CHARACTERIZATION OF FINITE GROUP INVARIANT DISTRIBUTIONS*

By Alexander Volfovsky and Edoardo M. Airoldi

Harvard University

Invariance properties of probability distributions have long been an area of interest in statistics and probability. In particular, much has been written about the famous de Finetti's theorem for infinitely exchangeable sequences and the Aldous-Hoover representation for infinitely exchangeable arrays. The key to both of these results is the existence of an infinite sequence or an infinite array. When this assumption fails, so do the aforementioned representation theorems. In this article we demonstrate this failure for the Aldous-Hoover representation and then develop a characterization theorem for the space of distributions that are invariant under a general finite group $G$. In particular, the results of Diaconis (1977) for finitely exchangeable sequences are a special case of our general result. We find sharp bounds on the total variation distance between distributions of finitely and infinitely partially exchangeable sequences and finitely and infinitely exchangeable arrays. Additionally we provide an example of the applicability of the theory for groups other than the symmetric group.

CONTENTS

1 Introduction ............................................. 2
2 Definitions and preliminaries .............................. 5
  2.1 Groups .............................................. 5
  2.2 Finding the orbits ................................... 6
3 Main results ............................................. 7
  3.1 Characterizing the space of $G$-invariant distributions .............................. 7
  3.2 Projection criterion ................................... 8
4 Examples of $G$-invariance ................................ 9
  4.1 Exchangeable sequences ................................ 9
  4.2 Partially exchangeable sequences ....................... 10
  4.3 Row-column exchangeable matrices ..................... 10

*This work was supported, in part, by ONR award N00014-14-1-0485, by ARO MURI award W911NF-11-1-0036 and NSF CAREER grant IIS-1149662 to Harvard University. AV is an NSF MSPR Fellow supported by NSF DMS-1402235. EMA is an Alfred P. Sloan Research Fellow.

AMS 2000 subject classifications: Primary 62E10; secondary 60E05

Keywords and phrases: Finite exchangeability, exchangeable matrices, networks, total variation bound, hypothesis testing
### 1. Introduction

The study of invariance properties of probability distributions has long been the focus of studies in probability (De Moivre, 1756; de Laplace, 1820; De Finetti, 1931; Haag, 1924; Hewitt and Savage, 1955; Aldous, 1985; Austin and Panchenko, 2013). The most common type of invariance is the notion of independent and identically distributed distributions. It is fundamental for classical results such as the weak and strong laws of large numbers and central limit theorems. Relaxation of the statistical notion of independence in this definition to simply requiring that the order of a sequence of random variables not affect their joint distribution gives rise to the notion of exchangeable random variable sequences. These variables are marginally identically distributed but they can exhibit different forms of dependence. Infinite exchangeability of a sequence of random variables is in turn related back to independent and identically distributed random variable sequences via a famous theorem due to de Finetti (De Finetti, 1931, 1972). Considering an infinitely exchangeable sequence $X_1, X_2, \ldots$ of binary random variables the theorem states that there is a unique probability measure $\mu$ such that

$$\Pr(X_1 = e_1, \ldots, X_n = e_n) = \int p^{\sum e_i}(1-p)^{n-\sum e_i}d\mu(p).$$

Generalizations of this famous theorem to higher dimensional arrays were given independently by Hoover (1989) and Aldous (1981): For example consider an infinitely row-column weakly exchangeable binary matrix $X$. The distribution of such a matrix is invariant under the permutation of the row and column indices of the matrix by the same permutation. Intuitively, such a matrix can represent relational data, where the shared index set of the rows and columns represents an actor, while the entries of the matrix represent the existence of a relationship among the actors. The Aldous-Hoover
representation for this matrix is

\[ X_{ij} = 1[W(U_i, U_j) \geq \lambda_{ij}], \]

where \( U_i \) and \( \lambda_{ij} \) are independent uniform \((0, 1)\) random variables and \( W \) is a measurable symmetric function from \([0, 1]^2 \to [0, 1]\). Intuitively, the \( U_i \) can be interpreted as actor attributes while the \( \lambda_{ij} \) as pairwise attributes.

The power of the above results is derived from the assumption that the index set of the array is infinite. Specifically, it is well known that de Finetti’s theorem does not hold for finitely exchangeable sequences (Diaconis, 1977; Diaconis and Freedman, 1980). Similarly, using a \( 2 \times 2 \) row-column weakly exchangeable matrix it is easy to show that the Aldous-Hoover representation theorem does not hold for finite arrays (the difference between strong and weak exchangeability is in how the indices are permuted: independently in the former and jointly in the latter). For \( X \) a \( 2 \times 2 \) weakly exchangeable matrix define \( p_{abcd} = \Pr(X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}) \) and let \( p_{abcd} = 0 \ \forall (a, b, c, d) \in \{(a, b, c, d) : b = c\} \). Letting the remaining probabilities have the same nonzero value we clearly define a non-trivial finite weakly exchangeable distribution. If an Aldous-Hoover representation exists then we can write

\[ \Pr(X \in A) = \int P_W(A) d\mu(W) \]

where \( P_W \) can be parametrized as follows:

\[ P_W = q_1^a (1 - q_1)^{1-a} q_2^d (1 - q_2)^{1-d} q_{12}^{b+c} (1 - q_{12})^{2 - b - c}, \]

for \( q_1 = W(U_1, U_1), q_2 = W(U_2, U_2) \) and \( q_{12} = W(U_1, U_2) \). Using the form of \( P_W \) in Equation (1) it is easy to verify that there does not exist a non-negative measure \( \mu \) that satisfies \( p_{abcd} = \int P_W d\mu(W) \) for \( (a, b, c, d) \in \{(a, b, c, d) : b = c\} \) (any non-negative solution would require that the measure \( \mu \) place weight 1 on \( q_{12} = 1 \) and on \( q_{12} = 0 \)).
representation as a sample from an infinite population, it is very common to observe the whole finite network such as in the study of trade between countries (Volfovsky and Hoff, 2013), friendship networks among students in the same class (Hoff et al., 2013) and interactions among monks in a monastery (Sampson, 1969). In these cases an assumption of exchangeability of the nodes is still desirable as there is no information in the individual labels, but the assumption of infinite exchangeability is inappropriate since there is no infinite network to relate the data to. Furthermore, we do not know of a treatment of a finite group invariance for groups not isomorphic to the symmetric group.

The remainder of this paper is devoted to fully characterizing a general class of distributions that are invariant under the operation of a finite group $G$. This extends the earlier results of Diaconis (1977) and Diaconis and Freedman (1980) from finite symmetric groups acting on spaces of sequences to arbitrary finite groups acting on general spaces (including arrays and networks). In the example above, the group in question is the symmetric group $S_2$ and the operation is defined by $g : (i, j) \rightarrow (gi, gj)$ while the space of interest is that of all $2 \times 2$ binary matrices. After characterizing the space of distributions, we extend the notion of extendibility to general finite groups and provide a closed form for a projection from distributions on the extended space to the original space. Using this new definition we provide the machinery for verifying extendibility via linear programming.

In the next section we develop the necessary group theoretic results for the general treatment of what we call $G$-invariant distributions and provide the computational tools necessary for computing the distributions of interest and for evaluating whether the distributions are projections of higher dimensional distributions. Section 3 presents the proofs of the main results of this article: the characterization theorem and the projection criterion. The projection criterion generalizes the classical notion of an extendible distribution to the general $G$-invariance case. In Section 4 we provide examples of $G$-invariant distributions. First we parse classical results on exchangeable sequences and partially exchangeable sequences in the terminology of our paper. The second part of the examples provides novel results for finitely exchangeable arrays as well as for invariances that are not given by the symmetric group. Section 5 develops total variation bounds for separately and jointly exchangeable arrays by exploiting the group operation and extending previously available results for exchangeable sequences to partially exchangeable sequences. Section 6 examines the introduction of a signed measure into the representation theorems of infinitely (row-column, partially) exchangeable spaces. A discussion follows.
2. Definitions and preliminaries. This section provides an overview of relevant group theory and computational group theory. These concepts provide a foundation for our formulation of $G$-invariance. Beyond the introduction to the notation of our article, these results demonstrate the difficulties associated with evaluating distributions that are invariant under groups other than the symmetric group. For a detailed introduction to group theory we refer the reader to Rotman (1995).

2.1. Groups. The primary object that we employ throughout this article is the finite group:

**Definition 1 (Finite group).** A group $G$ is an object consisting of a finite set $G$ together with a binary operation $\cdot$. $G$ is required to be closed under the group operation, to be associative, to have an identity element and to be closed under inverses. Finite groups are finitely generated.

The group most commonly associated with the study of invariance is $S_k$, the symmetric group of order $k$. The elements of the group are all possible permutations of subsets of the integers $\{1, \ldots, k\}$. The group operation of the symmetric group is composition and it is easy to see that composing two permutations gives a new permutation (the group is closed under the operation). Associativity and the existence of an identity element are trivial to verify. The generating set of a group is a (small) set of group elements such that any other element of the group can be constructed via the group action on elements of this set. The generating set of the symmetric group of order $k$ is the set of permutations of adjacent elements in the group, e.g. permuting 1 and 2, permuting 2 and 3, etc. There are $k + 1$ of those permutations. The group operation describes the internal structure of the group, while a group action defines the behavior of a set $\mathcal{X}$ under the elements of a group $G$. That is, the group action is a group homomorphism from the group $G$ to the symmetric group of the set $\mathcal{X}$. There are several useful notions associated with the action of a finite group $G$ on a space $\mathcal{X}$. They provide two different (but related) notions of equivalence classes under the group action.

**Definition 2 (Orbit).** An orbit of a point $x \in \mathcal{X}$ is the set $G(x) = \{gx : g \in G\}$ and is a $G$ invariant subset of $\mathcal{X}$ on which $G$ acts transitively.

**Definition 3 (Stabilizer).** The stabilizer (subgroup) of $x \in \mathcal{X}$ is the set of elements of $G$ that stabilizes $x$, that is $G_x = \{g \in G | gx = x\}$.

Let $\mathcal{X}$ be the collection of binary sequences of length $k$. Continuing the example of the symmetric group, define the group action of $S_k$ on $\mathcal{X}$ as
the permutation of the elements of $x \in \mathcal{X}$. Each orbit $G(x)$ then can be characterized by the number of 1s in $x$, while the stabilizer subgroup of an element $x$ is the group that permutes the 1s and 0s of the element separately. The orbit defines the equivalence classes of the set $\mathcal{X}$ under the action of $G$. Conjugacy defines an equivalence relation on the stabilizer subgroups. It is easy to see that if two elements of $\mathcal{X}$ belong to the same orbit then their stabilizer subgroups must be conjugate to each other. The following two results formalize this relationship:

**Burnside’s lemma.** The number of orbits of $\mathcal{X}$ under the action of a group $G$ is given by

$$1/|G| \sum_{g \in G} |\mathcal{X}^g| = 1/|G| \sum_{x \in \mathcal{X}} |G_x|.$$

Burnside’s lemma provides a count for the number of orbits by by considering (a) the average number of elements of $\mathcal{X}$ that is fixed by an element of $g \in G$, $|\mathcal{X}^g|$ or (b) the average size of the stabilizer subgroups in $G$, $|G_x|$.

**Orbit stabilizer lemma.** For a finite group $G$ and a set $\mathcal{X}$ we have that the size of the orbit $|G(x)|$ is equal to the index of the stabilizer subgroup $G_x$ in $G$, $[G : G_x]$. That is,

$$|G(x)| = [G : G_x] = |G|/|G_x|.$$

The orbit stabilizer lemma provides an intuition for this relationship as well as a means to calculate the size of the individual orbits, which will assist in determining the extreme distributions of finitely invariant distributions. An immediate observation due to the stabilizer lemma is that any element that is stabilized by the entire group must be the only element in its orbit.

### 2.2. Finding the orbits.

Several algorithms exist for computing the orbits of a group $G$ acting on a domain $\mathcal{X}$. The algorithms are intuitive though at a cost of frequently reducing to brute-force enumeration of elements. Computational improvements are available when the group $G$ has special properties such as being a permutation group (e.g. Seress (2003), Hulpke (2010)). Algorithm 1 assumes the generating set of the group $G$ is known and outputs all the orbits of $\mathcal{X}$ and is based on Algorithm I.4 of Hulpke (2010). This algorithm provides the technical tools necessary for implementing the new results presented in Theorems 1 and 2 of Section 3. A possible improvement to the algorithm is first calculating the conjugacy classes of the stabilizer...
subgroups, thus reducing both the number of elements of $X$ and $G$ over which the for loops must iterate.

**input**: a state space $\mathcal{X}$, a group $G$ with a generating set $\{g_1, \ldots, g_m\}$.

**output**: Collection of orbits $\{G(x)\}$.

Let $W = X$;

while $W \neq \emptyset$ do

  Choose $x \in W$ and let $G(x) = \{x\}$;

  for $y \in G(x)$ do

    for $i \in \{1, \ldots, m\}$ do

      $z = g_i y$;

      if $z \notin G(x)$ then

        Append $z$ to $G(x)$

      end

    end

  end

  $W \leftarrow W - G(x)$;

end

**return** Collection of orbits

**Algorithm 1**: Find orbits

### 3. Main results.

In this section we present the main results of the paper, namely, the general characterization for the space of $G$-invariant probability distributions as well as a projection theorem that generalizes the notion of extendibility to general groups.

#### 3.1. Characterizing the space of $G$-invariant distributions.

**Theorem 1** (Characterization Theorem). Let $G$ be a finite group acting on a finite state space $\mathcal{X}$. The collection of $G$-invariant measures on $\mathcal{X}$ is a convex set with extreme points $e_{G(x)}$ indexed by the orbits of $\mathcal{X}$ and placing weight $1/|G(x)|$ on each member of $G(x)$ and 0 everywhere else.

**Proof.** Convexity is a trivial consequence of the convexity of the general space of probability distributions on $\mathcal{X}$ when $|\mathcal{X}| < \infty$. To prove that $e_{G(x)}$ are the extreme measures we first demonstrate that measures placing weight on a single orbit are in fact $G$-invariant and correspond to our definition of $e_{G(x)}$. Note that the orbits of the group partition the state space $\mathcal{X}$ into disjoint subsets on which the action of the group $G$ is transitive. Combining the transitivity of the action with the fact that for $x \in \mathcal{X}$ we have $\Pr(x \in G(x)) = 1$, the distribution of $X$ must place equal weight on all entries in
$G(x)$, that is $\Pr(x \in G(x)) = \Pr(gx \in G(x)) \ \forall g \in G$. This is exactly the definition of $e_{G(x)}$. To see that the $e_{G(x)}$ are extreme we consider three orbits $G(x_1), G(x_2) \text{ and } G(x_3)$. Now, assume that $e_{G(x_1)} = \alpha e_{G(x_2)} + (1 - \alpha) e_{G(x_3)}$. Since $e_{G(x_1)}$ places all its weight on $G(x_1)$ then $e_{G(x_2)}$ and $e_{G(x_3)}$ must assign probability zero to all elements of $\mathcal{X} - G(x_1)$. By definition of $G$-invariance both must assign equal weight to elements in $G(x_1)$ which implies that $e_{G(x_2)} = e_{G(x_3)}$ as desired.

Since the $\{e_{G(x_i)} : \mathcal{X} = \cup_i G(x_i), G(x_i) \text{ disjoint}\}$ are extreme, all $G$-invariant distributions can be represented as a unique convex combinations of these measures.

Remark 1. If $\mathcal{X}$ is not finite the results of Theorem 1 still hold though the extreme measures all have measure zero and so the exchangeable distribution now has integral form $P(A) = \int e_{G(x)}(A)f(x)dx$ where $f$ integrates to 1 over $\mathcal{X}$. The only requirement for the above to be true is that $e_{G(x)}$ are measurable on $(\mathcal{X}, \Sigma)$ where $\Sigma$ is an appropriate $\sigma$-algebra.

A special case of Theorem 1 where $G$ is the symmetric group and $\mathcal{X}$ is the collection of binary sequences appears in Eaton (1989).

3.2. Projection criterion. The study of finitely exchangeable distributions on sequences discusses the question of extendibility of the distribution, that is whether the finitely exchangeable distribution (of a sequence of length $k$) can be viewed as the distribution of the first $k$ elements of a finitely exchangeable distribution for sequences of length $n > k$. Similarly, if a distribution is $G$-invariant, it is of interest to determine if there are any finite groups $H$ that induce a similar behavior on the state space $\mathcal{X}$ in the following sense: Can a $G$-invariant distribution be written as the restriction (projection) of an $H$-invariant distribution?

We develop a constructive approach to this problem. Consider $P_H$ a $H$-invariant distribution and a $G$ that is isomorphic to a subgroup of $H$. To find out if $P_G$ is a projection of $P_H$ we will construct the possible space of projected distributions $P_{G|H}$ by projecting the extreme $H$-invariant distributions down. Because of the convexity of the space of $H$-invariant distributions this projected space is also convex and so if $P_G$ can be written as a convex combination of the projected extreme points, it must be $H$-extendible. Finally, we require a few definitions. $\tilde{\mathcal{X}}$ is an extension of $\mathcal{X}$ if there exists a natural embedding $f$ of $\mathcal{X}$ in $\tilde{\mathcal{X}}$ that is onto but not one-to-one. We let $f(x)$ be the range of a point $x \in \mathcal{X}$. The projection is defined via the extreme measures on the space - that is those defined by the orbits of $H$ on $\tilde{\mathcal{X}}$. Recall the extreme distribution of a particular orbit defined in
Theorem 1, $e_{H(\tilde{x})}$. Its projection onto the space of $G$-invariant distributions is defined as follows:

$$P_{G|H(\tilde{x})}(x) = \sum_{y \in f(x)} e_{H(\tilde{x})}(y) = \frac{|f(x) \cap H(\tilde{x})|}{|H(\tilde{x})|}.$$  

Since all $H$-invariant distributions can be written as convex combinations of the summands on the right hand side above, this is a projection of $H$-invariant distributions into $G$-invariant distributions. As the projection is a linear map that takes convex regions into convex regions, we get the following theorem.

**Theorem 2 (Projection Theorem).** Given a finite group $G$ that is isomorphic to a subgroup of a finite group $H$ and their respective actions on the state spaces $X$ and its extension $\tilde{X}$, the projection (2) takes the extreme points of the space of $H$-invariant distributions onto the convex collection \{$P_{G|H(\tilde{x})} : H(\tilde{x})$ is an orbit\}.

An immediate consequence of this theorem is that a $G$-invariant distribution is $H$-extendible if and only if it falls in the convex region defined by the theorem. Since the region is convex verifying this result is an exercise in linear programming.

We note that the linear program has been shown to be relatively simple to solve for the case of exchangeable sequences ($G = S_k$ is the symmetric group acting on sequences of length $k$). On the other hand, for even slightly more complicated structures (such as partial exchangeability) solving the explicit linear program appears to be necessary (Di Cecco, 2009).

**4. Examples of $G$-invariance.** For each of the following examples we provide a brief description of an application and then provide an explicit description of the state space of the application as well as the invariance group that acts on this state space. Sections 4.1 and 4.2 discuss exchangeability of sequences while Sections 4.3 and 4.4 deal with matrices and structures with symmetries not related to the symmetric group.

**4.1. Exchangeable sequences.** Consider operational information collected on $k$ machines (on/off status). Since there is no reason to believe that the order of the status contains any information, in general these observations can be modeled as exchangeable. Let $x$ be binary sequence of length $k$. The distribution of $x$ is exchangeable if for any permutation $\pi$ of $\{1, \ldots, k\}$ we have $P(x_1 = e_1, \ldots, x_k = e_k) = P(x_{\pi(1)} = e_1, \ldots, x_{\pi(k)} = e_k)$. Translating this to the notation we have developed:
The space: $\mathcal{X} = \{x: x \in \{0,1\}^k\}$, all binary sequences of length $k$.
The group: $G = S_k$, the symmetric group of dimension $k$.

To apply Theorem 1 we must find the orbits of $\mathcal{X}$ under the group action. It is easy to see that each orbit must fix the number of zero and nonzero elements in the sequence. The specification of this example was previously developed in Eaton (1989) while Diaconis (1977) explicitly uses a counting argument to find the extreme measures. Going back to the example of on/off status of machines, it is easy to see that if we know that $l < k$ of the machines are on but we do not know which ones then the joint distribution of the status is finitely exchangeable since this is the distribution of an urn with $l$ ones and $k-l$ zeros from which one samples without replacement.

The above result is easily generalized to a sequence $x$ of length $k$ taking on values in $\{1, \ldots, s\}$. Note that the group acting on $\mathcal{X} = \{x: x \in \{1, \ldots, s\}^k\}$ is still the symmetric group of dimension $k$. Analogously to the orbits above we have that $G(x)$ fixes the number of 1s, 2s, etc. in the sequence.

4.2. Partially exchangeable sequences. Consider a school with $p$ classrooms. Students across classrooms are not exchangeable while students within a classroom are exchangeable. This is called hierarchical, or partial, exchangeability. To keep the example simple we consider $x$ binary as above, but with the following indexing: $x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2}, \ldots, x_{pn_p}$ where the first index element numbers the classrooms while the second numbers each child within a classroom. The partial exchangeability assumption implies that

$$P(x_{11} = e_{11}, \ldots, x_{1n_1} = e_{1n_1}, \ldots, x_{p1} = e_{p1}, \ldots, x_{pn_p} = e_{pn_p})$$
$$= P((x_{1\pi_1(1)} = e_{11}, \ldots, x_{1\pi_1(n_1)} = e_{1n_1}, \ldots, x_{p\pi_p(1)} = e_{p1}, \ldots, x_{p\pi_p(n_p)} = e_{pn_p}),$$

where $\pi_i$ is any permutation of the $\{1, \ldots, n_p\}$. In our notation:

The space: $\mathcal{X} = \{x: x \in \{0,1\}^{\sum_{i=1}^p n_i}\}$.
The group: $G = S_{n_1} \times \cdots \times S_{n_p}$, a direct product of symmetric groups.

Each $S_{n_i}$ acts on elements $\sum_{t=0}^{i-1} n_t + 1$ to $\sum_{t=0}^{i} n_t$ with $n_0 = 0$ of $x \in \mathcal{X}$. As above, the orbits must fix the number of zero and nonzero elements in the sequence, but now they do so in each of the subsequences.

4.3. Row-column exchangeable matrices.

4.3.1. Separately exchangeable matrices. Imagine we ask $m$ individuals $n$ unrelated yes or no questions. In this case, the questions as well as the
individuals can be considered exchangeable. More formally, consider a binary matrix $X \in \{0, 1\}^{m \times n}$. We say that the distribution of these matrices is separately row-column exchangeable if for $(g_1, g_2) \in S_m \times S_n$ we have $\Pr((g_1, g_2)X = A) = \Pr(X = A)$ where $(g_1, g_2)(X_{ij}) = X_{g_1(i), g_2(j)}$. While we can think of the group $G = S_m \times S_n$ acting on the two indices of the matrix as described above, a characterization in terms of two separate groups provides greater intuition about the properties of these invariant distributions. Specifically, consider $S_m$ and $S_n$ acting on $X$ sequentially by permuting the rows and columns, respectively. To more clearly see these operations, let

$$\Delta_m(S_n) = \{(g, \ldots, g) : g \in S_n\},$$

be the diagonal subgroup of $S_m^n$ and similarly define $\Delta_n(S_m)$. These two groups are isomorphic to $S_n$ and $S_m$, respectively. The associated group actions are related to whether the group is to act on the rows or the columns of the matrix $X$. For our $m \times n$ matrix we define the action of $\delta_m \in \Delta_n(S_m)$ as $\text{mat}_r(\delta_m \text{vec}(X))$ while the action of $\delta_n \in \Delta_m(S_n)$ is $\text{mat}_c(\delta_n \text{vec}(X^t))$. The operation $\text{vec}(\cdot)$ vectorizes a matrix while $\text{mat}(\cdot)$ reconstructs the matrix via the rows and $\text{mat}_c(\cdot)$ reconstructs the matrix via the columns (Kolda and Bader, 2009). The action permutes the index of the vectorized matrix and then reassembles the matrix. It is easy to see that the group actions commute and that a distribution that is invariant under the combined group action is exactly the separately exchangeable matrix distribution.

In this characterization a relationship between separately row-column exchangeable matrices and partially exchangeable sequences becomes apparent as the vectorized matrix is a partially exchangeable sequence (under a very specific subgroup).

4.3.2. Jointly exchangeable matrices. Now consider a (directed) graph $\mathcal{G}$ with $k$ vertices which can represent a directed network of connections among individuals, companies, cells, etc. This graph is represented by a binary matrix $A$ (the adjacency matrix of the graph) where $a_{ij} = 1$ means that there is a relationship from node $i$ to node $j$ and is 0 otherwise. Define $E = \{(i, j) : \exists$ a relationship from $i$ to $j\}$, a collection of ordered pairs representing the edges in the graph. We consider a finitely exchangeable graph (which is equivalent to a jointly finitely row-column exchangeable matrix) to be one where the index of the vertices does not carry information. Writing this in our notation we have:

The space: $\mathcal{X}_1 = \{A \in \{0, 1\}^{k \times k}\}$

or $\mathcal{X}_2 = \{\text{collection of ordered pairs of integers in } \{1, \ldots, k\}\}$. 


The group: $G = S_k$.

The group action on $X_1$ permutes the row and column indices jointly, while the the action of $G$ on $X_2$ is given by $g \rightarrow (gi, gj)$.

4.3.3. Computing the orbits of row-column exchangeable matrices. Unlike the previous examples, the computation of the orbits for both separately and jointly exchangeable matrices is a complicated task. We can characterize this difficulty by appealing to results from graph theory. First, note that separately row-column exchangeable matrices can be represented by a bipartite graph where each row and each column corresponds to a node. With that in mind, we note that the relative ease with which one finds the orbits of (partially) exchangeable sequences rests in the representation of each orbit in a very simple characterization - e.g. the number of ones for binary sequences. On the other hand, there does not appear to be an easily computable summary of a graph that can identify which orbit it belongs to. In effect such a quantity would be a complete graph invariant as it would be the same for all isomorphic graphs and different for all non-isomorphic graphs. Finding such an invariant that is computable quickly (for example, in polynomial time) would lead to a solution of the graph isomorphism problem, which remains open for general graphs (Babai and Luks, 1983).

4.4. Different symmetries. Scientists frequently rely on different invariance properties of collected data in order to make sound scientific observations. For example, in the study of geometric optics, Lakshminarayanan and Viana (2005) provide an application of the invariance given by the dihedral group. In this example, data is indexed by the dihedral group of order 4 (that is four rotational and four reflection symmetries). Lakshminarayanan and Viana (2005) are interested in testing hypotheses about equality of means of different combinations of their 8 columns of data. It is easy to see that the question is equivalent to establishing whether the distribution of the eight dimensional data is invariant under certain subgroups of the dihedral group. One could construct such a test by estimating the (sub-)group invariant distribution and comparing it to the empirical distribution of the sample. Further description of a testing framework is given in the Discussion.

5. Total variation distance. In this section we derive bounds for the total variation distance between finitely and infinitely exchangeable arrays. These bounds provide a quantitative summary on the difference between finite and infinite exchangeability for arrays, and taking limits in the dimensions recovers the classical Aldous-Hoover representation. Recall that
a distribution of a \( m \times n \) array \( X \) is row-column exchangeable if for all \((g_1, g_2) \in S_m \times S_n\) we have \( \Pr((g_1, g_2)X = A) = \Pr(X = A) \). For infinitely row-column exchangeable arrays, the results of Aldous (1981) and Hoover (1979) provide a representation theorem in the spirit of de Finetti's theorem for infinitely exchangeable sequences. Specifically, if \((X_{ij})\) is infinitely row-column exchangeable then \( X_{ij} = g(\alpha, u_i, v_j, \lambda_{ij}) \) for a measurable function \( g \) and \( \alpha, u_i, v_j, \lambda_{ij} \) all independent uniform(0,1) random variables. When \( g(\cdot, u_i, u_j, \cdot) \) is symmetric in the two middle arguments then we call the array jointly (or weakly) infinitely row-column exchangeable as the distribution of the array \( X \) is now exchangeable under the simultaneous (and identical) permutation of the rows and columns. This representation is frequently used when describing relational data where the object of interest is a square array \( X \).

It is natural to ask how close the finitely exchangeable distributions that we have defined are to the Aldous-Hoover representation of an infinitely row-column exchangeable array. We first demonstrate the result for joint exchangeability using available results for sequences from Diaconis and Freedman (1980). We then extend the results of Diaconis and Freedman (1980) from exchangeable sequences to partially exchangeable sequences and apply these to general infinitely exchangeable arrays.

**Theorem 3** (Diaconis and Freedman (1980) Theorem 13). Let \( P \) be an exchangeable probability on \((S^n, B^n)\). Then there exists a probability \( \mu \) on \((S^*, B^*)\) such that \( \|P_k - P_{\mu k}\| \leq 2\beta(k, n) \) for all \( k \leq n \) where \( \beta(k, n) = 1 - n^{-k}k!/n! \).

The space \( S \) is an abstract measurable space, \( S^* \) is the set of probabilities on \((S, B)\) and is endowed with the smallest appropriate \( \sigma \)-field \( B^* \). To apply this result we will demonstrate how one can consider the row and column exchangeability of an array in terms of exchangeability of sequences taking values in a general space.

**Theorem 4** (Total variation of jointly row-column exchangeable arrays). Finitely joint row-column (\( m \) dimensional) exchangeable (and \( r \) extendible) probabilities are bounded in total variation distance from infinitely joint row-column exchangeable probabilities by \( 2\beta(m, r) \). This bound is sharp.

**Proof.** We appeal to the singular value decomposition of a matrix. It is obvious that if a matrix \( X \) is weakly row-column exchangeable, then we must have that for \( X = UDV^t \) that \( U \) and \( V \) are row-exchangeable with the same permutation. As such we have that for \( g \in S_m \) acting on the space
of square matrices by permuting the row and column indices simultaneously and acting on $U$ and $V$ by permuting the row indices only that $\Pr(gX = A) = \Pr(X = A)$ and $\Pr(gU = U_A, gV = V_A, D = D) = \Pr(U = U_A, V = V_A, D = D)$ are equivalent (and $\Pr(X = A) = \Pr(U = A_u, V = A_v, D = D)$). To make this even more explicit let $[U; V]$ be the concatenated matrix with $m$ rows and $2m$ columns of the matrices $U$ and $V$. As such the joint row-column exchangeability simply means $[U; V]$ is row exchangeable. By letting the rows of $[U; V]$ define the state space of Theorem 3, the result follows.

**Remark 2.** The theorem replaces the discussion of the joint row-column exchangeability of a distribution on square binary matrices to the discussion of the joint row exchangeability of orthogonal matrices. As such, to reflect the notation of Theorem 3 we say that $S = S_1 \times S_2$ is the collection of concatenated rows of orthogonal matrices of all dimensions up to dimension $r$. The finitely exchangeable distributions on $S^m$ that correspond to jointly row-column distribution exchangeable distributions on binary square matrices have two constraints: (i) there is zero probability that a sequence from $S^m$ is not the concatenation of two orthogonal matrices, and (ii) the inner product of each jointly selected elements from $S_1$ and $S_2$ with respect to a diagonal matrix $D$ with ordered positive entries is either 1 or 0. As such, each jointly row-column exchangeable distribution has a counterpart in terms of an exchangeable distribution on $S^m$, but the reverse is not true.

To formulate this result for separately row-column exchangeable arrays we must first prove an extension of Theorem 3:

**Theorem 5.** Total variation bounds for partial exchangeability Let $P$ be a partially exchangeable probability on $(S_1^{n_1} \times S_2^{n_2}, B_1^{n_1} \times B_2^{n_2})$. Then there exists a probability $\mu$ on $(S_1^{n_1} \times S_2^{n_2}, B_1^{n_1} \times B_2^{n_2})$ such that $\|P_{k_1, k_2} - P_{\mu k_1, k_2}\| \leq 2\beta((k_1, k_2), (n_1, n_2))$ for all $k_1 \leq n_1, k_2 \leq n_2$ where $\beta((k_1, k_2), (n_1, n_2)) = 1 - \frac{k_1!}{(n_1 - k_1)!} \frac{n_2 - k_2!}{(n_2 - k_2)!}$. This bound is tight.

**Proof.** The proof is the same as the proof of Theorem 3 after making the obvious substitutions. Specifically, what was previously the urn $U(\omega)$ becomes the collection of urns with $\omega \in S_1^{n_1} \times S_2^{n_2}$. The distributions are now sampling with and without replacement from this collection of urns and any finitely partially exchangeable distributions will be a mixture of the distributions of sampling without replacement from these urns while the projection of infinitely partially exchangeable distributions will be a mixture of the distributions of sampling with replacement from these urns.
The bound and the tightness of the bound are given by the technical lemmas in the Appendix.

The theorem trivially extends to dimensions greater than 2.

**Theorem 6 (Total variation bounds).** Let $P_{r,q}^{m,n}$ be a distribution on $m \times n$ dimensional matrices that is finitely row-column exchangeable and is extendible to a row-column exchangeable distribution on $r \times q$ dimensional matrices. Let $P_{\mu,m,n}$ be the projection of an infinitely row-column exchangeable distribution onto the space of $m \times n$ dimensional matrices. Then for $\| \cdot \|$ representing the total variation norm we have

$$\|P_{r,q}^{m,n} - P_{\mu,m,n}\| \leq \beta\big((m, n), (r, q)\big).$$

The bound is tight.

**Proof.** The proof is similar to that of Theorem 4. The singular value of the matrix $X = UDV^t$ is again considered, but the separate exchangeability of the rows and columns implies that the rows of the left and right eigenvector matrices are separately permuted. Concatenating $U$ and $V$ along the rows into $Z = [U^t; V^t]^t$ forming a $m + n$ row matrix we see that separately row column exchangeability of $X$ is equivalent to the partial exchangeability of the rows of $Z$. The results then follows from Theorem 5.

The contribution of Theorem 6 is two-fold. As far as we know, this is the first time a form for the total variation bound for row-column exchangeability has been proposed in explicit form. More importantly, we now see that the rate of convergence of the total variation bound is bounded by $(m^2 - m)/q + (n^2 - n)/r$ for matrices, similar to the results of Diaconis and Freedman (1980) for sequences. Knowing this bound provides a justification for the assumption of infinite row-column exchangeability (allowing for an Aldous-Hoover type of latent variable representation) in high-dimensional statistics when both the subjects and features are not independent but can assumed to be exchangeable and network analysis where the labels of the nodes do not carry information.

**6. Signed measures.** Here we consider the introduction of signed measures into the representation theorems of de Finetti and Aldous-Hoover. In the case of finitely exchangeable sequences Paul Cartier originally observed that de Finetti’s theorem holds for finite sequences when the de Finetti measure is allowed to be signed. An outline of this result was provided in
Dellacherie and Meyer (1978). In recent work Kerns and Székely (2006) explored these results further and provided a succinct proof for abstract finite sequences. It is natural to consider extending these results to a more general setting.

Currently available in the literature are characterization theorems for infinitely partially exchangeable sequences and infinitely row-column exchangeable arrays. These theorems provide a representation of infinite exchangeability via mixtures of independent and identically distributed random variables. Since the extension to signed measures only makes sense in the presence of such a representation theorem, we restrict ourselves to partial and row-column exchangeability rather than the general group invariant setting. Despite our restricted focus here, it is important to note that infinite extendibility for other forms of invariance are likely to be useful in light of the example in Section 4.4. In that example the infinite extendibility of a dihedral invariant sequence would represent an object that is invariant under all actions on the unit circle and suggests a relationship to orthogonally invariant distributions. Since the application relates to geometry for optics, this appears to be an invariance of interest.

To only state a single general theorem for all forms of exchangeability we let $G$-exchangeable be a placeholder for a directed product of finitely many symmetric groups, where the action of $G$ is defined in accordance with the space of interest. For example the direct product of two symmetric groups can either act on the indices of a sequence or on the indices of a matrix - this covers both partial exchangeability and row-column exchangeability.

**Theorem 7 (Signed measure infinitely $G$-exchangeable representation).** If there exists a representation theorem for infinitely $G$-exchangeable random variables then let the space of infinite $G$-exchangeable distributions be $\mathcal{M}^*$ and let the projection to finite dimensions be $\mathcal{M}^{\text{proj}}$. For any choice of basis elements $\{m\}$ that span $\mathcal{M}^{\text{proj}}$, there exists an invertible linear map from the extreme points of Theorem 1 to $\{m\}$.

The proof of this theorem is constructive and is identical to that for the special case of exchangeable sequences presented in Kerns and Székely (2006).

This general result’s value is suggested by the specialized case of finitely exchangeable sequences in Kerns and Székely (2006). Specifically, they are able to relax the assumption of an infinitely exchangeable sequence for proving a posteriori Bayesian consistency (as in Berk (1970) and Bernardo and Smith (2009)) to only requiring the assumption of finite exchangeability. We postulate that in general theorems that similarly rely on a representation theorem
(that is some assumption of infinite $G$-exchangeability) so as to decompose
the likelihood can have that assumption relaxed due to Theorem 7.

7. Discussion. In this paper we presented a unifying approach for char-
acterizing finite $G$-invariant distributions. The Theorems of Section 3 pro-
vide the machinery for constructing distributions that are invariant under
any finite group operation, as well as checking whether a distribution is
invariant under a larger group. The results of Section 5 answer an open
question from Aldous (1985) with regards to a bound on the total variation
distance between the distributions of a finitely exchangeable array and an
infinitely exchangeable array.

Addressing the original motivation behind the paper, the tight bound for
the jointly row-column exchangeable distributions provides an insight into
the efficacy of the assumption of existence of an Aldous-Hoover represen-
tation when analyzing relational data. In particular, it is likely that when
the assumption of infinitely exchangeable nodes in a network is not satisfied
(such as with students in a classroom or countries in the world), inference
based on models that use this assumption will be approximately correct.
Similar results for sequences were previously used to justify the use of de
Finetti’s theorem for Bayesian causal inference (Rubin, 1978). The magni-
tude of the deviation of parameter estimates for the finite exchangeability-
infinit e exchangeability paradigm is a part of ongoing research by the au-
thors. We conjecture that any estimator that has a faster rate of convergence
than the total variation bounds should have similar properties whether one
assumes finite or infinite exchangeability.

The results of this paper also provide an intuition for testing for differ-
ent types of invariance. In Section 4.4 we alluded to a testing framework
for finite invariance that we describe in more detail here. For a random
variable $X$ distributed according to $P$ we are interested in testing the null
hypothesis that $P$ is invariant under a finite group $G$ against the hypothesis
that $P$ is not invariant under $G$. A test based on $T = \|P - P_G\|$ where $P$
is the unrestricted empirical distribution, $P_G$ is the finite group invariance
restricted empirical distribution, and $\| \cdot \|$ is total variation distance rejects
the null for large values of $T$. To construct a test for infinite exchangeability,
one can employ the projection criterion results of Theorem 2. Specifically,
one can test the null of infinite extendibility versus an alternative of finite
extendibility of a finitely exchangeable distribution using the test statistic
$T = \arg\max_{|H|} \{H : \exists P_G|H \text{ st } P_G \neq P_G|H\} \text{ where } P_G$ is the finitely ex-
changeable restricted empirical distribution. The test will reject for $T$ small.
For both tests the null distributions can be constructed via Monte Carlo
simulation.

APPENDIX A: TECHNICAL LEMMAS

To prove Theorem 5 we require an extension of the results of Freedman (1977) to problems with multiple urns. We follow Freedman’s notation and method. In particular, consider a collection of $d$ urns $F_1, \ldots, F_d$ where $F_i$ contains elements $\{f_{i1}, \ldots, f_{im_i}\}$. Let $F_i^{k_i}$ consist of the $k_i$-tuples of elements of $F_i$. Sampling with replacement from the collection of $d$ urns induces a uniform probability $M$ on $F_1^{k_1} \times \cdots \times F_d^{k_d}$ where

$$M(\{s_{11}, \ldots, s_{1k_1}\}, \ldots, \{s_{d1}, \ldots, s_{dk_d}\}) = \frac{1}{n_1 \cdots n_d}$$

Letting $G$ be the vectors $\{s_{11}, \ldots, s_{1k_1}\}, \ldots, \{s_{d1}, \ldots, s_{dk_d}\} \in F_1^{k_1} \times \cdots \times F_d^{k_d}$ for which all the components are unequal, sampling without replacement is given by

$$Q(\{s_{11}, \ldots, s_{1k_1}\}, \ldots, \{s_{d1}, \ldots, s_{dk_d}\}) = \frac{(n_1 - k_1)! \cdots (n_d - k_d)!}{n_1! \cdots n_d!}.$$

Lemma 8 (Sampling with and without replacement). The total variation distance between sampling with and without replacement from multiple urns is

$$\|M - Q\| = 1 - \frac{n_1! \cdots n_d!}{(n_1 - k_1)! \cdots (n_d - k_d)! n_1^{k_1} \cdots n_d^{k_d}}.$$

Proof. The proof is by construction. We note that $Q(G) = 1$ and the total variation norm is given by $Q(G) - M(G)$. To calculate $M(G)$ we note that the probability of any particular sequence is equal under $M$ and so we simply need to count the number of sequences in $G$. This is an easy combinatorial exercise and each urn has $\frac{n_i^{k_i}}{(n_i - k_i)!}$ sequences of length $k_i$ with unique entries. Thus $M(G) = \frac{n_1! \cdots n_d!}{(n_1 - k_1)! \cdots (n_d - k_d)! n_1^{k_1} \cdots n_d^{k_d}}$ and the lemma follows. \[\square\]

Lemma 9 (Tight bounds). The bound in Theorem 5 is tight. That is

$$\|P_{n_1, n_2, k_1, k_2} - P_{\mu, k_1, k_2}\| \geq \|P_{n_1, n_2, k_1, k_2} - M\| = 1 - \frac{n_1! n_2!}{(n_1 - k_1)!(n_2 - k_2)! n_1^{k_1} n_2^{k_2}}$$

where $M$ is as in Lemma 8.
Proof. Without loss of generality we let the distributions of interest be over two urns with balls labeled $1^1, \ldots, n_1^1$ and $1^2, \ldots, n_2^2$ respectively. We note that the equality in the statement is given by Lemma 8. Thus as in the original proof of tightness in Diaconis and Freedman (1980) we must show that for $G$ as defined in Lemma 8 we have $(p_{1}^{k_1}p_{2}^{k_2})(G) \leq M(G)$ for any distributions $p_1, p_2$ on the the sets $\{1^1, \ldots, n_1^1\}$ and $\{1^2, \ldots, n_2^2\}$ respectively. To see that this is true we note $M$ is a product of two pure power probabilities and that we can write the set $G = G_1 \times G_2$ where $G_1$ are all the sets of length $k_1$ of unique elements in the first urn and similarly for $G_2$. As in Diaconis and Freedman (1980) we note that the Schur convexity of the indicator functions $1_{G_1}$ and $1_{G_2}$ implies that $p_{i}^{k_i}(G_i) \leq M_i(G_i)$. This implies that $P_{\mu,k_1,k_2}(G) \leq M(G)$ as desired.

REFERENCES

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems, pages 33–40.

Aldous, D. (1985). Exchangeability and related topics. École d’Été de Probabilités de Saint-Flour XIII?1983, pages 1–198.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. Journal of Multivariate Analysis, 11(4):581–598.

Austin, T. and Panchenko, D. (2013). A hierarchical version of the de finetti and aldous-hoover representations. Probability Theory and Related Fields, pages 1–15.

Babai, L. and Luks, E. M. (1983). Canonical labeling of graphs. In Proceedings of the fifteenth annual ACM symposium on Theory of computing, pages 171–183. ACM.

Berk, R. H. (1970). Consistency a posteriori. The Annals of Mathematical Statistics, pages 894–906.

Bernardo, J. M. and Smith, A. F. (2009). Bayesian theory, volume 405. John Wiley & Sons.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and newman-girvan and other modularities. Proceedings of the National Academy of Sciences, 106(50):21068–21073.

Bonassi, F. V., Stern, R. B., Wechsler, S., and Peixoto, C. M. (2014). Exchangeability and the law of maturity. arXiv:1404.6572.

De Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio.

De Finetti, B. (1972). Probability, induction, and statistics.

de Laplace, P. S. (1820). Théorie analytique des probabilités. Courcier.

De Moivre, A. (1756). The doctrine of chances: or, A method of calculating the probabilities of events in play, volume 1. Chelsea Publishing Company.

Dellacherie, C. and Meyer, P.-A. (1978). Probability and potential. Paris: Hermann.

Di Cecco, D. (2009). On the extendibility of partially and markov exchangeable binary sequences. arXiv preprint arXiv:0908.4158.

Diaconis, P. (1977). Finite forms of de finetti’s theorem on exchangeability. Synthese, 36(2):271–281.

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. The Annals of Probability, pages 745–764.
Diaconis, P. and Janson, S. (2007). Graph limits and exchangeable random graphs. *arXiv preprint arXiv:0712.2749*.

Dixit, N. M., Srivastava, P., and Vishnoi, N. K. (2012). A finite population model of molecular evolution: Theory and computation. *Journal of Computational Biology*, 19(10):1176–1202.

Eaton, M. L. (1989). Group invariance applications in statistics. In *Regional conference series in Probability and Statistics*, pages i–133. JSTOR.

Freedman, D. (1977). A remark on the difference between sampling with and without replacement. *Journal of the American Statistical Association*, 72(359):681–681.

Haag, J. (1924). Sur un problème général de probabilités et ses diverses applications. In *Proceedings of the International Congress of Mathematicians, Toronto*, volume 1, pages 659–674.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, pages 470–501.

Hoff, P., Fosdick, B., Volfovsky, A., and Stovel, K. (2013). Likelihoods for fixed rank nomination networks. *Network Science*, 1(03):253–277.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Hoover, D. (1989). Tail fields of partially exchangeable arrays. *Journal of Multivariate Analysis*, 31(1):160–163.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*, 2.

Hulpke, A. (2010). Notes on computational group theory. Technical report, Colorado State University.

Kerns, G. J. and Székely, G. J. (2006). Definetti's theorem for abstract finite exchangeable sequences. *Journal of Theoretical Probability*, 19(3):589–608.

Kolod, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3):455–500.

Lakshminarayanan, V. and Viana, M. (2005). Dihedral representations and statistical geometric optics. i. spherocylindrical lenses. *JOSA A*, 22(11):2483–2489.

Rotman, J. J. (1995). *An introduction to the theory of groups*, volume 148. Springer.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, pages 34–58.

Sampson, S. F. (1969). *Crisis in a cloister*. PhD thesis, Ph. D. Thesis. Cornell University, Ithaca.

Seress, Á. (2003). *Permutation group algorithms*, volume 152. Cambridge University Press.

Volfovsky, A. and Hoff, P. D. (2013). Testing for nodal dependence in relational data matrices. *arXiv preprint arXiv:1306.5786*.