T \in [0, 1]^k \times A_1 \times \dots \times A_k \text{ such that } \sum_{i=1}^k w_i = 1 \right\}, \quad (2)$$

and assume that the model is identifiable. Let $k_0 \in \{1, \dots, k-1\}$.

We are willing to test if $(k - k_0)$ components in (1) have null coefficients. We assume that their labels are $k_0 + 1, \dots, k$. Denote Θ_0 the subset of Θ defined by

$$\Theta_0 := \{ \theta \in \Theta \text{ such that } w_{k_0+1} = \dots = w_k = 0 \}.$$

On the basis of an i.i.d sample X_1, \dots, X_n with distribution P_{θ_T} , $\theta_T \in \Theta$, we intend to perform tests of the hypothesis

$$\mathcal{H}_0 : \theta_T \in \Theta_0 \text{ against the alternative } \mathcal{H}_1 : \theta_T \in \Theta \setminus \Theta_0. \quad (3)$$

When considering the test (3), it is known that the generalized likelihood ratio test, based on the statistic

$$2 \log \lambda := 2 \log \frac{\sup_{\theta \in \Theta} \prod_{i=1}^n p_{\theta}(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_{\theta}(X_i)}, \quad (4)$$

is not valid, since the asymptotic approximation by χ^2 distribution does not hold in this case; the problem is due to the fact that the null value of θ_T is not in the interior of the parameter space Θ . We clarify now this problem.

For simplicity, consider a mixture of two known densities p_0 and p_1 with $p_0 \neq p_1$:

$$p_{\theta} = (1 - \theta)p_0 + \theta p_1 \text{ where } \theta \in \Theta := [0, 1]. \quad (5)$$

Given data X_1, \dots, X_n with distribution P_{θ_T} , $\theta_T \in [0, 1]$, consider the test problem

$$\mathcal{H}_0 : \theta_T = 0 \text{ against the alternative } \mathcal{H}_1 : \theta_T > 0. \quad (6)$$

The generalized likelihood ratio statistic for this test problem is

$$W_n(0) := 2 \log \frac{L(\hat{\theta})}{L(0)}, \quad (7)$$

Under suitable regularity conditions we can prove that the limit distribution of the statistic W_n in (7) is $0.5\delta_0 + 0.5\chi_1^2$, a mixture of the χ^2 -distribution and the Dirac measure at zero; see e.g [12] and [11].

Moreover, in the case of more than two components and $k - k_0 \geq 2$, the limit distribution of the GLR statistic (4) under \mathcal{H}_0 is complicate and not standard (not a χ^2 distribution) which poses some difficulty in determining the critical value that will give correct asymptotic size; see [11]. Gassiat and al [1] proposes for instance a likelihood ratio approach for mixtures and give the asymptotic properties of the test, but its numerical application is extremely complicated, especially under non gaussian mixtures. On the other hand, the likelihood ratio statistic can not be used to construct asymptotic confidence region for the parameter θ_T since its limit law is not the same when $\theta_T = 0$ and $\theta_T > 0$.

In the sequel, we propose a simple solution for testing the number of components of a parametric mixture model. This method consists in constructing a test statistic based on φ -divergences and their asymptotic properties. In the following section, we will provide the general framework that will be used to construct the test procedure, i.e the definitions, representation and properties of φ -divergences.

2 Some definition and notation in relation with minimum divergence inference

Let $\mathcal{P} := \{P_{\theta}, \theta \in \Theta\}$ be an identifiable parametric model on \mathbb{R}^s where Θ is a subset of \mathbb{R}^d . All measures in \mathcal{P} will be assumed to be measure equivalent sharing therefore the same support. The parameter space Θ does not need to be open in the present setting. It

may even happen that the model includes measures which would not be probability distributions; cases of interest cover the present setting, namely models including mixtures of probability distributions; see [3].

Let φ be a proper closed convex function from $] -\infty, +\infty[$ to $[0, +\infty[$ with $\varphi(1) = 0$ and such that its domain $\text{dom}\varphi := \{x \in \mathbb{R} \text{ such that } \varphi(x) < \infty\}$ is an interval with endpoints $a_\varphi < 1 < b_\varphi$ (which may be finite or infinite). For two measures P_α and P_θ in \mathcal{P} the φ -divergence between the two is defined by

$$\phi(\alpha, \theta) := \int_{\mathbb{R}^s} \varphi \left(\frac{dP_\alpha}{dP_\theta}(x) \right) dP_\theta(x).$$

In a broader context, the φ -divergences were introduced by [8] as “ f -divergences”. The basic property of φ -divergences states that when φ is strictly convex on a neighborhood of $x = 1$, then

$$\phi(\alpha, \theta) = 0 \text{ if and only if } \alpha = \theta.$$

We refer to [9] chapter 1 for a complete study of those properties.

2.1 Examples of φ -divergences

The Kullback-Leibler (KL), modified Kullback-Leibler (KL_m), χ^2 , modified χ^2 (χ_m^2), Hellinger (H), and L_1 divergences are respectively associated to the convex functions $\varphi(x) = x \log x - x + 1$, $\varphi(x) = -\log x + x - 1$, $\varphi(x) = \frac{1}{2}(x-1)^2$, $\varphi(x) = \frac{1}{2}(x-1)^2/x$, $\varphi(x) = 2(\sqrt{x}-1)^2$ and $\varphi(x) = |x-1|$. All these divergences except the L_1 one, belong to the class of the so called “power divergences” introduced in [7] (see also [9] chapter 2), a class which takes its origin from Rényi [10]. They are defined through the class of convex functions

$$x \in]0, +\infty[\mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)} \quad (8)$$

if $\gamma \in \mathbb{R} \setminus \{0, 1\}$, $\varphi_0(x) := -\log x + x - 1$ and $\varphi_1(x) := x \log x - x + 1$. So, the KL -divergence is associated to φ_1 , the KL_m to φ_0 , the χ^2 to φ_2 , the χ_m^2 to φ_{-1} and the Hellinger distance to $\varphi_{1/2}$.

Considering any φ -divergence with φ a differentiable function but the likelihood divergence defined through the divergence function φ_0 , when θ_T in $\text{int}\Theta$ is defined as the true parameter of the distribution of the i.i.d. sample (X_1, \dots, X_n) it is convenient to assume that

$$\begin{aligned} &\text{There exists a neighborhood } \mathcal{U} \text{ of } \theta_T \text{ for which} & (A) \\ &\phi(\theta, \theta') \text{ is finite whatever } \theta \text{ and } \theta' \text{ in } \mathcal{U}. \end{aligned}$$

We will only consider divergences defined through differentiable functions φ , which we assume to satisfy

- (RC) There exists a positive δ such that for all c in $[1 - \delta, 1 + \delta]$, we can find numbers c_1, c_2, c_3 such that $\varphi(cx) \leq c_1 \varphi(x) + c_2 |x| + c_3$, for all real x .

Condition **(RC)** holds for all power divergences including KL and KL_m divergences. For all divergences considered in this paper it will be assumed that for any α and θ in \mathcal{U}

$$\int \left| \varphi' \left(\frac{dP_\theta}{dP_\alpha} \right) \right| dP_\theta < \infty. \quad (9)$$

Note that this condition trivially holds for the likelihood divergence. Sufficient and simple conditions encompassing (9) can be assessed under standard requirements for nearly all divergences. We state the following Lemma (see Liese and Vajda (1987)[9] and Broniatowski and Kéziou (2006) [2], Lemma 3.2).

Lemma 1. *Assume that **RC** holds and $\phi(\theta, \alpha)$ is finite. Then (9) holds.*

2.2 Dual form of the divergence and dual estimators in parametric models

The following representation is the cornerstone of parametric inference through divergence based methods.

Theorem 1. *Let θ belong to Θ and let $\phi(\theta, \theta_T)$ be finite. Assume that **RC** holds together with Condition (A). Then*

$$\begin{aligned} \phi(\theta, \theta_T) &= \sup_{\alpha \in \mathcal{U}} \int \varphi' \left(\frac{dP_\theta}{dP_\alpha} \right) dP_\theta - \int \varphi^\# \left(\frac{dP_\theta}{dP_\alpha} \right) dP_{\theta_T} \\ &= \sup_{\alpha \in \mathcal{U}} \int h(\theta, \alpha, x) dP_{\theta_T} \end{aligned} \quad (10)$$

Furthermore the sup is reached at θ_T and uniqueness holds.

From (10), simple estimators for $\phi(\theta, \theta_T)$ and θ_T can be defined, plugging any convergent empirical measure in place of P_{θ_T} and taking the infimum in θ in the resulting estimator of $\phi(\theta, \theta_T)$.

In the context of simple i.i.d. sampling, introducing the empirical measure $P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ where the X_i 's are i.i.d. r.v.'s with common unknown distribution P_{θ_T} in \mathcal{P} , the natural estimator of $\phi(\theta, \theta_T)$ is

$$\begin{aligned} \phi_n(\theta, \theta_T) &:= \sup_{\alpha \in \mathcal{U}} \left\{ \int h(\theta, \alpha, x) dP_n(x) \right\} \\ &= \sup_{\alpha \in \mathcal{U}} \int \varphi' \left(\frac{dP_\theta}{dP_\alpha} \right) dP_\theta - \frac{1}{n} \sum_{i=1}^n \varphi^\# \left(\frac{dP_\theta}{dP_\alpha} (X_i) \right) \text{ when (A) holds.} \end{aligned}$$

Under (A), since

$$\inf_{\theta \in \Theta} \phi(\theta, \theta_T) = \phi(\theta_T, \theta_T) = 0$$

the resulting estimator of $\phi(\theta_T, \theta_T)$ is

$$\phi_n(\theta_T, \theta_T) := \inf_{\theta \in \Theta} \phi_n(\theta, \theta_T) = \inf_{\theta \in \Theta} \sup_{\alpha \in \mathcal{U}} \left\{ \int h(\theta, \alpha, x) dP_n(x) \right\}. \quad (11)$$

And the estimator of θ_T is obtained as

$$\hat{\theta} := \arg \inf_{\theta \in \mathcal{U}} \sup_{\alpha \in \mathcal{U}} \left\{ \int h(\theta, \alpha, x) dP_n(x) \right\}. \quad (12)$$

Also, as stated in theorem 3.2 in [3]:

Theorem 2. *Under some derivability assumptions on $\varphi \left(\frac{dP_\theta}{dP_\alpha} \right)$ (conditions A.0 to A.2 in [3]),*

$$\text{If } \theta = \theta_T, \text{ then } \frac{2n}{\varphi''(1)} \phi_n(\theta, \theta_T) \xrightarrow{d} \chi_{(d)}^2 \text{ for } d = \dim(\Theta) \quad (13)$$

This last result of convergence of the estimated φ -divergence is of great interest in the problem we are taking on and serves as the basis for the test procedure that we propose.

3 A simple solution for testing finite mixture models

3.1 Testing between mixtures of fully characterized components

Let us consider a set of signed measures defined by

$$p_\theta = (1 - \theta)p_0 + \theta p_1, \quad \theta \in \mathbb{R} \quad (14)$$

where p_0 and p_1 are two known density (belonging or not to the same parametric family).

The mixture (5) is obviously contained in (14) and the case $\theta_T = 0$ is in this framework an interior point of the parameter space \mathbb{R} .

We observe a random sample X_1, \dots, X_n of distribution p_T . We are willing to test:

$$H_0 : p_T = p_0 \text{ vs } H_1 : p_T = p_\theta \neq p_0 \quad (15)$$

which can be reduced to

$$H_0 : \theta = 0 \text{ vs } H_1 : \theta \neq 0 \quad (16)$$

whenever $p_0 \neq p_1$ is met. The latter condition ensures the identifiability of the model and enables to consider different parametric families for p_0 and p_1 . Conversely, Chen [6], for instance, assumes that $0 < \theta < 1$, and tests the equality of the parameters of p_0 and p_1 inside a unique family \mathcal{F} .

In the following, we will thus assume that $p_0 \neq p_1$.

3.2 Test statistics

The choice of the test statistic is driven by the result given in Theorem 2. Accordingly, let ϕ be any divergence associated with convex finite functions and such that 0 is an interior point of the space parameter defined by:

$$\Theta := \left\{ \alpha \in \mathbb{R} : \int \left| \varphi' \left(\frac{dP_0}{dP_\alpha} \right) \right| dP_0 < \infty \right\} \quad (17)$$

Then the statistic $2n\phi_n(0, \theta_T)$ can be used as a tes statistic for (20) and

$$2n\phi_n(0, \theta_T) \longrightarrow \chi_{(1)}^2 \quad \text{when } H_0 \text{ holds.} \quad (18)$$

Also, this (18) also holds when testing whether the true distribution is a k_0 -component mixture or a k -component mixture as in (3). In this case, the test statistic $2n\phi_n(\Theta_0, \theta_T)$ converges to a $\chi_{(k-k_0)}^2$ distribution when H_0 holds.

While many divergences meet the former properties, we will restrict in the sequel ourselves to two generators.

Chi-square divergence

The first divergence that we consider is the χ^2 -divergence. The corresponding φ function $\varphi_2(x) := \frac{1}{2}(x-1)^2$ is defined and convex on whole \mathbb{R} ; an example when \mathcal{P} may contain signed finite measures and not be restricted to probability measures is considered in [2] in relation with a two components mixture model defined in (14) and where θ is allowed to assume values in an open neighborhood Θ of 0, in order to provide a test for (20), with θ an interior point of Θ .

Modified Kullback-Leibler divergence

The second divergence that we retain is generated by a function described below, namely

$$\varphi_c(x) := (x + e^c - 1) \cdot \log(x + e^c - 1) + 1 - (x + e^c - 1) + (1 - c) \cdot (e^c - 1) - c \cdot x \geq 0$$

$$x \in]1 - e^c, \infty[, \quad c \in \mathbb{R}, \quad (19)$$

which has been derived within the recent general framework of Broniatowski and Stummer [5]. It is straightforward to see that φ_c is strictly convex and satisfies $\varphi_c(1) = 0 = \varphi'_c(1)$. For the special choice $c = 0$, (19) reduces to the omnipresent Kullback-Leibler divergence generator

$$\varphi_0(x) := x \log x - x + 1 \geq 0, \quad x \in]0, \infty[.$$

According to (19), in case of $c > 0$ the domain $]1 - e^c, \infty[$ of φ_c covers also negative numbers (see [4] for insights on divergence-generators with general real-valued domain); thus, the same facts holds for the new generator than for the χ^2 and this opens the gate to considerable comfort in testing mixture-type hypotheses against corresponding marginal-type alternatives, as we derive in the following. We denote KL_c the corresponding divergence functional for which $KL_c(Q, P)$ is well defined whenever P is a probability measure and Q is a signed measure.

It can be noted that, depending on the type of model considered, the validity of the test can be subject to constraints over the parameters of the densities. Indeed, the convergence of $I = \int \left| \phi' \left(\frac{p_0}{p_\theta} \right) \right| dP_0$ is not always guaranteed. This kind of considerations may guide the choice of the test statistic. For instance, in some cases, including

scaling models, conditions that are required for the χ^2 -divergence, do not apply to the KL_c -divergence.

For instance, consider a Gaussian mixture model with different variances:

$$p_0 \sim \mathcal{N}(\mu, \sigma_0^2), \quad p_1 \sim \mathcal{N}(\mu, \sigma_1^2)$$

The convergence of I with the χ^2 requires either $\sigma_1^2 > \sigma_0^2$ or $\sigma_0^2 > \sigma_1^2 > \frac{1}{2}\sigma_0^2$. On the other hand, the convergence is always ensured with the KL_c -divergence.

The same observations can be made for lognormal, exponential and Weibull densities.

3.3 Generalization to parametric distributions with unknown parameters

In the previous section, the densities of each component were supposed to be known. We will now generalize to the case where the components belong to parametric families with unknown parameter. For the sake of simplicity, we will present the generalization for a two component mixture, but it is valid as well for more k -component mixtures with $k \geq 2$.

Let us assume $p_0 \in \mathcal{F}_0 = \{p_0(\cdot | \lambda_0) : \lambda_0 \in \Lambda_0\}$ and $p_1 \in \mathcal{F}_1 = \{p_1(\cdot | \lambda_1) : \lambda_1 \in \Lambda_1\}$, with Λ_0 and Λ_1 compact subsets of \mathbb{R}^d , $d \geq 1$.

We will consider aggregated tests of composite hypotheses. Indeed the null hypothesis of homogeneity of the population is rejected when there exists at least one couple of parameters $(\lambda_0^*, \lambda_1^*) \in \Lambda_0 \times \Lambda_1$ with $\lambda_1^* \neq \lambda_0^*$ for which the simple hypothesis $H_0(\lambda_0^*, \lambda_1^*)$ is rejected. The condition $\{\lambda_1^* \neq \lambda_0^*\}$ is only required when p_0 and p_1 belong to the same parametric family.

The test (15) can be reformulated as follows:

$$H_0 = \cap_{\lambda_0 \in \Lambda_0} H_0(\lambda_0, \Lambda_1) \quad \text{vs} \quad H_1 = \cup_{\lambda_0 \in \Lambda_0} H_1(\lambda_0, \Lambda_1) \quad (20)$$

where

$$\begin{aligned} \forall \lambda_0 \in \Lambda_0, \quad & H_0(\lambda_0, \Lambda_1) : p_T = p_0(\cdot | \lambda_0) \\ \text{vs} \quad & H_1(\lambda_0, \Lambda_1) : \cup_{\lambda_1 \in \Lambda_1 \setminus \lambda_0} \{p_T = p_\theta(\cdot | \lambda_0, \lambda_1) \neq p_0(\cdot | \lambda_0)\} \end{aligned} \quad (21)$$

Let $\phi_n(0, \lambda_T | \lambda_0, \lambda_1)$ be the estimated divergence used to construct the test statistic for (15) for fixed λ_0 and λ_1 . The test statistic for (20) is derived from:

$$\sup_{\lambda_0 \in \Lambda_0} \sup_{\lambda_1 \in \Lambda_1 \setminus \lambda_0} \phi_n(0, \lambda_T | \lambda_0, \lambda_1)$$

where the parameter spaces Λ_0 and Λ_1 can be discretized in $\Lambda_{0,n}$ and $\Lambda_{1,n}$ for the sake of computational complexity.

In order to facilitate the computation of the test statistic, the statistic 3.3 will be computed as follows:

$$\sup_{\alpha \in \Theta} \left\{ \sup_{\lambda_0 \in \Lambda_0} \sup_{\lambda_1 \in \Lambda_1 \setminus \lambda_0} \int \varphi' \left(\frac{dP_{0,\lambda_0}}{dP_{\alpha,(\lambda_0,\lambda_1)}} \right) dP_{0,\lambda_0} - \frac{1}{n} \sum_{i=1}^n \varphi^\# \left(\frac{dP_{0,\lambda_0}}{dP_{\alpha,(\lambda_0,\lambda_1)}}(X_i) \right) \right\}$$

4 Numerical example

We here present the performances of the test procedure on different kinds of two-component mixtures. In each case, the distributions of the components are such that the mixture is not bimodal.

4.1 Lognormal mixture

We first consider a Lognormal mixture. The two components belonging to the same parametric family, we can compare the performances of the divergence based test with the modified likelihood ratio test proposed by Chen [6].

The alternate hypotheses are the following:

$$H_0 : p_T = p_0 \sim \text{LN}(\lambda_0, 1) \quad \text{vs} \quad H_1 : p_T = p_\theta \sim (1 - \theta)\text{LN}(\lambda_0, 1) + \theta\text{LN}(\lambda_1, 1),$$

where $\lambda_0 \in \Lambda_0 = [0.4, 1.6]$ and $\lambda_1 \in \Lambda_1 = [1.4, 2.6]$.

The critical region is computed numerically through Monte Carlo simulations under H_0 . The power of the test is also computed numerically when the realizations are drawn from the mixture model with $\theta = 0.2$, $\lambda_0 = 1$ and $\lambda_1 = 2$ for the χ^2 and KL_c test statistics and Chen's modified likelihood ratio.

The results in table 3 show that both χ^2 and KL_c outperform the modified likelihood ratio test and the test based on the KL_c divergence achieves in this case the greatest power.

Lognormal Mixture $\text{LN}(1, 1)$ vs $0.8\text{LN}(1, 1) + 0.2\text{LN}(2, 1)$						
n=250 observations						
	χ^2 test statistic		KL_c test statistic		Chen's modified lik ratio	
First kind risk	0.05	0.10	0.05	0.10	0.05	0.10
Power	0.22	0.41	0.50	0.65	0.12	0.18

Table 1: Power of the test for a two-component logNormal mixture

4.2 Gamma mixture

We test the following hypothesis

$$H_0 : p_T = p_0 \sim \mathcal{G}(\lambda_0, 1) \quad \text{vs} \quad H_1 : p_T = p_\theta \sim (1 - \theta)\mathcal{G}(\lambda_0, 1) + \theta\mathcal{G}(\lambda_1, 2),$$

where $\lambda_0 \in \Lambda_0 = [1.4, 2.6]$ and $\lambda_1 \in \Lambda_1 = [4.4, 5.6]$. The realizations are drawn from the mixture model with $\theta = 0.2$, $\lambda_0 = 2$ and $\lambda_1 = 5$

Here again, both divergence based statistics achieves higher power than the modified likelihood ratio test.

Gamma Mixture $\mathcal{G}(2, 1)$ vs $0.8\mathcal{G}(2, 1) + 0.2\mathcal{G}(5, 2)$						
n=250 observations						
	χ^2 test statistic		KL_c test statistic		Chen's modified lik ratio	
First kind risk	0.05	0.10	0.05	0.10	0.05	0.10
Power	0.31	0.46	0.35	0.45	0.13	0.22

Table 2: Power of the test for a two-component Gamma mixture

4.3 Weibull and Lognormal mixture

We here consider the case where the two components are from different parametric families. We want to test

$$H_0 : p_T = p_0 \sim IN(\lambda_0, 0.2) \quad vs \quad H_1 : p_T = p_\theta \sim (1 - \theta)IN(\lambda_0, 0.2) + \theta\mathcal{W}(\lambda_1, 2),$$

where $\lambda_0 \in \Lambda_0 = [0.4, 1.6]$ and $\lambda_1 \in \Lambda_1 = [2.4, 3.6]$. The realizations are drawn from the mixture model with $\theta = 1, \lambda_0 = 1$ and $\lambda_1 = 3$

The results in the following table show that the test based on the KL_c divergence performs better than the χ^2 statistic.

Lognormal and Weibull Mixture $IN(\lambda_0, 0.2)$ vs $0.8IN(\lambda_0, 0.2) + 0.2\mathcal{W}(\lambda_1, 2)$						
n=250 observations						
	χ^2 test statistic		KL_c test statistic			
First kind risk	0.05	0.10	0.05	0.10		
Power	0.28	0.47	0.34	0.57		

Table 3: Power of the test for the mixture of a lognormal and a Weibull distribution

Concerning the choice of the test statistic, we might note that the KL_c performs better when the two alternate distributions differ mainly in their central tendency, while the χ^2 might be preferred when the difference lays in the tails.

References

1. Azais, J.M., Gassiat, E., Mercadier, C.: Asymptotic distribution and local power of the log-likelihood ratio test formixtures: bounded and unbounded cases. *Bernoulli* 12(5), 775–799 (2006).
2. Broniatowski, M., Keziou, A.: Minimization of ϕ -divergences on sets of signed measures. *Studia Sci. Math. Hungar.*, 43(4), 403–442 (2006).
3. Broniatowski, M., Keziou, A: Parametric estimation and tests through divergences and the duality technique. *Journal of Multivariate Analysis* 100, 16–36 (2009)

4. Broniatowski, M., Stummer, W.: Some universal insights on divergences for statistics, machine learning and artificial intelligence. In: Nielsen, F. (ed.) *Geometric Structures of Information*, pp. 149—211. Springer Nature, Switzerland (2019)
5. Broniatowski, M., Stummer, W.: A bare simulation approach to finding minimum distances. Preprint (2019)
6. Chen, H., Chen, J., Kalbfleisch: A modified likelihood ratio test for homogeneity in finite mixture models. *J.R. Statist. Soc. B* 63(1), 19–29 (2001)
7. Cressie, N. and Read, T. R. C.: Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B*, 46(3), 440–464 (1984).
8. Csiszár, I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8, 85–108 (1963)
9. Liese, F. and Vajda, I.: *Convex statistical distances*, vol 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig (1987).
10. Rényi, A.: On measures of entropy and information. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, vol. 1 pp. 547-561 Univ. California Press, Berkeley, Calif (1961).
11. Self, S. G. and Liang, K.-Y.: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.*, 82(398), 605–610 (1987).
12. Titterton, D. M., Smith, A. F. M., and Makov, U. E.: *Statistical analysis of finite mixture distributions*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester (1985).