A CRAMÉR–WOLD THEOREM FOR MIXTURES

RICARDO FRAIMAN, LEONARDO MORENO, AND THOMAS RANSFORD

Abstract. We show how a Cramér–Wold theorem for a family of multivariate probability distributions can be used to generate a similar theorem for mixtures (convex combinations) of distributions drawn from the same family.

Using this abstract result, we establish a Cramér–Wold theorem for mixtures of multivariate Gaussian distributions. According to this theorem, two such mixtures can be distinguished by projecting them onto a certain predetermined finite set of lines, the number of lines depending only on the total number Gaussian distributions involved and on the ambient dimension. A similar result is also obtained for mixtures of multivariate t-distributions.

1. Introduction

The Cramér–Wold device is the name given to a general technique for analyzing multivariate probability distributions via their lower-dimensional projections. The name originates from a classical theorem of Cramér and Wold [3] to the effect that a probability measure in Euclidean d-dimensional space is uniquely determined by its one-dimensional projections in all directions.

For certain classes of measures, one can do better. Indeed, in some cases, just finitely many projections suffice to determine the measure. As an example, we cite the case of elliptic measures in $\mathbb{R}^d$ (which includes that of Gaussian measures), where just $(d^2+d)/2$ suitably chosen projections suffice (see [5]).

The central object of study in this article is that of *mixtures*, namely convex combinations of probability measures. For example, suppose that we already have a Cramér–Wold device for a certain family of probability measures: can we then deduce a similar device for mixtures of measures taken from that family? We give an affirmative answer to this question in certain cases, including the very important one of Gaussian mixtures.

Here, in more detail, is a road-map of the article.

2020 *Mathematics Subject Classification.* 60B11.

Key words and phrases. Projection, identifiability, mixture, Gaussian distribution, t-distribution.

Fraiman and Moreno were supported by grant FCE-3-2022-1-172289, Agencia Nacional de Investigación e Innovación, Uruguay. Ransford was supported by NSERC Discovery Grant RGPIN–2020–04263.

Section 2 begins with a short introduction to the notion of the Cramér–Wold device. We then establish a rather general abstract result, Theorem 2.1, which shows how a Cramér–Wold theorem for a class of measures can be used to generate a similar theorem for mixtures drawn from the same class.

To apply Theorem 2.1, it is necessary to check that certain families of probability distributions are linearly independent, or equivalently, that they are identifiable. There is an extensive literature on techniques for proving identifiability. In Section 3, we review results concerning the identifiability of two particular families: the Gaussian distributions and the t-distributions. We also derive what we believe to be a new criterion for identifiability, based on a deep theorem in complex analysis due to Borel.

In Section 4, these ideas are applied to the important special case of Gaussian measures and Gaussian mixtures. In particular, we derive a Cramér–Wold theorem for Gaussian mixtures, Theorem 4.1, from the previously known one for Gaussian measures. It says that, if the one-dimensional projections of two mixtures of ℓ and m normal distributions on $\mathbb{R}^d$ coincide on a certain pre-determined set of $(1/2)(l+m-1)(d^2+d-2)+1$ lines, then the two distributions are the same. A similar result, Theorem 4.2, is also established for mixtures of multivariate t-distributions.

We conclude in Section 5 with some remarks about potential applications of these results.

2. A Cramér–Wold theorem for mixtures

2.1. **Introduction.** Given a Borel probability measure P on $\mathbb{R}^d$ and a vector subspace H of $\mathbb{R}^d$, we write P_H for the projection of P onto H, namely the Borel probability measure on H given by

$$P_H(B) := P(\pi_H^{-1}(B)),$$

where $\pi_H : \mathbb{R}^d \to H$ is the orthogonal projection of $\mathbb{R}^d$ onto H.

According to a well-known theorem of Cramér and Wold [3], if P, Q are two Borel probability measures on $\mathbb{R}^d$, and if $P_L = Q_L$ for all lines L, then $P = Q$.

There have been numerous refinements of the Cramér–Wold theorem. To help describe these, it is convenient to introduce the following terminology. Let $\mathcal{P}$ be a family of Borel probability measures on $\mathbb{R}^d$ and let $\mathcal{H}$ be a family of vector subspaces of $\mathbb{R}^d$ (not necessarily all of the same dimension). We say that $\mathcal{H}$ is a *Cramér–Wold system* for $\mathcal{P}$ if, for every pair $P, Q \in \mathcal{P}$,

$$P_H = Q_H \ (\forall H \in \mathcal{H}) \quad \Rightarrow \quad P = Q.$$

In this terminology, the original Cramér–Wold theorem says simply that the family of all lines in $\mathbb{R}^d$ is a Cramér–Wold system for the family of Borel probability measures on $\mathbb{R}^d$. Here are some other examples.

- If $\mathcal{P}$ is the family of compactly supported Borel probability measures on $\mathbb{R}^2$, then any infinite set of lines in $\mathbb{R}^2$ is a Cramér–Wold system for $\mathcal{P}$ (Rényi [17, Theorem 1]).
- If $\mathcal{P}$ is the family of probability measures on $\mathbb{R}^d$ whose supports contain at most k points, and if $\mathcal{H} = \{H_1, \dots, H_{k+1}\}$ is any family of $k+1$ subspaces of $\mathbb{R}^d$ such that $H_i^\perp \cap H_j^\perp = \{0\}$ whenever $i \neq j$, then $\mathcal{H}$ is a Cramér–Wold system for $\mathcal{P}$ (Heppes, [11, Theorem 1′]).
- If $\mathcal{P}$ is the family of probability measures P on $\mathbb{R}^d$ whose moments $m_n := \int_{\mathbb{R}^d} \|x\|^n \, dP(x)$ are finite and satisfy $\sum_n m_n^{-1/n} = \infty$, and if $\mathcal{H}$ is any family of subspaces such that $\cup_{H \in \mathcal{H}} H$ has positive Lebesgue measure in $\mathbb{R}^d$, then $\mathcal{H}$ is a Cramér–Wold system for $\mathcal{P}$ (Cuesta-Albertos *et al* [4, Corollary 3.2]).
- If $\mathcal{P}$ is the family of elliptical distributions on $\mathbb{R}^d$ and if $\mathcal{L}$ is the set of lines $\langle e_i + e_j \rangle$ $(1 \le i \le j \le d)$, where $\{e_1, \dots, e_d\}$ is any basis of $\mathbb{R}^d$, then $\mathcal{L}$ is a Cramér–Wold system for $\mathcal{P}$ (Fraiman *et al*, [5, Theorem 1]).

Further results of this kind may be found in [1, 6, 7, 8, 10].

2.2. **Mixtures.** Let $d \ge 1$ and let $\mathcal{P}$ be a family of Borel probability measures on $\mathbb{R}^d$.

We say that $\mathcal{P}$ is *linearly dependent* if there exist distinct $P_1, \dots, P_n \in \mathcal{P}$ and non-zero $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that $\sum_{j=1}^n \lambda_j P_j = 0$. Otherwise $\mathcal{P}$ is *linearly independent*.

A $\mathcal{P}$-*mixture* is a convex combination of measures from $\mathcal{P}$, in other words, a measure of the form $\sum_{j=1}^n \lambda_j P_j$, where $P_1, \dots, P_n \in \mathcal{P}$ and $\lambda_1, \dots, \lambda_n \ge 0$ with $\sum_{j=1}^n \lambda_j = 1$.

Our aim in this section is to establish the following Cramér–Wold theorem for $\mathcal{P}$-mixtures.

Theorem 2.1. *Let $\mathcal{P}$ be a family of Borel probability measures on $\mathbb{R}^d$, let $\mathcal{H}$ be a collection of vector subspaces of $\mathbb{R}^d$, and let $\ell, m \ge 1$. Suppose that:*

(i) *For each $H \in \mathcal{H}$, the distinct measures in the set $\{P_H : P \in \mathcal{P}\}$ are linearly independent.*

(ii) *Each partition of $\mathcal{H}$ into $\ell+m-1$ subsets contains at least one Cramér–Wold system for $\mathcal{P}$.*

Let P and Q be convex combinations of ℓ and m measures from $\mathcal{P}$ respectively. If $P_H = Q_H$ for all $H \in \mathcal{H}$, then $P = Q$.

Proof. We argue by contradiction. Suppose that $P_H = Q_H$ for all $H \in \mathcal{H}$, but that $P \neq Q$. Then the difference $P - Q$ can be written as $P - Q = \sum_{j=1}^k \lambda_j P_j$, where $P_1, \dots, P_k \in \mathcal{P}$ are distinct, where $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ are non-zero, and where $k \le \ell + m$. Set

$$\mathcal{H}_j := \{H \in \mathcal{H} : (P_1)_H = (P_j)_H\} \quad (j = 2, \dots, k).$$

Clearly no $\mathcal{H}_j$ is Cramér–Wold system for $\mathcal{P}$. If we had $\cup_{j=2}^k \mathcal{H}_j = \mathcal{H}$, then we could construct a partition of $\mathcal{H}$ into $k-1$ sets (perhaps some of them empty) none of which is a Cramér–Wold system for $\mathcal{P}$, contradicting the assumption (ii) on $\mathcal{H}$. We conclude that $\cup_{j=2}^{k}\mathcal{H}_j \neq \mathcal{H}$, so there exists an $H_0 \in \mathcal{H}$ such that $(P_1)_{H_0} \neq (P_j)_{H_0}$ $(j = 2, \dots, k)$.

By assumption (i), the distinct measures in the set $\{(P_j)_{H_0} : j = 1, \dots d\}$ are linearly independent. In particular, since $(P_1)_{H_0} \neq (P_j)_{H_0}$ $(j = 2, \dots, k)$, it follows that $(P_1)_{H_0}$ is not in the span of $\{(P_j)_{H_0} : j = 2, \dots d\}$. On the other hand, we have

$$\sum_{j=1}^{k} \lambda_j (P_j)_{H_0} = \Big(\sum_{j=1}^{k} \lambda_j P_j\Big)_{H_0} = (P - Q)_{H_0} = P_{H_0} - Q_{H_0} = 0.$$

Since $\lambda_1 \neq 0$, this is a contradiction. □

3. Linear independence and identifiability

3.1. **Introduction.** Condition (i) in Theorem 2.1 begs the question as to how one determines whether a set of measures on a subspace H of $\mathbb{R}^d$ is linearly independent.

In the statistical literature, the notion of linear independence of measures is synonymous with that of identifiability. We say that a family $\mathcal{P}$ of Borel probability measures on $\mathbb{R}^d$ is *identifiable* if it is impossible to express any $\mathcal{P}$-mixture as two different convex combinations of elements of $\mathcal{P}$. It is easy to see that $\mathcal{P}$ is identifiable if and only if it is linearly independent (see e.g. [19, Theorem 3.1.1]). There is an extensive literature concerning techniques for proving identifiability/linear independence. A useful background reference is the book [19].

3.2. **Linear independence of Gaussian and t-distributions.** We shall need two results in particular. The first concerns the family of Gaussian distributions, with densities

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \quad (x \in \mathbb{R}). \tag{1}$$

Theorem 3.1. *The family of Gaussian distributions $\{f_{\mu,\sigma} : \mu \in \mathbb{R}, \sigma > 0\}$ on $\mathbb{R}$ is linearly independent.*

Proof. This result is well known. See for example [19, Example 3.1.4]. □

The second result treats the identifiability/linear independence of the family of t-distributions. Recall that a *t-distribution* on $\mathbb{R}$ with ν degrees of freedom is a Borel measure with density of the form

$$f_{\nu,\mu,\sigma}(x) = c_{\nu,\mu,\sigma}\Big(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\Big)^{-(\nu+1)/2},$$

where ν is a positive integer, $\mu \in \mathbb{R}$ and $\sigma > 0$. The constant $c_{\nu,\mu,\sigma}$ is chosen to ensure that $\int_{\mathbb{R}} f_{\nu,\mu,\sigma}(x)\,dx = 1$. This distribution has mean μ (if $\nu > 1$) and variance $\sigma^2\nu/(\nu-2)$ (if $\nu > 2$).

Theorem 3.2. *The family of t-distributions* $\{f_{\nu,\mu,\sigma} : \nu \in \mathbb{Z}^+, \mu \in \mathbb{R}, \sigma > 0\}$ *is linearly independent.*

Special cases of this result have previously been obtained by Otiniano *et al.* (case $\mu = 0$), see [16, Proposition 4], and by Ho–Nguyen (case ν odd), see [12, Theorem 3.4(b)]. However, we have not been able to find the general case elsewhere in the literature, so we present a proof here.

Proof. Suppose, on the contrary, that there is a linear relation of the form

$$\sum_{j=1}^n \lambda_j f_{\nu_j,\mu_j,\sigma_j}(x) = 0 \quad (x \in \mathbb{R}), \tag{2}$$

where the triples (ν_j, μ_j, σ_j) are distinct and the coefficients $\lambda_j \neq 0$. Reordering the triples, we may further suppose that $\nu_1 \geq \nu_j$ for all $j \geq 2$.

If all the ν_j are odd integers, then we may argue as follows. Each function f_{ν_j,μ_j,σ_j} extends to a function in the complex plane $\mathbb{C}$

$$f_{\nu_j,\mu_j,\sigma_j}(z) = c\Big(1 + \frac{(z-\mu_j)^2}{\nu_j\sigma_j^2}\Big)^{-(\nu_j+1)/2},$$

which is holomorphic on $\mathbb{C}$ except for poles at the two points $\mu_j \pm i\sigma_j\sqrt{\nu_j}$. Moreover, the order of the poles at these points is exactly $(\nu_j + 1)/2$. In particular, it follows that, if $j \geq 2$, then either f_{ν_j,μ_j,σ_j} has no singularity at $\mu_1 + i\sigma_1\sqrt{\nu_1}$, or it has a pole there of order strictly less than ν_1. Thus

$$\frac{f_{\nu_j,\mu_j,\sigma_j}(z)}{f_{\nu_1,\mu_1,\sigma_1}(z)} \to 0 \quad \text{as } z \to (\mu_1 + i\sigma_1\sqrt{\nu_1}) \quad (2 \leq j \leq n).$$

Now, by the identity principle, the relation (2) extends to

$$\sum_{j=1}^n \lambda_j f_{\nu_j,\mu_j,\sigma_j}(z) = 0 \quad (z \in \mathbb{C},\, z \neq \mu_j \pm i\sigma_j\sqrt{\nu_j}).$$

Dividing both sides by $f_{\nu_1,\mu_1,\sigma_1}(z)$ and then letting $z \to \mu_1 + i\sigma_1\sqrt{\nu_1}$, we deduce that $\lambda_1 = 0$, contrary to hypothesis.

If some of the ν_j are even integers, then the argument needs to be adjusted slightly, since the presence of square roots means that the holomorphic extensions of f_{ν_j,μ_j,σ_j} are multivalued. The problem is fixed as follows. We construct a finite set of non-intersecting half-lines terminating at the points $\mu_j \pm i\sigma_j\sqrt{\nu_j}$ and not crossing the real axis. The complement of these half-lines is then a simply connected domain D containing $\mathbb{R}$ whose boundary contains each point $\mu_j \pm i\sigma_j\sqrt{\nu_j}$. Each function f_{ν_j,μ_j,σ_j} has a single-valued holomorphic extension to D, and we can now argue exactly as before. □

3.3. **Linear independence of measures via Borel's theorem.** Theorems 3.1 and 3.2 above are just two examples of a large class of theorems asserting the identifiability of various families of distributions. The proofs of many of these results are very similar, and indeed our proof of Theorem 3.2

relies on the same ideas. There is even an abstract formulation of the general method due to Teicher [18, Theorem 2].

In this subsection we present a further general result on identifiability/linear independence, based on a completely different principle, namely a classic result in complex analysis due to Borel [2]. We have not seen this theorem used elsewhere in the statistical literature, and we believe that the following application may be of interest.

Theorem 3.3. *Let $\mathcal{P}$ be a family of Borel probability measures on $\mathbb{R}^d$. Suppose that the characteristic function of every measure in $\mathcal{P}$ is the the restriction to $\mathbb{R}^d$ of a nowhere-vanishing holomorphic function on $\mathbb{C}^d$. Then $\mathcal{P}$ is linearly independent.*

Remarks. (i) This result immediately yields an alternative proof of Theorem 3.1, since the characteristic functions of the Gaussian distributions (1) have the form $\phi(\xi) = \exp(i\mu\xi - \sigma^2\xi^2/2)$.

(ii) There is a well-known criterion for the characteristic function of P to be the restriction to $\mathbb{R}^d$ of a holomorphic function on $\mathbb{C}^d$: this happens if and only if the moments $m_n := \int_{\mathbb{R}^d} \|x\|^n \, dP(x)$ are finite and satisfy $m_n^{1/n} = o(n)$ as $n \to \infty$ (see e.g. [14, Theorem 4.2.2]).

(iii) Suppose that characteristic function of P is the restriction to $\mathbb{R}^d$ of a holomorphic function f on $\mathbb{C}^d$. A sufficient condition for f to be zero-free on $\mathbb{C}^d$ is that P be infinitely divisible. This follows from [13, Theorem 3.1] and [15, Theorem 4].

(iv) The following simple example shows that the nowhere-vanishing condition in Theorem 3.3 cannot be omitted. Let P_0 and P_1 be the Dirac measures on $\mathbb{R}$ concentrated at 0 and 1 respectively, and let $P_2 := (1/3)P_0 + (2/3)P_1$. The characteristic functions are the P_j are given by $\phi_{P_0}(\xi) = 1$ and $\phi_{P_1}(\xi) = e^{i\xi}$ and $\phi_{P_2} = (1/3)(1 + 2e^{i\xi})$. All three are nowhere-zero on $\mathbb{R}$ and all three extend to be holomorphic on $\mathbb{C}$. However, the set $\{P_0, P_1, P_2\}$ is clearly linearly dependent. Theorem 3.3 does not apply in this case, because the holomorphic extension of ϕ_{P_2} has zeros in $\mathbb{C}$ (namely at $i \log 2 + 2\pi n$ for each integer n).

As mentioned above, the proof of Theorem 3.3 is based on a theorem of Borel. The precise version that we need is due to Green [9, p.98]:

Lemma 3.4. *Let $g_1, \dots, g_n : \mathbb{C}^d \to \mathbb{C}$ be holomorphic functions such that*

$$\exp(g_1) + \cdots + \exp(g_n) \equiv 0.$$

Then, for some distinct j, k, the function $g_j - g_k$ is constant.

Proof of Theorem 3.3. We argue by contradiction. Suppose that $\mathcal{P}$ is linearly dependent, so there exist distinct $P_1, \dots, P_n \in \mathcal{P}$ and non-zero scalars $\lambda_1, \dots, \lambda_n$ such that $\sum_{j=1}^n \lambda_j P_j = 0$. Then the characteristic functions ϕ_{P_j} of the P_j satisfy $\sum_{j=1}^n \lambda_j \phi_{P_j}(\xi) = 0$ for all $\xi \in \mathbb{R}^d$. By assumption, each ϕ_{P_j} is the restriction to $\mathbb{R}^d$ of a nowhere-vanishing holomorphic function on $\mathbb{C}^d$.

Thus we may write $\lambda_j \phi_{P_j} = \exp g_j|_{\mathbb{R}^d}$, where $g_j : \mathbb{C}^d \to \mathbb{C}$ is a holomorphic function on $\mathbb{C}^d$, and $\sum_{j=1}^n \exp g_j(\xi) = 0$ for all $\xi \in \mathbb{R}^d$. By the identity principle, a holomorphic function on $\mathbb{C}^d$ that vanishes on $\mathbb{R}^d$ is identically zero on $\mathbb{C}^d$. Therefore $\sum_{j=1}^n \exp g_j(\zeta) = 0$ for all $\zeta \in \mathbb{C}^d$. We now invoke Lemma 3.4, to deduce that $g_j - g_k$ is constant for some pair of distinct indices j, k. This implies that ϕ_{P_j}/ϕ_{P_k} is constant on $\mathbb{R}^d$. As P_j, P_k are probability measures, we have $\phi_{P_j}(0) = \phi_{P_k}(0) = 1$. Therefore $\phi_{P_j} = \phi_{P_k}$ on $\mathbb{R}^d$. By the uniqueness theorem for characteristic functions, it follows that $P_j = P_k$. This contradicts the fact that P_j and P_k are distinct measures. □

Remark. Theorem 3.3 actually implies a stronger form of itself, as follows. Given $A \in M_d(\mathbb{R})$ (the set of $d \times d$ matrices) and $b \in \mathbb{R}^d$, let us write $P_{A,b}$ for the Borel probability measure on $\mathbb{R}^d$ defined by

$$P_{A,b}(B) := P\big(\{x \in \mathbb{R}^d : Ax + b \in B\}\big). \tag{3}$$

Suppose that the characteristic function of each measure in $\mathcal{P}$ is the restriction to $\mathbb{R}^d$ of a nowhere-vanishing function on $\mathbb{C}^d$. Then, not only is $\mathcal{P}$ linearly independent, but even the distinct measures in the family $\{P_{A,b} : P \in \mathcal{P}, A \in M_d(\mathbb{R}), b \in \mathbb{R}^d\}$ are linearly independent. Indeed, a simple calculation shows that

$$\phi_{P_{A,b}}(\xi) = e^{ib\cdot\xi}\phi_P(A^T\xi) \quad (\xi \in \mathbb{R}^d),$$

so, if ϕ_P is the restriction to $\mathbb{R}^d$ of a nowhere-vanishing holomorphic function on $\mathbb{C}^d$, then the same is true of $\phi_{P_{A,b}}$. Theorem 3.3 now gives the result.

4. Gaussian mixtures

4.1. **Introduction.** A Gaussian measure P on $\mathbb{R}^d$ is one whose density has the form

$$\frac{1}{(2\pi\det(\Sigma))^{d/2}}\exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big) \quad (x \in \mathbb{R}^d), \tag{4}$$

where $\mu \in \mathbb{R}^d$, and where Σ is a real $d \times d$ positive-definite matrix. A *Gaussian mixture* is a measure on $\mathbb{R}^d$ that is a finite convex combination of Gaussian measures.

Mixtures of multivariate Gaussian distributions have several nice properties. In particular, in Titterington et al. [19], it is shown that Gaussian kernel density estimators can approximate any continuous density given enough kernels (universal consistency). It is well known that Gaussian mixtures are weak*-dense in the space of all Borel probability measures on $\mathbb{R}^d$. They also have numerous applications in statistics; for more on this, see Section 5 below.

4.2. **A Cramér–Wold theorem for Gaussian mixtures.** In this section we consider the problem of testing for equality for two Gaussian mixtures by looking at a finite number of projections. The basic theorem underlying this approach has two ingredients. One is the abstract result Theorem 2.1. The other is a characterization of Cramér–Wold systems for Gaussian measures in $\mathbb{R}^d$ (and, more generally, for elliptical distributions) established in [5, Theorems 1 and 2], which we now recall.

Let S be a set of vectors in $\mathbb{R}^d$. Then the corresponding set of lines $\{\langle x\rangle : x \in S\}$ is a Cramér–Wold system for the Gaussian measures in $\mathbb{R}^d$ if and only if S has the property that the only real symmetric $d \times d$ matrix A satisfying $x^T Ax = 0$ for all $x \in S$ is the zero matrix. A set S with this property is called a *symmetric-matrix uniqueness set* (or *sm-uniqueness set* for short).

It was shown in [5] that an sm-uniqueness set for $\mathbb{R}^d$ spans $\mathbb{R}^d$ and that it contains at least $(d^2 + d)/2$ vectors. We shall call S a *strong sm-uniqueness set* if every subset of S containing $(d^2 + d)/2$ vectors is an sm-uniqueness set. We can now state our Cramér–Wold theorem for Gaussian mixtures.

Theorem 4.1. *Let P and Q be convex combinations of ℓ and m Gaussian measures on $\mathbb{R}^d$ respectively. Let S be a strong sm-uniqueness set for $\mathbb{R}^d$ containing at least $(1/2)(\ell + m - 1)(d^2 + d - 2) + 1$ vectors. If $P_{\langle x\rangle} = Q_{\langle x\rangle}$ for all $x \in S$, then $P = Q$.*

Proof. We apply Theorem 2.1 with $\mathcal{P}$ equal to the set of Gaussian measures on $\mathbb{R}^d$ and $\mathcal{H} := \{\langle x\rangle : x \in S\}$. All we need to do is to check that $\mathcal{P}$ and $\mathcal{H}$ satisfy the hypotheses (i) and (ii) in Theorem 2.1.

Concerning hypothesis (i), the projection of a multivariate Gaussian measure onto a line is just a Gaussian measure on that line. We have already seen in Theorem 3.1 that the set of all one-dimensional Gaussian measures is a linearly independent family. So hypothesis (i) holds.

As for hypothesis (ii), we argue as follows. If S is partitioned into $\ell+m-1$ sets, then one of them, S_0 say, must contain at least $(d^2 + d)/2$ vectors (otherwise S would contain at most $(1/2)(\ell + m - 1)(d^2 + d - 2)$ vectors, contrary to assumption). As S is a strong sm-uniqueness set, it follows that S_0 is an sm-uniqueness set. In summary, if S is partitioned into $\ell + m - 1$ sets, then at least one of them is an sm-uniqueness set. In other words, if $\mathcal{H}$ is partitioned into $\ell+m-1$ sets, then at least one of them is a Cramér–Wold system for the family of Gaussian measures. Thus hypothesis (ii) holds, and we are done. □

4.3. **A Cramér–Wold theorem for t-mixtures.** We now establish an analogue of Theorem 4.1 for mixtures of multivariate t-distributions, thereby allowing heavy-tailed distributions. A t-distribution on $\mathbb{R}^d$ is a measure with density of the form

$$f_{\nu,\mu,\Sigma}(x) = c_{\nu,\mu,\Sigma}\Big(1 + \frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{\nu}\Big)^{-(\nu+d)/2} \quad (x \in \mathbb{R}^d),$$

where ν is a positive integer, μ is a vector in $\mathbb{R}^d$, and where Σ is a positive-definite $d \times d$ matrix. Once again, the constant $c_{\nu,\mu,\Sigma}$ is chosen to ensure that $\int_{\mathbb{R}^d} f_{\nu,\mu,\Sigma}(x)\,dx = 1$.

Theorem 4.2. *Let P and Q be convex combinations of ℓ and m multivariate t-distributions on $\mathbb{R}^d$ respectively. Let S be a strong sm-uniqueness set for $\mathbb{R}^d$ containing at least $(1/2)(\ell+m-1)(d^2+d-2)+1$ vectors. If $P_{\langle x\rangle} = Q_{\langle x\rangle}$ for all $x \in S$, then $P = Q$.*

Proof. This is virtually identical to the proof of Theorem 4.1. Once again, we apply Theorem 2.1, and we need to check that the hypotheses (i) and (ii) of that theorem hold.

Hypothesis (i) holds because the one-dimensional projection of a multivariate t-distribution is a univariate t-distribution, and by Theorem 3.2 the family of all univariate t-distributions is a linearly independent set.

Hypothesis (ii) holds for the same reason that it did before. Indeed, [5, Theorem 1] applies to all elliptical distributions, which includes t-distributions as well as Gaussian ones. □

4.4. **Strong sm-uniqueness sets.** Theorems 4.1 and 4.2 beg the question as to whether there exist strong sm-uniqueness sets of arbitrarily large cardinality. The following result provides an affirmative answer, and suggests a realistic method for generating them.

Theorem 4.3. *Let $d \ge 2$, let $k \ge (d^2+d)/2$, and let*

$$V := \Big\{(v_1,\dots,v_k) \in (\mathbb{R}^d)^k : \{v_1,\dots,v_k\} \text{ is a strong sm-uniqueness set}\Big\}.$$

Then V is an open subset of $\mathbb{R}^{dk}$, and $\mathbb{R}^{dk} \setminus V$ has Lebesgue measure zero.

Thus, if $v_1,\dots,v_k$ are independent random vectors in $\mathbb{R}^d$ with distributions given by densities on $\mathbb{R}^d$, then, with probability one, the set $\{v_1,\dots,v_k\}$ is a strong sm-uniqueness set for $\mathbb{R}^d$. To test whether a specific family $\{v_1,\dots,v_k\}$ is a strong sm-uniqueness set, one can use the following criterion for sm-uniqueness sets, which is also an ingredient in the proof of Theorem 4.3.

Let $d \ge 2$ and set $D := (d^2+d)/2$. Given $x = (t_1,\dots,t_d) \in \mathbb{R}^d$, let $\widehat{x}$ be the upper triangular $d \times d$ matrix with entries $\widehat{x}_{ij} := t_i t_j$ $(1 \le i \le j \le d)$, but viewed as a column vector in $\mathbb{R}^D$.

Lemma 4.4. *A D-tuple $(x_1,\dots,x_D)$ of vectors in $\mathbb{R}^d$ is an sm-uniqueness set if and only if $\widehat{x}_1,\dots,\widehat{x}_D$ are linearly independent vectors in $\mathbb{R}^D$, in other words, if and only if the determinant of the $D \times D$ block matrix $(\widehat{x}_1|\widehat{x}_2|\cdots|\widehat{x}_D)$ does not vanish.*

Proof. This is [5, Corollary 5]. □

Proof of Theorem 4.3. By Lemma 4.4, the set $\mathbb{R}^{dk} \setminus V$ can be expressed as the union

$$\bigcup_{1 \le j_1 < j_2 < \cdots < j_D \le k} \Big\{ ((v_1, \dots, v_k) \in (\mathbb{R}^d)^k : \det\big(\widehat{v}_{j_1} | \widehat{v}_{j_2} | \cdots | \widehat{v}_{j_D}\big) = 0 \Big\}.$$

For each choice of $(j_1, j_2, \dots, j_D)$, the map $(v_1, \dots, v_k) \mapsto \det(\widehat{v}_{j_1} | \widehat{v}_{j_2} | \cdots | \widehat{v}_{j_D})$ is a polynomial in the entries of $(v_1, \dots, v_k)$ that is not identically zero, so its zero set is a closed subset of $\mathbb{R}^{dk}$ of Lebesgue measure zero. As $\mathbb{R}^{dk} \setminus V$ is a finite union of such sets, it too is a closed subset of $\mathbb{R}^{dk}$ of Lebesgue measure zero. □

5. Concluding remarks

Gaussian-mixture models have been shown to be very effective in modeling different real data, see for instance Titterington et al. [19] for a deep study of their properties. They provide flexible and general models, and relevant applications can be found in the literature in different fields like density estimation, machine learning and clustering, among others. The estimation of these models is quite involved, in particular for high-dimensional data, typically using Markov-chain Monte-Carlo methods in a Bayesian framework. We believe that the Cramér–Wold device may well have a role to play in this circle of ideas, in particular through Theorems 3.1 and 3.2. We intend to explore this in a future paper.

References

[1] C. Bélisle, J.-C. Massé, and T. Ransford. When is a probability measure determined by infinitely many projections? *Ann. Probab.*, 25(2):767–786, 1997.

[2] E. Borel. Sur les zéros des fonctions entières. *Acta Math.*, 20:357–396, 1897.

[3] H. Cramér and H. Wold. Some theorems on distribution functions. *J. London Math. Soc.*, 11(4):290–294, 1936.

[4] J. A. Cuesta-Albertos, R. Fraiman, and T. Ransford. A sharp form of the Cramér-Wold theorem. *J. Theoret. Probab.*, 20(2):201–209, 2007.

[5] R. Fraiman, L. Moreno, and T. Ransford. A Cramér-Wold theorem for elliptical distributions. *J. Multivariate Anal.*, 196:Paper No. 105176, 13, 2023.

[6] R. Fraiman, L. Moreno, and T. Ransford. A quantitative Heppes theorem and multivariate Bernoulli distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 85(2):293–314, 2023.

[7] R. Fraiman, L. Moreno, and T. Ransford. Application of the Cramér-Wold theorem to testing for invariance under group actions. *TEST*, 33(2):379–399, 2024.

[8] W. M. Gilbert. Projections of probability distributions. *Acta Math. Acad. Sci. Hungar.*, 6:195–198, 1955.

[9] M. L. Green. Holomorphic maps into complex projective space omitting hyperplanes. *Trans. Amer. Math. Soc.*, 169:89–103, 1972.

[10] K. Gröchenig and P. Jaming. The Cramér-Wold theorem on quadratic surfaces and Heisenberg uniqueness pairs. *J. Inst. Math. Jussieu*, 19(1):117–135, 2020.

[11] A. Heppes. On the determination of probability distributions of more dimensions by their projections. *Acta Math. Acad. Sci. Hungar.*, 7:403–410, 1956.

[12] N. Ho and X.-L. Nguyen. On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electron. J. Stat.*, 10(1):271–307, 2016.

[13] R. A. Horn and F. W. Steutel. On multivariate infinitely divisible distributions. *Stochastic Process. Appl.*, 6(2):139–151, 1977/78.
[14] R. G. Laha and V. K. Rohatgi. *Probability theory.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-Chichester-Brisbane, 1979.
[15] E. Lukacs and O. Szász. On analytic characteristic functions. *Pacific J. Math.*, 2:615–625, 1952.
[16] C. E. G. Otiniano, P. N. Rathie, and L. C. S. M. Ozelim. On the identifiability of finite mixture of skew-normal and skew-t distributions. *Statist. Probab. Lett.*, 106:103–108, 2015.
[17] A. Rényi. On projections of probability distributions. *Acta Math. Acad. Sci. Hungar.*, 3:131–142, 1952.
[18] H. Teicher. Identifiability of finite mixtures. *Ann. Math. Statist.*, 34:1265–1269, 1963.
[19] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1985.

Centro de Matemática, Facultad de Ciencias, Universidad de la República, Uruguay.
Email address: `rfraiman@cmat.edu.uy`

Instituto de Estadística, Departamento de Métodos Cuantitativos, FCEA, Universidad de la República, Uruguay.
Email address: `mrleo@iesta.edu.uy`

Département de mathématiques et de statistique, Université Laval, Québec City (Québec), Canada G1V 0A6.
Email address: `ransford@mat.ulaval.ca`