Electronic Journal of Statistics
Vol. 5 (2011) 329–341
ISSN: 1935-7524
DOI: 10.1214/11-EJS609

Identifiability of the proportion of null hypotheses in skew-mixture models for the p-value distribution

Subhashis Ghosal*

Department of Statistics
North Carolina State University
e-mail: subhashis_ghoshal@ncsu.edu
url: http://www4.stat.ncsu.edu/~ghoshal

and

Anindya Roy†

Department of Mathematics and Statistics,
University of Maryland Baltimore County
e-mail: anindya@umbc.edu
url: http://www.math.umbc.edu/~anindya

Abstract: In many multiple testing procedures, accurate modeling of the p-value distribution is a key issue. Mixture distributions have been shown to provide adequate models for p-value densities under the null and the alternative hypotheses. An important parameter of the mixture model that needs to be estimated is the proportion of true null hypotheses, which under the mixture formulation becomes the probability mass attached to the value associated with the null hypothesis. It is well known that in a general mixture model, especially when a scale parameter is present, the mixing distribution need not be identifiable. Nevertheless, under our setting for mixture model for p-values, we show that the weight attached to the null hypothesis is identifiable under two very different types of conditions. We consider several examples including univariate and multivariate mixture models for transformed p-values. Finally, we formulate an abstract theorem for general mixtures and present other examples.

AMS 2000 subject classifications: Primary 62E10; secondary 62G99.
Keywords and phrases: Identifiability, mixture models, multiple testing, skew-normal distribution.

Received September 2010.

1. Introduction

Many multiple testing procedures depend critically on the distribution of the p-values associated with the multiple hypotheses. Following Storey (2002), the p-value density can be represented as a mixture of a null component and an

*Research partially supported by NSF grant number DMS-0803540.
†Research partially supported by NSF grant number DMS-0803531.

alternative component. Under ideal conditions, the null component is the uniform density. The density of a p-value under an alternative usually has no fixed functional form, and hence non-parametric mixtures are useful to model this. The tremendous flexibility of mixture models makes it an attractive modeling tool; see McLachlan and Basford (1988) and Lindsay (1995). Tang et al. (2007) successfully used mixtures of beta density to model the p-value density in the independent case. Moreover, as argued in Ghosal et al. (2008), such mixtures can, in addition, easily impose natural shape restrictions on p-value densities. Recently, Ghosal and Roy (2011) have used skew-normal mixtures to model the p-value density in the probit scale and used the mixture model to estimate the proportion of true null hypotheses. The skew-normal mixture, in addition to being very flexible, has the added advantage that it can be easily generalized to the dependent situation to model the joint behavior of the p-values.

In order to study statistical procedures based on a mixture model, it is essential to establish identifiability of the mixing measure. If $\mathcal{F}$ denotes a class of distributions and $\mathcal{M}$ denotes a class of probability measures on $\mathcal{F}$, then a mixture of $\mathcal{F}$ is defined as any distribution expressible as $H_\mu(x) = \int_{\mathcal{F}} F(x)d\mu(F)$ for some $\mu \in \mathcal{M}$; see Lindsay (1995). The family of mixtures, $\{H_\mu : \mu \in \mathcal{M}\}$ is called *identifiable* if for any $\mu, \mu^* \in \mathcal{M}$ and $H_\mu = H_{\mu^*}$ imply $\mu = \mu^*$. If the class of distributions is parameterized by a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^k$ as $F(x) = \Psi(x;\theta)$, then the mixing measure μ can be defined through a distribution G over the parameter space Θ and the mixture will be written as $H_G(x) = \int_\theta \Psi(x;\theta)dG(\theta)$. If $\Psi(x;\theta)$ admits a Lebesgue density $\psi(x;\theta)$, then the corresponding G-mixture density is given by $h_G(x) = \int_\Theta \psi(x;\theta)dG(\theta)$. We say that a mixture h_G of kernel $\psi(\cdot;\theta)$ can be deconvoluted if it is possible to uniquely recover G from h_G.

Identifiability makes an inference problem meaningful, while the lack of it may lead to absurd conclusions unless the issue is resolved by appropriate constraints. There is a large literature on identifiability of finite or countable mixtures, i.e., when the distribution G is supported only on at most countable number of points. Teicher (1961) and Yakowitz and Spragins (1968) provided sufficient conditions for identifiability of finite mixtures. Patil and Bildikar (1966) and Tallis (1969) investigated identifiability of countable mixtures; see Chandra (1977), Chapter 3 of Titterington et al. (1985), Chapter 8 of Prakasa Rao (1992), and Chapter 2 of Lindsay (1995) for a more complete list of references. The study of identifiability for continuous G is more involved. The seminal papers Teicher (1960) and Teicher (1961) provided some results on identifiability of general mixtures. Blum and Susarla (1977) provided necessary and sufficient conditions for identifiability of general mixtures in the one parameter case. They used functional analytic methods to investigate identifiability. Tallis (1969) and Tallis and Chesson (1982) gave conditions based on techniques from integral equation theory. Nevertheless, most verification of identifiability proceed only on a case-by-case basis.

In this paper, we are primarily concerned about identifiability issues of skew-normal mixtures that are relevant in multiple hypothesis testing. The skew-normal density $q(y;\mu,\omega,\lambda)$ with location parameter μ, scale parameter ω and shape parameter λ is defined as

$$q(x;\mu,\omega,\lambda) = 2\frac{1}{\omega}\phi\left(\frac{x-\mu}{\omega}\right)\Phi\left(-\lambda\frac{x-\mu}{\omega}\right), \tag{1.1}$$

where ϕ denotes the standard normal density and the Φ denotes the corresponding cumulative distribution function (c.d.f.). In (1.1), we have used a slight reparameterization by switching the skewness parameter λ to $-\lambda$. The skew-normal family of distributions and other related skewed distributions such as skew-elliptical and skew-symmetric distributions, have recently become popular tools for modeling and have found a wide variety of applications; see Genton (2004). The skew-normal distribution was introduced by Azzalini (1985) and generalized in the multivariate situation by Azzalini and Dalla Valle (1996) and others. The potential of the skew-normal distribution as a flexible tool for modeling increases many-fold when one considers mixtures of such distributions. Ghosal and Roy (2011) considered mixtures of skew-normal distributions to model probit-transformed p-value distribution arising in general multiple testing problem. Clearly, under probit transformation, the null p-value density of standard uniform transforms into the standard normal density, which corresponds to the parameter value $(0, 1, 0)$ in the skew-normal family.

In general, mixtures may not be identifiable. For instance, location-scale mixture of normal densities, one of the most commonly used mixtures, is not identifiable; see Lindsay (1995), Page 54. This result renders the family of unrestricted skew-normal mixtures non-identifiable as well. Even though the entire mixing distribution may not be identifiable, some key features of it may be still identifiable. In multiple testing problems, a key estimand is the proportion of true null hypotheses among all hypotheses that are being tested. In the context of skew-normal mixture models for the probit transformed p-values, the true null proportion corresponds to the weight given to the point $(\mu, \omega, \lambda) = (0, 1, 0)$. We present some conditions on the mixing distribution for univariate skew-normal mixtures under which the point mass at $(0, 1, 0)$ can be identified. Further we extend the result to multivariate skew-normal mixtures, which are appropriate tools for dependent p-values. Abstract generalizations with further examples are also presented.

Our results on identifiability of the proportion of the true null hypotheses are given under two very different scenarios for the range of parameters under the alternative hypotheses. For the first type, the densities under the alternative have tails thicker than that of the null density $\phi(x)$ and the null value $(0, 1, 0)$ is a boundary point for the possible parameter values under the alternative. In this situation, we use a technique based on the characteristic function (c.f.) to identify the proportion of true null hypotheses. For the second type of mixtures, the null value $(0, 1, 0)$ is also a boundary point for the possible parameter values under the alternative, but the densities under the alternative have tails thinner than that of the null density $\phi(x)$. In this case, the ratios of the densities under the alternative and the null are studied to obtain the corresponding identifiability results.

2. Univariate skew-normal mixtures

We begin with an identifiability result for univariate skew-normal mixtures. Consider a skew-normal mixture model with weight π_0 attached to the distinguished value $(0, 1, 0)$ corresponding to the standard normal distribution. In a multiple testing problem, the mixture density may represent the overall density of probit transformed p-values, where null hypotheses hold true randomly with probability π_0. The following describes a setting where the true null proportion π_0 is uniquely identified from such a mixture.

Theorem 2.1. *Consider a skew-normal mixture of the type*

$$f(x) = \pi_0 \phi(x) + (1 - \pi_0) \int q(x; \mu, \omega, \lambda) dG(\mu, \omega, \lambda), \tag{2.1}$$

where G is concentrated on the region

$$\Theta_1 = \{(\mu, \omega, \lambda) : \omega^2 \geq 1 + \lambda^2, (\omega, \lambda) \neq (1, 0)\}.$$

Then f uniquely determines π_0.

Proof. Let $\hat{f}(t) = \int e^{itx} f(x)dx$ denote the c.f. of f. If X has density $q(x; 0, 1, \lambda)$, then X can be represented as

$$X \stackrel{d}{=} -\frac{\lambda}{\sqrt{1 + \lambda^2}}|Y_0| + \frac{1}{\sqrt{1 + \lambda^2}}Y_1, \tag{2.2}$$

where Y_0, Y_1 are independent standard normal; see Dalla Valle (2004), Proposition 1.2.3.

Note that $\mathrm{E}(e^{is|Y_0|}) = 2e^{-s^2/2}\Phi(is)$ for all $s \in \mathbb{R}$, where Φ is the unique entire function which agrees with the standard normal c.d.f. on $\mathbb{R}$. This may be shown by evaluating $2\int_0^\infty e^{isy}\phi(y)dy$ by direct contour integration. Alternatively, by observing that the half-normal distribution has finite moment generating function (m.g.f.) everywhere, $\mathrm{E}(e^{z|Y_0|})$ is an entire function which agrees with the function $2\int_0^\infty e^{zy}\phi(y)dy = 2e^{z^2/2}\Phi(z)$ for $z \in \mathbb{R}$, and hence must agree everywhere on $z \in \mathbb{C}$.

Thus the c.f of $q(x; 0, 1, \lambda)$ is given by

$$\hat{q}(t; 0, 1, \lambda) = e^{-t^2/2}\Phi(-it\lambda/\sqrt{1 + \lambda^2}).$$

Shifting the location by μ and scaling by ω, it follows that the c.f. of $q(x; \mu, \omega, \lambda)$ is

$$\hat{q}(t; \mu, \sigma, \lambda) = e^{it\mu - \omega^2 t^2/2}\Phi(-it\omega\lambda/\sqrt{1 + \lambda^2}).$$

Therefore, we obtain

$$e^{t^2/2}\hat{f}(t) = \pi_0 + (1 - \pi_0)\int_{\Theta_1} e^{it\mu - (\omega^2 - 1)t^2/2}\Phi\left(-\frac{it\omega\lambda}{\sqrt{1 + \lambda^2}}\right)dG(\mu, \omega, \lambda). \tag{2.3}$$

We shall show that the second term in (2.3) goes to zero as $|t| \to \infty$, identifying π_0 uniquely from f. It suffices to show that for every (μ, ω, λ) with $\omega^2 \geq 1 + \lambda^2$, $(\omega, \lambda) \neq (1, 0)$, we have

$$e^{it\mu - (\omega^2 - 1)t^2/2} \Phi\left(-\frac{it\omega\lambda}{\sqrt{1+\lambda^2}}\right) \to 0 \text{ as } |t| \to \infty \tag{2.4}$$

and the expression in (2.4) is uniformly bounded by a constant. If it can be shown that

$$e^{it\mu - (\omega^2 - 1)t^2/2} \Phi(-it\omega\lambda/\sqrt{1+\lambda^2})$$

is the c.f. of a continuous random variable, then (2.4) holds by the Riemann-Lebesgue lemma while the second assertion holds by the absolute boundedness of a c.f. by 1.

To complete the proof, we use the representation (2.2) for a general μ, ω:

$$X \stackrel{d}{=} \mu - \omega \frac{\lambda}{\sqrt{1+\lambda^2}} |Y_0| + \omega \frac{1}{\sqrt{1+\lambda^2}} Y_1.$$

Since $\omega/\sqrt{1+\lambda^2} \geq 1$ by the given condition, Y_1 can be represented as $Z_1 + Z_2$, where $Z_1 \sim N(0, \frac{\omega^2}{1+\lambda^2} - 1)$, $Z_2 \sim N(0, 1)$ and they are independent. Hence $e^{it\mu - (\omega^2 - 1)t^2/2} \Phi(-it\omega\lambda/\sqrt{1+\lambda^2})$ is the c.f of the continuous random variable $\mu - \frac{\omega\lambda}{\sqrt{1+\lambda^2}} |Y_0| + Z_1$; note that at least one of $\frac{\omega\lambda}{\sqrt{1+\lambda^2}} |Y_0|$ and Z_1 is non-degenerate since $(\omega, \lambda) \neq (1, 0)$. □

Theorem 2.1 implies the following result for normal mixtures.

Corollary 2.2. *A normal mixture of the type*

$$f(x) = \pi_0 \phi(x) + (1 - \pi_0) \int \frac{1}{\omega} \phi\left(\frac{x - \mu}{\omega}\right) dG(\mu, \omega),$$

where G is concentrated on the region $\mathbb{R} \times (1, \infty)$, uniquely determines π_0.

The conclusion is not unexpected since $N(0, 1)$ may not be written as mixtures of $N(0, \omega^2)$ with $\omega > 1$. For the skew-normal family, the corresponding natural lower bound for ω seems to be $\sqrt{1+\lambda^2}$, since this guarantees that the variance of the distribution is more than 1.

For modeling probit transformed p-values, a more useful region for the mixing parameter in the multiple testing context is the complementary region $\omega < \sqrt{1+\lambda^2}$. This is due to the fact that the conditions $\mu \leq 0$, $1 < \omega < \sqrt{1+\lambda^2}$, $\lambda > 0$ ensure that the density of the original p-values is decreasing [cf., Ghosal and Roy (2011)], which is a natural shape restriction in the testing context. Indeed, the required condition rules out the normal case $\lambda = 0$. For a precise characterization of the decreasing p-value density, see Theorem 2 of Ghosal and Roy (2011). The normal mixture model may, however, be useful in the case when test statistics are modeled directly. A referee pointed out that p-values for two sided tests can sometimes lose valuable information. If there is imbalance in the distribution of the direction of alternative in a two-sided t-test, then that information can be retained by preserving the sign of the t-statistic while considering the probit transform of the p-values. Since the standard normal distribution is invariant under sign change, a mixture model like in Corollary 2.2 may be appropriate for such signed transformed p-values.

For identifiability purposes, we can work with a larger set of parameters than those ensuring a decreasing p-value density under the alternative. Identifiability of π_0 is guaranteed if the p-value density under the alternative attains the minimum value 0 as the p-values approach 1, since then the weight π_0 attached to the uniform can be easily identified from the height of the mixture density for the original p-values at one; see Ghosal et al. (2008). For testing against a one-sided alternative hypothesis in a monotone likelihood ratio family, the p-value density at 1 is usually zero. One-sided alternatives will be relevant whenever the direction of activity is known beforehand. For a two-sided alternative, the p-value density under the alternative is generally not zero at 1, but is usually a small number η. In the later situation, identifiability can hold only approximately in the sense that the value of π_0 can be asserted only within a range of values of span η. The condition that p-value density at 1 is zero in terms of probit transformed p-value x is equivalent to showing that the density ratio $\int q(x;\mu,\omega,\lambda)dG(\mu,\omega,\lambda)/\phi(x) \to 0$ as $x \to \infty$. This motivates the following result.

Theorem 2.3. *Consider a skew-normal mixture of the type as in (2.1), where G is concentrated on the region*

$$\Theta_2 = \{(\mu,\omega,\lambda) \neq (0,1,0) : \mu \leq 0, \omega^2 \leq 1+\lambda^2, \lambda \geq 0\}.$$

Further assume that $\int_{\{\lambda>0\}} \lambda^{-1} dG(\mu,\omega,\lambda) < \infty$ and $\int_{\{\lambda=0\}} \omega^{-1} dG(\mu,\omega,\lambda) < \infty$. Then f uniquely determines π_0.

Proof. Since $\Phi(-t) \leq (\sqrt{2\pi}t)^{-1}e^{-t^2/2}$ for all $t>0$, a bound on the ratio of the density $q(x;\mu,\omega,\lambda)$ to $\phi(x)$ is given by

$$\begin{aligned} &2\omega^{-1}\exp[-\frac{1}{2}x^2(\omega^{-2}-1) + x\omega^{-1}\mu - \frac{1}{2}\omega^{-2}\mu^2]\Phi(-\lambda\omega^{-1}x + \lambda\omega^{-1}\mu) \\ &\leq 2\omega^{-1}\exp[-\frac{1}{2}x^2\{\omega^{-2}(1+\lambda^2)-1\}]\frac{\omega}{\lambda x} \leq \frac{2}{\lambda x} \end{aligned} \tag{2.5}$$

for all $x>0$, assuming that $\lambda>0$. The bound is G-integrable by assumption. If $\lambda=0$, the skewness factor is redundant and the density ratio is bounded by

$$\omega^{-1}\exp[-\frac{1}{2}x^2(\omega^{-2}-1)+x\omega^{-1}\mu-\frac{1}{2}\omega^{-2}\mu^2] \leq \omega^{-1}\exp[-\frac{1}{2}x^2\{\omega^{-2}-1\}] \leq \omega^{-1},$$

which is also G-integrable by assumption.

If $\lambda=0$, necessarily $\omega^2<1$ or $\mu<0$. In either case,

$$\exp[-\frac{1}{2}x^2(\omega^{-2}-1) + x\omega^{-1}\mu] \to 0 \text{ as } x \to \infty.$$

For $\lambda>0$, we use estimate (2.5) to reach the same conclusion. Therefore the density ratio goes to 0 as $x \to \infty$ for any fixed $(\mu,\omega,\lambda) \in \Theta_2$. □

Remark 2.4. In the Bayesian context, a popular method of inducing prior distribution on densities is using a mixture model and assigning a prior distribution on the mixing distribution G. Ferguson (1983) and Lo (1984) pioneered this idea for Bayesian density estimation. When G is given a Dirichlet process prior [Ferguson (1973)] with $\mathrm{E}(G) = G_0$, the condition on G in Theorem 2.3 can be met by requiring that $G_0(\Theta_2) = 1$, $\int_{\{\lambda>0\}} \lambda^{-1} dG_0(\mu,\omega,\lambda) < \infty$ and $\int_{\{\lambda=0\}} \omega^{-1} dG_0(\mu,\omega,\lambda) < \infty$.

It is interesting to observe that the situations in Theorems 2.1 and 2.3 are diametrically opposite in that in the former case, the c.f. under the alternative has thinner tail than the c.f. under the null, while in the latter case, the density under the alternative has thinner tail than that under the null. According to a well-known "uncertainty principle" [Hardy (1933)], a function and its Fourier transform cannot both have thinner tails than the standard normal. If the mixing distributions gives weight to both Θ_1 and Θ_2, then neither the technique of controlling the ratios of the c.f.'s, nor that of the ratios of the densities used in the proofs, will work. This is the primary reason why we need G to give weights only to one type of alternatives. The following remark clarifies the comment further.

Remark 2.5. Consider a mixture

$$f(x) = \frac{1}{2} \times 2\phi(2x) + \frac{1}{2} \times \frac{1}{2}\phi(x/2) \tag{2.6}$$

of two densities from the normal scale family, where one of the densities in the mixture, $2\phi(2x)$ has a thinner tail than the null density $\phi(x)$ and the other density, $\frac{1}{2}\phi(x/2)$ has thicker tail than $\phi(x)$. Then f can be written as a mixture of $N(0,1)$ and a symmetric unimodal density, and hence the null proportion π_0 may not be identified from f. More precisely, we may write

$$f(x) = \frac{1}{4}\phi(x) + \frac{3}{4}h(x) \tag{2.7}$$

for some symmetric unimodal density h.

Observe that $g(x) = 2\phi(2x) + \frac{1}{2}\phi(x/2) - \frac{1}{2}\phi(x) \geq 0$ for all x, and $g'(x) > 0$ for $x < 0$ and $g'(x) < 0$ for $x > 0$. Also note that $\int g(x)dx = \frac{3}{2}$. Therefore, with the symmetric unimodal probability density $h(x) = \frac{2}{3}g(x)$, representation (2.7) holds.

3. Multivariate skew-normal mixtures

In this section, we consider analogs of the results of the last section for multivariate skew-normal mixtures.

Let $\boldsymbol{R}$ be an $m \times m$ positive definite correlation matrix and let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)' \in \mathbb{R}^m$. Let $\boldsymbol{\Delta} = \mathrm{diag}((1+\lambda_1^2)^{-1/2}, \dots, (1+\lambda_m^2)^{-1/2})$. The standardized multivariate skew-normal density with matrix of correlation parameters $\boldsymbol{R}$ and skewness vector $\boldsymbol{\lambda}$ is given by

$$\frac{2}{(2\pi)^{m/2}\det(\boldsymbol{\Delta})\sqrt{\det(\boldsymbol{R}+\boldsymbol{\lambda}\boldsymbol{\lambda}')}}\exp[-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda}\boldsymbol{\lambda}')^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{x}] \times\Phi\left(-\frac{\boldsymbol{\lambda}'\boldsymbol{\Delta}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda}\boldsymbol{\lambda}')^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{x}}{\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{R}^{-1}\boldsymbol{\lambda}}}\right).$$

When a location vector $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_m)'$ and scale parameters $\omega_1,\ldots,\omega_m>0$ are introduced, a multivariate skew-normal density $q(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{D},\boldsymbol{\lambda},\boldsymbol{R})$ is given by

$$\frac{2}{(2\pi)^{m/2}\det(\boldsymbol{\Gamma})\sqrt{\det(\boldsymbol{R}+\boldsymbol{\lambda}\boldsymbol{\lambda}')}} \times\exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda}\boldsymbol{\lambda}')^{-1}\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]\Phi\left(-\frac{\boldsymbol{\lambda}'\boldsymbol{R}^{-1}\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{R}^{-1}\boldsymbol{\lambda}}}\right), \tag{3.1}$$

where $\boldsymbol{D}=\mathrm{diag}(\omega_1,\ldots,\omega_m)$ and $\boldsymbol{\Gamma}=\boldsymbol{D}\boldsymbol{\Delta}=\boldsymbol{\Delta}\boldsymbol{D}$. A random vector $\boldsymbol{X}=(X_1,\ldots,X_m)'$ following the density $q(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{D},\boldsymbol{R},\boldsymbol{\lambda})$ can be represented as

$$\begin{aligned} X_j &= \mu_j+\omega_j Z_j,\\ Z_j &= -\frac{\lambda_j}{\sqrt{1+\lambda_j^2}}|Y_0|+\frac{1}{\sqrt{1+\lambda_j^2}}Y_j, \quad j=1,\ldots,m, \end{aligned} \tag{3.2}$$

where $Y_0\sim N(0,1)$ and independently $(Y_1,\ldots,Y_m)'$ follows the m-variate normal distribution $N_m(\mathbf{0},\boldsymbol{R})$ with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{R}$; see Dalla Valle (2004).

In the multiple hypothesis testing situation considered by Ghosal and Roy (2011), null or alternative hypotheses hold true independently of each other with probability π_0 and $1-\pi_0$ respectively. Let $H_1,\ldots,H_m$ stand for hypothesis indicators, where 0 stands for a true null and 1 for a false null. For any $\boldsymbol{H}=(H_1,\ldots,H_m)$, let $\boldsymbol{\mu}_{\boldsymbol{H}}$ (respectively, $\boldsymbol{\lambda}_{\boldsymbol{H}}$) be the vector obtained from $\boldsymbol{\mu}$ (respectively, $\boldsymbol{\lambda}$) by replacing the jth component by 0 whenever $H_j=0$. Similarly, let $\boldsymbol{D}_{\boldsymbol{H}}$ (respectively, $\boldsymbol{\Delta}_{\boldsymbol{H}}$, $\boldsymbol{\Gamma}_{\boldsymbol{H}}$) be the diagonal matrix obtained from $\boldsymbol{D}$ (respectively, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$) by replacing the jth diagonal entry by 1 whenever $H_j=0$. Given all hypothesis indicators $H_1,\ldots,H_m$, the joint density of probit p-values $(X_1,\ldots,X_m)'$ is assumed to be $q(\boldsymbol{x};\boldsymbol{\mu}_{\boldsymbol{H}},\boldsymbol{D}_{\boldsymbol{H}},\boldsymbol{\lambda}_{\boldsymbol{H}},\boldsymbol{R})$, and $(\mu_j,\omega_j,\lambda_j)$ are i.i.d. following a joint distribution G. The correlation matrix $\boldsymbol{R}$ is kept fixed in the mixing. Thus the multivariate skew-normal mixture density f can be written as

$$\pi_0^m\phi_m(\boldsymbol{x};\mathbf{0},\boldsymbol{R})+\sum_{(H_1,\ldots,H_m)\neq(0,\ldots,0)}\pi_0^{m-n_{\boldsymbol{H}}}(1-\pi_0)^{n_{\boldsymbol{H}}} \times\int q(\boldsymbol{x};\boldsymbol{\mu}_{\boldsymbol{H}},\boldsymbol{D}_{\boldsymbol{H}},\boldsymbol{\lambda}_{\boldsymbol{H}},\boldsymbol{R})\prod_{j=1}^m dG(\mu_j,\omega_j,\lambda_j), \tag{3.3}$$

where $n_{\boldsymbol{H}}$ stands for the number of false null hypotheses.

Below, we use the following orderings: for vectors $\boldsymbol{x},\boldsymbol{y}$, let $\boldsymbol{x}<\boldsymbol{y}$ or $\boldsymbol{x}\leq\boldsymbol{y}$ stand for componentwise ordering and for matrices, let $\boldsymbol{A}\geq\boldsymbol{B}$ mean that $\boldsymbol{A}-\boldsymbol{B}$ is non-negative definite while $\boldsymbol{A}>\boldsymbol{B}$ stand for $\boldsymbol{A}\geq\boldsymbol{B}$ and $\boldsymbol{A}\neq\boldsymbol{B}$.

Theorem 3.1. *Consider a skew-normal mixture of the type (3.3), where G is concentrated on a region Θ_1 such that $\boldsymbol{\Gamma_H R \Gamma_H} > \boldsymbol{R}$ for all $\boldsymbol{H} \neq \boldsymbol{0}$ and $(\boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{\lambda}) \in \Theta_1^m$. Then f uniquely determines π_0.*

Proof. We use the c.f. based argument as in the case of univariate mixtures and apply to every term in the sum (3.3) by showing that the ratio of c.f's $\exp(\|\boldsymbol{t}\|^2/2)\hat{q}(\boldsymbol{t}; \boldsymbol{\mu_H}, \boldsymbol{D_H}, \boldsymbol{\lambda_H}, \boldsymbol{R}) \to 0$ as $\|\boldsymbol{t}\| \to \infty$ along a line. For a given sequence $\boldsymbol{H}$ of hypotheses indicators, a random variable $\boldsymbol{X}$ having density $q(\boldsymbol{x}; \boldsymbol{\mu_H}, \boldsymbol{D_H}, \boldsymbol{\lambda_H}, \boldsymbol{R})$ can be represented using (3.2) as

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu_H} - |Y_0|\boldsymbol{\Gamma_H \lambda_H} + \boldsymbol{\Gamma_H Y},$$

Clearly, $\boldsymbol{X}$ is the convolution of $N_m(\boldsymbol{0}, \boldsymbol{R})$ with another variable since the variance-covariance matrix of $\boldsymbol{\Gamma_H Y}$ is $\boldsymbol{\Gamma_H R \Gamma_H} \geq \boldsymbol{R}$. It remains to show that the other variable in the convolution is continuous. Then the Riemann-Lebesgue lemma will apply on $\boldsymbol{t}$ approaching infinity along at least one line.

If at least one $\lambda_i \neq 0$ for some i with $H_i = 1$, then the first term $|Y_0|\boldsymbol{\Gamma_H \lambda_H}$ is non-degenerate. If all $\lambda_i = 0$, the skew-normal density is reduced to a normal and $\boldsymbol{\Gamma_H Y}$ contains a $N_m(\boldsymbol{0}, \boldsymbol{\Gamma_H R \Gamma_H} - \boldsymbol{R})$ variable, which is a non-degenerate normal variable. □

Remark 3.2. The condition $\boldsymbol{\Gamma_H R \Gamma_H} > \boldsymbol{R}$ for all $\boldsymbol{H}$ in particular implies that $\omega_i^2 > 1 + \lambda_i^2$ for all i. These two conditions are, of course, equivalent in the independent case.

As in the univariate case, identifiability can be established under a diametrically opposite condition on parameters using density considerations.

Theorem 3.3. *Consider a multivariate skew-normal mixture of the type as in (3.3), where G is concentrated on a region Θ_2 such that*

$$(\boldsymbol{R} + \boldsymbol{\lambda_H \lambda_H'})^{-1}\boldsymbol{\Gamma_H^{-1}\mu_H} \leq \boldsymbol{0}, \boldsymbol{\Gamma_H}(\boldsymbol{R} + \boldsymbol{\lambda_H \lambda_H'})\boldsymbol{\Gamma_H} < \boldsymbol{R} \tag{3.4}$$

for all $\boldsymbol{H}$ and $(\boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{\lambda}) \in \Theta_2^m$. Further assume that $\int \omega^{-1}\sqrt{1+\lambda^2}dG(\mu, \omega, \lambda) < \infty$. Then f uniquely determines π_0.

Proof. We shall show that for every H, the ratio

$$\int_{\Theta_2^m} \frac{q(\boldsymbol{x}; \boldsymbol{\mu_H}, \boldsymbol{D_H}, \boldsymbol{\lambda_H}, \boldsymbol{R})}{\phi_m(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{R})} \prod_{j=1}^m dG(\mu_j, \omega_j, \lambda_j)$$

converges to 0 as $\|\boldsymbol{x}\|$ tends to infinity along some line. We establish this by showing that

(i) $q(\boldsymbol{x}; \boldsymbol{\mu_H}, \boldsymbol{D_H}, \boldsymbol{\lambda_H}, \boldsymbol{R})/\phi_m(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{R})$ is uniformly bounded by a G-integrable function;

(ii) for every fixed $(\boldsymbol{\mu_H}, \boldsymbol{\omega_H}, \boldsymbol{\lambda_H})$, $q(\boldsymbol{x}; \boldsymbol{\mu_H}, \boldsymbol{D_H}, \boldsymbol{\lambda_H}, \boldsymbol{R})/\phi_m(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{R}) \to 0$ as $\|\boldsymbol{x}\|$ tends to infinity along some line.

For (i), observe that

$$\begin{aligned}
&\frac{q(\boldsymbol{x};\boldsymbol{\mu_H},\boldsymbol{D_H},\boldsymbol{\lambda_H},\boldsymbol{R})}{\phi_m(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R})}\\
=\;& \frac{2\sqrt{\det(\boldsymbol{R})}}{\det(\boldsymbol{\Gamma_H})\sqrt{\det(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')}}\\
&\times \exp[-\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu_H})'\boldsymbol{\Gamma_H}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')^{-1}\boldsymbol{\Gamma_H}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_H})-\boldsymbol{x}'\boldsymbol{R}^{-1}\boldsymbol{x}\}]\\
&\times \Phi\left(-\frac{\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\Gamma_H}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_H})}{\sqrt{1+\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\lambda_H}}}\right)\\
\le\;& 2\prod_{j=1}^{m}\max\left(\frac{\sqrt{1+\lambda_j}}{\omega_j},1\right)\times\exp[-\frac{1}{2}\boldsymbol{x}'\{\boldsymbol{\Gamma_H}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')^{-1}\boldsymbol{\Gamma_H}^{-1}-\boldsymbol{R}^{-1}\}\boldsymbol{x}]\\
&\times\exp[\boldsymbol{\mu_H}'\boldsymbol{\Gamma_H}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')^{-1}\boldsymbol{\Gamma_H}^{-1}\boldsymbol{x}-\frac{1}{2}\boldsymbol{\mu_H}'\boldsymbol{\Gamma_H}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')^{-1}\boldsymbol{\Gamma_H}^{-1}\boldsymbol{\mu_H}]\\
\le\;& 2\prod_{j=1}^{m}\left(1+\frac{\sqrt{1+\lambda_j}}{\omega_j}\right)
\end{aligned}$$

whenever $\boldsymbol{x}\ge\boldsymbol{0}$, by the given conditions. By assumption,

$$\int\prod_{j=1}^{m}\frac{\sqrt{1+\lambda_j}}{\omega_j}\prod_{j=1}^{m}dG(\mu_j,\omega_j,\lambda_j)=\left(\int\frac{\sqrt{1+\lambda}}{\omega}dG(\mu,\omega,\lambda)\right)^m<\infty,$$

proving (i).

Now to prove (ii), fix any $(\boldsymbol{\mu},\boldsymbol{\omega},\boldsymbol{\lambda})\in\Theta_2^m$. Since

$$(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')^{-1}=\boldsymbol{R}^{-1}-\frac{\boldsymbol{R}^{-1}\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}}{1+\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\lambda_H}}, \tag{3.5}$$

$(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')^{-1}\boldsymbol{\Gamma_H}^{-1}\boldsymbol{\mu_H}\le\boldsymbol{0}$ on Θ_2^m and $\boldsymbol{\lambda_H}\ge\boldsymbol{0}$, we have that

$$\begin{aligned}
\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\Gamma_H}^{-1}\boldsymbol{\mu_H} &\le \frac{\boldsymbol{R}^{-1}\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}}{1+\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\lambda_H}}\boldsymbol{\Gamma_H}^{-1}\boldsymbol{\mu_H} \qquad (3.6)\\
&= \frac{\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\lambda_H}}{1+\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\lambda_H}}\boldsymbol{\lambda_H}'\boldsymbol{R}^{-1}\boldsymbol{\Gamma_H}^{-1}\boldsymbol{\mu_H}\le 0.
\end{aligned}$$

Choose $\boldsymbol{x}=a\boldsymbol{\Gamma_H}\boldsymbol{R}\boldsymbol{1}$, where $\boldsymbol{1}=(1,\dots,1)'$ and $a\to\infty$. Then $\boldsymbol{\lambda_H}\boldsymbol{R}^{-1}\boldsymbol{\Gamma_H}^{-1}\boldsymbol{x}=a\boldsymbol{\lambda_H}'\boldsymbol{1}\to\infty$ unless $\boldsymbol{\lambda_H}=\boldsymbol{0}$. Hence for $\boldsymbol{\lambda_H}\ne\boldsymbol{0}$, in view of (3.6) and $\Phi(-t)\le(\sqrt{2\pi}t)^{-1}e^{-t^2/2}$ for all $t>0$, we obtain the bound

$$\frac{q(\boldsymbol{x};\boldsymbol{\mu_H},\boldsymbol{D_H},\boldsymbol{\lambda_H},\boldsymbol{R})}{\phi_m(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R})}\le C\exp[-\frac{1}{2}\boldsymbol{x}'\{\boldsymbol{\Gamma_H}^{-1}(\boldsymbol{R}+\boldsymbol{\lambda_H}\boldsymbol{\lambda_H}')^{-1}\boldsymbol{\Gamma_H}^{-1}-\boldsymbol{R}^{-1}\}\boldsymbol{x}]\frac{1}{a}\to 0,$$

where C stands for a constant (depending on $(\boldsymbol{\mu},\boldsymbol{\omega},\boldsymbol{\lambda})\in\Theta_2^m$). Thus, (ii) holds in this case.

If $\boldsymbol{\lambda_H} = \mathbf{0}$, we can ignore the skewness factor in the expression for skew-normal density, i.e., the density becomes a normal density. Further, $\boldsymbol{\Delta_H} = \boldsymbol{I}$, i.e., $\boldsymbol{\Gamma_H} = \boldsymbol{D_H}$. As $\boldsymbol{\Gamma_H R \Gamma_H} < \boldsymbol{R}$, there exists $\boldsymbol{\xi}$ such that $\boldsymbol{\xi}'(\boldsymbol{D_H^{-1} R^{-1} D_H^{-1}} - \boldsymbol{R}^{-1})\boldsymbol{\xi} > 0$. Choosing $\boldsymbol{x} = a\boldsymbol{\xi}$ and letting $a \to \infty$, we obtain (ii). □

Remark 3.4. The ideal opposite of the condition in Theorem 3.1 is given by $\boldsymbol{\Gamma_H R \Gamma_H} < \boldsymbol{R}$ for all $\boldsymbol{H} \neq \mathbf{0}$, which is weaker than the condition assumed in Theorem 3.3. Using (3.5), it is possible to prove assertion (ii) in the proof only under $\boldsymbol{\Gamma_H R \Gamma_H} < \boldsymbol{R}$. However, it seems that ensuring positivity of $\boldsymbol{\lambda}'_{\boldsymbol{H}} \boldsymbol{R}^{-1} \boldsymbol{\Gamma}_{\boldsymbol{H}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_H})$ for some $\boldsymbol{x}$ not depending on the latent parameter $(\boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{\lambda})$ is a challenge. The positivity is essential in applying the estimate of Mills ratio $\Phi(-t) \leq t^{-1}\phi(t)$. A successful substitution of $\Phi(-t)$ will allow the resulting exponential factor to combine with the factor already present, thus leading to ordering condition $\boldsymbol{\Gamma_H R \Gamma_H} < \boldsymbol{R}$ in view of (3.5). In the special case when $\boldsymbol{R}^{-1}$ is a positive operator, i.e., $\boldsymbol{R}^{-1}\boldsymbol{x} \geq \mathbf{0}$ for all $\boldsymbol{x} \geq \mathbf{0}$, the condition $\boldsymbol{\Gamma_H R \Gamma_H} < \boldsymbol{R}$ will suffice in (3.4). Further, the condition on $\boldsymbol{\mu}$ can be simplified to $\boldsymbol{\mu} \leq \mathbf{0}$.

4. Abstraction and further examples

The basic idea behind the two types of identifiability theorems can be put in an abstract form for an arbitrary parametric family in $\mathbb{R}^m$ forming a univariate mixture model of Section 2. Abstraction of the multivariate mixture model of Section 3 will be more challenging, and will possibly depend on the availability of a decomposition like (3.2). For the purpose of simplicity and transparency of the conditions imposed, below we restrict to the univariate situation.

Let $h_\theta(x)$, $\theta \in \Theta$, be a parametric family of densities with c.f. $\hat{h}_\theta(t)$ and let θ_0 be a distinguished point in Θ. Consider a mixture of the type $f(x) = \pi_0 h_{\theta_0}(x) + (1 - \pi_0) \int h_\theta(x) dG(\theta)$.

Theorem 4.1. *Let $\theta_0 \notin \Theta_1 \subset \Theta$ and G be concentrated on Θ_1. Suppose that $\hat{h}_\theta(t)/\hat{h}_{\theta_0}(t) \to 0$ as $\|t\| \to \infty$ and $|\hat{h}_\theta(t)/\hat{h}_{\theta_0}(t)| \leq B(\theta)$ for all $\theta \in \Theta_1$, where $\int_{\Theta_1} B(\theta) dG(\theta) < \infty$. Then f uniquely identifies π_0.*

Let $\theta_0 \notin \Theta_2 \subset \Theta$ and G be concentrated on Θ_2. Suppose that $h_\theta(x)/h_{\theta_0}(x) \to 0$ as $\|x\| \to \infty$ along some line, and $|h_\theta(x)/h_{\theta_0}(x)| \leq B(\theta)$ for all x on that line, where $\int_{\Theta_1} B(\theta) dG(\theta) < \infty$. Then f uniquely identifies π_0.

The proof of the theorem follows from the arguments used in the proofs of Theorems 2.1 and 2.3.

Example 4.2. Let h_θ be the gamma family with shape parameter θ and scale parameter 1. Let $\theta_0 = 1$ and $\Theta_1 = (1, \infty)$. Then $f(x) = \pi_0 h_{\theta_0}(x) + (1 - \pi_0) \int_1^\infty h_\theta(x) dG(\theta)$ uniquely identifies π_0.

To prove, observe that $\hat{h}_\theta(t) = (1 - it)^{-\theta}$. Thus, for any $\theta > 1$,

$$|\hat{h}_\theta(t)/\hat{h}_1(t)| = |(1 - it)^{-(\theta-1)}| = (1 + t^2)^{(1-\theta)/2} \leq 1,$$

and $(1+t^2)^{(1-\theta)/2} \to 0$ as $|t| \to \infty$. Observe that $(1-it)^{-(\theta-1)}$ is the c.f. of the gamma density with shape parameter $\theta - 1 > 0$. Therefore, the ratio of c.f.s is uniformly bounded by 1 and converges to 0 as $|t| \to \infty$ by the Riemann-Lebesgue lemma as well.

For $\theta_0 = 1$, $\Theta_2 = (0,1)$, $f(x) = \pi_0 h_{\theta_0}(x) + (1-\pi_0)\int_0^1 h_\theta(x) dG(\theta)$ also uniquely identifies π_0. To see this, observe that $h_\theta(x)/h_1(x) = x^{\theta-1}/\Gamma(\theta) \leq x^{\theta-1}$, since $1/\Gamma(\theta) \leq 1$ for all $0 < \theta < 1$. If $x \geq 1$, then $x^{\theta-1} \leq 1$ and $x^{\theta-1} \to 0$ as $x \to \infty$.

Example 4.3. Let h_θ be the Cauchy scale family $h_\theta(x) = [\pi\theta(1 + x^2/\theta^2)]^{-1}$, $\theta > 0$. Let $\theta_0 = 1$ and observe that $\hat{h}_\theta(t) = e^{-\theta|t|}$.

If $\Theta_1 = (1,\infty)$, then $\hat{h}_\theta(t)/\hat{h}_1(t) = e^{-(\theta-1)|t|}$. Clearly $e^{-(\theta-1)|t|}$ is bounded by 1 and tends to 0 as $|t| \to \infty$. Note that the ratio is also the c.f. of the Cauchy density with scale parameter $\theta - 1 > 0$.

If $\Theta_2 = (0,1)$, the ratio of densities is given by $\theta^{-1}(1+x^2)/(1+x^2/\theta^2) \to 0$ as $|x| \to \infty$ and the ratio is uniformly bounded by θ^{-1}. Thus if $\int \theta^{-1} G(d\theta) < \infty$, then f uniquely identifies π_0.

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist. 11*, 171–178. MR0808153

Azzalini, A. and A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika 83*, 715–726. MR1440039

Blum, J. R. and V. Susarla (1977). Estimation of a mixing distribution function. *Ann. Probab. 5*, 200–209. MR0445703

Chandra, S. (1977). On mixtures of probability distributions. *Scand. J. Statist. 4*, 105–112. MR0474583

Dalla Valle, A. (2004). The skew-normal distribution. In M. G. Genton (Ed.), *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*, pp. 3–24. Boca Raton: Chapman and Hall/CRC. MR2155320

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Stat. 1*, 209–230. MR0350949

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In M. e. a. Rizvi (Ed.), *Recent Advances in Statistics*, pp. 287–302. New York: Academic Press. MR0736538

Genton, M. G. (2004). *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*. Vol. 9, Lecture Notes –Monograph Series. Chapman and Hall/CRC: Chapman and Hall/CRC. MR2155324

Ghosal, S. and A. Roy (2011). Predicting false discovery proportion under dependence. *To appear: J. Amer. Statist. Assoc. 106*, Preprint available at `www4.stat.ncsu.edu/~ghoshal/papers/skew.pdf`.

Ghosal, S., A. Roy, and Y. Tang (2008). Posterior consistency of Dirichlet mixtures of beta densities in estimating positive false discovery rates. In N. Balakrishnan, E. Peña, and M. Silvapulle (Eds.), *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab*

K. Sen, pp. 105–115. Beachwood, OH: Institute of Mathematical Statistics. MR2462198

Hardy, G. H. (1933). A theorem concerning Fourier transforms. *J. London Math. Soc. 8*, 227–231.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. Vol. 5, NSF-CBMS Regional Conference Series in Probability and Statistics. Hayward, CA: Institute of Mathematical Statistics.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates. I: Density estimates. *Ann. Stat. 12*, 339–348. MR0733519

McLachlan, G. J. and K. E. Basford (1988). *Mixture Models*. New York: Marcel-Dekker. MR0926484

Patil, G. P. and S. Bildikar (1966). Identifiability of countable mixtures of discrete probability distributions using methods of infinite matrices. *Proc. Cambridge Philos. Soc. 62*, 485–494. MR0202169

Prakasa Rao, B. L. S. (1992). *Identifiability in Stochastic Models: Characterization of Probability Distributions*. New York: Academic Press.

Storey, J. D. (2002). A direct approach to false discovery rates. *J. Roy. Statist. Soc. B 64*, 479–498. MR1924302

Tallis, G. M. (1969). The identifiability of mixtures of distributions. *J. Appl. Probab. 6*, 389–398. MR0264734

Tallis, G. M. and P. Chesson (1982). Identifiability of mixtures. *J. Austral. Math. Soc. Ser. A 32*, 339–348. MR0652411

Tang, Y., S. Ghosal, and A. Roy (2007). Nonparametric Bayesian estimation of positive false discovery rates. *Biometrics 63*, 1126–1134. MR2414590

Teicher, H. (1960). On the mixture of distributions. *Ann. Math. Stat. 31*, 55–73. MR0121825

Teicher, H. (1961). On identifiability of mixtures. *Ann. Math. Stat. 32*, 244–248. MR0120677

Titterington, D. M., A. F. M. Smith, and V. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.

Yakowitz, S. J. and J. D. Spragins (1968). On the identifiability of finite mixtures. *Ann. Math. Statist. 39*, 209–214. MR0224204