

A SAMPLE DOCUMENT

BY ROBERT VANDERMEULEN[†] AND CLAYTON SCOTT[§]

*University of Michigan: Electrical and Computer Engineering^{†§} and
Statistics[§]*

The abstract should summarize the contents of the paper. It should be clear, descriptive, self-explanatory and not longer than 200 words. It should also be suitable for publication in abstracting services. Please avoid using math formulas as much as possible.

This is a sample input file. Comparing it with the output it generates can show you how to produce a simple document of your own.

1. Introduction. A finite mixture model \mathcal{P} is a probability measure over a space of probability measures where $\mathcal{P}(\{\mu_i\}) = w_i > 0$ for some finite collection of probability measures μ_1, \dots, μ_m with $\sum_{i=1}^m w_i = 1$. A realization from this mixture model first randomly selects some mixture component $\mu \sim \mathcal{P}$ and then draws from μ . Mixture models have seen extensive use in statistics and machine learning.

A central theoretical question concerning mixture models is that of identifiability. A mixture model is said to be *identifiable* if there is no other mixture model which defines the same distribution over the observed data. Classically mixture models were concerned with the case where the observed data X_1, X_2, \dots are iid with X_i distributed according to some unobserved random measure μ_i with $\mu_i \stackrel{iid}{\sim} \mathcal{P}$. This situation is equivalent to $X_i \stackrel{iid}{\sim} \sum_{j=1}^m w_j \mu_j$. If we impose no restrictions on the mixture components μ_1, \dots, μ_m one could easily concoct many choices of μ_j and w_j which yield an identical distribution on X_i . As a consequence most previous work on identifiability assumes some sort of structure on μ_1, \dots, μ_m , such as Gaussianity [3, 7]. In this work we consider an alternative scenario where we make no assumptions on μ_1, \dots, μ_m and instead have access to groups of samples which are known to come from the same component. We will call these groups of samples “random groups.” Mathematically a random group is a random element \mathbf{X}_i where $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,n})$ with $X_{i,1}, \dots, X_{i,n} \stackrel{iid}{\sim} \mu_i$ and $\mu_i \stackrel{iid}{\sim} \mathcal{P}$.

In this setting identifiability is now concerned with the distribution over \mathbf{X}_i and the value of n , the number of samples in each random group. We call a mixture of measures \mathcal{P} n -identifiable if it is the simplest mixture

MSC 2010 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: sample, L^AT_EX 2_ε

model (in terms of number of mixture components) which yields the observed distribution on \mathbf{X}_i . We also introduce a new concept which is stronger than identifiability. We call \mathcal{P} n -determined if it is the *only* mixture model which yields the observed distribution on \mathbf{X}_i . In this paper we show that a mixture model with m components is always $(2m - 1)$ -identifiable and $2m$ -determined. Furthermore we show that any mixture model with linearly independent components is 3-identifiable and a mixture model with jointly irreducible components is 2-identifiable. These results hold for any mixture model over any space and cannot be improved.

2. Previous Work. In classical mixture model theory identifiability is achieved by making assumptions about the mixture components. Some assumptions which yield identifiability are Gaussian or binomial mixture components [7, 18]. If one makes no assumptions on the mixture components then one must leverage some other type of structure in order to achieve identifiability. An example of such structure exists in the context of multiview models. In a multiview model samples have the form $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,n})$ and the distribution of \mathbf{X}_i is defined by $\sum_{i=1}^m w_i \prod_{j=1}^n \mu_i^j$. In [1] it was shown that if μ_i^j are probability distributions on \mathbb{R} with μ_1^j, \dots, μ_m^j linearly independent for all j and $n \geq 3$, then the model is identifiable.

The setting which we investigate is a special case of the multiview model where $\mu_i^j = \mu_i^{j'}$ for all i, j, j' . If the sample space of the μ_i is finite then this problem is exactly the topic modelling problem with a finite number of topics and one topic for each document. In topic modelling each μ_i is a “topic”, and the sample space is a finite collection of words. This setting is well studied and it has been shown that one can recover the true topics provided certain assumptions on the topics [1, 2, 4]. This problem was studied for arbitrary topics in [17]. In this paper the authors introduce an algorithm that recovers any mixture of m topics provided $2m - 1$ words per document. They also show, in a result analogous to our own, that this $2m - 1$ value cannot be improved. Our proof techniques are quite different than those used in [17], hold for arbitrary sample spaces, and are less complex.

3. Problem Setup. We will be treating this problem in as general of a setting as possible. For any measurable space we define δ_x as the Dirac measure at x . For \odot a set, σ -algebra, or measure, we denote $\odot^{\times a}$ to be the standard a -fold product associated with that object. For any natural number k we define $[k] \triangleq \mathbb{N} \cap [1, k]$. Let Ω be a set containing more than one element. This set is the sample space of our data. Let \mathcal{F} be a σ -algebra over Ω . Assume $\mathcal{F} \neq \{\emptyset, \Omega\}$. We denote the space of probability measures

over this space as $\mathcal{D}(\Omega, \mathcal{F})$, which we will shorten to \mathcal{D} . We will equip \mathcal{D} with the σ -algebra $2^{\mathcal{D}}$ so that each Dirac measure over \mathcal{D} is unique. Define $\Delta(\mathcal{D}) \triangleq \text{span}(\delta_x : x \in \mathcal{D})$. This is the ambient space where our mixtures of probability measures live. Let $\mathcal{P} = \sum_{i=1}^m w_i \delta_{\mu_i}$ be a probability measure in $\Delta(\mathcal{D})$. Let $\mu \sim \mathcal{P}$ and $X_1, \dots, X_n \stackrel{iid}{\sim} \mu$. Here \mathbf{X} is a random group sample, which was described in the introduction. We will denote $\mathbf{X} = (X_1, \dots, X_n)$.

We will now derive the probability law of \mathbf{X} . Let $A \in \mathcal{F}^{\times n}$. We have

$$\begin{aligned} \mathbb{P}(\mathbf{X} \in A) &= \sum_{i=1}^m \mathbb{P}(\mathbf{X} \in A | \mu = \mu_i) \mathbb{P}(\mu = \mu_i) \\ &= \sum_{i=1}^m w_i \mu_i^{\times n}(A). \end{aligned}$$

The second equality follows from Lemma 3.10 in [12]. So the probability law of \mathbf{X} is

$$(1) \quad \sum_{i=1}^m w_i \mu_i^{\times n}.$$

We want to view the probability law of \mathbf{X} as a function of \mathcal{P} in a mathematically rigorous way, which requires a bit of technical buildup. Let \mathcal{V} be a vector space. We will now construct a version of the integral for \mathcal{V} -valued functions over \mathcal{D} . Let $\mathcal{Q} \in \Delta(\mathcal{D})$. From the definition of $\Delta(\mathcal{D})$ it follows that \mathcal{Q} admits the representation

$$\mathcal{Q} = \sum_{i=1}^r \alpha_i \delta_{\nu_i}.$$

From the well-ordering principle there must exist some representation with minimal r and we define this r as the *order* of \mathcal{Q} . We can show that the representation of any $\mathcal{Q} \in \Delta(\mathcal{D})$ is unique up to permutation of its indices.

LEMMA 3.1. *Let $\mathcal{Q} \in \Delta(\mathcal{D})$ and admit minimal representations $\mathcal{Q} = \sum_{i=1}^r \alpha_i \delta_{\nu_i} = \sum_{i=1}^r \alpha'_i \delta_{\nu'_i}$. There exists some permutation $\psi : [r] \rightarrow [r]$ such that $\nu_{\psi(i)} = \nu'_i$ and $\alpha_{\psi(i)} = \alpha'_i$ for all i .*

Henceforth when we define an element of $\Delta(\mathcal{D})$ with a summation we will assume that the summation is a minimal representation.

DEFINITION 3.1. *We call \mathcal{P} a mixture of measures if it is a probability measure in $\Delta(\mathcal{D})$. We will say that \mathcal{P} has m mixture components if it has order m .*

Any minimal representation of a mixture of measures \mathcal{P} with m components satisfies $\mathcal{P} = \sum_{i=1}^m w_i \delta_{\mu_i}$ with $w_i > 0$ for all i and $\sum_{i=1}^m w_i = 1$. So any mixture of measures is a convex combination of Dirac measures at elements in \mathcal{D} .

For a function $f : \mathcal{D} \rightarrow \mathcal{V}$ define

$$\int f(\nu) d\mathcal{Q}(\nu) = \sum_{i=1}^r \alpha_i f(\nu_i),$$

where $\sum_{i=1}^r \alpha_i \delta_{\nu_i}$ is a minimal representation of \mathcal{Q} . This integral is well defined as a consequence of Lemma 3.1.

For a measurable space (Ψ, \mathcal{G}) we define $\mathcal{M}(\Psi, \mathcal{G})$ as the space of all finite signed measures over (Ψ, \mathcal{G}) . Let $\lambda_n : \mathcal{M}(\Omega, \mathcal{F}) \rightarrow \mathcal{M}(\Omega^{\times n}, \mathcal{F}^{\times n})$; $\nu \mapsto \nu^{\times n}$. We introduce the operator $V_n : \Delta(\mathcal{D}) \rightarrow \mathcal{M}(\Omega^{\times n}, \mathcal{F}^{\times n})$

$$V_n(\mathcal{Q}) = \int \lambda_n(\nu) d\mathcal{Q}(\nu) = \int \nu^{\times n} d\mathcal{Q}(\nu).$$

For a minimal representation $\mathcal{Q} = \sum_{i=1}^r \alpha_i \delta_{\nu_i}$, we have

$$V_n(\mathcal{Q}) = \sum_{i=1}^r \alpha_i \nu_i^{\times n}.$$

From this definition we have that $V_n(\mathcal{P})$ is simply the law of \mathbf{X} which we derived earlier. For sake of explicitness we mention that a pair of mixtures of measures are equal iff they are the same measure.

DEFINITION 3.2. *We call a mixture of measures, \mathcal{P} , n -identifiable if there does not exist a different mixture of measures \mathcal{P}' , with order no greater than the order of \mathcal{P} , such that $V_n(\mathcal{P}) = V_n(\mathcal{P}')$.*

DEFINITION 3.3. *We call a mixture of measures, \mathcal{P} , n -determined if there exists no other mixture of measures \mathcal{P}' such that $V_n(\mathcal{P}) = V_n(\mathcal{P}')$.*

Definition 3.2 and 3.3 are the central objects of interest in this paper. Given a mixture of measures, $\mathcal{P} = \sum_{i=1}^m w_i \delta_{\mu_i}$ then $V_n(\mathcal{P})$ is equal to $\sum_{i=1}^m w_i \mu_i^{\times n}$, the measure from which \mathbf{X} is drawn. If \mathcal{P} is not n -identifiable then we know that there exists a different mixture of measures which is no more complex (in terms of number of mixture components) than \mathcal{P} which is not discernible from \mathcal{P} given the data. Practically speaking this means we need more samples in each random group \mathbf{X} in order for the full richness of \mathcal{P} to be manifested in \mathbf{X} . n -determination is analogous to n -identifiability,

but with the requirement that our mixture of measures be the *only* mixture of measures (of any order) which admits the distribution over the data, not just the simplest.

4. Results. The first result is a bound on the n -identifiability of all mixtures of measures with m or fewer components. This bound cannot be uniformly improved.

THEOREM 4.1. *Let (Ω, \mathcal{F}) be a measurable space. Mixtures of measures with m components are $(2m - 1)$ -identifiable.*

THEOREM 4.2. *Let (Ω, \mathcal{F}) be a measurable space with $\mathcal{F} \neq \{\emptyset, \Omega\}$. For all m , there exists a mixture of measures with m components which is not $(2m - 2)$ -identifiable.*

The following lemmas convey the unsurprising fact that n -identifiability is, in some sense, monotonic.

LEMMA 4.1. *If a mixture of measures is n -identifiable then it is q -identifiable for all $q > n$.*

LEMMA 4.2. *If a mixture of measures is not n -identifiable then it is not q -identifiable for any $q < n$.*

Viewed alternatively these results say that $n = 2m - 1$ is the smallest value for which V_n is injective over the set of mixtures of measures with m or fewer components.

We also present an analogous bound for n -determined-ness. This bound also cannot be improved.

THEOREM 4.3. *Let (Ω, \mathcal{F}) be a measurable space. Mixtures of measures with m components are $2m$ -determined.*

THEOREM 4.4. *Let (Ω, \mathcal{F}) be a measurable space with $\mathcal{F} \neq \{\emptyset, \Omega\}$. For all m , there exists a mixture of measures with m components which is not $(2m - 1)$ -determined.*

Again n -determined-ness is monotonic in the number of samples per group.

LEMMA 4.3. *If a mixture of measures is n -determined then it is q -determined for all $q > n$.*

LEMMA 4.4. *If a mixture of measures is not n -determined then it is not q -determined for any $q < n$.*

This collection of results can be interpreted in an alternative way. Consider some pair of mixtures of measures $\mathcal{P}, \mathcal{P}'$. If $n \geq 2m$ and either mixture of measures is of order m or less, then $V_n(\mathcal{P}) = V_n(\mathcal{P}')$ implies $\mathcal{P} = \mathcal{P}'$. Furthermore $n = 2m$ is the smallest value of n for which the previous statement is true.

Finally we also include two results which are analogous to previously shown results for the discrete setting. We note that our proof techniques are markedly different than the previous proofs for the discrete case.

THEOREM 4.5. *If $\mathcal{P} = \sum_{i=1}^m w_i \mu_i$ is a mixture of measures where μ_1, \dots, μ_m are linearly independent then \mathcal{P} is 3-determined.*

This bound is tight as a consequence of Theorem 4.2 with $m = 2$.

The discrete version of this theorem was first proven in [1] by making use of Kruskal's Theorem [13]. Kruskal's Theorem demonstrates that order 3 tensors over \mathbb{R}^d admit unique decompositions (up to scaling and permutation) given certain linear independence assumptions. Our proof makes no use of Kruskal's Theorem and demonstrates that n -identifiability for linearly independent mixture components need not be attached to the discrete version in any way. Furthermore Kruskal's theorem only implies the identifiability of the mixture components, not determined-ness. Though it is not an identifiability result, [2] contains an efficient algorithm for recovering linearly independent mixture components for discrete sample spaces so long as there are 3 samples in each random group.

Our final result is related to the “separability condition” found in [9]. The separability condition in the discrete case requires that, for each mixture component μ_i , there exists $B_i \in \mathcal{F}$ such that $\mu_i(B_i) > 0$ and $\mu_j(B_i) = 0$ for all $i \neq j$. There exists a generalization of the separability condition, known as *joint irreducibility*.

DEFINITION 4.1. *A collection of probability measures μ_1, \dots, μ_m are said to be jointly irreducible if $\sum_{i=1}^m w_i \mu_i$ being a probability measure implies $w_i \geq 0$.*

In other words, any probability measure in the span of μ_1, \dots, μ_m must be a convex combination of those measures. It was shown in [6] that separability implies joint irreducibility, but not visa-versa. In that paper it was also shown that joint irreducibility implies linear independence, but the converse does not hold.

THEOREM 4.6. *If $\mathcal{P} = \sum_{i=1}^m w_i \mu_i$ is a mixture of measures where μ_1, \dots, μ_m are jointly irreducible then \mathcal{P} is 2-determined.*

A straightforward consequence of [4] is that the previous theorem holds for a discrete setting. In that setting joint irreducibility implies the separability condition. Because joint separability is more general one could not arrive at the previous theorem by applying a simple binning argument to the discrete version of our result. Again the determined-ness of the result is, as far as we know, totally new.

5. Tensor Products of Hilbert Spaces. Our proofs will rely heavily on the geometry of tensor products of Hilbert spaces which we will introduce in this section.

5.1. Overview of Tensor Products. First we introduce tensor products of Hilbert spaces. To our knowledge there does not exist a rigorous construction of the tensor product Hilbert space which is both succinct and intuitive. Because of this we will simply state some basic facts about tensor products of Hilbert spaces and hopefully instill some intuition for the uninitiated by way of example. A thorough treatment of tensor products of Hilbert spaces can be found in [11].

Let H and H' be Hilbert spaces. From these two Hilbert spaces the “simple tensors” are elements of the form $h \otimes h'$ with $h \in H$ and $h' \in H'$. We can treat the simple tensors as being the basis for some inner product space H_0 , with the inner product of simple tensors satisfying

$$\langle h_1 \otimes h'_1, h_2 \otimes h'_2 \rangle = \langle h_1, h_2 \rangle \langle h'_1, h'_2 \rangle.$$

The tensor product of H and H' is the completion of H_0 and is denoted $H \otimes H'$. To avoid potential confusion we note that the notation just described is standard in operator theory literature. In some literature our definition of H_0 is denoted as $H \otimes H'$ and our definition of $H \otimes H'$ is denoted $H \widehat{\otimes} H'$.

As an illustrative example we consider the tensor product $L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$. It can be shown that there exists an isomorphism between $L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$ and $L^2(\mathbb{R}^2)$ which maps the simple tensors to separable functions [11], $f \otimes f' \mapsto f(\cdot)f'(\cdot)$. We can demonstrate this isomorphism with a simple example. Let $f, g, f', g' \in L^2(\mathbb{R})$. Taking the $L^2(\mathbb{R}^2)$ inner product of $f(\cdot)f'(\cdot)$ and $g(\cdot)g'(\cdot)$ gives us

$$\begin{aligned}
\int \int (f(x)f'(y)) (g(x)g'(y)) dx dy &= \int f(x)g(x) dx \int f'(y)g'(y) dy \\
&= \langle f, g \rangle \langle f', g' \rangle \\
&= \langle f \otimes f', g \otimes g' \rangle.
\end{aligned}$$

Beyond tensor product we will need to define tensor power. To begin we will first show that tensor products are, in some sense, associative. Let H_1, H_2, H_3 be Hilbert spaces. Proposition 2.6.5 in [11] states that there is a unique unitary operator, $U : (H_1 \otimes H_2) \otimes H_3 \rightarrow H_1 \otimes (H_2 \otimes H_3)$, which satisfies the following for all $h_1 \in H_1, h_2 \in H_2, h_3 \in H_3$,

$$U((h_1 \otimes h_2) \otimes h_3) = h_1 \otimes (h_2 \otimes h_3).$$

This implies that for any collection of Hilbert spaces, H_1, \dots, H_n , the Hilbert space $H_1 \otimes \dots \otimes H_n$ is defined unambiguously regardless of how we decide to associate the products. In the space $H_1 \otimes \dots \otimes H_n$ we define a simple tensor as a vector of the form $h_1 \otimes \dots \otimes h_n$ with $h_i \in H_i$. In [11] it is shown that $H_1 \otimes \dots \otimes H_n$ is the closure of the span of these simple tensors. To conclude this primer on tensor products we introduce the following notation. For a Hilbert space H we denote $H^{\otimes n} = \underbrace{H \otimes H \otimes \dots \otimes H}_{n \text{ times}}$ and for $h \in H$,

$$h^{\otimes n} = \underbrace{h \otimes h \otimes \dots \otimes h}_{n \text{ times}}.$$

5.2. Tensor Rank. A tool we will use frequently in our proofs is *tensor rank*, which behave similarly to matrix rank.

DEFINITION 5.1. Let $h \in H^{\otimes n}$ where H is a Hilbert space. The rank of h is the smallest natural number r such that $h = \sum_{i=1}^r h_i$ where h_i are simple tensors.

5.3. Some Results for Tensor Product Spaces. We derive some technical results concerning tensor product spaces which will be useful for the rest of the paper. These lemmas are similar to or are straightforward extensions of previous results which we needed to modify for our particular purposes. Let $(\Psi, \mathcal{G}, \gamma)$ be a σ -finite measure space. We have the following lemma which connects tensor power of a L^2 space to the L^2 space of the product measure. The proofs of this lemma and the next one are straightforward but technical, and can be found in the appendix.

LEMMA 5.1. *There exists a unitary transform $U : L^2(\Psi, \mathcal{G}, \gamma)^{\otimes n} \rightarrow L^2(\Psi^{\times n}, \mathcal{G}^{\times n}, \gamma^{\times n})$ such that, for all $f_1, \dots, f_n \in L^2(\Psi, \mathcal{G}, \gamma)$, $U(f_1 \otimes \dots \otimes f_n) = f_1(\cdot) \dots f_n(\cdot)$.*

The following lemma is used in the proof of Lemma 5.1 as well as the proof of Theorem 4.2.

LEMMA 5.2. *Let $H_1, \dots, H_n, H'_1, \dots, H'_n$ be a collection of Hilbert spaces and U_1, \dots, U_n a collection of unitary operators with $U_i : H_i \rightarrow H'_i$ for all i . There exists a unitary operator $U : H_1 \otimes \dots \otimes H_n \rightarrow H'_1 \otimes \dots \otimes H'_n$ satisfying $U(h_1 \otimes \dots \otimes h_n) = U_1(h_1) \otimes \dots \otimes U_n(h_n)$ for all $h_1 \in H_1, \dots, h_n \in H_n$.*

A statement of the following lemma for \mathbb{R}^d can be found in [8]. We present our own proof for the Hilbert space setting.

LEMMA 5.3. *Let $n > 1$ and let h_1, \dots, h_n be elements of a Hilbert space such that no elements are zero and no pairs of elements are collinear. Then $h_1^{\otimes n-1}, \dots, h_n^{\otimes n-1}$ are linearly independent.*

The following lemma is a Hilbert space version of a well known property for positive semi-definite matrices.

LEMMA 5.4. *Let h_1, \dots, h_m be elements of a Hilbert space. The rank of $\sum_{i=1}^m h_i^{\otimes 2}$ is the dimension of $\text{span}(h_1, \dots, h_m)$.*

6. Proofs of Theorems. With the tools developed in the previous sections we can now prove our theorems. First we introduce one additional piece of notation. For a function f on a domain \mathcal{X} we define $f^{\times k}$ as simply the product of the function k times on the domain $\mathcal{X}^{\times k}$, $\underbrace{f(\cdot) \dots f(\cdot)}_{k \text{ times}}$. For a measure the notation continues to denote the standard product measure.

In these proofs we will be making extensive use of various L^2 spaces. These spaces will be equivalence classes of functions which are almost everywhere equal with respect to the measure associated with that space. When considering elements of these spaces, equality will always mean almost everywhere equality with respect to the measure associated with that space. When performing integrals or other manipulations of elements in L^2 spaces, we will be performing operations which do not depend on the representative of that equivalence class.

PROOF OF THEOREM 4.1. We will proceed by contradiction. Suppose there exist two different mixtures of measures $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i} \neq \mathcal{P}' =$

$\sum_{j=1}^l b_j \delta_{\nu_j}$, such that

$$(2) \quad \sum_{i=1}^m a_i \mu_i^{\times 2m-1} = \sum_{j=1}^l b_j \nu_j^{\times 2m-1}$$

and $l \leq m$. From our assumption on representation we know $\mu_i \neq \mu_j$ and $\nu_i \neq \nu_j$ for all $i \neq j$. We will also assume that $\mu_i \neq \nu_j$ for all i, j . Were this not true we could simply subtract the smaller of the common terms from both sides of (2) and normalize to yield another pair of distinct mixtures of measures with fewer components and no shared terms, \mathcal{Q} and \mathcal{Q}' . Let \mathcal{Q} have m' components and \mathcal{Q}' have l' with $m' \geq l'$. If $m \neq m'$ then we can apply Lemma 4.2 to give us $V_{2m'-1}(\mathcal{Q}) = V_{2m'-1}(\mathcal{Q}')$ and proceed as usual. We will use the following lemma to embed the mixture components in a Hilbert space.

LEMMA 6.1. *Let $\gamma_1, \dots, \gamma_n$ be finite measures on a σ -algebra (Ψ, \mathcal{G}) . There exists a finite measure π and non-negative functions $f_1, \dots, f_n \in L^1(\Psi, \mathcal{G}, \pi) \cap L^2(\Psi, \mathcal{G}, \pi)$ such that, for all i and all $B \in \mathcal{G}$*

$$\gamma_i(B) = \int_B f_i d\pi.$$

From Lemma 6.1 there exists a finite measures ξ and non-negative functions $p_1, \dots, p_m, q_1, \dots, q_l \in L^1(\Omega, \mathcal{F}, \xi) \cap L^2(\Omega, \mathcal{F}, \xi)$ such that, for all $B \in \mathcal{F}$, $\mu_i(B) = \int_B p_i d\xi$ and $\nu_j(B) = \int_B q_j d\xi$ for all i, j . Clearly no two of these derivatives are equal. If one of the derivatives were a scalar multiple of another, for example $p_1 = \alpha p_2$ for some $\alpha \neq 1$, it would imply

$$\mu_1(\Omega) = \int p_1 d\xi = \int \alpha p_2 d\xi = \alpha.$$

This is not true so no pair of these derivatives are collinear.

We can use the following lemma to extend this new representation to a product measure.

LEMMA 6.2. *Let (Ψ, \mathcal{G}) be a measurable space, γ and π a pair of finite measures on that space, and f a nonnegative function in $L^1(\Psi, \mathcal{G}, \pi)$ such that, for all $A \in \mathcal{G}$, $\gamma(A) = \int_A f d\pi$. Then for all n , for all $B \in \mathcal{G}^{\times n}$ we have*

$$\gamma^{\times n}(B) = \int_B f^{\times n} d\pi^{\times n}.$$

Thus for any $R \in \mathcal{F}^{\times 2m-1}$ we have

$$\begin{aligned} \int_R \sum_{i=1}^m a_i p_i^{\times 2m-1} d\xi^{\times 2m-1} &= \sum_{i=1}^m a_i \mu_i^{\times 2m-1}(R) \\ &= \sum_{j=1}^l b_j \nu_j^{\times 2m-1}(R) \\ &= \int_R \sum_{j=1}^l b_j q_j^{\times 2m-1} d\xi^{\times 2m-1}. \end{aligned}$$

The following lemma is simply Proposition 2.23 in [10], but it is worth mentioning explicitly.

LEMMA 6.3. *Let $(\Psi, \mathcal{G}, \gamma)$ be a measure space and $f, g \in L^1(\Psi, \mathcal{G}, \gamma)$. Then $f = g$ γ -almost everywhere iff, for all $A \in \mathcal{G}$, $\int_A f d\gamma = \int_A g d\gamma$.*

From this lemma it follows that

$$\sum_{i=1}^m a_i p_i^{\times 2m-1} = \sum_{j=1}^l b_j q_j^{\times 2m-1}.$$

Applying the U^{-1} operator from Lemma 5.1 to the previous equation yields

$$\sum_{i=1}^m a_i p_i^{\otimes 2m-1} = \sum_{j=1}^l b_j q_j^{\otimes 2m-1}.$$

Since $l+m \leq 2m$ Lemma 5.3 states that $p_1^{\otimes 2m-1}, \dots, p_m^{\otimes 2m-1}, q_1^{\otimes 2m-1}, \dots, q_l^{\otimes 2m-1}$ are all linearly independent and thus $a_i = 0$ and $b_j = 0$ for all i, j , a contradiction. \square

PROOF OF THEOREM 4.2. To prove this theorem we will construct a pair of mixture of measures, $\mathcal{P} \neq \mathcal{P}'$ which both contain m components and satisfy $V_{2m-2}(\mathcal{P}) = V_{2m-2}(\mathcal{P}')$. From our definition of (Ω, \mathcal{F}) we know there exists $F \in \mathcal{F}$ such that F, F^C are nonempty. Let $x \in F$ and $x' \in F^C$. It follows that δ_x and $\delta_{x'}$ are different probability measures on (Ω, \mathcal{F}) . The theorem follows from the next lemma.

LEMMA 6.4. *Let (Ψ, \mathcal{G}) be a measurable space and γ, γ' be distinct probability measures on that space. Let $\varepsilon_1, \dots, \varepsilon_t$ be t distinct values in $[0, 1]$, then*

$$\sum_{i=1}^l \beta_i (\varepsilon_i \gamma + (1 - \varepsilon_i) \gamma')^{\times t-2} = \sum_{j=l+1}^t \beta_j (\varepsilon_j \gamma + (1 - \varepsilon_j) \gamma')^{\times t-2}$$

where $\beta_i > 0$ for all i , $\sum_{i=1}^l \beta_i = \sum_{j=l+1}^t \beta_j = 1$, and $l, t-l \geq \lfloor \frac{t}{2} \rfloor$.

Applying Lemma 6.4 to the measures δ_x and $\delta_{x'}$, with $t = 2m$ we get and letting $\mu_i = \varepsilon_i \delta_x + (1 - \varepsilon_i) \delta_{x'}$, we get

$$\sum_{i=1}^m \beta_i \mu_i^{\times 2m-2} = \sum_{j=m+1}^{2m} \beta_j \mu_j^{\times 2m-2}.$$

If we let $\mathcal{P} = \sum_{i=1}^m \beta_i \mu_i$ and $\mathcal{P}' = \sum_{j=m+1}^{2m} \beta_j \mu_j$, we have that $V_{2m-2}(\mathcal{P}) = V_{2m-2}(\mathcal{P}')$. \square

We now prove Lemma 6.4.

PROOF OF LEMMA 6.4. From Lemma 6.1, there exists a finite measure π and non-negative functions $f, f' \in L^1(\Psi, \mathcal{G}, \pi) \cap L^2(\Psi, \mathcal{G}, \pi)$ such that, for all A , $\gamma(A) = \int_A f d\pi$ and $\gamma'(A) = \int_A f' d\pi$.

Let H_2 be the Hilbert space generated from the span of f, f' . Let $(f_i)_{i=1}^t$ be non-negative functions in $L^1(\Psi, \mathcal{G}, \pi) \cap L^2(\Psi, \mathcal{G}, \pi)$ with $f_i = \varepsilon_i f + (1 - \varepsilon_i) f'$. Clearly f_i is a pdf over π for all i and there are no pairs in this collection which are collinear. Since H_2 is isomorphic to \mathbb{R}^2 there exists a unitary operator $U : H_2 \rightarrow \mathbb{R}^2$. From Lemma 5.2 there exists a unitary operator $U_{t-2} : H_2^{\otimes t-2} \rightarrow \mathbb{R}^{2 \otimes t-2}$ with $U_{t-2}(h_1 \otimes \cdots \otimes h_{t-2}) = U(h_1) \otimes \cdots \otimes U(h_{t-2})$. Because U is unitary the set $U_{t-2}(\text{span}(\{h^{\otimes t-2} : h \in H_2\}))$ maps exactly to the set $\text{span}(x^{\otimes t-2} : x \in \mathbb{R}^2)$. An order r tensor, A_{i_1, \dots, i_r} , is *symmetric* if $A_{i_1, \dots, i_r} = A_{\psi(i_1), \dots, \psi(i_r)}$ for any i_1, \dots, i_r and permutation ψ . A consequence of Lemma 4.2 in [8] is that $\text{span}(\{x^{\otimes t-2} : x \in \mathbb{R}^2\}) \subset S^{t-2}(\mathbb{C}^2)$, the space of all symmetric order $t-2$ tensors over \mathbb{C}^2 .

From Proposition 3.4 in [8] it follows that the dimension of $S^{t-2}(\mathbb{C}^2)$ is $\binom{2+t-2-1}{t-2} = t-1$. From this we get that $\dim(\text{span}(\{h^{\otimes t-2} : h \in H_2\})) \leq t-1$.

The bound on the dimension of $\text{span}(\{h^{\otimes t-2} : h \in H_2\})$ implies that $(f_i^{\otimes t-2})_{i=1}^t$ are linearly dependent. Conversely Lemma 5.3 implies that removing a single vector from $(f_i^{\otimes t-2})_{i=1}^t$ yields a set of vectors which are linearly independent. It follows that there exists $(\alpha_i)_{i=1}^t$ with $\alpha_i \neq 0$ for all i and

$$\sum_{i=1}^t \alpha_i f_i^{\otimes t-2} = 0.$$

Without loss of generality we will assume that $\alpha_i < 0$ for $i \in [l]$ with $l \leq \lfloor \frac{t}{2} \rfloor$. From this we have

$$(3) \quad \sum_{i=1}^l -\alpha_i f_i^{\otimes t-2} = \sum_{j=l+1}^t \alpha_j f_j^{\otimes t-2}.$$

From Lemma 5.1 we have

$$\sum_{i=1}^l -\alpha_i f_i^{\times t-2} = \sum_{j=l+1}^t \alpha_j f_j^{\times t-2}$$

and thus

$$\begin{aligned} \int \sum_{i=1}^l -\alpha_i f_i^{\times t-2} d\pi^{\times t-2} &= \int \sum_{j=l+1}^t \alpha_j f_j^{\times t-2} d\pi^{\times t-2} \\ \Rightarrow \sum_{i=1}^l -\alpha_i &= \sum_{j=l+1}^t \alpha_j. \end{aligned}$$

Let $r = \sum_{i=1}^l -\alpha_i$. We know $r > 0$ so dividing both sides of (3) by r gives us

$$\sum_{i=1}^l -\frac{\alpha_i}{r} f_i^{\otimes t-2} = \sum_{j=l+1}^t \frac{\alpha_j}{r} f_j^{\otimes t-2}$$

where the left and the right side are convex combinations. Let $(\beta_i)_{i=1}^t$ positive numbers with $\beta_i = \frac{-\alpha_i}{r}$ for $i \in \{1, \dots, l\}$ and $\beta_j = \frac{\alpha_j}{r}$ for $j \in \{l+1, \dots, t\}$. This gives us

$$\sum_{i=1}^l \beta_i f_i^{\otimes t-2} = \sum_{j=l+1}^t \beta_j f_j^{\otimes t-2}.$$

If t is divisible by two then we can do the following,

$$\sum_{i=1}^l \beta_i f_i^{\otimes \frac{t}{2}-1} \otimes f_i^{\otimes \frac{t}{2}-1} = \sum_{j=l+1}^t \beta_j f_j^{\otimes \frac{t}{2}-1} \otimes f_j^{\otimes \frac{t}{2}-1}.$$

Consider the elements in the last inequality as order two tensors in $(L^2(\Psi, \mathcal{G}, \pi)^{\otimes \frac{t}{2}-1}) \otimes (L^2(\Psi, \mathcal{G}, \pi)^{\otimes \frac{t}{2}-1})$. From Lemma 5.3 and Lemma 5.4 we have that the RHS

of the previous equation has rank at least $\frac{t}{2}$ and since $l \leq \frac{t}{2}$ it follows that $l = \frac{t}{2}$. If t is not divisible by 2 then we can apply Lemma 5.1 and integrate out a term to get

$$\begin{aligned} \int_{\Psi} \sum_{i=1}^l \beta_i f_i^{\times t-3} f_i(x) d\pi(x) &= \int_{\Psi} \sum_{j=l+1}^t \beta_j f_j^{\times t-3} f_j(y) d\pi(y) \\ \Rightarrow \sum_{i=1}^l \beta_i f_i^{\times t-3} &= \sum_{j=l+1}^t \beta_j f_j^{\times t-3}. \end{aligned}$$

Applying Lemma 5.1 again we get

$$\sum_{i=1}^l \beta_i f_i^{\otimes \frac{t-1}{2}-1} \otimes f_i^{\otimes \frac{t-1}{2}-1} = \sum_{j=l+1}^t \beta_j f_j^{\otimes \frac{t-1}{2}-1} \otimes f_j^{\otimes \frac{t-1}{2}-1}.$$

From Lemma 5.3 and Lemma 5.4 we have that the RHS of the previous equation has rank at least $\frac{t-1}{2}$ and thus $l \geq \frac{t-1}{2}$. From this we have that $t-l, l \geq \lfloor \frac{t}{2} \rfloor$ whether t be even or odd. Applying Lemma 5.1 we have

$$\sum_{i=1}^l \beta_i f_i^{\times t-2} = \sum_{j=l+1}^t \beta_j f_j^{\times t-2}.$$

From Lemma 6.2 we have,

$$\sum_{i=1}^l \beta_i (\varepsilon_i \gamma + (1 - \varepsilon_i) \gamma')^{\times t-2} = \sum_{j=l+1}^t \beta_j (\varepsilon_j \gamma + (1 - \varepsilon_j) \gamma')^{\times t-2}.$$

□

PROOF OF THEOREM 4.3. Let $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$ and $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$ be mixtures of measures such that $\mathcal{P}' \neq \mathcal{P}$. We will proceed by contradiction. Suppose that $\sum_{i=1}^m a_i \mu_i^{\times 2m} = \sum_{j=1}^l b_j \nu_j^{\times 2m}$. From Theorem 4.1 we know that \mathcal{P} is $2m-1$ identifiable and therefore $2m$ -identifiable by Lemma 4.1. Because of this it follows that $l > m$. From Lemma 6.1 there exists a finite measure ξ and non-negative functions $p_1, \dots, p_m, q_1, \dots, q_l \in L^1(\Omega, \mathcal{F}, \xi) \cap L^2(\Omega, \mathcal{F}, \xi)$ such that, for all $B \in \mathcal{F}$, $\mu_i(B) = \int_B p_i d\xi$ and $\nu_j(B) = \int_B q_j d\xi$ for all i, j . Using Lemmas 6.2 and 6.3 we have

$$\sum_{i=1}^m a_i p_i^{\times 2m} = \sum_{j=1}^l b_j q_j^{\times 2m}.$$

By Lemma 5.1 we have

$$\sum_{i=1}^m a_i p_i^{\otimes 2m} = \sum_{j=1}^l b_j q_j^{\otimes 2m},$$

and therefore

$$\sum_{i=1}^m a_i p_i^{\otimes m} \otimes p_i^{\otimes m} = \sum_{j=1}^l b_j q_j^{\otimes m} \otimes q_j^{\otimes m}.$$

Consider the elements in the last inequality as order two tensors in $(L^2(\Omega, \mathcal{F}, \xi)^{\otimes m}) \otimes (L^2(\Omega, \mathcal{F}, \xi)^{\otimes m})$. Since no pair of vectors in p_1, \dots, p_m are collinear, from Lemma 5.3 and Lemma 5.4 we know that the LHS has rank m . On the other hand, no pair of vectors q_1, \dots, q_l are collinear either, so Lemma 5.3 says that there are at least $m+1$ linearly independent elements in $\{q_i^{\otimes m}\}_{i=1}^l$ so by Lemma 5.4 the RHS has rank $m+1$ a contradiction. \square

PROOF OF THEOREM 4.4. To prove this theorem we will construct a pair of mixture of measures, $\mathcal{P} \neq \mathcal{P}'$ which both contain m components and satisfy $V_{2m-1}(\mathcal{P}) = V_{2m-1}(\mathcal{P}')$. From our definition of (Ω, \mathcal{F}) we know there exists $F \in \mathcal{F}$ such that F, F^C are nonempty. Let $x \in F$ and $x' \in F^C$. It follows that δ_x and $\delta_{x'}$ are different probability measures on (Ω, \mathcal{F}) . Applying Lemma 6.4 to the measures δ_x and $\delta_{x'}$, with $t = 2m+1$ we get and letting $\mu_i = \varepsilon_i \delta_x + (1 - \varepsilon_i) \delta_{x'}$, we get

$$\sum_{i=1}^m \beta_i \mu_i^{\times 2m-1} = \sum_{j=m+1}^{2m} \beta_j \mu_j^{\times 2m-1}.$$

If we let $\mathcal{P} = \sum_{i=1}^m \beta_i \mu_i$ and $\mathcal{P}' = \sum_{j=m+1}^{2m+1} \beta_j \mu_j$, we have that $V_{2m-1}(\mathcal{P}) = V_{2m-1}(\mathcal{P}')$. \square

For the next theorem we will need to define Hilbert-Schmidt operators.

DEFINITION 6.1. *Let H, H' be Hilbert spaces and $T \in \mathcal{L}(H, H')$. T is called a Hilbert-Schmidt operator if $\sum_{x \in J} \|T(x)\|^2 < \infty$ for an orthonormal basis $J \subset H$. We denote the set of Hilbert-Schmidt operators in $\mathcal{L}(H, H')$ by $\mathcal{HS}(H, H')$.*

This definition does not depend on the choice of orthonormal basis: the sum $\sum_{x \in J} \|T(x)\|^2 < \infty$ will always yield the same value regardless of the choice of orthonormal basis J . The set of Hilbert-Schmidt operators is a subspace in the set of bounded linear operators.

PROOF OF THEOREM 4.5. Let $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$ be a mixture of measures with linearly independent components. Let $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$ be a mixture of measures with $l \leq m$ and $V_3(\mathcal{P}) = V_3(\mathcal{P}')$. From Lemma 6.1 there exists a finite measure ξ and non-negative functions $p_1, \dots, p_m, q_1, \dots, q_l \in L^1(\Omega, \mathcal{F}, \xi) \cap L^2(\Omega, \mathcal{F}, \xi)$ such that, for all $B \in \mathcal{F}$, $\int_B p_i d\xi = \mu_i(B)$ and $\int_B q_j d\xi = \nu_j$ for all i, j . Let $A \in \mathcal{F}^{\times 2}$ be arbitrary. We have

$$\begin{aligned} \sum_{i=1}^m a_i \mu_i^{\times 3}(A \times \Omega) &= \sum_{j=1}^l b_j \nu_j^{\times 3}(A \times \Omega) \\ \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times 2}(A) &= \sum_{j=1}^l b_j \nu_j^{\times 2}(A). \end{aligned}$$

Since A was arbitrary this implies

$$\sum_{i=1}^m a_i \mu_i^{\times 2} = \sum_{j=1}^l b_j \nu_j^{\times 2}.$$

From Lemma 6.3 we have that

$$\sum_{i=1}^m a_i p_i^{\times 2} = \sum_{j=1}^l b_j q_j^{\times 2}.$$

From Lemma 5.1 we have

$$\sum_{i=1}^m a_i p_i^{\otimes 2} = \sum_{j=1}^l b_j q_j^{\otimes 2}.$$

We will now show that $q_j \in \text{span}(p_1, \dots, p_m)$ for all j . Suppose that $q_t \notin \text{span}(p_1, \dots, p_m)$. Then there exists z such that $z \perp p_1, \dots, p_m$ but $z \not\perp q_t$. Now we have

$$\begin{aligned} \sum_{i=1}^m a_i p_i^{\otimes 2} &= \sum_{j=1}^l b_j q_j^{\otimes 2} \\ \Rightarrow \left\langle \sum_{i=1}^m a_i p_i \otimes p_i, z \otimes z \right\rangle &= \left\langle \sum_{j=1}^l b_j q_j \otimes q_j, z \otimes z \right\rangle \\ \Rightarrow \sum_{i=1}^m a_i \langle p_i \otimes p_i, z \otimes z \rangle &= \sum_{j=1}^l b_j \langle q_j \otimes q_j, z \otimes z \rangle \\ \Rightarrow \sum_{i=1}^m a_i \langle p_i, z \rangle^2 &= \sum_{j=1}^l b_j \langle q_j, z \rangle^2. \end{aligned}$$

We know that the LHS of the last equation is zero but the RHS is not, a contradiction.

We will find the following well known property of tensor products to be useful for continuing the proof ([11] Proposition 2.6.9).

LEMMA 6.5. *Let H, H' be Hilbert. There exists a unitary operator $U : H \otimes H' \rightarrow \mathcal{HS}(H, H')$, such that for any simple tensor $h \otimes h' \in H \otimes H'$, $U(h \otimes h') = \langle h, \cdot \rangle h'$.*

Because p_1, \dots, p_m are linearly independent we can do the following: for each $k \in [m]$ let $z_k \in \text{span}(p_1, \dots, p_m)$ be such that $z_k \perp \{p_i : i \neq k\}$ and $\langle z_k, p_k \rangle = 1$. Considering elements of $L^2(\Omega, \mathcal{F}, \xi)^{\otimes 3}$ as $L^2(\Omega, \mathcal{F}, \xi)^{\otimes 2} \otimes L^2(\Omega, \mathcal{F}, \xi)$, from Lemma 6.5 we have

$$\begin{aligned} \sum_{i=1}^m a_i p_i^{\otimes 2} \langle p_i, z_k \rangle &= \sum_{j=1}^l b_j q_j^{\otimes 2} \langle q_j, z_k \rangle \\ (4) \quad \Rightarrow a_k p_k^{\otimes 2} &= \sum_{j=1}^l b_j q_j^{\otimes 2} \langle q_j, z_k \rangle. \end{aligned}$$

The LHS of (4) is rank 1 and thus by Lemma 5.4 the RHS must have only one nonzero summand, which we will index with $\varphi(k)$. So we have

$$a_k p_k^{\otimes 2} = b_{\varphi(k)} q_{\varphi(k)}^{\otimes 2} \langle q_{\varphi(k)}, z_k \rangle$$

and from Lemma 6.2 we have

$$(5) \quad a_k \mu_k^{\times 2} = b_{\varphi(k)} \nu_{\varphi(k)}^{\times 2} \langle q_{\varphi(k)}, z_k \rangle.$$

Since $\mu_k^{\times 2}$ and $\nu_{\varphi(k)}^{\times 2}$ are both probability measures it follows that $\mu_k = \nu_{\varphi(k)}$. Since $\mu_i \neq \mu_j$ and $\nu_i \neq \nu_j$ for all $i \neq j$ we see that φ must be bijection. We can assert without loss of generality that $\varphi(k) = k$ and $\mu_k = \nu_k$ for all k . It directly follows that $l \geq m$. It also follows that $l = m$. To see this, consider a hypothetical q_{m+1} . From our previous argument it follows that q_{m+1} is perpendicular to z_1, \dots, z_m , which implies that q_{m+1} is not in the span of p_1, \dots, p_m , a contradiction.

Let $A \in \mathcal{F}$ be arbitrary. Since $V_3(\mathcal{P}) = V_3(\mathcal{P}')$ we have

$$\begin{aligned} \sum_{i=1}^m a_i \mu_i^{\times 3} &= \sum_{j=1}^m b_j \mu_j^{\times 3} \\ \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times 3}(\Omega \times \Omega \times A) &= \sum_{j=1}^m b_j \mu_j^{\times 3}(\Omega \times \Omega \times A) \\ \Rightarrow \sum_{i=1}^m a_i \mu_i(A) &= \sum_{j=1}^m b_j \mu_j(A) \end{aligned}$$

and thus $\sum_{i=1}^m a_i \mu_i = \sum_{j=1}^m b_j \mu_j$. Since μ_1, \dots, μ_m are linearly independent we have that $a_i = b_i$ for all i finally $\mathcal{P} = \mathcal{P}'$. \square

PROOF OF THEOREM 4.6. Let $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$ be a mixture of measures with jointly irreducible components. Consider a mixture of measures $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$ with $V_2(\mathcal{P}) = V_2(\mathcal{P}')$. From Lemma 6.1 there exists a finite measure ξ and non-negative functions $p_1, \dots, p_m, q_1, \dots, q_l \in L^1(\Omega, \mathcal{F}, \xi) \cap L^2(\Omega, \mathcal{F}, \xi)$ such that, for all $B \in \mathcal{F}$, $\int_B p_i d\xi = \mu_i(B)$ and $\int_B q_j d\xi = \nu_j$ for all i, j . From Lemma 6.2 we have

$$\sum_{i=1}^m a_i p_i \times p_i = \sum_{j=1}^l b_j q_j \times q_j.$$

From Lemma 5.1 we have

$$(6) \quad \sum_{i=1}^m a_i p_i \otimes p_i = \sum_{j=1}^l b_j q_j \otimes q_j.$$

Suppose for a moment that \mathcal{P}' contains a mixture component which does not lie in $\text{span}(\mu_1, \dots, \mu_m)$. Without loss of generality we will assume that $\nu_1 \notin \text{span}(\mu_1, \dots, \mu_m)$. It follows that $\nu_1, \mu_1, \dots, \mu_m$ are a linearly independent set of vectors and thus q_1, p_1, \dots, p_m are linearly independent. It follows that we can find some z such that $\langle z, q_1 \rangle \neq 0$ and $z \perp \mu_i$ for all i .

From (6) this we have the following

$$\begin{aligned}
\left\langle \sum_{i=1}^m a_i p_i \otimes p_i, z \otimes z \right\rangle &= \left\langle \sum_{j=1}^l b_j q_j \otimes q_j, z \otimes z \right\rangle \\
\Rightarrow \sum_{i=1}^m a_i \langle p_i \otimes p_i, z \otimes z \rangle &= \sum_{j=1}^l b_j \langle q_j \otimes q_j, z \otimes z \rangle \\
\Rightarrow \sum_{i=1}^m a_i \langle p_i, z \rangle^2 &= \sum_{j=1}^l b_j \langle q_j, z \rangle^2.
\end{aligned}$$

All the summands on both sides of the last equation are nonnegative. By our construction of z the LHS of the previous equation is zero and the first summand on the RHS is nonnegative, a contradiction. Thus, each component in \mathcal{P}' must lie in the span of the components of \mathcal{P} .

Now we have, for all j , $q_j = \sum_{i=1}^m c_i^j p_i$. From joint irriducibility we have that $c_i^j \geq 0$ for all i and j . Now suppose that there exists r, s, s' such that $c_s^r, c_{s'}^r > 0$. From the linear independence of p_1, \dots, p_m we can find a z such that $\langle p_s, z \rangle = 1$ and $z \perp p_q$ for all $q \neq s$. Applying Lemma 6.5 to (6). We have

$$\begin{aligned}
\sum_{i=1}^m a_i p_i \langle p_i, \cdot \rangle &= \sum_{j=1}^l b_j q_j \langle q_j, \cdot \rangle \\
\Rightarrow \sum_{i=1}^m a_i p_i \langle p_i, z \rangle &= \sum_{j=1}^l b_j q_j \langle q_j, z \rangle \\
\Rightarrow a_s p_s &= \sum_{j=1}^l b_j \left[\sum_{t=1}^m c_t^j p_t \right] \left\langle \sum_{u=1}^m c_u^j p_u, z \right\rangle \\
\Rightarrow a_s p_s &= \sum_{j=1}^l b_j \left[\sum_{t=1}^m c_t^j p_t \right] c_s^j \\
&= \sum_{j=1}^l b_j \left[\sum_{t=1}^m c_t^j p_t \right] c_s^j \\
&= \sum_{t=1}^m \sum_{j=1}^l b_j c_t^j c_s^j p_t \\
&= \sum_{t=1}^m p_t \sum_{j=1}^l b_j c_t^j c_s^j.
\end{aligned}$$

Let $\alpha_t = \sum_{j=1}^l b_j c_t^j c_s^j$ and note that each summand is nonnegative. Now we have

$$a_s p_s = \sum_{t=1}^m \alpha_t p_t.$$

We know that $\alpha_{s'} > 0$ since $b_r c_s^r c_{s'}^r > 0$. This violates the linear independence of p_1, \dots, p_m . Now we have that for all i there exists j such that $p_i = q_j$. From the minimality of the representation of mixtures of measures it follows that $l = m$ and without loss of generality we can assert that $p_i = q_i$ for all i . Because $p_1 \otimes p_1, \dots, p_m \otimes p_m$ are linearly independent it follows that $a_i = b_i$ for all i and thus $\mathcal{P} = \mathcal{P}'$. \square

7. Discussion. The following are a few observations related to the results.

7.1. Possible Statistical Estimator or Test. One interesting observation from the proof of Theorem 4.3 is that, if $\mathcal{P} = \sum_{i=1}^m a_i \mu_i$ is a mixture of measures with p_i being the pdf for μ_i then if $n > m$ the rank of $\sum_{i=1}^m a_i p_i^{\otimes n} \otimes p_i^{\otimes n}$ will be exactly m . This suggests a statistical test for number of mixture components. The form of this tensor is amenable to spectral methods since it is a positive semi-definite tensor of order 2, which is akin to positive semi-definite matrix. Embedding the data with kernel mean mapping, using a universal kernel [14], seems like a promising approach to constructing such a test or estimator.

7.2. Identifiability and the Value $2n - 1$. The value $2n - 1$ seems to carry some significance for identifiability beyond the setting we proposed. This value can also be found in results concerning metrics on trees [15], hidden Markov models [16], and frame theory, with applications to signal processing [5]. All of these results are related to identifiability of an object or the injectivity of an operator. We can offer no further insight as to why this value recurs, but it appears to be an algebraic phenomenon.

8. Meta-Algorithm. Here we will present a few algorithms for recovery of the mixture components. blah BLAH. For our first algorithm we will let $\sum_{i=1}^m w_i \mu_i$ be an arbitrary mixture of measures with p_1, \dots, p_m being densities representing μ_1, \dots, μ_m . We will also assume that p_1, \dots, p_m have distinct norms. Let

$$C = \sum_{i=1}^m w_i p_i^{\otimes m-1} \langle p_i^{\otimes m-1}, \cdot \rangle.$$

Here C is a PSD operator $L^2(\Omega, \mathcal{F}, \xi)$ to itself. Let C^\dagger be the pseudo-inverse of C and $W = \sqrt{C^\dagger}$. Now W is an operator which whitens $\sqrt{w_1}p_1^{\otimes m-1}, \dots, \sqrt{w_m}p_m^{\otimes m-1}$. That is $W\sqrt{w_1}p_1^{\otimes m-1}, \dots, W\sqrt{w_m}p_m^{\otimes m-1}$ are orthonormal vectors. Now we can construct the following tensor

$$\sum_{i=1}^m w_i p_i \otimes W p_i^{\otimes m-1} \otimes p_i \otimes W p_i^{\otimes m-1}$$

Which can again be represented as a PSD operator

$$\begin{aligned} T &\triangleq \sum_{i=1}^m w_i p_i \otimes W p_i^{\otimes m-1} \langle p_i \otimes W p_i^{\otimes m-1}, \cdot \rangle \\ &= \sum_{i=1}^m \sqrt{w_i} p_i \otimes W p_i^{\otimes m-1} \langle \sqrt{w_i} p_i \otimes W \sqrt{w_i} p_i^{\otimes m-1}, \cdot \rangle. \end{aligned}$$

For $i \neq j$ it follows that $\sqrt{w_i} p_i \otimes W p_i^{\otimes m-1} \perp \sqrt{w_j} p_j \otimes W p_j^{\otimes m-1}$. To see this

$$\begin{aligned} \langle p_i \otimes W \sqrt{w_i} p_i^{\otimes m-1}, p_j \otimes \sqrt{w_j} W p_j^{\otimes m-1} \rangle &= \langle p_i, p_j \rangle \langle W \sqrt{w_i} p_i^{\otimes m-1}, \sqrt{w_j} W p_j^{\otimes m-1} \rangle \\ &= \langle p_i, p_j \rangle 0 \\ &= 0. \end{aligned}$$

Also note that

$$\begin{aligned} \|p_i \otimes W \sqrt{w_i} p_i^{\otimes m-1}\|^2 &= \langle p_i \otimes W \sqrt{w_i} p_i^{\otimes m-1}, p_i \otimes W \sqrt{w_i} p_i^{\otimes m-1} \rangle \\ &= \langle p_i, p_i \rangle \langle W \sqrt{w_i} p_i^{\otimes m-1}, W \sqrt{w_i} p_i^{\otimes m-1} \rangle \\ &= \|p_1\|^2. \end{aligned}$$

If p_1, \dots, p_m have distinct norms then it follows that $p_1 \otimes W \sqrt{w_1} p_1^{\otimes m-1}, \dots, p_m \otimes W \sqrt{w_m} p_m^{\otimes m-1}$ are eigenvectors of T with distinct eigenvalues. Given an eigenvector of T , $p_i \otimes W \sqrt{w_i} p_i^{\otimes m-1}$ we need only transform it into a linear operator $p_i \langle W \sqrt{w_i} p_i^{\otimes m-1}, \cdot \rangle$ to recover p_i , scaled by some constant. From the reconstructed p_1, \dots, p_m one can recover the mixture proportions. Because $p_1^{\otimes m-1}, \dots, p_m^{\otimes m-1}$ are linearly independent and we have access to $\sum_{i=1}^m w_i p_i^{\otimes m-1}$ we need only solve for the coefficients in the last summation.

We can construct a similar algorithm with linearly independent components and 4 samples per group. If p_1, \dots, p_m are linearly independent and have distinct norms then we can whiten the mixture components by taking the square root of the inverse of the operator $\sum_{i=1}^m w_i p_i \langle p_i, \cdot \rangle$ and proceeding similarly.

References.

- [1] Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, 37(6A):3099–3132, 12 2009.
- [2] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15:2773–2832, 2014.
- [3] Joseph Anderson, Mikhail Belkin, Navin Goyal, Luis Rademacher, and James Voss. The more, the merrier: the blessing of dimensionality for learning large gaussian mixtures. In *Proceedings of The 27th Conference on Learning Theory*, pages 1135–1164, 2014.
- [4] Sanjeev Arora, Rong Ge, Ravindran Kannan, and Ankur Moitra. Computing a non-negative matrix factorization – provably. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC ’12, pages 145–162, New York, NY, USA, 2012. ACM.
- [5] Radu Balan, Pete Casazza, and Dan Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345 – 356, 2006.
- [6] Gilles Blanchard and Clayton Scott. Decontamination of mutually contaminated models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, AISTATS 2014, Reykjavik, Iceland, April 22-25, 2014*, pages 1–9, 2014.
- [7] C. Bruni and G. Koch. Identifiability of continuous mixtures of unknown gaussian distributions. *Ann. Probab.*, 13(4):1341–1357, 11 1985.
- [8] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279, 2008.
- [9] David L. Donoho and Victoria Stodden. When does non-negative matrix factorization give a correct decomposition into parts? In Sebastian Thrun, Lawrence K. Saul, and Bernhard Schölkopf, editors, *NIPS*. MIT Press, 2003.
- [10] Gerald B. Folland. *Real analysis: modern techniques and their applications*. Pure and applied mathematics. Wiley, 1999.
- [11] R.V. Kadison and J.R. Ringrose. *Fundamentals of the theory of operator algebras. V1: Elementary theory*. Pure and Applied Mathematics. Elsevier Science, 1983.
- [12] Olav Kallenberg. *Foundations of modern probability*. Probability and its applications. Springer, New York, Berlin,, Paris, 2002. Sur la 4e de couv. : This new edition contains four new chapters as well as numerous improvements throughout the text.
- [13] Joseph B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18(2):95 – 138, 1977.
- [14] Charles A. Micchelli, Yuesheng Xu, Haizhang Zhang, and Gabor Lugosi. Universal kernels. *J. Machine Learning Research*, 7:2651–2667, 2006.
- [15] Lior Pachter and David E Speyer. Reconstructing trees from subtree weights. *Applied Mathematics Letters*, 17:615–621, 2004.
- [16] Azaria Paz. *Introduction to Probabilistic Automata (Computer Science and Applied Mathematics)*. Academic Press, Inc., Orlando, FL, USA, 1971.
- [17] Yuval Rabani, Leonard J. Schulman, and Chaitanya Swamy. Learning mixtures of arbitrary distributions over large discrete domains. *ArXiv e-prints*, 2013.
- [18] Henry Teicher. Identifiability of finite mixtures. *Ann. Math. Statist.*, 34(4):1265–1269, 12 1963.

- [19] N. Young. *An Introduction to Hilbert Space*. Cambridge mathematical textbooks. Cambridge University Press, 1988.

APPENDIX A: ADDITIONAL PROOFS

Some of the proofs use Hilbert-Schmidt operators. See Definition 6.1 for the definition of Hilbert-Schmidt operator.

PROOF OF LEMMA 3.1. Because both representations are minimal it follows that $\alpha'_i \neq 0$ for all i and $\nu'_i \neq \nu'_j$ for all $i \neq j$. From this we know $\mathcal{Q}(\{\nu'_i\}) \neq 0$ for all i . Because $\mathcal{Q}(\{\nu'_i\}) \neq 0$ for all i it follows that for any i there exists some j such that $\nu'_i = \nu_j$. Let $\psi : [r] \rightarrow [r]$ be a function satisfying $\nu'_i = \nu_{\psi(i)}$. Because the elements ν_1, \dots, ν_r are also distinct, ψ must be injective and thus a permutation. Again from this distinctness we get that, for all i , $\mathcal{Q}(\{\nu'_i\}) = \alpha'_i = \alpha_{\psi(i)}$ and we are done. \square

PROOF OF LEMMA 4.1 AND 4.3. We will proceed by contradiction. Let $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$ be n -identifiable/determined, let $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$ be a different mixture of measures, with $l \leq m$ for the n -identifiable case, and

$$\sum_{i=1}^m a_i \mu_i^{\times q} = \sum_{j=1}^l b_j \nu_j^{\times q}$$

for some $q > n$. Let $A \in \mathcal{F}^{\times n}$ be arbitrary. We have

$$\begin{aligned} \sum_{i=1}^m a_i \mu_i^{\times q} &= \sum_{j=1}^l b_j \nu_j^{\times q} \\ \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times q} (A \times \Omega^{\times q-n}) &= \sum_{j=1}^l b_j \nu_j^{\times q} (A \times \Omega^{\times q-n}) \\ \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times n} (A) &= \sum_{j=1}^l b_j \nu_j^{\times n} (A). \end{aligned}$$

This implies that \mathcal{P} is not n -identifiable/determined, a contradiction. \square

PROOF OF LEMMA 4.2 AND 4.4. Let a mixture of measures $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$ not be n -identifiable/determined. It follows that there exists a different mixture of measures $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$, with $l \leq m$ for the n -identifiability case, such that

$$\sum_{i=1}^m a_i \mu_i^{\times n} = \sum_{j=1}^l b_j \nu_j^{\times n}.$$

Let $A \in \mathcal{F}^{\times q}$ be arbitrary, we have

$$\begin{aligned} \sum_{i=1}^m a_i \mu_i^{\times n} (A \times \Omega^{\times n-q}) &= \sum_{j=1}^l b_j \nu_j^{\times n} (A \times \Omega^{\times n-q}) \\ \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times q} (A) &= \sum_{j=1}^l b_j \nu_j^{\times q} (A) \end{aligned}$$

and therefore \mathcal{P} is not q -identifiable/determined. \square

PROOF OF LEMMA 5.1. Example 2.6.11 in [11] states that for any two σ -finite measure spaces $(S, \mathcal{S}, m), (S', \mathcal{S}', m')$ there exists a unitary operator $U : L^2(S, \mathcal{S}, m) \otimes L^2(S', \mathcal{S}', m') \rightarrow L^2(S \times S', \mathcal{S} \times \mathcal{S}', m \times m')$ such that, for all f, g ,

$$U(f \otimes g) = f(\cdot)g(\cdot).$$

Because $(\Psi, \mathcal{G}, \gamma)$ is a σ -finite measure space it follows that $(\Psi^{\times m}, \mathcal{G}^{\times m}, \gamma^{\times m})$ is a σ -finite measure space for all $m \in \mathbb{N}$. We will now proceed by induction. Clearly the lemma holds for $n = 1$. Suppose the lemma holds for $n - 1$. From the induction hypothesis we know that there exists a unitary transform $U_{n-1} : L^2(\Psi, \mathcal{G}, \gamma)^{\otimes n-1} \rightarrow L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1})$ such that for all simple tensors $f_1 \otimes \cdots \otimes f_{n-1} \in L^2(\Psi, \mathcal{G}, \gamma)^{\otimes n-1}$ we have $U_{n-1}(f_1 \otimes \cdots \otimes f_{n-1}) = f_1(\cdot) \cdots f_{n-1}(\cdot)$. Combining U_{n-1} with the identity map via Lemma 5.2 we can construct a unitary operator $T_n : L^2(\Psi, \mathcal{G}, \gamma)^{\otimes n-1} \otimes L^2(\Psi, \mathcal{G}, \gamma) \rightarrow L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1}) \otimes L^2(\Psi, \mathcal{G}, \gamma)$, which maps $f_1 \otimes \cdots \otimes f_{n-1} \otimes f_n \mapsto f_1(\cdot) \cdots f_{n-1}(\cdot) \otimes f_n$.

From the aforementioned example there exists a unitary transform $K_n : L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1}) \otimes L^2(\Psi, \mathcal{G}, \gamma) \rightarrow L^2(\Psi^{\times n-1} \times \Psi, \mathcal{G}^{\times n-1} \times \mathcal{G}, \gamma^{\times n-1} \times \gamma)$ which maps simple tensors $g \otimes g' \in L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1}) \otimes L^2(\Psi, \mathcal{G}, \gamma)$ as $K_n(g \otimes g') = g(\cdot)g'(\cdot)$. Defining $U_n(\cdot) = K_n(T_n(\cdot))$ yields our desired unitary transform. \square

PROOF OF LEMMA 5.2. Proposition 2.6.12 in [11] states that there exists a continuous linear operator $\tilde{U} : H_1 \otimes \cdots \otimes H_n \rightarrow H'_1 \otimes \cdots \otimes H'_n$ such that $\tilde{U}(h_1 \otimes \cdots \otimes h_n) = U_1(h_1) \otimes \cdots \otimes U_n(h_n)$ for all $h_1 \in H_1, \dots, h_n \in H_n$. Let \hat{H} be the set of simple tensors in $H_1 \otimes \cdots \otimes H_n$ and \hat{H}' be the set of simple tensors in $H'_1 \otimes \cdots \otimes H'_n$. Because U_i is surjective for all i , clearly $\tilde{U}(\hat{H}) = \hat{H}'$. The linearity of \tilde{U} implies that $\tilde{U}(\text{span}(\hat{H})) = \text{span}(\hat{H}')$. Because $\text{span}(\hat{H}')$ is dense in $H'_1 \otimes \cdots \otimes H'_n$ the continuity of \tilde{U} implies that $\tilde{U}(H_1 \otimes \cdots \otimes H_n) = H'_1 \otimes \cdots \otimes H'_n$ so \tilde{U} is surjective. All that remains to be shown is that \tilde{U}

preserves the inner product. By the continuity of inner product we need only show that $\langle h, g \rangle = \langle \tilde{U}(h), \tilde{U}(g) \rangle$ for $h, g \in \text{span}(\hat{H})$. With this in mind let $h_1, \dots, h_N, g_1, \dots, g_M \in \hat{H}$. We have the following

$$\begin{aligned} \left\langle \tilde{U} \left(\sum_{i=1}^N h_i \right), \tilde{U} \left(\sum_{j=1}^M g_j \right) \right\rangle &= \left\langle \sum_{i=1}^N \tilde{U}(h_i), \sum_{j=1}^M \tilde{U}(g_j) \right\rangle \\ &= \sum_{i=1}^N \sum_{j=1}^M \langle \tilde{U}(h_i), \tilde{U}(g_j) \rangle \\ &= \sum_{i=1}^N \sum_{j=1}^M \langle h_i, g_j \rangle \\ &= \left\langle \sum_{i=1}^N h_i, \sum_{j=1}^M g_j \right\rangle. \end{aligned}$$

We have now shown that \tilde{U} is unitary which completes our proof. \square

PROOF OF LEMMA 5.3. We will proceed by induction. For $n = 2$ the lemma clearly holds. Suppose the lemma holds for $n - 1$ and let h_1, \dots, h_n satisfy the assumptions in the lemma statement. Let $\alpha_1, \dots, \alpha_n$ satisfy

$$(7) \quad \sum_{i=1}^n \alpha_i h_i^{\otimes n-1} = 0.$$

To finish the proof we will show that α_1 must be zero which can be generalized to any α_i without loss of generality. Applying Lemma 6.5 to (7) we get

$$(8) \quad \sum_{i=1}^n \alpha_i h_i^{\otimes n-2} \langle h_i, \cdot \rangle = 0.$$

Because h_1 and h_n are linearly independent we can choose z such that $\langle h_1, z \rangle \neq 0$ and $z \perp h_n$. Plugging z into (8) yields

$$\sum_{i=1}^{n-1} \alpha_i h_i^{\otimes n-2} \langle h_i, z \rangle = 0$$

and therefore $\alpha_1 = 0$ by the inductive hypothesis. \square

PROOF OF LEMMA 5.4. Let $\dim(\text{span}(h_1, \dots, h_m)) = l$ and let $h = \sum_{i=1}^m h_i^{\otimes 2}$. Without loss of generality assume that h_1, \dots, h_l are linearly independent and nonzero. From Lemma 6.5 there exists a unitary transform $U : H \otimes H \rightarrow \mathcal{HS}(H, H)$ which, for any simple tensor $x \otimes y$, we have $U(x \otimes y) = x \langle y, \cdot \rangle$.

First we will show that the rank is greater than or equal to l by contradiction. Suppose that $g = \sum_{i=1}^{l'} x_i \otimes y_i = h$ with $l' < l$. Since $l' < l$ there must exist some j such that $h_j \notin \text{span}(x_1, \dots, x_{l'})$. Let $z \perp x_1, \dots, x_{l'}$ and $z \not\perp h_j$. Now we have

$$\langle z \otimes z, h \rangle = \sum_{i=1}^m \langle z, h_i \rangle^2 \geq \langle z, h_j \rangle^2 > 0,$$

but

$$\langle z \otimes z, g \rangle = \sum_{i=1}^{l'} \langle z, x_i \rangle \langle z, y_i \rangle = 0,$$

a contradiction.

For the other direction, observe that $U(h)$ is a compact Hermitian operator and thus admits an eigen-decomposition ([19] Theorem 8.15). From this we have that $U(h) = \sum_{i=1}^{\infty} \lambda_i \langle \psi_i, \cdot \rangle \psi_i$ with $(\psi_i)_{i=1}^{\infty}$ orthonormal and $\lambda_i \geq 0$ for all i . Clearly the span of $U(h) \leq l$ and thus this decomposition has exactly l nonzero terms. From this we can let $U(h) = \sum_{i=1}^l \lambda_i \langle \psi_i, \cdot \rangle \psi_i$ and applying U^{-1} we have that $h = \sum_{i=1}^l \lambda_i \psi_i^{\otimes 2}$. From this it follows that the rank of h is less than or equal to l and we are done. \square

PROOF OF LEMMA 6.1. Let $\pi = \sum_{i=1}^n \gamma_i$. Because π is σ -finite for all i we can define $f_i = \frac{d\gamma_i}{d\pi}$, where the derivatives are Radon-Nikodym derivatives. Let f_k be arbitrary. We will first show that $f_k \leq 1$ π -almost everywhere. Suppose there exists a non π -null set $A \in \mathcal{G}$ such that $f_i(A) > 1$. Then we would have

$$\begin{aligned} \gamma_k(A) &= \int_A f_k d\pi \\ &> \int_A 1 d\pi \\ &= \sum_{i=1}^n \gamma_i(A) \\ &\geq \gamma_k(A) \end{aligned}$$

a contradiction. From this we have

$$\begin{aligned} \int f_k^2 d\pi &\leq \int 1 d\pi \\ &\leq \sum_{i=1}^n \gamma_i(\Psi) \\ &< \infty. \end{aligned}$$

From our construction it is clear that $f_i \geq 0$ ξ -almost everywhere so we can assert $f_i \geq 0$ without issue. \square

PROOF OF LEMMA 6.2. The fact that f is non-negative and integrable implies that the map $S \mapsto \int_S f^{\times n} d\pi^{\times n}$ is a bounded measure on $(\Psi^{\times n}, \mathcal{G}^{\times n})$ (see [10] Exercise 2.12).

Let $R = R_1 \times \cdots \times R_n$ be a rectangle in $\mathcal{G}^{\times n}$. Let $\mathbb{1}_S$ be the indicator function for a set S . Integrating over R and using Tonelli's theorem we get

$$\begin{aligned} \int_R f^{\times n} d\pi^{\times n} &= \int \mathbb{1}_R f^{\times n} d\pi^{\times n} \\ &= \int \left(\prod_{i=1}^n \mathbb{1}_{R_i}(x_i) \right) \left(\prod_{j=1}^n f(x_j) \right) d\pi^{\times n}(x_1, \dots, x_n) \\ &= \int \cdots \int \left(\prod_{i=1}^n \mathbb{1}_{R_i}(x_i) \right) \left(\prod_{j=1}^n f(x_j) \right) d\pi(x_1) \cdots d\pi(x_n) \\ &= \int \cdots \int \left(\prod_{i=1}^n \mathbb{1}_{R_i}(x_i) f(x_i) \right) d\pi(x_1) \cdots d\pi(x_n) \\ &= \prod_{i=1}^n \left(\int \mathbb{1}_{R_i}(x_i) f(x_i) d\pi(x_i) \right) \\ &= \prod_{i=1}^n \gamma(R_i) \\ &= \gamma^{\times n}(R). \end{aligned}$$

Any product probability measure is uniquely determined by its measure over the rectangles (this is a consequence of Lemma 1.17 in [12] and the definition of product σ -algebra) therefore, for all $B \in \mathcal{G}^{\times n}$,

$$\gamma^{\times n}(B) = \int_B f^{\times n} d\pi^{\times n}.$$

\square

EECS BUILDING
1301 BEAL AVENUE
ANN ARBOR MI, 48105
E-MAIL: rvdm@umich.edu
clayscot@umich.edu