The Annals of Probability
1985, Vol. 13, No. 4, 1341–1357

IDENTIFIABILITY OF CONTINUOUS MIXTURES OF UNKNOWN GAUSSIAN DISTRIBUTIONS

BY C. BRUNI AND G. KOCH

Università di Roma

The problem of the identifiability of the mixing distribution and of the unknown parameters for a continuous mixture of Gaussian distributions is considered. Relevance of the problem under various analytical, statistical, and applicative points of view is stressed. Uniqueness of the mixing distribution and of the mean and variance functions for the mixed Gaussian distribution is proved. Furthermore, their continuous dependence on the mixture itself is proved under suitable topologies. These results also extend to the multidimensional case and to the case of non-Gaussian distributions, and/or signed mixing measure.

1. Problem statement. Given a mixture of a possibly uncountable family of Gaussian distributions, with unknown mean values and variances, we consider here the problem of recovering the mixing distribution and the mean value and variance functions as well.

Formally, we consider the equation:

$$f(x) = \int_D N(x\colon \lambda(y))\mu(dy), \tag{1.1}$$

where $x \in \mathbb{R}^p$, D is a compact subset of $\mathbb{R}^n$, μ is a probability measure on D, $\lambda = (m, \Sigma)$ denotes the pair of mean value vector function m and variance matrix function Σ defined on D, and N denotes the p-dimensional Gaussian density:

$$N(x\colon \lambda(y)) = \frac{1}{(2\pi)^{p/2}|\Sigma(y)|^{1/2}}\exp\left\{-\frac{1}{2}(x - m(y))^T\Sigma^{-1}(y)(x - m(y))\right\}. \tag{1.2}$$

With no loss of generality, D may be assumed to be connected by adding to it suitable subsets in $\mathbb{R}^n$ with zero μ measure. The problem is the one of identifiability of λ, μ; that is, given an f with the representation (1.1), whether it is possible to uniquely (and continuously) associate to it a pair (λ, μ).

Various motivations appear to be relevant for this problem, under analytical, statistical, applicative points of view.

Within an analytical context, the problem (restricted to the identifiability of μ) reduces to the solution of a classical Fredholm integral equation with the assumptions that the kernel is Gaussian and the solution is a probability measure. The restriction to a kernel with some specific structure is crucial. As is well known, this equation does not admit unique solution in the general case [7]. On the other hand, within that structure the kernel is allowed to be unknown, and the mixing distribution is not necessarily assumed to have a density with respect to Lebesgue measure.

Received January 1983; revised April 1985.
AMS 1980 *subject classifications*. Primary 60E05; secondary 45B05.
Key words and phrases. Identifiability of measures, mixture of Gaussian distributions, Fredholm equations.

In respect to identifying a distribution, the problem looks like that of investigating whether $\{N(x; \lambda(y)),\ x \in \mathbb{R}^n\}$ constitutes a determining class for the measure μ [3], [20] with the additional difficulty that here λ is unknown.

As far as the statistical relevance is concerned, we recall that the problem of identifiability of mixtures of Gaussian distributions has been widely investigated, especially in the case of finite mixtures [8], [23]–[26], [29]. It is also currently of actual interest, and the papers [11], [15] are evidence of that. In the infinite mixture case, compactness of the support D of mixing distribution is necessary for identifiability [23]; results for non-Gaussian mixed distributions are reported in [1], [9], [11], [12], [15], [19], [22], [28].

On a probabilistic ground, we note that (1.1) defines the probability density of a random variable X, such that its conditional density with respect to another variable Y is Gaussian. The identifiability problem essentially means recovering the marginal distribution μ of Y, and the specific Gaussian density N that describes conditioning of X on Y (or equivalently its first two moments λ as functions of Y) from the knowledge of just the marginal density of X.

The positive solution we give to the problem implies, among other results that, under broad regularity assumptions on λ and within obvious equivalence classes, given random variables X, Y_1, Y_2 if X is Gaussian conditioned on either Y_1 or Y_2, then Y_1 and Y_2 have the same distribution (and the densities of X conditioned on either Y_1 or Y_2 coincide).

From an applicative point of view, one faces just this problem when one wishes to draw the distribution of a given population (particles, cells, fishes, etc.) according to a certain character, but it is experimentally observable only for the distribution of the same population according to another character (and of course one knows something about the conditional distribution of one character with respect to the other, for instance that it has to be Gaussian).

This latter remark is indeed the key to appreciation of the relevance of the problem in a broad applicative context. With little imagination one can figure out quite a long series of different applicative fields to which the same problem can be tailored.

For instance in [17] the issue is the determination of adsorption energy distribution from that one of sites covered at equilibrium at a certain pressure.

In [5] the fluorescence distribution of a cell sample obtained through cytofluorometric procedures is exploited to recover the distribution of the same sample according to its DNA content.

Other applications are conceivable in the context of pattern recognition, for instance in reconstruction of the subject in blurred pictures (astronomy, aerospace recognition, etc.).

For other examples of application of the same problem in the finite mixture case, we refer to [6], [8].

In the next section some technical tools are provided. Regularity assumptions on λ and equivalence classes for the solution are dealt with by introducing the appropriate sets of functions to which λ is assumed to belong. Also, a lemma is proved, which will be crucial in the proof of the main result: Assuming

$$\int_D N(x; \lambda(y))\mu(dy) = \int_D N(x; \lambda'(y))\mu'(dy), \tag{1.3}$$

then by repeated differentiation of both sides of (1.3) and heavily exploiting the asymptotic behavior of the Gaussian kernel, we prove that the nth derivatives of both sides of (1.3) are sums of terms such that equality holds termwise.

In Section 3 we give the main result, that is uniqueness and continuity of the solution of (1.1) in the scalar case ($p = n = 1$). As far as uniqueness is concerned, it is shown that using the terms in the derivatives of (1.3), a set of functions can be defined that is a basis for the space of continuous functions along suitable curves in $\mathbb{R}^2$ parametrized by y through λ, λ'. Subsequent analysis and use of the lemma lead first to establishing the equality $\lambda = \lambda'$, and then $\mu = \mu'$.

Continuity is achieved by classical compactness arguments.

In Section 4 the extension to the multidimensional case is sketched; the case of non-Gaussian and/or noncontinuous kernels is discussed; the possible embedding of μ measures in the larger class of signed measures is also mentioned.

2. Notation and preliminary results. In the scalar case, let D be a compact subset of $\mathbb{R}$. Let $\lambda = (\lambda_1, \lambda_2)$ denote the two-component function defined on D, whose components represent, respectively, the mean value function m and the variance function σ^2 of the family of Gaussian densities $\{N(\cdot\,; \lambda(y)),\ y \in D\}$.

We now introduce some sets of functions λ to account for the regularity conditions and the specific hypotheses requested in the proof of the main results.

The first set is defined as follows:

$$\Lambda_1 = \big\{\lambda \in C^1(D)\colon |\lambda_1(y)| \le K_1 < \infty, 0 < s_1 \le \lambda_2(y) \le s_2 < \infty,\\ |\dot\lambda_1(y)| + |\dot\lambda_2(y)| \le K_2 < \infty, \forall\, y \in D\big\}$$

and clearly accounts for uniformity and regularity of λ, as well as the very meaning of λ_2 (variance function).

A second set within which the basic uniqueness result can be proved is:

$$\Lambda_2 = \text{any subset of } \Lambda_1 \text{ with the property}$$
$$\forall\, \lambda, \lambda' \in \Lambda_2, \qquad \forall\, y, y' \in D\colon \lambda(y) = \lambda'(y') \Rightarrow y = y'.$$

We observe that this property is crucial. Should it not hold, identifiability cannot be guaranteed any more. Indeed, if there exist y, y', $y \ne y'$ such that $\lambda(y) = \lambda'(y')$ then identity (1.3) holds with μ, μ' Dirac measures, respectively, in y, y'.

The Λ_2-type sets can be easily chosen sufficiently large and adapted to specific applications. For instance if D does not contain the origin, we may define:

$$\Lambda_2 = \big\{\lambda\colon \lambda_1(y) = C_1 y^\alpha,\\ \lambda_2(y) = C_2|y|^\beta,\ \alpha, \beta \text{ fixed: } C_1' \le C_1 \le C_1'', 0 < C_2' \le C_2 \le C_2''\big\}.$$

By restricting the choice for one component, the other one can vary in a larger set, for instance, as it occurs in the following case:

$$\Lambda_2 = \big\{\lambda\colon \lambda_1(y) = y;\ \lambda_2 \in C^1(D), 0 < s_1 \le \lambda_2(y) \le s_2, |\dot\lambda_2(y)| \le C_3\big\}.$$

One might also account for discontinuities in λ, at the expenses of some restriction on the measure μ, as will be pointed out in Section 4.

As far as continuity of the identifiable pair (λ, μ) with respect to f is concerned, classical results of functional analysis can be exploited provided we take compact subsets of Λ_2. A possibility is considering

$$\Lambda_3 = \text{any subset of } \Lambda_1 \text{ with property}$$
$$\exists\, \psi\colon \mathbb{R} \to \mathbb{R}^+,\ \psi \text{ continuous at } 0,\ \psi(x) = 0 \Rightarrow x = 0$$

such that $\forall\ \lambda, \lambda' \in \Lambda_3$, $\forall\ y, y' \in D$:

$$|\lambda(y) - \lambda'(y')| \geq \psi(y - y').$$

Notice that any subset Λ_3 is a subset of a Λ_2 set.

Finally, as will be shown in the next section, a weaker identifiability result that is identifiability modulo some suitable change of variable on D may be established if Λ_2 is substituted by the following set:

$$\Lambda_4 = \{\lambda \in \Lambda_1\colon \lambda(y) = \lambda(y') \Rightarrow y = y';$$
$$\lambda_1, \lambda_2 \text{ monotonic}, |\dot{\lambda}_1(y)| + |\dot{\lambda}_2(y)| > 0, \forall\ y \in D\}.$$

Notice that a Λ_4-type set does not necessarily contain, nor is contained in, a Λ_2-type set.

On all these sets, the topology is that induced by the usual sup norm.

Let us denote by $\mathscr{M}(D)$ the set of signed measures with support in D and bounded variation equipped with the vague topology [3]. Let $\mathscr{P}(D) \subset \mathscr{M}(D)$ be the set of probability measures. On $\mathscr{P}(D)$ vague topology becomes the weak topology, i.e., for $\{\mu_n\}$, $\mu \in \mathscr{P}(D)$

$$\mu_n \to \mu \quad \text{if} \quad \int_D \phi(y)\mu_n(dy) \to \int_D \phi(y)\mu(dy), \qquad \forall\ \phi \in C(D).$$

We define the operator T: $\Lambda_1 \times \mathscr{M}(D) \to H(\mathbb{R})$, where $H(\mathbb{R})$ is the space of integrable analytic functions on $\mathbb{R}$ with the sup norm:

$$f = T(\lambda, \mu)\colon f(x) = \int_D N(x; \lambda(y))\mu(dy). \tag{2.1}$$

It is easily seen that T is separately uniformly continuous with respect to λ and μ, and is therefore continuous on $\Lambda_1 \times \mathscr{M}(D)$.

For $\mu \in \mathscr{M}(D)$, $\phi \in C(D)$, $\phi \circ \mu$ denotes the element in $\mathscr{M}(D)$ whose Radon–Nikodym derivative with respect to μ is ϕ.

As a preliminary result, we establish the following lemma:

LEMMA. *Let* $\lambda, \lambda' \in \Lambda_1$, $\mu, \mu' \in \mathscr{P}(D)$. *If*

$$T(\lambda, \mu) = T(\lambda', \mu'), \tag{2.2}$$

then

$$T(\lambda, \psi_{h,k}) = T(\lambda', \psi'_{h,k}), \qquad h = 0, 1, \dots k,\ k = 0, 1, 2, \dots, \tag{2.3}$$

where $\psi_{h,k}$ *and* $\psi'_{h,k}$ *denote*

$$\psi_{h,k} = \frac{\lambda_1^h}{\lambda_2^k} \circ \mu, \qquad \psi'_{h,k} = \frac{\lambda_1'^h}{\lambda_2'^k} \circ \mu', \qquad h = 0, 1, \dots k,\ k = 0, 1, 2, \dots. \tag{2.4}$$

PROOF. Equation (2.3), for $h = k = 0$, is already guaranteed by (2.2). Thus by induction the lemma is proved as soon as we show that (2.3) implies:

$$T(\lambda, \psi_{h,k+1}) = T(\lambda', \psi'_{h,k+1}), \tag{2.5$'$}$$

$$T(\lambda, \psi_{h+1,k+1}) = T(\lambda', \psi'_{h+1,k+1}). \tag{2.5$''$}$$

By differentiating both sides of (2.3) we get:

$$\begin{aligned} &x[(T(\lambda, \psi_{h,k+1}))(x) - (T(\lambda', \psi'_{h,k+1}))(x)] \\ &= (T(\lambda, \psi_{h+1,k+1}))(x) - (T(\lambda', \psi'_{h+1,k+1}))(x). \end{aligned} \tag{2.6}$$

We first observe that (2.5′) and (2.6) imply (2.5″). It is then enough to prove (2.5′). By contradiction, let us assume that there exists a sequence $\{x_i\}$ dense in $\mathbb{R}$ and diverging to $+\infty$, such that

$$(T(\lambda, \psi_{h,k+1}))(x_i) \neq (T(\lambda', \psi'_{h,k+1}))(x_i), \qquad i = 1, 2, \dots . \tag{2.7}$$

If this were not the case, analyticity would imply (2.5′).

Denoting by $|\nu|$ the total variation measure of $\nu \in \mathscr{M}(D)$, we make a preliminary observation that Eq. (2.7) clearly excludes the possibility that $|\psi_{h,k}|$ and $|\psi'_{h,k}|$ both vanish.

Assume now that one of them, for instance $|\psi'_{h,k}|$ vanishes. Recalling (2.7), Eq. (2.6) yields

$$x_i = A(x_i), \tag{2.8}$$

where

$$A(x_i) = [(T(\lambda, \psi_{h+1,k+1}))(x_i)][(T(\lambda, \psi_{h,k+1}))(x_i)]^{-1} \tag{2.9}$$

and the contradiction will follow from the fact that $A(x_i)$ is bounded as $x_i \to +\infty$. To show this, let us define

$$\bar{\lambda}_2 = \mu - \operatorname{ess\,sup}\{\lambda_2(y)\} = \inf\{t\colon \lambda_2(y) \le t,\ \mu\text{-a.e. in } D\}, \tag{2.10}$$

$$\begin{aligned} \bar{\lambda}_1 &= \mu - \operatorname{ess\,sup}\{\lambda_1(y)\colon \lambda_2(y) = \bar{\lambda}_2\} \\ &= \inf_{\varepsilon>0} \inf\{t\colon \lambda_1(y) \le t,\ \mu\text{-a.e. on the set in } D \text{ where } \lambda_2(y) \ge \bar{\lambda}_2 - \varepsilon\}, \end{aligned} \tag{2.11}$$

$$\tilde{\lambda}_2 = |\lambda_1 \circ \mu| - \operatorname{ess\,sup}\{\lambda_2(y)\} = \inf\{t\colon \lambda_2(y) \le t,\ |\lambda_1 \circ \mu|\text{-a.e. in } D\}, \tag{2.12}$$

$$\begin{aligned} \tilde{\lambda}_1 &= |\lambda_1 \circ \mu| - \operatorname{ess\,sup}\{\lambda_1(y)\colon \lambda_2(y) = \tilde{\lambda}_2\} \\ &= \inf_{\varepsilon>0} \inf\{t\colon \lambda_1(y) \le t,\ |\lambda_1 \circ \mu|\text{-a.e. in the set in } D \\ &\qquad\qquad \text{where } \lambda_2(y) \ge \tilde{\lambda}_2 - \varepsilon\}. \end{aligned} \tag{2.13}$$

Note that if $\bar{\lambda}_1 \neq 0$, then $\bar{\lambda}_1 = \tilde{\lambda}_1$ and $\bar{\lambda}_2 = \tilde{\lambda}_2$.

For $\varepsilon_1, \varepsilon_2 > 0$ arbitrarily small we have that

$$\lim_{x_i \to \infty} \frac{(T(\lambda, \psi_{r,s}))(x_i)}{C_{r,s}(x_i; \varepsilon)} = 1 \quad \text{uniformly in } \varepsilon = (\varepsilon_1, \varepsilon_2), \tag{2.14}$$

$$r = 0, 1 \dots s,\ s = 0, 1 \dots,$$

where

$$(2.15)\quad C_{r,s}(x_i;\varepsilon)=\int_{\tilde{D}}N(x_i;\lambda(y))\psi_{r,s}(dy),\qquad r=1,\dots s,\ s=1,2,\dots,$$

where $\tilde{D}=D\cap\{y\colon \tilde{\lambda}_1-\varepsilon_1<\lambda_1(y)\le\tilde{\lambda}_1,\ \tilde{\lambda}_2-\varepsilon_2<\lambda_2(y)\le\tilde{\lambda}_2\}$ and $C_{0,s}(x_i,\varepsilon)$, $s=0,1,2,\dots$ is defined in the same way but with $\bar{\lambda}_1,\bar{\lambda}_2$ in the place of $\tilde{\lambda}_1,\tilde{\lambda}_2$, respectively.

Indeed, exploiting the Gaussian nature of the kernel in (2.1) it is possible to show that:

$$(2.16)\qquad (T(\lambda,\psi_{r,s}))(x_i)=C_{r,s}(x_i;\varepsilon)\left(1+O\left(\frac{1}{x_i}\right)\right),$$

where $O(1/x_i)$ converges to zero uniformly in ε. Moreover, because of the continuity of λ_1 and the definition of $\bar{\lambda}_1$ and $\tilde{\lambda}_1$, in the integration set of (2.15) λ_1 is either strictly positive or strictly negative, $|\psi_{r,s}|$-a.e.. From positivity of N and μ, it follows that, for all x_is, $C_{r,s}(x_i;\varepsilon)$ is different from zero and definitively keeps the same sign for decreasing ε_1.

It follows from above that

$$(2.17)\qquad \lim_{x_i\to\infty}A(x_i)=\lim_{x_i\to\infty}\frac{C_{h+1,k+1}(x_i;\varepsilon)}{C_{h,k+1}(x_i;\varepsilon)}$$

and, recalling the definition (2.15) and the positivity (or negativity) of λ_1 in the integration set, the limit in (2.17) is seen to be constrained in the interval $[\bar{\lambda}_1-\varepsilon_1,\bar{\lambda}_1]$ (or $[\bar{\lambda}_1,\bar{\lambda}_1-\varepsilon_1]$) for $h=0$, or in the interval $[\tilde{\lambda}_1-\varepsilon_1,\tilde{\lambda}_1]$ (or $[\tilde{\lambda}_1,\tilde{\lambda}_1-\varepsilon_1]$) for $h>0$.

Finally, we assume both $|\psi_{h,k}|$, $|\psi'_{h,k}|$ nonvanishing. Recalling (2.7), Eq. (2.6) yields

$$(2.18)\qquad x_i=A(x_i)B(x_i),$$

where

$$(2.19)\quad B(x_i)=\frac{1-[(T(\lambda',\psi'_{h+1,k+1}))(x_i)][(T(\lambda,\psi_{h+1,k+1}))(x_i)]^{-1}}{1-[(T(\lambda',\psi'_{h,k+1}))(x_i)][(T(\lambda,\psi_{h,k+1}))(x_i)]^{-1}}$$

and the contradiction will follow from the fact that $B(x_i)$ is also bounded as $x_i\to+\infty$.

To show this, we define $\bar{\lambda}'_2,\bar{\lambda}'_1,\tilde{\lambda}'_2,\tilde{\lambda}'_1,C'_{r,s}$, respectively, similar to $\bar{\lambda}_2,\bar{\lambda}_1,\tilde{\lambda}_2,\tilde{\lambda}_1,C_{r,s}$ but for the primed quantities.

Let us consider the quantities $\psi_{r,s}(\bar{D}_\varepsilon)$, $\psi'_{r,s}(\bar{D}'_\varepsilon)$ where

$$(2.20)\qquad \bar{D}_\varepsilon=D\cap\{y\colon \bar{\lambda}_1-\varepsilon_1<\lambda_1(y)\le\bar{\lambda}_1,\bar{\lambda}_2-\varepsilon_2<\lambda_2(y)\le\bar{\lambda}_2\}$$

and similarly for $\bar{D}'_\varepsilon$. Assume at first $\bar{\lambda}_1\ne 0$. Then

$$(2.21)\qquad \lim_{x_i\to+\infty}B(x_i)=\lim_{x_i\to+\infty}\frac{1-C'_{h+1,k+1}(x_i;\varepsilon)C^{-1}_{h+1,k+1}(x_i;\varepsilon)}{1-C'_{h,k+1}(x_i;\varepsilon)C^{-1}_{h,k+1}(x_i;\varepsilon)}.$$

For ε small enough, $\psi'_{r,s}(\overline{D}'_\varepsilon)$ is different from zero (as already shown for $C_{r,s}$) because of the assumption $\overline{\lambda}_1 \neq 0$.

If both $\psi_{r,s}(\overline{D}_\varepsilon)$, $\psi'_{r,s}(\overline{D}'_\varepsilon)$ are nonpositive, or nonnegative, the following inequalities hold for x_i sufficiently large:

$$\begin{aligned}\frac{\psi'_{r,s}(\overline{D}'_\varepsilon)N(x_i;\overline{\lambda}'-\varepsilon)}{\psi_{r,s}(\overline{D}_\varepsilon)N(x_i;\overline{\lambda})} &\leq \frac{C'_{r,s}(x_i;\varepsilon)}{C_{r,s}(x_i;\varepsilon)}\\ &\leq \frac{\psi'_{r,s}(\overline{D}'_\varepsilon)N(x_i;\overline{\lambda}')}{\psi_{r,s}(\overline{D}_\varepsilon)N(x_i;\overline{\lambda}-\varepsilon)},\end{aligned} \tag{2.22}$$

where $\overline{\lambda} - \varepsilon = (\overline{\lambda}_1 - \varepsilon_1, \overline{\lambda}_2 - \varepsilon_2)$ and similarly for $\overline{\lambda}' - \varepsilon$. If $\psi_{r,s}(\overline{D}_\varepsilon)$, $\psi'_{r,s}(\overline{D}'_\varepsilon)$ have different signs, inequalities in (2.22) are reversed. The asymptotic behaviours for $x_i \to +\infty$ of the lower and upper bound in (2.22), in the case $\overline{\lambda}_2 \neq \overline{\lambda}'_2$, are given, respectively, by $\exp\{x_i^2(\overline{\lambda}'_2 - \overline{\lambda}_2)(2\overline{\lambda}_2\overline{\lambda}'_2)^{-1}\}\rho^{v}_{r,s}(x_i;\varepsilon)$ and $\exp\{x_i^2(\overline{\lambda}'_2 - \overline{\lambda}_2)(2\overline{\lambda}_2\overline{\lambda}'_2)^{-1}\}\rho^{u}_{r,s}(x_i;\varepsilon)$ where $\rho^{v}_{r,s}$, $\rho^{u}_{r,s}$ converge to 1, for each x_i, as $\varepsilon_2 \downarrow 0$. In the case $\overline{\lambda}_2 = \overline{\lambda}'_2$, but $\overline{\lambda}_1 \neq \overline{\lambda}'_1$, the same asymptotic behaviours are given by $\exp\{x_i(\overline{\lambda}'_1 - \overline{\lambda}_1)(\overline{\lambda}_2)^{-1}\}\delta^{v}_{r,s}(x_i;\varepsilon)$ and $\exp\{x_i(\overline{\lambda}'_1 - \overline{\lambda}_1)(\overline{\lambda}_2)^{-1}\}\delta^{u}_{r,s}(x_i;\varepsilon)$, where $\delta^{v}_{r,s}$, $\delta^{u}_{r,s}$ converge again to 1, for each x_i, as $\varepsilon \downarrow 0$. Finally, if $\overline{\lambda}_2 = \overline{\lambda}'_2$, $\overline{\lambda}_1 = \overline{\lambda}'_1$ the ratio $C'_{r,s}(x_i;\varepsilon)C^{-1}_{r,s}(x_i;\varepsilon)$ turns out to be essentially independent of x_i, and as $\varepsilon \downarrow 0$ converges to a bounded value independent of r, s. In conclusion, in all cases the limit (2.21) is bounded.

If $\overline{\lambda}_1 = 0$, but $\overline{\lambda}'_1 \neq 0$, we simply exchange primed by unprimed quantities and repeat the argument. If $\overline{\lambda}_1 = \overline{\lambda}'_1 = 0$ we introduce the quantities $\overline{\overline{\lambda}}_1$, $\tilde{\overline{\lambda}}_1$ substituting the ess sup with the ess inf in definitions (2.11), (2.13) and similarly for $\overline{\overline{\lambda}}'_1$, $\tilde{\overline{\lambda}}'_1$. Then a similar procedure can be set up to prove that $A(x_i)$, $B(x_i)$ are bounded quantities as $x_i \to -\infty$. The situation in which $\overline{\lambda}_1 = \overline{\lambda}'_1 = \overline{\overline{\lambda}}_1 = \overline{\overline{\lambda}}'_1 = 0$ essentially corresponds to $\lambda_1 \equiv \lambda'_1 \equiv 0$, which case was already considered at the beginning. □

3. Main results. We are now in a position to state the main result of this work, which appears to be a relevant step forth in the problem of identifiability of nonfinite Gaussian mixtures.

THEOREM 1. *Given* $\lambda, \lambda' \in \Lambda_2$, $\mu, \mu' \in \mathscr{P}(D)$, *if*

$$T(\lambda,\mu) = T(\lambda',\mu'), \tag{3.1}$$

then $\mu = \mu'$ *and* $\lambda = \lambda'$ (μ-*a.e.*).

PROOF. Obviously there are no results on λ, λ' outside the support of μ. On the other hand the assumptions on λ, λ' could be in fact relaxed by only considering their behavior on the support of μ.

Let us consider the curves in $\mathbb{R}^2$:

$$\begin{aligned}\mathscr{C} &= \{(\xi,\eta): \xi = \lambda_1(y),\ \eta = \lambda_2(y),\ y \in D\},\\ \mathscr{C}' &= \{(\xi,\eta): \xi = \lambda'_1(y),\ \eta = \lambda'_2(y),\ y \in D\},\end{aligned} \tag{3.2}$$

and a function $\phi \in C([-k_1, k_1] \times [s_1, s_2])$ such that

$$\phi(\xi,\eta) = 1, \qquad (\xi,\eta) \in \mathscr{C}, \tag{3.3}$$

$$\phi(\xi,\eta) = \frac{\lambda_2^{r}(y)}{(\lambda_1(y)+2k_1)^s}\,\frac{(\lambda_1'(y)+2k_1)^s}{\lambda_2'^{r}(y)}, \tag{3.4}$$

$(\xi,\eta) \in \mathscr{C}'$, r, s nonnegative integers.

Notice that (3.3), (3.4) do not contradict each other due to the assumption $\lambda, \lambda' \in \Lambda_2$.

In $C([-k_1, k_1] \times [s_1, s_2])$ we now consider the algebra generated by the functions $1/\eta$, ξ/η. It is clearly formed by the functions $\{\xi^h/\eta^k,\ h = 0, 1, \dots k;\ k = 0, 1, \dots\}$ which separate the points and contain the identity function. Then the Stone–Weierstrass theorem (see for instance [21]) guarantees that the linear subspace generated by the algebra is dense in $C([-k_1, k_1] \times [s_1, s_2])$. Therefore ϕ can be represented as:

$$\phi(\xi,\eta) = \sum_{O \le h \le k}^{\infty} \alpha_{h,k} \frac{\xi^h}{\eta^k} \tag{3.5}$$

for suitable coefficients $\{\alpha_{h,k}\}$, and because of the very definition of ϕ:

$$\phi(\xi,\eta)|_{\mathscr{C}} = \sum_{O \le h \le k}^{\infty} \alpha_{h,k} \frac{\lambda_1^h(y)}{\lambda_2^k(y)} = 1, \tag{3.6}$$

$$\begin{aligned}\phi(\xi,\eta)|_{\mathscr{C}'} &= \sum_{\substack{O(h,k)\\ h \le k}}^{\infty} \alpha_{h,k} \frac{\lambda_1'^h(y)}{\lambda_2'^k(y)}\\ &= \frac{\lambda_2^{r}(y)}{(\lambda_1(y)+2k_1)^s}\,\frac{(\lambda_1'(y)+2k_1)^s}{\lambda_2'^{r}(y)}.\end{aligned} \tag{3.7}$$

On the other hand, because of the above lemma, we have

$$T\left(\lambda, \frac{\lambda_1^h}{\lambda_2^k} \circ \mu\right) = T\left(\lambda', \frac{\lambda_1'^h}{\lambda_2'^k} \circ \mu'\right), \qquad h = 0, 1, \dots k,\ k = 0, 1, \dots, \tag{3.8}$$

which integrated over $\mathbb{R}$ yields

$$\int_D \frac{\lambda_1^h(y)}{\lambda_2^k(y)} \mu(dy) = \int_D \frac{\lambda_1'^h(y)}{\lambda_2'^k(y)} \mu'(dy), \qquad h = 0, 1, \dots k,\ k = 0, 1, \dots \tag{3.9}$$

By linear combination of (3.9) with coefficients $\alpha_{h,k}$ and exploiting the uniform convergence of the series, from (3.6), (3.7) we get

$$1 = \int_D \frac{\lambda_2^{r}(y)}{(\lambda_1(y)+2k_1)^s}\,\frac{(\lambda_1'(y)+2k_1)^s}{\lambda_2'^{r}(y)} \mu'(dy) \tag{3.10}$$

for r, s arbitrary nonnegative integers.

This amounts to saying that $\lambda = \lambda'$, μ'-a.e. The same result clearly can be shown to hold also μ-a.e. Then (3.9) becomes:

$$(3.11)\quad \int_D \frac{\lambda_1^h(y)}{\lambda_2^k(y)}\mu(dy) = \int_D \frac{\lambda_1^h(y)}{\lambda_1^k(y)}\mu'(dy), \qquad h = 0,1,\dots k,\ k = 0,1,\dots .$$

Let us now consider the algebra in $C(D)$ generated by $1/\lambda_2$, λ_1/λ_2. Clearly, it is formed by the functions $\{\lambda_1^h/\lambda_2^k, h = 0,1,\dots k,\ k = 0,1\dots\}$ and again satisfies the assumptions of the Stone–Weierstrass theorem. Therefore it is a determining class and (3.11) implies $\mu = \mu'$. □

REMARK. Let us define on $\Lambda_2 \times \mathscr{P}(D)$ the equivalence

$$(3.12)\qquad (\lambda,\mu) \sim (\lambda',\mu') \Leftrightarrow \mu = \mu' \quad \text{and} \quad \lambda = \lambda', \qquad \mu\text{-a.e.}$$

and let A_2 denote the space of equivalence classes in $\Lambda_2 \times \mathscr{P}(D)$. Let $\mathscr{R}(A_2)$ be the range of T restricted to A_2.

Then Theorem 1 states a one-to-one correspondence between A_2 and $\mathscr{R}(A_2)$.

Let now A_3 denote the space of equivalence classes in $\Lambda_3 \times \mathscr{P}(D)$, equipped with the product topology induced by the uniform convergence in Λ_3 and the weak convergence in $\mathscr{P}(D)$. We have the following result:

THEOREM 2. *The operator T^{-1}: $\mathscr{R}(A_3) \to A_3$ is continuous.*

PROOF. We first note that T^{-1} is well defined on $\mathscr{R}(A_3)$, since $A_3 \subset A_2$. Now $\mathscr{P}(D)$ is compact because D is such. Λ_3 on its own is compact, since it is a closed subset of Λ_1 which is compact due to the Ascoli–Arzelà theorem. The compactness of A_3 in the product topology follows. This and continuity of T: $A_3 \to \mathscr{R}(A_3)$ imply continuity of T^{-1} on $\mathscr{R}(A_3)$. □

Theorem 1 states a uniqueness result in a set of λ functions, with the property that any two functions belonging to the set itself cannot take the same value for different values of their arguments. The following theorem yields a weaker uniqueness result in which this property is loosened at the expense of some monotonicity assumption.

THEOREM 3. *Given $\lambda, \lambda' \in \Lambda_4$, $\mu, \mu' \in \mathscr{P}(D)$, if*

$$(3.1)\qquad T(\lambda,\mu) = T(\lambda',\mu'),$$

then there exists a strictly monotonic C^1 function h from D onto D, uniquely defined on the support of μ, such that

$$(3.13)\qquad \mu'(dh(y)) = \mu(dy),$$

$$(3.14)\qquad \lambda'(h(y)) = \lambda(y), \qquad \mu\text{-a.e.}$$

PROOF. The proof will be developed for the case of nondecreasing (component wise) λ, λ'; but it can be easily adapted with minor adjustments to the more general monotonic case. Once again, there are no results (nor assumptions) on λ (λ') outside the support of μ (μ').

Because of the assumptions on Λ_4, $\mathscr{C}$ and $\mathscr{C}'$ defined in (3.2) turn out to be simple regular curves parametrically represented by λ, λ', respectively. Consider now the compact set $\Gamma \subset D$ defined by:

$$\Gamma = \{ y \text{ in the support of } \mu \colon \exists\ y' \text{ in the support of } \mu' \colon \lambda(y) = \lambda'(y') \}. \tag{3.15}$$

Due to the definition of Λ_4, a unique y' corresponds in the sense of (3.15) to each $y \in \Gamma$.

First of all, we observe that Γ cannot be empty, since otherwise λ, λ' would belong to a Λ_2-type set and therefore, due to Theorem 1, (3.13), (3.14) would hold with $h(y) = y$, so that $\Gamma \equiv D$.

Due to monotonicity assumptions for λ, λ', the quantities $\bar{\lambda}_1, \bar{\lambda}_2$, defined in (2.11), (2.10), reduce, respectively, to the μ-ess sup $\lambda_1(y)$ and the μ-ess sup $\lambda_2(y)$ in D. Similarly $\bar{\bar{\lambda}}_1, \bar{\bar{\lambda}}_2$ coincide, respectively, to the μ-ess inf $\lambda_1(y)$ and the μ-ess inf $\lambda_2(y)$ in D. Similar notation will be used for primed quantities.

A first remark is that $\bar{\lambda} \equiv (\bar{\lambda}_1, \bar{\lambda}_2)$ coincides with $\bar{\lambda}' \equiv (\bar{\lambda}'_1, \bar{\lambda}'_2)$, since otherwise (3.1) would not hold as $x \to \infty$. This implies that $\bar{y} = \mu$-ess sup y in D, for which $\lambda(\bar{y}) = \bar{\lambda}$, belongs to Γ (and corresponds to $\bar{y}' = \mu'$-ess sup y in D).

Denote now by $\bar{\bar{y}}, \bar{\bar{y}}'$, respectively, the μ-ess inf y and μ'-ess inf y in D [for which $\lambda(\bar{\bar{y}}) = \bar{\bar{\lambda}}$, $\lambda(\bar{\bar{y}}') = \bar{\bar{\lambda}}'$].

(i) Assume first that $\bar{\bar{\lambda}} = \bar{\bar{\lambda}}'$ (in which case $\bar{\bar{y}} \in \Gamma$ and corresponds to $\bar{\bar{y}}'$), or else that $\bar{\bar{y}} \notin \Gamma$ and $\bar{\bar{y}}'$ does not correspond to any point in Γ. This amounts to assuming that either $\mathscr{C}, \mathscr{C}'$ coincide at their initial points on the support of μ, μ' or that neither one of these two initial points belongs to the other curve.

From the theory of simple curves, the existence follows, on each connected component of Γ, of a strictly monotonic C^1 function, such that (3.14) holds (see for instance [16]). Moreover, due to monotonicity properties of λ, λ', a strictly monotonic C^1 function $h\colon D \to D$ can be built, such that it coincides with those already found when restricted to each component of Γ.

Clearly λ and $\lambda'(h(\cdot))$ are such that if $\lambda(y) = \lambda'(h(\tilde{y}))$ then $y = \tilde{y}$ and therefore they belong to a Λ_2-type set.

By the change of variable $y' = h(y)$ in the right hand side of (3.1), on the basis of Theorem 1, the equalities (3.13), (3.14) follow. Uniqueness of such h on the support of μ follows from monotonicity assumptions on λ, λ'.

Incidentally, notice that (3.14) implies that Γ indeed is the whole support of μ.

(ii) Suppose now that $\bar{\bar{y}} \notin \Gamma$, but $\bar{\bar{y}}'$ corresponds to some point y^* in Γ. Then $y^* > \bar{\bar{y}}$. By repeating the procedure in (i), we can establish the existence and uniqueness of the requested $h\colon [y^*, \bar{y}] \to D$. By an obvious change of variable, from (3.1) we get

$$\begin{aligned}\int_D N(x; \lambda(y))\mu(dy) &= \int_D N(x; \lambda'(y'))\mu'(dy') \\ &= \int_{[y^*, \bar{y}]} N(x; \lambda'(h(y))\mu'(dh(y)).\end{aligned}$$

Let us introduce $\mu^* \in \mathscr{P}(D)$ such that

$$\mu^*(dy) = \mu'(dh(y)), \qquad y \in [y^*, \bar{y}],$$
$$\mu^*(D - [y^*, \bar{y}]) = 0.$$

Then, no matter how we extend $h \in C^1$ on $D - [y^*, \bar{y}]$, we have

$$T(\lambda, \mu) = \int_D N(x; \lambda(y))\mu(dy) = \int_D N(x; \lambda'(h(y))\mu^*(dy) = T(\lambda'(h(\cdot)), \mu^*).$$

By Theorem 1, $\mu = \mu^*$ and in particular we get

$$\mu(D - [y^*, \bar{y}]) = 0.$$

This contradicts $\bar{\bar{y}} < y^*$, since, by definition, $\bar{\bar{y}} = \mu - \operatorname{ess\,inf} y$ in D.

(iii) In case $\bar{\bar{y}} \in \Gamma$, but $\bar{\bar{y}}'$ does not correspond to any point in Γ, it is enough to exchange primed and unprimed quantities to again arrive at a contradiction. □

REMARK. Let us restrict Λ_4 to its subset Λ_5 defined by

$$\Lambda_5 = \{\lambda \in \Lambda_1 \colon \lambda(y) = \lambda(y') \Rightarrow y = y';\ \lambda_1, \lambda_2 \text{ monotonic};$$
$$|\dot{\lambda}_1(y)| + |\dot{\lambda}_2(y)| \geq K_3 > 0, \forall\, y \in D,$$
$$\exists\, \psi \colon \mathbb{R} \to \mathbb{R}^+,\ \psi \text{ continuous at } 0,\ \psi(x) = 0 \Rightarrow x = 0$$
$$\text{such that } \|\lambda(y) - \lambda(y')\| \geq \psi(y - y')\}.$$

Denoting by A_5 the space of equivalence classes in $\Lambda_5 \times \mathscr{P}(D)$ with respect to the equivalence defined by (3.13), (3.14), we may repeat the compactness argument already used in Theorem 2 to prove continuity of T^{-1}: $\mathscr{R}(A_5) \to A_5$.

REMARK. The results obtained in this section clearly contain as a particular case those results already known for finite Gaussian mixtures (in which case μ is restricted to be atomic) [8], [11], [15], [23]–[26], [29]. Of course, in this case no continuity properties for λ are requested. Moreover, Theorem 3 holds without introducing monotonicity assumptions.

4. Additional results.

(a) *Choice of topologies.* The choice of the topology of $H(\mathbb{R})$ (induced by the sup norm) is unessential, in the sense that the continuity properties of T and T^{-1} are preserved under other suitable topologies.

For instance, let us substitute $H(\mathbb{R})$ by the space $H_q(\mathbb{R})$ of integrable analytic functions on $\mathbb{R}$, with the norm

$$\|f\|_q^2 = \int_{-\infty}^{+\infty} f^2(x) q^2(x)\, dx, \tag{4.1}$$

where $\int_{-\infty}^{+\infty} q^2(x)\, dx = Q < \infty$ and $\inf_{x \in I} q^2(x) > 0$, for any finite interval I.

It is immediately verified that convergence in $H(\mathbb{R})$ implies convergence in $H_q(\mathbb{R})$. Furthermore, convergence in $\mathscr{R}(A_3)$ as a subset of $H_q(\mathbb{R})$ implies convergence in $\mathscr{R}(A_3)$ as a subset of $H(\mathbb{R})$, $\mathscr{R}(A_3)$ being a set of continuous uniformly bounded functions, uniformly vanishing at infinity. These implications allow to preservation of continuity of T and T^{-1} when the norm (4.1) is adopted and therefore support possible (numerical) inversion algorithms for T based on minimization of quadratic error indices [5].

On the other hand, the choice of the weak topology on $\mathscr{P}(D)$ is crucial as far as the continuity of T^{-1} is concerned.

Indeed one can easily verify that the choice of a stronger topology on $\mathscr{P}(D)$, like that induced by the total variation norm, destroys the continuity of T^{-1}. Thus, algorithms aimed at the numerical inversion of T appear to be ill-conditioned when analyzed in the light of the stronger kind of topology [2], [17], [27], [30]. On the contrary, no ill-conditioning phenomenon arises when the weak topology is adopted [5]; and the latter appears to be sufficiently fine for most applicative problems. In that respect, we remark that the choice of a parametrization for the unknown μ, and consequently the convergence of μs in the sense of convergence of the sets of parameters thus attached to them, corresponds to a choice for the topology itself. It may thus be understood that an improper choice for unknown parameters (like the samples of a possible density for μ) can lead to ill-conditioned problems and call for regularization procedures [10], [17], [18].

(b) *Multidimensional case.* The main results of Section 3, namely Theorems 1 and 2, hold also in the multidimensional case ($p > 1$, $n > 1$), modulo some suitable modifications in the definition of Λ_i, $i = 1, 2, 3$. The situation in which $n > 1$ is indeed already included in the formulation of the theorems, since in their proof (as well as in the proof of the lemma) the dimensionality of y plays no role.

As far as the general case is concerned, definitions and proofs can be extended following the same kind of arguments, at the expenses of an increase in the formal complexity. Here, to give some hints on how one can proceed, we sketch the extension of the proofs for the case $p = 2$, $n > 1$.

The vector $\lambda(y)$ is now five-dimensional, and its components include the elements of the mean value vector $m(y)$ and of the variance matrix $\Sigma(y)$:

$$m(y) = (\lambda_1(y)\ \lambda_2(y))^T, \tag{4.2}$$

$$\Sigma(y) = \begin{pmatrix} \lambda_3(y) & \lambda_4(y) \\ \lambda_4(y) & \lambda_5(y) \end{pmatrix}. \tag{4.3}$$

The set Λ_1 is defined as

$$\Lambda_1 = \Big\{ \lambda \in C^1(D) \colon |\lambda_i(y)| \le k_1 < \infty,\ i = 1, \dots 5,\ 0 \le \lambda_i(y),\ i = 3, 5;$$
$$|\lambda_3(y)\lambda_5(y) - \lambda_4^2(y)| \ge s_1 > 0;\ \sum_{i=1}^{5} |\dot{\lambda}_i(y)| \le k_2 < \infty,\ \forall\, y \in D \Big\}$$

while Λ_2, Λ_3 are defined as above. Denoting the inverse of $\Sigma(y)$ by

$$\Sigma^{-1}(y) = \begin{pmatrix} a(y) & b(y) \\ b(y) & c(y) \end{pmatrix} \tag{4.4}$$

and introducing the notations

$$\begin{aligned}\psi_h &= \left(a^{h_1}b^{h_2}c^{h_3}(\lambda_1 a + \lambda_2 b)^{h_4}(\lambda_1 b + \lambda_2 c)^{h_5}\right)\circ \mu,\\ \psi'_h &= \left(a^{h_1}b^{h_2}c^{h_3}(\lambda'_1 a + \lambda'_2 b)^{h_4}(\lambda'_1 b + \lambda'_2 c)^{h_5}\right)\circ \mu',\end{aligned} \tag{4.5}$$

where h is the five-dimensional vector with integer components h_i, the thesis (2.3) of the lemma becomes

$$T(\lambda, \psi_h) = T(\lambda', \psi'_h). \tag{4.6}$$

To prove (4.6) we again proceed by induction. Equation (4.6) is already guaranteed by (2.2) for $h = 0$.

Let us denote by $e(i)$, $i = 1, 2, \dots 5$, the five-dimensional vector with the ith component equal to 1 and the other ones equal to 0. By differentiating both sides of (4.6) we get

$$\begin{aligned} &x_1\left[\left(T(\lambda, \psi_{h+e(1)})\right)(x) - \left(T(\lambda', \psi'_{h+e(1)})\right)(x)\right]\\ &\quad + x_2\left[\left(T(\lambda, \psi_{h+e(2)})\right)(x) - \left(T(\lambda', \psi'_{h+e(2)})\right)(x)\right]\\ &\qquad = \left(T(\lambda, \psi_{h+e(4)})\right)(x) - \left(T(\lambda', \psi'_{h+e(4)})\right)(x)\end{aligned} \tag{4.7}$$

and

$$\begin{aligned} &x_1\left[\left(T(\lambda, \psi_{h+e(2)})\right)(x) - \left(T(\lambda', \psi'_{h+e(2)})\right)(x)\right]\\ &\quad + x_2\left[\left(T(\lambda, \psi_{h+e(3)})\right)(x) - \left(T(\lambda', \psi'_{h+e(3)})\right)(x)\right]\\ &\qquad = \left(T(\lambda, \psi_{h+e(5)})\right)(x) - \left(T(\lambda'\psi'_{h+e(5)})\right)(x).\end{aligned} \tag{4.8}$$

By setting one of the components of x equal to zero, and processing Eqs. (4.7), (4.8) (in which the other component of x is the independent variable) in the same way that we used for Eq. (2.6) in the lemma, the proof of (4.6) by induction can be carried out.

The formulation of Theorem 1 is unchanged. To prove it, we define $\mathscr{C}, \mathscr{C}'$ according to (3.2) in $\mathbb{R}^5$, and the set $E = \{\xi \in \mathbb{R}^5\colon |\xi_i| \le k_1 < \infty,\ i = 1, 2, \dots 5;\ 0 \le \xi_i,\ i = 3, 5;\ |\xi_3\xi_5 - \xi_4^2| \ge s_1 > 0\}$. We consider $\phi \in C(E)$ such that

$$\phi(\xi) = 1, \qquad \xi \in \mathscr{C} \subset E,$$

$$\phi(\xi) = \prod_{i=1}^{5} \frac{(\lambda'_i(y) + 2k_1)^{r_i}}{(\lambda_i(y) + 2k_1)^{r_i}}, \qquad \xi \in \mathscr{C}' \subset E,\ r_i \text{ nonnegative integers.}$$

Again, the algebra generated by the functions in $C(E)$

$$\left\{\frac{\xi_3}{\xi_3\xi_5 - \xi_4^2}, \frac{-\xi_4}{\xi_3\xi_5 - \xi_4^2}, \frac{\xi_5}{\xi_3\xi_5 - \xi_4^2}, \frac{\xi_1\xi_3 - \xi_2\xi_4}{\xi_3\xi_5 - \xi_4^2}, \frac{-\xi_1\xi_4 + \xi_2\xi_5}{\xi_3\xi_5 - \xi_4^2}\right\}$$

is such that its linear span is dense in $C(E)$. Thus the same argument already exploited in Theorem 1 leads us to the conclusion. No modification is needed for Theorem 2 and its proof.

(c) *Non-Gaussian kernels.* The results in Section 3 offer a useful tool to completely solve the identifiability problem for mixtures of a wider class of distributions, namely that one obtained by convex combinations of a finite number of Gaussian distributions. This class is of paramount relevance in application, since it includes a large variety of possible shapes (asymmetric, multimodal, etc.).

Let us consider the problem of solving the equation

$$f(x) = \sum_{i=1}^{\nu} \alpha_i \int_D N(x; \lambda^{(i)}(y))\mu^{(i)}(dy), \tag{4.9}$$

where ν is a known integer; $\alpha_i \geq 0$, $i = 1, \dots \nu$; $\sum_{i=1}^{\nu}\alpha_i = 1$; $\mu^{(i)} \in \mathscr{P}(D)$, $i = 1, \dots \nu$ (and possibly $\mu^{(i)} = \mu$, $i = 1, \dots \nu$). Each $\lambda^{(i)}$ is assumed to belong to a Λ_2-type set $\Lambda_2^{(i)}$, with the constraint $\mathscr{R}_i \cap \mathscr{R}_j = \varnothing$, $i, j = 1, 2, \dots \nu$, $i \neq j$ where $\mathscr{R}_i$ is the union of the ranges of elements of $\Lambda_2^{(i)}$. One might then consider ν disjoint closed sets $D_1, \dots D_\nu$ and ν strictly monotonic C^1 functions $h^{(i)}$, from D_i onto D, such that Eq. (4.9), by suitable changes of variable, can be rewritten as

$$\begin{aligned} f(x) &= \sum_{i=1}^{\nu} \alpha_i \int_{D_i} N(x; \lambda^{(i)}(h^{(i)}(z))\mu^{(i)}(dh^{(i)}(z)) \\ &= \int_{\tilde{D}} N(x; \tilde{\lambda}(z))\tilde{\mu}(dz), \end{aligned} \tag{4.10}$$

where $\tilde{D}$ is connected, $\tilde{D} \supset \bigcup_{i=1}^{\nu} D_i$, and $\tilde{\lambda}, \tilde{\mu}(dz)$ are such that their restrictions to D_i are, respectively, $\lambda^{(i)}(h^{(i)})$ and $\alpha_i\mu^{(i)}(dh^{(i)}(z))$. Due to the above construction Eq. (4.6) must be solved with respect to $\tilde{\lambda}, \tilde{\mu}$ where $\tilde{\mu} \in \mathscr{P}(\tilde{D})$, $\tilde{\mu}(\tilde{D} - \bigcup_{i=1}^{\nu} D_i) = 0$ and $\tilde{\lambda} \in \tilde{\Lambda}_2$, $\tilde{\Lambda}_2$ being the Λ_2-type set which follows from the original assumptions on the $\lambda^{(i)}$s. Theorem 1 now guarantees uniqueness of the solution $(\tilde{\lambda}, \tilde{\mu})$, from which, by the inverse transformations defined by the $h^{(i)}$s, the unique solution of the original problem (4.9) is easily obtained.

Further results on continuity of the inverse operator in (4.9) and uniqueness of its solution in the sense of Theorem 3 can be similarly established by suitable modifications of the hypotheses on the $\lambda^{(i)}$s.

(d) *Kernels with piecewise continuous parameter functions.* The case of kernels in which the parameter function satisfies the basic assumptions only in a piecewise form can be handled by arguments similar to those of point (c) above, provided μ is assumed to be nonatomic (at least in the points of discontinuity for λ).

(e) *Signed measures.* The result in Theorem 1 can be extended to the case of $\mu, \mu' \in \mathscr{M}(D)$. Indeed, let the Hahn decomposition for μ, μ' be

$$\mu = \mu^+ - \mu^-, \qquad \mu' = \mu'^+ - \mu'^-$$

with supports D^+, D^-, D'^+, D'^-, respectively. Equation (3.1) may then be rewritten as

$$\int_{D^+} N(x;\lambda(y))\mu^+(dy) + \int_{D'^-} N(x;\lambda'(y))\mu'^-(dy)$$
$$= \int_{D'^+} N(x;\lambda'(y))\mu'^+(dy) + \int_{D^-} N(x;\lambda(y))\mu^-(dy).$$

By normalization

$$(4.11)\qquad \begin{aligned} &a\int_{D^+} N(x;\lambda(y))\mu_1^+(dy) + 1 - a\int_{D'^-} N(x;\lambda'(y))\mu_1'^-(dy) \\ &\quad = a'\int_{D'^+} N(x;\lambda'(y))\mu_1'^+(dy) + (1-a')\int_{D^-} N(x;\lambda(y))\mu_1^-(dy), \end{aligned}$$

where $a = \|\mu^+\|/(\|\mu^+\| + \|\mu'^-\|)$ and $a' = \|\mu'^+\|/(\|\mu'^+\| + \|\mu^-\|)$, with $\mu_1^+, \mu_1^-, \mu_1'^+, \mu_1'^- \in \mathscr{P}(D)$, and $\|\mu\|$ denoting the total variation of μ.

It is possible now to consider a C^1 strictly monotonic function h from D onto D, with D_1 closed and $D_1 \cap D = \varnothing$ and define by it a change of variable in the second integral on both sides of (4.11). Then each side of (4.11) takes the form of summation in (4.10):

$$(4.12)\qquad \int_{\tilde{D}} N(x;\tilde{\lambda}(z))\tilde{\mu}(dz) = \int_{\tilde{D}} N(x;\tilde{\lambda}'(z))\tilde{\mu}'(dz),$$

where $\tilde{D} = D \cup D_1$, $\tilde{\mu}(dz) = \mu_1^+(dz) + \mu_1'^-(dh(z))$, $\tilde{\mu}'(dz) = \mu_1'^+(dz) + \mu_1^-(dh(z))$, and $\tilde{\lambda}$ ($\tilde{\lambda}'$) is such that its restriction to D and D_1 is, respectively, λ (λ') and $\lambda'(h)$ ($\lambda(h)$).

Note that neither one of the supports of $\tilde{\mu}, \tilde{\mu}'$ contains $(D^- \cap D'^-) \cup (D_1^+ \cap D_1'^+)$ where $D_1^+, D_1'^+$ are the inverse images of D^+, D'^+ under h. In $\tilde{D}/(D^- \cap D'^-) \cup (D_1^+ \cap D_1'^+)$ $\tilde{\lambda}, \tilde{\lambda}'$ are such that $\tilde{\lambda}(z) = \tilde{\lambda}'(z') \Rightarrow z = z'$; thus they can be assumed to belong to a Λ_2-type set.

Theorem 1 now guarantees $\tilde{\mu} = \tilde{\mu}'$ and $\tilde{\lambda} = \tilde{\lambda}'$ ($\tilde{\mu}$-a.e.). Therefore $\mu = \mu'$ and $\lambda = \lambda'$ (μ-a.e.).

5. Concluding remarks. We have shown that the problem of identifiability of mixtures of Gaussian distributions is well posed in a quite general setting: Given a density that is a mixture of Gaussian densities with possibly unknown means and variances, it uniquely corresponds to a mixing distribution as well as to a pair of functions for their mean and variance. Moreover, this correspondence is continuous in both ways. Note that in [13], [14] a similar problem is discussed with reference to mixtures of exponential distributions evidencing difficulties due to lack of continuity and ill-conditioning.

Some of the assumptions we introduced (like compactness of D, or absence of multiple points for λ) are essential conditions for uniqueness and continuity results. Other ones (such as Gaussianness of the kernel, continuity of its parameter functions, positivity of mixing measures) can be relaxed, as mentioned in the previous section.

As far as actual identification of the pair (λ, μ), which within a prescribed subset of $\Lambda_2 \times \mathscr{P}(D)$ corresponds to an experimentally given f, the effect of an additional error term in Eq. (1.1) cannot be disregarded. Besides the usual measurement error, this term might have to account for approximations in the kernel structure and/or for errors on f due to its determination as a frequency distribution of a finite population sample [4].

In this context the identification problem becomes an estimation problem for (λ, μ). An error functional is to be defined on the basis of the statistics of the error term; then its minimum point is to be looked for within the prescribed (λ, μ) set [see Section 4(a) above]. Now difficulties may well arise since this set is not guaranteed to be convex (particularly with respect to parameters in λ), so that uniqueness of the global minimum point might not hold any more [4].

Investigation on how to properly set the above-mentioned problem and how to solve it constitutes the object of future research work.

Acknowledgment. The authors are indebted to J. C. Willems and one of the referees for some stimulating comments on an earlier version of the paper.

REFERENCES

[1] AHMED, K. E. and AL-HUSSAINI, E. K. (1982). Remarks on the nonidentifiability of mixtures of distributions. *Ann. Inst. Statist. Math.* **34** 543–544.

[2] BAISCH, H. ET AL. (1982). A comparison of mathematical methods for the analysis of DNA histograms obtained by flow cytometry. *Cell Tissue Kinet.* **15** 235–249.

[3] BILLINGSLEY, P. (1968). *Convergence of probability measures.* Wiley, New York.

[4] BRUNI, C., CURZI, L., KOCH, G. and ROSSI, C. (1981). Biological aspects and mathematical analysis of the inverse problem in flow cytometry. IASI Report R.16.

[5] BRUNI, C., KOCH, G. and ROSSI, C. (1983). On the inverse problem in cytofluorometry: Recovering DNA distribution from FMF data. *Cell Biophys.* **5** 5–19.

[6] CHRISTENSEN, P. R. ET AL. (1980). Gamma-ray multiplicity moments from ^{86}Kr reactions on ^{144,154}Sm at 490 MeV. *Nuclear Phys. A* **349** 217–257.

[7] DELVES, L. M. and WALSH, J. (1974). *Numerical Solutions of Integral Equations.* Oxford Univ. Press.

[8] EVERITT, B. S. and HAND, D. J. (1981). *Finite Mixture Distributions.* Chapman and Hall, London.

[9] FRASER, M. D., HSU, Y. and WALKER, J. J. (1981). Identifiability of finite mixtures of Von Mises distributions. *Ann. Statist.* **9** 1130–1137.

[10] FRIED, J. (1977). Analysis of deoxyribonucleic acid histograms from flow cytofluorometry. *J. Histology and Cytology* **25** 924–951.

[11] GUPTA, R. C. (1982). On the determination of mixing densities in reliability studies. *J. Statist. Plann. Inference* **7** 101–105.

[12] HARRIS, C. M. and SINGPURWALLA, N. D. (1968). Life distribution derived from stochastic hazard functions. *IEEE Trans. Reliability* **17** 70–79.

[13] HECKMAN, J. and SINGER, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52** 271–320.

[14] HECKMAN, J. and SINGER, B. (1984). The identifiability of the proportional hazard model. *Rev. Econom. Stud.* **51** 231–241.

[15] KENT, J. T. (1983). Identifiability of finite mixtures for directional data. *Ann. Statist.* **11** 984–988.

[16] KLINGENBERG, W. (1978). *A Course in Differential Geometry.* Springer-Verlag, Berlin.

[17] MERZ, P. H. (1980). Determination of adsorption energy distribution by regularization and a characterization of certain adsorption isotherms. *J. Comput. Phys.* **34** 64–85.

[18] NEDELMAN, J. (1982). Error analysis in flow microfluorometry. *Cell Biophys.* **4** 1–14.

[19] PADGETT, W. J. and TSOKOS, C. P. (1978). On Bayes estimation of reliability for mixtures of life

distributions. *SIAM J. Appl. Math.* **34** 692–703.

[20] RANA, I. K. (1983). A Pompeiu problem for locally finite measures on R^n. *Rend. Mat.* (7) **3** 291–300

[21] RUDIN, W. (1973). *Functional Analysis.* McGraw-Hill, New York.

[22] TALLIS, G. M. and CHESSON, P. (1982). Identifiability of mixtures. *J. Austral. Math. Soc. Ser. A* **32** 339–348.

[23] TEICHER, H. (1960). On the mixture of distributions. *Ann. Math. Statist.* **31** 55–73.

[24] TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.

[25] TEICHER, H. (1963). Identifiability of finite mixtures. *Ann. Math. Statist.* **34** 1265–1269.

[26] TEICHER, H. (1967). Identifiability of mixtures of product measures. *Ann. Math. Statist.* **38** 1300–1302.

[27] VAN DER LINDEN, P. M. (1980). An iterative method of analysis for DNA histograms. *Flow Cytometry* **IV** 133–137.

[28] XEKALAK, E. and PANARETOS, J. (1983). Identifiability of compound Poisson distributions. *Scand. Actuar. J.* **66** 39–45.

[29] YAKOWITZ, S. T. and SPRAGINS, J. D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.* **39** 209–214.

[30] ZIETZ, S. and NICOLINI, C. (1978). Flow microfluorometry and cell kinetics: A review. In: *Biomathematics and Cell Kinetics* (A. J. Valleron and P. D. M. Macdonald, eds.) Elsevier/North-Holland, Amsterdam.

DIPARTIMENTO DI INFORMATICA E SISTEMISTICA
UNIVERSITY OF ROME-LA SAPIENZA
VIA EUDOSSIANA 18
00184 ROME
ITALY

DIPARTIMENTO DI MATEMATICA
"G. CASTELNUOVO"
UNIVERSITY OF ROME-LA SAPIENZA
PIAZZALE A. MORO 5
00185 ROME
ITALY