Finite Exchangeability, Ergodic Decomposition and Exponential Families

Vince Velkey, Peter Orbanz

June 24, 2025

Abstract

We establish a connection between exchangeability and maximum entropy distributions by studying the mixing measure in de Finetti's theorem. The main technical tool borrowed is the convex core of a measure, originally invented to study kl-projections in information theory. Along the way, we recover a representation result for finite exchangeable sequences and its relation in the infinite length limit to a standard proof of de Finetti's theorem. Our method can be applied for more general symmetric structures, we discuss an abstract proof recipe. We apply the method for finite exchangeable graphs, using results from the theory of graph limits. Using the convex core, we draw the analogous connection to maximal entropy distributions, which are the exponential random graph distributions. We collect the many applications of the method in one joint framework by studying the action of a direct limit of compact groups on a compact metric space.

Keywords: finite exchangeability; maximum entropy distributions; kl-projections; de Finetti's theorem, ergodic decomposition

1 Introduction

Throughout, we fix a probability space $(\Omega, \mathbb{P})$. A random sequence $\mathbf{Y} = (Y_1(\omega), Y_2(\omega), \dots)$ with values in a measurable space $(S, \mathcal{S})$ is said to be exchangeable if its distribution $\mathcal{L}_{\mathbf{Y}}$ is invariant under finite permutations of its elements. The remarkable theorem of de Finetti, see [Kal05] Theorem 11.10, characterises such distributions (as long as S as is a Polish space) as mixtures of laws of independent and identically distributed sequences (i.i.d.). That is, the random infinite sequence $\mathbf{Y}$ is exchangeable if and only if there exists a probability measure μ on the space $\mathcal{P}(S)$ of probability measures on S such that

$$\mathbb{P}(\mathbf{Y} \in A) = \int_{\mathcal{P}(S)} \theta^n(A)\, \mu(d\theta), \quad A \in \mathcal{S}^n. \tag{1}$$

In this case, μ is necessarily unique.

This result also has an important sampling perspective. To sample an instance from the exchangeable law $\mathcal{L}_{\mathbf{Y}}$, we first need to sample a latent random measure $\Theta \sim \mu$, and then $Y_{1:n}|\Theta$ from Θ^∞, so $\mathbf{Y}$ is conditionally i.i.d. Then, using a law of large numbers argument on the samples, we can recover de Finetti's law of large numbers:

$$\lim_{n\to\infty} \frac{1}{n} \sum_i \delta_{Y_i(\omega)}(\cdot) \xrightarrow{\text{weakly}} \Theta(\omega) \qquad \mu - \text{ almost surely.} \tag{2}$$

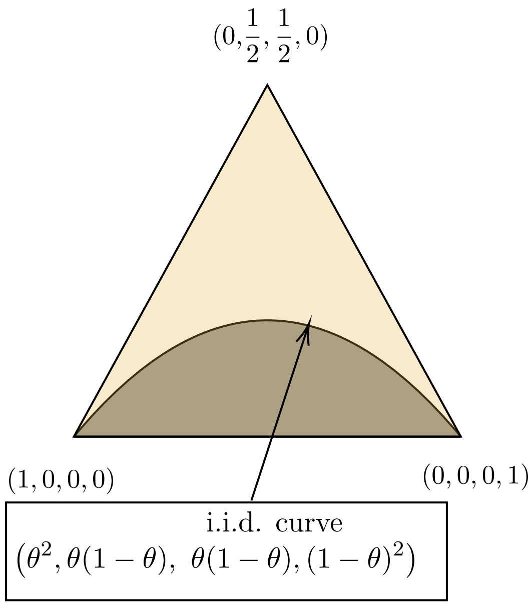

Figure 1: The exchangeable simplex

Note in particular, that exchangeable laws $\mathcal{L}$ on S^∞ and probability measures μ on $\mathcal{P}(S)$ are in a one-to-one correspondence. We write P_μ for the exchangeable law corresponding to the mixing measure μ. Also, by the de Finetti's law of large numbers, we can recover, at least in theory, the law of μ from empirical averages of samples from P_μ, according to the two-step sampling procedure above. For a given n, there are numerous infinite exchangeable sequences with the same n-dimensional marginal, so the law of $(Y_1, \ldots Y_n)$ does not specify the mixing measure μ. That is, the mapping $\mu \to P_\mu^n$ where P_μ^n stands for the n-dimensional marginal of P_μ, is not injective. In particular, there is no hope to recover μ from samples from this finite marginal only. One of the focuses of this work is this non-recoverability phenomenon, i.e. to study the set $\mathcal{R}_{\mathbf{Y}}^{n,+}$ of possible mixing measures for the marginal $(Y_1, \ldots Y_n)$. Our result, **Theorem 6.1**, which applies in the classical case when the state space S is binary (or finite), says that it is in some sense as large as possible: for any measure μ that satisfies some neccessary support conditions, there is a possible mixing measure $\mu_0 \in \mathcal{R}_{\mathbf{Y}}^{n,+}$ that is a member of an exponential family with base measure μ. To obtain this and related results, we utilise a connection with the theory of information projections as outlined in [Csi75], [CM01], [CM03] and other works. Indeed, $\mu_0 \in \mathcal{R}_{\mathbf{Y}}^{n,+}$ for a given μ will be taken as its kl−projection on $\mathcal{R}_{\mathbf{Y}}^{n,+}$.

Such finite dimensional marginals of infinite exchangeable sequences are not the only random finite sequences that are exchangeable. Indeed, there are finite exchangeable laws that cannot be written as mixtures of product laws [Dia77]. We recall the intuitive geometric viewpoint of [Dia77] where exchangeable distributions over length-2 binary sequences are depicted in barycentric coordinates representing the probabilities of the sequences $00, 01, 10, 11$ (Figure 1). The dark area, which is the convex hull of the curve of all product laws, corresponds to the set of mixture sequences $\mathcal{E}_\infty^2$. **Theorem 6.1** can be intuitively understood in this geometric framework, which we will discuss at the end of this section.

There are two main approaches in the literature to study finite exchangeability. In the more traditional approach, e.g. [Dia77], [DF80], and further [GK21] the authors study the set $\mathcal{E}_m^n(S)$ of finite exchangeable distributions on n variables that are initial segments of

m-long exchangeable sequences. Such distributions are called m-extendible. They derive approximations in terms of n and m on some notions of distance to the set $\mathcal{E}_\infty^n(S)$ of infinitely extendible sequences, which is precisely the set of distributions in n variables that are mixtures of i.i.d ones [KY18] (i.e. for which the non-recoverability problem above applies). This approach then yields approximate equations of the form (1) for extendible finite exchangeable distributions.

The other approach [Res85] [KS06] [JKY16], see also [KY18] provides an exact integral representation via a signed mixing measure, which is less interpretable [Dia23], but still useful in statistical applications [BRW14], [KS06]. That is, for a finite exchangeable sequence $(X_1, \dots X_n)$ with $X_i \in S$, there is a signed measure ν on $\mathcal{P}(S)$ such that

$$\mathbb{P}(\mathbf{Y} \in A) = \int\limits_{\mathcal{P}(S)} \pi^n(A)\, \nu(d\pi), \quad A \in \mathcal{S}^n. \tag{3}$$

There are many more random structures beyond sequences that are invariant under permutations of some index set and exhibit a de Finetti-style representation. The theory of Aldous, Hoover and Kallenberg (e.g. [Ald81], [Hoo79], [Kal89] and further [Kal05] and references therein) of exchangeable arrays provides a general treatment of many of these.

The theory of graph limits (e.g. [LS06], [BCL$^+$06], [BCL$^+$08] and further [Lov12] and references therein) gives a parallel perspective to the Aldous-Hoover-Kallenberg theory [DJ08] [Aus08]. Based on this connection, several parallel exchangeability and limit theories have been developed for other combinatorial objects [Jan11a], like bipartite graphs, hypergraphs, posets, permuatations, trees etc, see [LS10], [Jan11b], [ES12], [HKM$^+$13], [Stu21], [ET22]. Most of these results are largely analogous, but a unified perspective seems still to be missing [Jan11a].

Finite exchangeability seems to not have gained much attention in either perspectives of the theory, with the exception of [BS98], [Mat95], [SBC00] and more recently [Sad20] [Leo18] in the spirit of the first and second approaches respectively.

In the first part of the paper, **Sections 2-5** we develop a functional analytic framework where we can treat many of these exchangeable objects, both infinite and finite, together. The framework includes the second approach to finite exchangeability via signed mixtures. We first identify a proof recipe for integral representation results of the form (1) and (3) in **Section 2** by considering exchangeable binary sequences. In **Section 3**, we apply the proof recipe for exchangeable graphs, using results from the theory of graph limits. We deduce a signed measure representation for graphs, **Theorem 3.8**, analogous to (3). In **Section 4** we show some further applications of the proof recipe to deduce de Finetti theorems for other infinite exchangeable objects and the corresponding signed measure representation for their finite counterpart. Then in **Section 5** we treat these structures jointly by studying the more abstract setup of the action of a direct limit $\mathbb{G}$ of compact groups on a compact metric space Z, and $\mathbb{G}$-invariant distributions on Z. An abstract integral representation is immediately available from the ergodic decomposition theorem, which writes $\mathbb{G}$-invariant distributions as mixtures of ergodic ones. The de Finetti integral representations satisfy further important properties, like the weak compactness of the set of ergodic distributions, or the laws of large numbers. We identify structural properties of the action that imply these properties and hold for the actions of the symmetric group corresponding to the various exchangeability notions discussed.

In the second part, starting with **Section 6**, we discuss the non-recoverability of infinitely extendible sequences via the connection to information projections. **Section 6** is mostly self-contained, the framework of the first part motivates the connection and gives a way to deduce similar results for exchangeable graphs and more. Tha graph case is discussed in **Section 7**, where the exponential family mixture densities take the form of exponential random graph distributions [CD13].

In the rest of this section, we discuss in more detail the geometric intepretation of finite exchangeability following [Dia77] and [KS06]. This provides useful intuition throughout and especially in the second part.

1.1 Geometry of finite exchangeable binary sequences

Consider the space $\mathcal{P}_n^2$ of probability distributions on n-long binary sequences. This can be represented by the $2^n - 1$ dimensional simplex Δ^{2^n} embedded in 2^n-dimensional Euclidean space The coordinate vector $\mathbf{p} = (p_0, p_1, ...p_{2^n-1})$ corresponds to the distribution $D_\mathbf{p}$ where p_j is the probability under $D_\mathbf{p}$ of the n-long sequence given by the binary representation of j. That is, if $\mathbf{Z} = (Z_1, \ldots Z_n) \sim D_\mathbf{p}$ and $j = a_1 \cdot 2^0 + a_1 2^1 + \ldots a_n 2^{n-1}$, then

$$p_j = \mathbb{P}(Z_1 = a_1, \ldots Z_n = a_n).$$

Let $\mathcal{E}_n^2 \subset \mathcal{P}_n^2$ be the convex subset of exchangeable distributions. For $0 \leq k \leq n$, let Ω_n^k be the set of sequences with k many coordinates taking value 1. By definition, a distribution is exchangeable, if and only if it assigns the same mass to each element of Ω_n^k. It is then observed in [KS06] that the uniform distributions $\boldsymbol{h}_k$ on Ω_n^k (the so-called urn distributions) are the extreme points of $\mathcal{E}_n^k$ and are linearly independent. It then follows that $\mathcal{E}_n^2$ is also a simplex, of dimension n.

In the case $n = 2$, the simplex $\mathcal{P}_2^2$ is a tetrahedron and $\mathcal{E}_2^2$ is a triangle inside it, see Figure 2. Here e.g. the coordinate vector $(0, 0, 0, 1)$ stands for the distribution that assigns mass 1 to the sequence $(1, 1)$, whilst $(0, \frac{1}{2}, \frac{1}{2}, 0)$ stands for the one that assign equal mass of $\frac{1}{2}$ to the sequences $(0, 1)$ and $(1, 0)$.

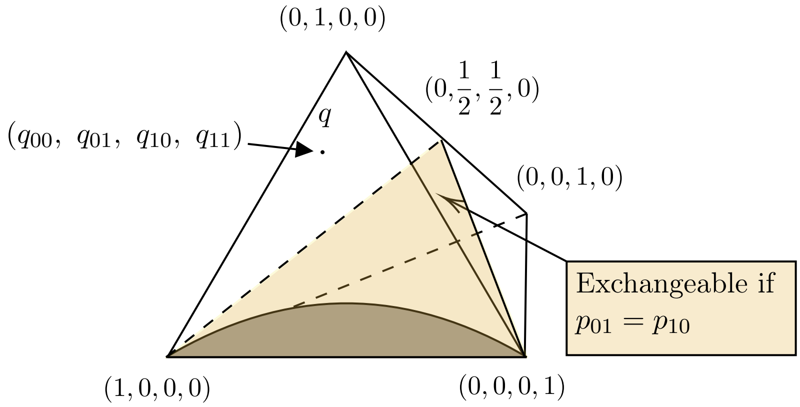

Figure 2: Simplex of distributions on sequences

Since each sequence in Ω_n^k receives the same mass under exchangeable distributions, it is convenient to project the coordinates corresponding to these sequences to a single coordinate and thus embed the exchangeable simplex directly in $\mathbb{R}^{n+1}$. In the $n = 2$ case this corresponds to the projection $(q_{00}, q_{01}, q_{10}, q_{11}) \to (q_{00}, q_{01}, q_{11})$. This convention is followed from now on.

In the exchangeable simplex, the i.i.d. laws can be parametrised by the function

$$\mathbf{c}^n \colon [0, 1] \to \mathcal{E}_n^2 \subset \mathbb{R}^{n+1}$$
$$\mathbf{c}^n(\theta) = (\theta^n, \theta^{n-1}(1-\theta), \ldots, (1-\theta)^n),$$

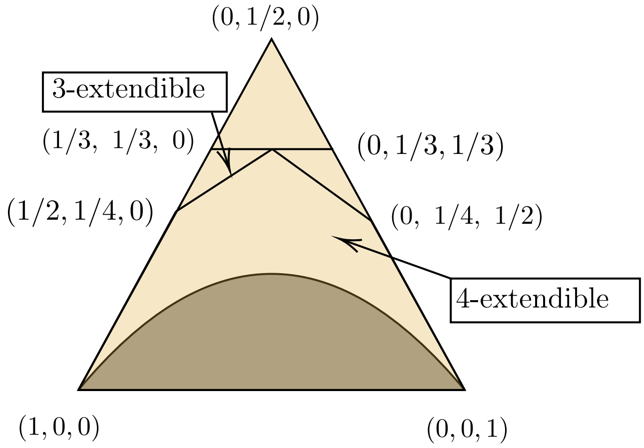

Figure 3: Extendible finite exchangeable distributions

which we call the i.i.d. curve. For $n = 2$ this is the parabola in Figure 1 and 2 above.

For an finite exchangeable sequence $\mathbf{X}$, we write $p_{\mathbf{X}}$ for the corresponding vector of coordinates in the exchangeable simples and call it the law vector of $\mathbf{X}$. Note then that for a finite exchangeable sequence with coordinate vector $p_{\mathbf{X}}$, equation(3) can be summarised for the binary case in this parametrisation by the equation

$$p_{\mathbf{X}} = \int_{[0,1]} \mathbf{c}^n(\theta)\,\nu(d\theta), \tag{4}$$

where ν is a signed measure on $[0,1]$. If $\mathbf{X}$ is infinitely extendible, de Finetti's theorem yields the same equation $p_{\mathbf{X}} = \int_{[0,1]} \mathbf{c}^n(\theta)\,\mu(d\theta)$ with a (positive) probability measure μ. Such integrals against a probability measures are generalised convex combinations in the sense that we have $p_{\mathbf{X}} \in \overline{\mathrm{conv}}(\mathbf{c}^n([0,1]))$, where $\overline{\mathrm{conv}}$ stands for the closed convex hull of a set. Note however, that $\mathbf{c}^n([0,1])$ is a compact subset of $\mathbb{R}^{n+1}$ and so its convex hull is again compact. So it suffices to write $\mathrm{conv}(\mathbf{c}^n([0,1])$. We write $\mathcal{C}^n$ for this area, which is the dark area under the i.i.d. curve in Figure 1 and 2 in the case $n = 2$. The results of [Dia77], [DF80], [KY18], [GK21] and many more is that the set of m-extendible exchangeable distributions approach the i.i.d. curve in various state spaces S and disctance notions, like total variation [DF80] or the kl-divergence [GK21]. In the exchangeable triangle (which again corresponds to exchangeable distributions on length-2 binary sequences), the sets of 3- and 4-extendible distributions are depicted in Figure 3, see [Dia77] for further discussion.

We now turn to stating our result **Theorem 6.1** in this context. For a measure ν on $[0,1]$, we write s_μ for its support and $\mathrm{int}(A)$ for the interior of a Borel set A. Note again that the value of $\int_{[0,1]} \mathbf{c}^n(\theta)\,\mu(d\theta)$ lies in the convex set $\mathcal{C}^n_\mu = \mathrm{conv}(\mathbf{c}^n(s_\mu)) \subset \mathcal{C}^n$. Figure 4 illustrates this set in the case where μ is supported on two disjoint intervals.

It is then apparent that $p_{\mathbf{X}} \in \mathcal{C}^n_\mu$ is obviously neccessary for μ to satisfy (4) and thus be a possible mixing measure for $\mathbf{X}$. The content of **Theorem 6.1** is that this is in a sense enough, up to an exponential family with base measure ν. We write $\mathcal{E}_{\mu,f}$ for the exponential family with base measure μ and sufficient statistic f, i.e.

$$\mathcal{E}_{\mu,f} = \left\{Q_\theta \colon \frac{dQ_\theta}{d\mu}(x) = e^{\langle\theta, f(x)\rangle - \Lambda_f(\theta)},\ \theta \in \mathrm{dom}(\Lambda_f)\right\}, \tag{5}$$

where $\Lambda_f(\theta) = \int_X e^{\langle\theta, f(x)\rangle} d\mu$ and $\mathrm{dom}(\Lambda_f) = \{\theta \colon \Lambda_f(\theta) < \infty\}$.

Theorem. Let $\mathbf{X}$ be a finite exchangeable distribution and μ be a measure on $[0,1]$ with infinite support. Then $\mathrm{int}(\mathcal{C}^n)$ is not empty and there is a $\mu_0 \in \mathcal{E}_{\nu,\mathbf{c}^n}$ that is a possible mixing measure for $\mathbf{X}$ if and only if

$$p_{\mathbf{X}} \in \operatorname{int}(\mathcal{C}^n_\mu). \tag{6}$$

This result is thus almost optimal in the sense that provided μ satisfies the strictly neccessary support condition $p_{\mathbf{X}} \in \mathcal{C}^n_\mu$, strengthened by taking the interior, there is an element in its exponential family that is a possible mixing measure for $\mathbf{X}$. So in this sense, the set of possible mixing measures is 'as large as possible'.

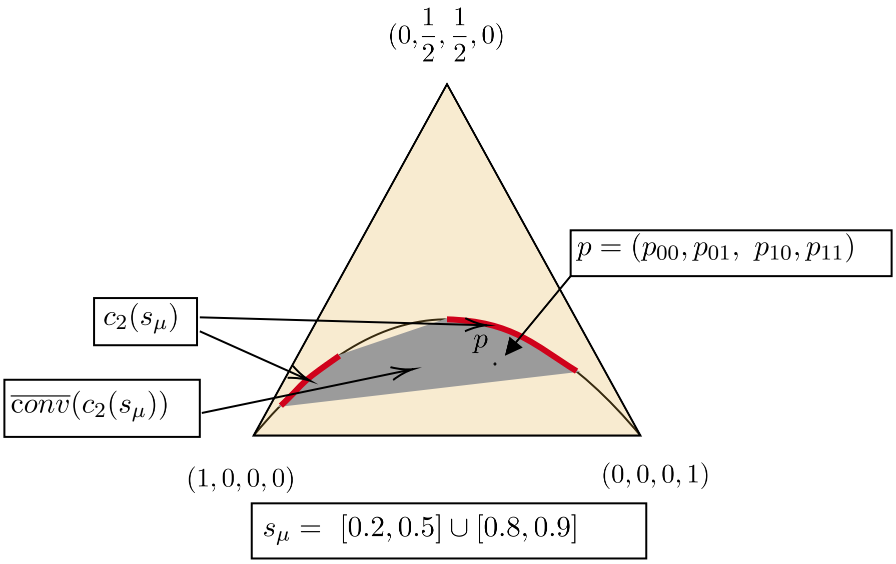

Figure 4: Theorem 6.1 support condition

All of the statements in this subsection have straightforward equivalents for sequences with values in a finite state space S, but a more cumbersome notation is needed [KS06]. The intuition also carries over to more general S as well, but the infinite dimensional nature of the simplices result in some technical difficulties. In particular, **Theorem 6.1** does not cover this case as we crucially need that the function $\mathbf{c}^n$ has finite number of components.

2 Binary sequences

In this section, we develop our method for proving the representation result of [KS06] for finite exchangeable sequences. We start with binary sequences and identify the key patterns of the proof that we adapt to other settings in the next sections. We first racall some standard results and definitions in the form we will refer to them.

Definition 2.1. The *total variation* norm of a signed measure λ on the measureable space $(S, \mathcal{S})$ is defined by

$$\|\lambda\|_{TV} = \lambda^+(S) + \lambda^-(S), \tag{7}$$

where λ has the Jordan decomposition $\lambda = \lambda^+ - \lambda^-$.

Theorem (Riesz Representation theorem)**.** Let $C(K)$ be the vector space of real valued continuous functions on the compact Hausdorff topological space K. Then for each linear functional $\phi \in C(K)^*$, there is a unique signed measure λ on $(K, \mathcal{B}(K))$ such that

$$\int_K f(x)\,\lambda(dx) = \phi(f)$$

for each $f \in C(K)$. Moreover,

$$\|\phi\|^* = \|\lambda\|_{TV}.$$

Thus, the space $(\mathcal{M}(S), \|\cdot\|_{TV})$ of signed measures on S in the total variation norm and $(C(K)^*, \|\cdot\|^*)$ are isometrically isomorphic.

Theorem (Hahn-Banach Theorem)**.** Let X be a real normed vector space and $Y \leq X$ a subspace. If $\phi \in Y^*$ is a linear functional on Y, there is $\varphi \in X^*$ with $\|\varphi\|^* = \|\phi\|^*$ such that

$$\varphi|_Y = \phi.$$

Theorem (Stone-Weierstrass Theorem)**.** Let K be a compact Hausdorff space and $\mathcal{A}$ be the subalgebra of $C(K)$ that contains a non-zero constant function. Then $\mathcal{A}$ is dense if and only if it seperates the points of K.

2.1 Exchangeable binary sequences

Let $\mathbf{Y} = (Y_1, Y_2, \dots)$ be an infinite exchangeable binary sequence. Let

$$p_k^m = \mathbb{P}(Y_1 = 0, \dots Y_k = 0, Y_{k+1} = 1 \dots Y_m = 1). \tag{8}$$

We call $(p_k^m)_{m \in \mathbb{N}, k \leq m}$ the *law sequence* of $\mathbf{Y}$ and $(p_k^m)_{k \leq m}$ its m-th level. Note that a law-sequence uniquely specifies a binary exchangeable distribution.

The parametrisation of a Bernoulli random variable Z by $P(Z = 0) = \theta$ with $\theta \in [0, 1]$ leads to a natural parametrisation of mixtures of iid sequences by integrals on $[0, 1]$ in the usual Borel σ-algebra. In this parametrisation de Finetti's theorem translates to

Fact 2.2 (de Finetti)**.** An infinite binary sequence $\mathbf{Y}$ is exchangeable, if and only if there exists a unique probability measure μ on $[0, 1]$, such that for all $m \in \mathbb{N}$ and $k \leq m$,

$$p_k^m = \int_0^1 \theta^k (1 - \theta)^{(m-k)} \mu(d\theta). \tag{9}$$

In this case, $\frac{Y_1 + \dots Y_m}{m} \to \Theta$ almost surely, where the random $\Theta \in [0, 1]$ has law μ.

If $\mathbf{X} = (X_1, \dots X_n)$ is a finite exchangeable sequence, we can define its law-sequence $(p_k^m)_{k \leq m \leq n}$ similarly, up to levels $m \leq n$, which again characterises the distribution of $\mathbf{X}$. In the above parametrisation, the signed measure representation result, Theorem 3 then translates to the following, noted in this form in [Jay86].

Fact 2.3 (Ressel)**.** A random finite binary sequence $(X_1, \dots X_n)$ is exchangeable, if and only if there exists a signed measure ν on $[0, 1]$, such that for $k \leq m \leq n$,

$$p_k^m = \int_0^1 \theta^k (1 - \theta)^{m-k} \nu(d\theta). \tag{10}$$

We give a functional analytic perspective to this result which leads to a straightforward characterisation of the set of all possible signed measures for a given $\mathbf{X}$, which seems to have not been noted before, c.f. [Leo18].

Proof. Consider for $k \leq m \leq n$ the integrands

$$c_k^m(\theta) := \theta^k (1 - \theta)^{m-k}. \tag{11}$$

These are polynomial functions on the compact parameter space $[0,1]$ and thus continuous. The scaled versions $b_k^m(\theta) := \binom{m}{k}\theta^k(1-\theta)^{m-k}$ are the well-known Bernstein polynomials. For a fixed m, they are known to be a basis for the vector space P_m of polynomials of degree at most m. It then follows that the same is true for $(c_k^m(\theta))_{k\leq m}$. Fix a finite exchangeable $\mathbf{X}$ with law sequence $(p_k^m)_{k\leq m\leq n}$. Let's define the linear functional T_X on P_m by

$$T_X(c_k^m(\theta)) = p_k^m \tag{12}$$

on the basis and then extend it linearly. By the Hahn-Banach theorem, this can be extended to a T_X' defined on the space of all continuous functions $C([0,1])$ on $[0,1]$ that has the same dual space norm $\|T_X'\|^* = \|T_X\|^*$. By the Riesz representation theorem, there is a unique signed measure ν on $[0,1]$ with

$$\int f(\theta)\nu(d\theta) = T_X'(f)$$

for each $f \in C([0,1])$ and $\|\nu\|_{TV} = \|T_X'\|^*$. This ν then satisfies (10). □

Note also, again by the Riesz representation theorem, the set $\mathcal{R}_X$ of all possible such signed measures for $\mathbf{X}$ is isometrically isomorphic to the set of linear functionals $\lambda : C([0,1]) \to \mathbb{R}$ with $\lambda(c_k^m(\theta)) = p_k^m$. This is an affine subspace of $C([0,1])^*$ of codimension $n+1$. We thus conclude the following.

Proposition 2.4. The set $\mathcal{R}_X$ of signed measures satisfying (10) is an affine set, with codimension $n+1$ in the space of all signed measures on $[0,1]$.

Since polynomials are dense in $C([0,1])$ and the $c_k^m(\theta)$ span all other polynomials, it follows that for an infinite exchangeable $\mathbf{Y}$ as above, there is at most one such T that satisfies all the constraints in (9). Then it is neccessarily a positive linear map, again by density since it maps all of $c_k^m(\theta)$ to positive values. De Finetti's theorem then says that such a T exists and then the corresponding positive measure μ on $[0,1]$ via the Riesz theorem is the unique mixing measure. We briefly sketch a proof based on the method of moments, following [Fel71], p.228.

Proof of De Finetti's Theorem, sketch. For an infinite exchangeable sequence $\mathbf{Y}$ with law sequence $(p_k^m)_{k\leq m}$, consider, for each $m \in \mathbb{N}$ the probability measure

$$\mu_m = \sum_{k=1}^{m} \binom{m}{k} p_k^m \delta_{\frac{k}{n}}.$$

Straightforward calculations, spelt out in [Fel71] show that $\mu_m(c_k^n) \to p_k^m$ as $m \to \infty$ for each $k, n \in \mathbb{N}$ with $k \leq n$. Now the set $\mathcal{P}([0,1])$ of probability measures is weakly compact, so there is a subsequence m_i such that $\mu_{m_i} \rightharpoonup \mu$ for some probability measure μ. Since $\mu_{m_i}(c_k^n) \to \mu(c_k^n)$. Hence $\mu(c_k^n) = p_k^n$ for each $k \leq n$ and μ satisfies (9). Uniqueness follows from the above considerations.

For the law of large numbers result, note that $\mu_m(f) = \mathbb{E}_{\mathbf{Y}}(f\left(\frac{Y_1+\cdots+Y_m}{m}\right))$ for any measurable function f on $[0,1]$. The random sequence $(\mathbf{M}_m)_{m\in\mathbb{N}}$ with $\mathbf{M}_m = \frac{Y_1+\cdots+Y_m}{m}$ is well-known to form a reverse martingale, see **Section 5** in this text and e.g. [Kin78]. It follows from the reverse martingale convergence theorem, that there is a random element $\Theta \in [0,1]$ such that $\mathbf{M}_n \to \Theta$ almost surely. Then, for a continuous g, it follows that $g(\mathbf{M}_n) \to g(\Theta)$ almost surely. Since any $g \in C([0,1])$ is also bounded, we also have that $\mu_m(g) = \mathbb{E}_{\mathbf{Y}}(g(\mathbf{M}_m)) \to \mathbb{E}_\Theta(g(\Theta))$. It thus follows that Θ has law μ. □

Remark 2.5. Let's write $B_{m,f}(\theta) = \sum_{k=1}^{m} f(\frac{k}{m})\binom{n}{k}\theta^k(1-\theta)^{m-k}$ for the m^{th} Bernstein approximating polynomial of the continuous function f. Then observe [Fel71] that the probability measures μ_n and μ are linked by $\mu(B_{m,f}) = \mu_m(f)$. This relates the rate of convergence of μ_m to that of the Bernstein approximating polynomials via

$$|\mu(f) - \mu_m(f)| = |\mu(f) - \mu(B_{m,f})| \leq \|f - B_{m,f}\|, \tag{13}$$

which is optimal for a general μ. The convergence of the Bernstein approximating polynomials depend on the modulus of continuity of f and have been extensively studied, [BC89]. Under some regularity assumptions on μ, stronger results have been recently proven in terms of various metrizations of weak convergence, see [Dö15], [MPS16] and [DF20].

Above, we used 5 key properties of the parametrisation of the i.i.d. Bernoulli variables by $\theta \in [0,1]$. We identify these and how they should be adapted to to obtain similar results for other symmetric random structures.

1. The parameter space $[0,1]$ is compact and Hausdorff. Generally, we need to take parameters for the ergodic distributions from a compact Hausdorff space K to apply the Riesz representation theorem.
2. The functions c_i^n are continuous. In the general case, we need to identify the sufficient statistic for the symmetry notion as a function of the parameters and we need to prove that they are continuous for Riesz's representation theorem to apply.
3. We also needed that these functions are linearly independent as vectors in $C(K)$ to obtain the signed measure representation.
4. We needed that they are dense in K. This last one was used to obtain the representation theorem in the limit $m \to \infty$.
5. Finally, we used that the averages $\frac{Y_1+\cdots+Y_m}{m}$ have a reverse martingale structure to conclude the law of large numbers result.

In the next section we show how this proof recipe can be used for exchangeable graphs.

3 Integral representation of exchangeable graphs

Throughout, by a finite simple graph we mean a graph $G = (V_G, E_G)$ with vertex set $V_G = [n] = \{1, \ldots, n\}$ for some n and an edge set E_G without loops and multiple edges. We write $\mathcal{L}_n$ for the set of simple graphs on n vertices and $\mathcal{L} = \cup_n \mathcal{L}_n$. A random graph $\mathbb{F} \in \mathcal{L}_n$ is finitely exchangeable, if its distribution is invariant under permutations of the vertex set. That is, for $G_1, G_2 \in \mathcal{L}_n$ that are isomorphic,

$$P(\mathbb{F} = G_1) = P(\mathbb{F} = G_2). \tag{14}$$

Following [DJ08], we write $\mathcal{U}_n$ for the set of unlabelled graphs on n vertices, i.e. the set of such graphs with isomorphic graphs identified and $\mathcal{U} = \cup_n \mathcal{U}_n$ An infinite graph $\mathbb{H}$ is defined similarly, with vertex set $\mathbb{N}$. A random infinite graph is exchangeable if the restriction $\mathbb{H}|_n$ to the vertex set $[n]$ is exchangeable for each $n \in \mathbb{N}$.

Definition 3.1. A random two-array $(X_{ij})_{i,j\in\mathbb{N}}$ is jointly exchangeable if

$$(X_{ij}) \stackrel{d}{=} (X_{\sigma(i)\sigma(j)}) \tag{15}$$

for every permutation $\sigma \in \mathbb{S}$.

The the special case of interest here is when (X_{ij}) is a random symmetric binary array, but with 0-s on the diagonal. In this case, the array (X_{ij}) form a random adjacency matrix of an infinite random graph. Joint exchangeability means in that case that the distribution of the graph is exchangeable. The following is a special case of a theorem of Aldous and Hoover, see [Ald81], [Hoo79], [Kal05] and [Aus12] for this special case.

Fact 3.2 (Aldous-Hoover)**.** Let $(X_{ij})_{i,j\in\mathbb{N}}$ be a symmetric binary, jointly exchangeable array with 0-s on the diagonal. Then there exists a measurable $f\colon [0,1]\times[0,1]^2\times[0,1]\to\{0,1\}$, symmetric in the last two coordinates, such that

$$(X_{ij}) \overset{d}{=} f(U, U_i, U_j, U_{\{i,j\}}). \tag{16}$$

In this section, we prove a signed measure representation for finitely exchangeable graphs, similar to **Fact 2.3** and also an integral representation version of **Fact 3.2** with a law of large numbers, using our proof recipe from **Section 2**. The aim is to make the analogy with the sequence case fully explicit via our proof recipe. To begin with, we state De Finetti's theorem in an analogous version following [Aus12].

Fact 3.3. Suppose (X_i) is an infinite exchangeable binary sequence. Then there is a measurable function $T\colon [0,1]\times[0,1]\to\{0,1\}$, such that

$$(X_i) \overset{d}{=} (T(U, U_i)) \tag{17}$$

where (U_i) and U are independent $U([0,1])$ random variables.

We show how we can recover the usual form of de Finetti's theorem from this form. Let $Y_i = T(U, U_i)$. Then, following the law-sequence notation for exchangeable sequences, we have

$$\begin{aligned} p_k^n &= P(Y_1 = Y_2 = \dots Y_k = 1, Y_{k+1} = 0, \dots Y_n = 0) \\ &= \int \left(\int_{[0,1]^n} \prod_{i=1}^k T(u,u_i) \prod_{i=k+1}^n (1 - T(u,u_i))\, du_1 \dots du_n \right) du. \end{aligned}$$

Writing $\theta(u) = \int_{[0,1]} T(u,u_i)du_i$, which is a measurable function $[0,1]\to[0,1]$, we obtain

$$p_k^n = \int (c_k^n \circ \theta)(u) du.$$

Taking image measures under p, we recover usual integral form (9).

We now adapt this argument to graphs. Take an infinite exchangeable random graph $\mathbb{H}$. For such, the analogous statistic to the n-th level law-sequence for binary sequences is the distribution of $\mathbb{H}|_n$, i.e. the set $\{p_G := P(\mathbb{H}|_{[n]} = G) \colon G \in \mathcal{L}_n\}$. For f as in Theorem 3.2, we write

$$w(u,u_i,u_j) = \int_0^1 f(u,u_i,u_j,u_{\{i,j\}}) du_{\{i,j\}} \tag{18}$$

Then, from (16) using (18), we have that

$$P(\mathbb{H}|_{[n]} \cong G) = \int_{[0,1]} \theta_G(u) du, \text{ where} \tag{19}$$

$$\theta_G(u) = \int_{[0,1]^{V_G}} \left(\prod_{\{i,j\}\in E_G} w(u,u_i,u_j) \prod_{\{i,j\}\notin E_G} (1 - w(u,u_i,u_j)) \right) du_1 \dots du_n. \tag{20}$$

Because nodes of the graph with a shared edge are coupled, $\theta_G(u)$ cannot be separated in a form like $c_k^n \circ p(u)$ in the case of sequences and we cannot resort to a simple image measure argument as above.

The usual solution is to view $w(u, u_i, u_j)$ as a parametric function in $\mathcal{W}$, the set of symmetric measurable functions $W\colon [0,1]^2 \to [0,1]$, with parameter u. The set $\mathcal{W}$, called the *graphon space*, is extensively studied in the graph limits literature, see [Lov12] for a comprehensive treatment. We use properties of the graphon space to derive a complete analogue of the integral representation (9) of de Finetti's theorem, via a reparametrisation of (20) in the form

$$P(\mathbb{H}|_{[n]} \cong G) = \int_{\mathcal{W}} f_G(W)\,\mu(dW), \text{ where} \tag{21}$$

$$f_G(W) = \int_{[0,1]^V} \left(\prod_{\{i,j\}\in E} W(u_i, u_j) \prod_{\{i,j\}\notin E} (1 - W(u_i, u_j)) \right) du_1 \dots du_n, \tag{22}$$

with a suitable notion of integrals of graphons that we specify shortly. Here we recover

$$t_{\text{ind}}(G, W) = f_G(W), \tag{23}$$

the *induced homomorphism density* of G in W ([Lov12], (7.4)). These functions are the direct analogues of the polynomials c_k^n in the graph setting. For reasons that we make clear soon, it is more convenient to work with the *homomorphism densities*

$$t(F, W) = \int_{[0,1]^V} \left(\prod_{\{i,j\}\in E} W(u_i, u_j) \right) du_1 \dots du_n. \tag{24}$$

They are seen ((7.4) and (7.5) in [Lov12]) to be related by the formula

$$t(F, W) = \sum_{\substack{F\subseteq F' \\ F'\in\mathcal{U}_n}} t_{\text{ind}}(F', W), \tag{25}$$

and its usual Möbius inverse

$$t_{\text{ind}}(F, W) = \sum_{\substack{F\subseteq F' \\ F'\in\mathcal{U}_n}} (-1)^{e(F')-e(F)} t(F', W). \tag{26}$$

For the homomorphism densities, the corresponding law sequence statistic that specifies exchangeable graph distributions is

$$\{p'_G := \mathbb{P}(G \subset \mathbb{H}) : G \in \mathcal{L}_n\},$$

where the relation $\subset$ denotes inclusion as subgraphs, that is, $E_G \subset E_{\mathbb{H}}$. In this statistic, the desired integral representation becomes

$$p'_G = \mathbb{P}(G \subset \mathbb{H}) = \int_W t(G, W)\,\mu(dW). \tag{27}$$

In the theory of graph limits, graphons are the limiting objects of dense graphs in the large vertex set limit. The functions $t(F, \cdot)$ and $t_{\text{ind}}(F, \cdot)$ are originally defined on finite graphs. For graphs F and G, the value of $t(F,G)$ is the proportion of maps $V_F \to V_G$ that are graph homomorphisms, i.e.

$$t(F,G) = \frac{\hom(F,G)}{n^k}, \tag{28}$$

where $n = |V_G|$ and $k = |V_F|$. If $k \le n$, we can similarly define $t_{\mathrm{inj}}(F,G)$ as the proportion of injective maps $V_F \to V_G$ that are homomorphisms, and $t_{\mathrm{ind}}(F,G)$ as the proportion of those injective maps that also preserve non-adjacency, see [Lov12]. Then $t(F,\cdot)$ and $t_{\mathrm{ind}}(F,\cdot)$ have the above defined extensions to graphons. For $t_{\mathrm{inj}}(F,\cdot)$, we recall from e.g. [LS06] that

$$|t_{\mathrm{inj}}(F,G) - t(F,G)| \le \frac{|V_F|}{2|V_G|}. \tag{29}$$

In particular, in the limit $|V_G| \to \infty$, the functions $t_{\mathrm{inj}}(F,\cdot)$ and $t(F,\cdot)$ coincide and they have the same extensions to graphons.

In the graphon space, a graph G on the vertex set $[n]$ can be represented by a symmetric *stepfunctions* W_G such that $W_G(x,y) = \mathbb{I}_{ij\in E(G)}$ for $x \in [\frac{i}{n}, \frac{i+1}{n}) \times [\frac{j}{n}, \frac{j+1}{n})$. It can then be checked ([Lov12] 7.2) that

$$t(F,G) = t(F,W_G). \tag{30}$$

The main reason that the functions $t(F,\cdot)$ are more convenient than $t_{\mathrm{ind}}(F,\cdot)$, is that $t_{\mathrm{ind}}(F,G)$ may not be equal to $t_{\mathrm{ind}}(F,W_G)$ and so we cannot represent finite graphs by graphons suitably under $t_{\mathrm{ind}}(F,\cdot)$.

We now specify the right σ-algebra on $\mathcal{W}$ for the integrals in (21) or (27).

Definition 3.4 ([Lov12], Definition 8.13)**.** The *cut norm* on $\mathcal{W}$ is given by

$$\|W\|_\square = \sup_{S,T\subset[0,1]} \left| \int_{S\times T} W(x,y)\,dx\,dy \right|, \tag{31}$$

where the supremum is taken over all measurable subsets S and T.

Definition 3.5 ([Lov12], Definition 8.17)**.** Let Ψ denote the set of invertible measure preserving maps $[0,1] \to [0,1]$. The *cut distance* of two graphons is given by

$$\delta_\square(U,W) = \inf_{\phi\in\Psi} \|U - W^\phi\|_\square \tag{32}$$

where $W^\phi(x,y) = W(\phi(x),\phi(y))$.

A sequence of finite simple graphs is said to *converge*, if (W_{G_n}) converges in $\mathcal{W}$ in the cut distance. This is the sense in which graphons are limiting objects of finite simple graphs. One of the main motivations for the cut distance from our perspective is the following result from [BCL$^+$08] see also [DJ08], and [Lov12] Corollary 10.34.

Fact 3.6 (Borgs, Chayes, Lovász, T. Sós, Vesztergombi)**.** We have $\delta_\square(U,W) = 0$ if and only if $t(G,W) = t(G,U)$ for every finite graph G.

Note in the integral (27) we only access $W \in \mathcal{W}$ via the functions $t(G,\cdot)$. So for uniqueness statements, we need to identify W-s for which those functions coincide. **Fact 3.6** grants us that this can be done by working in the quotient space $\overline{\mathcal{W}}$, where W and U with $\delta_\square(W,U) = 0$ are identified. From now on, we abuse notation and by a graphon W we mean its equivalence class in $\overline{\mathcal{W}}$. It can be seen that $\delta_\square$ is a metric on $\overline{\mathcal{W}}$ and we interpret (21) and (27) in the Borel σ-algebra or this metric structure on $\overline{\mathcal{W}}$. The precise statements that we prove then are the following. The first is implicit in e.g. [DJ08] and [OS16], but we believe it has only been stated in this form in [OR15]. The second we believe to be new.

Fact 3.7 (de Finetti's theorem for exchangeable graphs)**.** A random infinite graph $\mathbb{H}$ is exchangeable, if and only if there is a Borel probability measure on $\overline{\mathcal{W}}$, such that

$$P(G \subset \mathbb{H}|_{[n]}) = \int_{\overline{\mathcal{W}}} t(G, W)\, \mu(dW) \tag{33}$$

holds for each finite simple graph G. In this case, the measure μ is unique and $W_{\mathbb{H}|_n}$ converges almost surely in $\delta_\square$ to a random element W with law μ.

Theorem 3.8 (Integral representation of finite exchangeable graphs)**.** Let $\mathbb{F}$ be a finitely exchangeable random graph on the vertex set $[n]$. Then there exists a Borel signed measure ν on $\overline{\mathcal{W}}$ such that for each $G \in \mathcal{L}_n$,

$$P(G \subset \mathbb{F}) = \int_{\overline{\mathcal{W}}} t(G, W)\, \nu(dW). \tag{34}$$

Moreover, the set $\mathcal{R}_{\mathbb{F}}$ of signed measures that satisfy (34) is an affine subspace of the dual space $C(\overline{\mathcal{W}})^*$ of codimension m, where $m = |\mathcal{U}_n|$ is the number of graphs on n vertices up to isomorphism.

To prove these statements, we quote results from graph limits theory to see that the requirements on our proof recipe are satisfied with the parameter space $\overline{\mathcal{W}}$ and the functions $\{t(G, \cdot) \colon G \in \mathcal{L}\}$. First of all, if G_1 and G_2 are isomorphic, we have $t(G_1, W) = t(G_2, W)$ for any W, since the definition of $t(G, W)$ depends only on the edge structure. In particular, any infinite random graph $\mathbb{H}$ that satisfies (33) is exchangeable. From now on, we will identify these functions for isomorphic G_i and thus index them by unlabelled graphs.

Compactness As shown in [LS07], Szemerédi's Regularity Lemma [Sze75] can be phrased as the compactness result we need (see also as Theorem 9.23 in [Lov12]).

Fact 3.9 (Lovász, Szegedy)**.** The metric space $(\overline{\mathcal{W}}, \delta_\square)$ is compact.

Continuity Continuity of the functions $t(G, \cdot)$ follows directly from the Counting Lemma of [LS06], also as Lemma 10.23 in [Lov12].

Fact 3.10 (Lovász, Szegedy)**.** Let F be a simple graph and let $W, U \in \mathcal{W}$. Then

$$|t(F, W) - t(F, U)| \leq e(F)\delta_\square(W, W'). \tag{35}$$

So the functions $t(F, \cdot)$, are actually Lipschitz.

Linear independence The following was proven in [DGKR15], where the authors study the differential theory of functions in $C(\overline{W})$. We provide a proof without the differential machinery developed there.

Fact 3.11. The set $\{t(G, \cdot) \colon G \in \mathcal{U}_n\}$ is linearly independent as vectors in $C(\overline{\mathcal{W}})$ for any n.

We deduce this from the linear independence result in [ELS79] (Proposition 5.44 (c) in [Lov12]) for the classical homomorphism densities in graphs rather than in graphons.

Fact 3.12 (Erdős, Lovász, Spencer)**.** Let $F_1, \dots F_k$ be nonisomorphic simple graphs with no isolated nodes. Then there are simple graphs $G_1, \dots G_k$ such that the matrix $[t(F_i, G_j)]_{i,j=1}^k$ is nonsingular.

Proof of Fact 3.11. For $F_i \in \mathcal{U}_n$, let $F_i' \in \cup_{k=0}^n \mathcal{U}_k$ be the graph with all the isolated nodes removed from F_i. Note $t(F_i, W) = t(F_i', W)$ for any graphon W. It then follows from Proposition 3.12 that there are simple graphs $G_1, \dots G_m$ so that the functions $t(F_i, \cdot) \colon \overline{\mathcal{W}} \to [0,1]$, when restricted to (the $\delta_\Box$ equivalence classes of) the graphons $W_{G_1}, \dots W_{G_m}$ are linearly independent. It then follows that $t(F_i, \cdot)$ are certainly linearly independent on the whole domain $\overline{\mathcal{W}}$. □

At this point, we can already conclude **Theorem 3.8**.

Proof of Theorem 3.8. As for sequences, we can define a linear map T on $\mathrm{span}(\{t(G, \cdot) : G \in \mathcal{U}_n\})$ with $T(t(G, \cdot)) = p_G'$ noting the functions $\{t(G, \cdot) \colon G \in \mathcal{U}_n\}$ are linearly independent. Since they are also continuous, we then extend to a bounded linear map on $C(\overline{\mathcal{W}})$ by the Hahn-Banach theorem. We then conclude by the Riesz representation theorem. □

Density The following was also observed in [DGKR15]. We provide the proof for its importance in the next section.

Fact 3.13 (Diao, Guillot, Khare, Rajanatnam)**.** The set $\mathcal{A}_{\text{hom}} = \mathrm{span}(\{t(G, \cdot) \colon G \in \mathcal{U}\})$ is dense in $C(\overline{\mathcal{W}})$.

Proof. We use the Stone-Weierstrass theorem. For this, we need to show that $\mathcal{A}_{\text{hom}}$ is an algebra that separates points in $\overline{\mathcal{W}}$ and contains a nonzero constant function. Note that for graph G_0 with no edges, $t(G_0, \cdot)$ is the constant 1 function. For the algebra structure of $\mathcal{A}_{\text{hom}}$, we can directly check from the definition (24) of $t(G, \cdot)$ that

$$t(F_1 F_2, W) = t(F_1, W) t(F_2, W), \tag{36}$$

where for the finite simple graphs F_1 and F_2, the graph $F_1 F_2$ is their disjoint union. See also [Lov12], (7.6). It follows that for $f_i, f_j \in \mathcal{A}_{\text{hom}}$ also $f_i \cdot f_j \in \mathcal{A}_{\text{hom}}$ as well and $\mathcal{A}_{\text{hom}}$ is an algebra.

Finally, we need to show that $\mathcal{A}_{\text{hom}}$ separates the points in $\overline{\mathcal{W}}$. This was the motivation for taking the quotient with respect to $\delta_\Box$ and indeed, this is the restatement of **Theorem 3.6**. □

It thus follows that for an infinitely exchangeable graph $\mathbb{H}$, there is at most one measure $\mu \in C(\overline{\mathcal{W}})^*$ that satisfies (33) for all $G \in \mathcal{U}$. Moreover, such a μ is a probability measure.

Martingale Structure For both the existence and the law of large numbers result, we will use the following, pointed out in [DJ08] and [Lov12] Proposition 11.14. See **Section 5** in this text for a proof and a general perspective.

Fact 3.14. For an infinite exchangeable random graph $\mathbb{H}$, and a graph F with m vertices, the sequence $(t_{\text{inj}}(F, \mathbb{H}|_n))_{m \le n}$ forms a reverse martingale.

We are now ready for the proof of **Fact 3.7**

Proof. We have seen that any infinite random graph satisfying (33) must be exchangeable. The density result **Fact 3.13** shows that such a μ, if exists, is neccessarily unique.

For the existence, take an infinite exchangeable graph $\mathbb{H}$ with $p_G = \mathbb{P}(G = \mathbb{H}|_n)$ and $p'_G = \mathbb{P}(G \subset \mathbb{H})$ for a finite simple graph G. Consider the measure μ_n on $\overline{\mathcal{W}}$ given by

$$\mu_n = \sum_{G \in \mathcal{L}_n} p_G \delta_{W_G}. \tag{37}$$

We will show that $\mu_n(t(F, \cdot)) \to p'_F$ and then conclude, by a similar method of moments argument to the sequence case by passing to a subsequence, noting that $\mathcal{P}(\overline{\mathcal{W}})$ is weakly compact, since $\overline{\mathcal{W}}$ is compact.

Fix a finite simple graph F, since $t_{\text{inj}}(F, \mathbb{H}|_n)$ is a reverse martingale, there is a random $Z_F \in [0, 1]$ such that $t_{\text{inj}}(F, \mathbb{H}|_n) \to Z_F$ almost surely. By the approximation (29), neccessarily $t_{\text{inj}}(F, \mathbb{H}|_n) - t(F, \mathbb{H}|_n) \to 0$ almost surely and so $t(F, \mathbb{H}|_n) \to Z_F$ too almost surely. Since the functions $t_{\text{inj}}(F, \cdot$ and $t(F, \cdot)$ are bounded, all of these hold in expectation as well. Note that $\mu_n(f) = \mathbb{E}_{\mathbb{H}}(f(W_{\mathbb{H}|_n}))$ for any integrable function f on $\overline{\mathcal{W}}$. Using $t(F, G) = t(F, W_G)$ for any finite simple graph G, it follows that

$$\mu_n(t(F, \cdot)) = \mathbb{E}_{\mathbb{H}}(t(F, \mathbb{H}|_n)) \to \mathbb{E}(Z_F).$$

Via the treatment in **Section 5**, it will be immediately clear that $\mathbb{E}(Z_F) = \mathbb{P}(F \subset \mathbb{H}) = p'_F$, but we rephrase here for completeness. By the martingale structure, $\mathbb{E}_{\mathbb{H}}(t_{\text{inj}}(F, \mathbb{H}|_n)) = \mathbb{E}_{\mathbb{H}}(t_{\text{inj}}(F, \mathbb{H}|_k))$ for $n, k \geq m = |V_F|$ and so $\mathbb{E}(Z_F) = \mathbb{E}(t_{\text{inj}}(F, \mathbb{H}|_m))$. Now an injective map $[m] \to [m]$ is a permutation. Thus, since $\mathbb{H}$ is exchangeable,

$$\mathbb{E}_{\mathbb{H}}(t_{\text{inj}}(F, \mathbb{H}|_m)) = \mathbb{E}_{\mathbb{H}} \left(\frac{1}{|\mathbb{S}_m|} \sum_{\sigma \in \mathbb{S}_m} \mathbb{I}(F \subset (\sigma \cdot \mathbb{H}|_m)) \right) = \mathbb{E}_{\mathbb{H}} \left(\mathbb{I}(F \subset (\mathbb{H}|_m)) \right) = \mathbb{P}(F \subset \mathbb{H})$$

where we denote by $\sigma \cdot \mathbb{H}|_n$ the graph obtained by permuting the vertices by σ.

Finally, for the law of large numbers, we have already noted that $t(F, W_{\mathbb{H}|_n})$ converges almost surely for a finite simple graph F. Since there are countably many such F, this holds jointly for all finite simple graphs. It follows that $(W_{\mathbb{H}_n})_n$ is almost surely Cauchy: for otherwise, there would exist a continuous function f such that $f(W_{\mathbb{H}_n})$ is not convergent with nonzero probability. But by the density result **Fact 3.13**, on the probability 1 event that all the $t(F, \cdot)$, $F \in \mathcal{L}$, all such functions are convergent.

Since $\overline{\mathcal{W}}$ is complete, this sequence has an almost sure random limit $W \in \overline{\mathcal{W}}$. It also follows from above that $p'_F = \mathbb{E}_W(t(F, W))$ and so W has law μ. □

In the proof, we have implicitly deduced the following key result of [BCL+06], [BCL+08], also Theorem 11.5 in [Lov12], which characterises convergence in the graphon space, using the continuity and density of the homomorphism densities.

Fact 3.15 (Borgs, Chayes, Lovász, T.Sós, Vesztergombi)**.** Let W_n be a sequence of graphons in $\overline{\mathcal{W}}$ and let $W \in \overline{\mathcal{W}}$. Then $t(F, W_n)$ converges for all finite simple graphs F if and only if W_n is a Cauchy sequence in the $\delta_\square$ distance. Furthermore, $t(F, W_n) \to t(F, W)$ for all finite simple graphs F if and only if $\delta_\square(W_n, W) \to 0$.

The classical proof of this fact, is stronger, inasmuch it relates the convergence rate of the homomorphism densities to that of graphons via the so called Inverse Counting Lemma (Lemma 10.32 in [Lov12]). In turn, those ideas lead to a more quantitative proof of the density result **Fact 3.13**, by analogous approximating functions to the Bernstein approximating polynomials. Just like in **Remark 2.5**, their convergence rate is closely tied to that of the empirical measures μ_n.

Theorem 3.16. For a continuous function $f : \overline{\mathcal{W}} \to \mathbb{R}$, the sequence of functions $B_{n,f} \in \mathcal{A}_{\text{hom}}$ given by

$$B_{n,f}(W) = \sum_{F \in \mathcal{L}_n} f(W_F) t_{\text{ind}}(F, W) \tag{38}$$

converges in $C(\overline{\mathcal{W}})$. Moreover an infinite exchangeable distribution, the mixing measure μ and empirical measure $\mu_n = \sum_{G \in \mathcal{L}_n} p_G \delta_{W_G}$ are related by

$$\mu_n(f) = \mu(B_{n,f}) \tag{39}$$

for each $f \in C(\overline{\mathcal{W}})$ and in particular, $|\mu(f) - \mu_n(f)| \leq \|f - B_{n,f}\|$.

For the proof, we will use the *Second Sampling Lemma* from [BCL$^+$08] ([Lov12] Lemma 10.16). It follows from **Fact 3.7** that each $W_0 \in \overline{\mathcal{W}}$ corresponds to an infinite exchangeable graph distribution, that has the mixing measure δ_{W_0} in (33). These distributions are the ergodic exchangeable graph distributions [DJ08] and are studied in detail in [Lov12], where they are called W-random distributions. We write $\mathcal{G}(k, W_0)$ for the subgraph on $[k]$ generated from this distribution for a given $W_0 \in \overline{\mathcal{W}}$. With this notation, the lemma says the following.

Fact 3.17 (Borgs, Chayes, Lovász, T.Sós, Vesztergombi)**.** For $W_0 \in \overline{\mathcal{W}}$, and $n \geq 1$, with probability at least $1 - \exp(-k/(2\log(k))$, we have

$$\delta_\square(W_{\mathcal{G}(n,W_0)}, W_0) \leq \frac{22}{\sqrt{\log(n)}}$$

Proof of Theorem 3.16. Take an infinite exchangeable graph $\mathbb{H}$ with mixing measure μ in (33). Then we have that

$$p_G = \mathbb{P}(\mathbb{H}|_n = G) = \mu(t_{\text{ind}}(G, \cdot))$$

for $G \in \mathcal{L}$. It immediately follows that μ_n defined above satisfies $\mu_n(f) = \mu(B_{n,f})$ for each $f : \overline{\mathcal{W}} \to \mathbb{R}$ measurable. Observe that $\mathbb{P}(\mathcal{G}(k, W_0) = F) = t_{\text{ind}}(F, W_0)$ for each finite simple graph F on the vertex set $[k]$ and $W_0 \in \overline{\mathcal{W}}$. So the function $B_{n,f}$ satisfies $B_{n,f}(W_0) = \mathbb{E}(f(W_{\mathcal{G}(n,W_0)}))$ for each $W_0 \in \overline{\mathcal{W}}$ and measurable function $f : \overline{\mathcal{W}} \to \mathbb{R}$. In particular, for a fixed $f \in C(\overline{\mathcal{W}})$, we can calculate as follows using **Fact 3.17**. We write $r_n = \exp(-n/(2\log(n))$ for the probability that the approximation does not hold. Then

$$|B_{n,f}(W_0) - f(W_0)| = |\mathbb{E}f(W_{\mathcal{G}(n,W_0)}) - f(W_0)| \leq r_n \cdot 2\|f\| + (1 - r_n)\omega_f \left(\frac{22}{\sqrt{\log(n)}} \right),$$

where ω_f is a *modulus of continuity* for f that satisfies $|f(W) - f(U)| \leq \omega_f(\delta_\square(W, U))$. Note by uniform continuity, such w_f exists which is left continuous at 0 and so we conclude $\sup_{W_0 \in \overline{\mathcal{W}}} |B_{n,f}(W_0) - f(W_0)| \to 0$ as $n \to \infty$.

□

Remark 3.18. As mentioned in [BC89], the analogous standard argument in the sequence case via Chebyshev's theorem in place of the Second Sampling Lemma **Fact 3.17**, leads to a nearly optimal bound. It would be interesting to see whether the bound in the Second Sampling Lemma is optimal in this sense.

4 Further analogous results

In the previous sections, we explicitly connected the theories of exchangeable binary sequences and exchangeable graphs via a shared functional analytic framework to prove de Finetti-style integral representation theorems and related results. In this section, we briefly sketch some immediate applications of our framework for other closely related theories.

4.1 Exchangeable sequences in a finite state space

Consider an exchangeable sequence $\mathbf{Y}$ with values in the set $[k]$ for some $k \in \mathbb{N}$. In this case, de Finetti's theorem can be formulated as an integral on the parameter space given by the simplex $\Delta_k = \{\theta \in [0,1]^k : \sum_i \theta_i = 1\}$. For a vector $\mathbf{m} \in [k]^n$ and $i \in [k]$, let us write $\lambda_{\mathbf{m}}(i)$ for the number of coordinates of $\mathbf{m}$ that takes the value i and then $\lambda_{\mathbf{m}} = (\lambda_{\mathbf{m}}(0), \dots \lambda_{\mathbf{m}}(k-1))/n$, which is an element of Δ_k.

Fact 4.1 (de Finetti)**.** A random infinite sequence with values in $[k]$ is exchangeable, if and only if

$$p_{\mathbf{m}} = \mathbb{P}((Y_1, \dots Y_n) = \mathbf{m}) = \int_{\Delta_k} \prod_i^k \theta_i^{\lambda_{\mathbf{m}}(i)} \mu(d\theta)$$

holds for a measure μ. In this case, μ is unique and the sequence of empirical average measures $(\mu_n)_n$ given by $\mu_n = \sum_{\mathbf{m} \in [k]^n} p_{\mathbf{m}} \delta_{\lambda_{\mathbf{m}}}$ converges weakly to μ.

Now the parameter space Δ_k is compact, it is easy to check that multivariate polynomials $c_{\mathbf{m}} \colon \Delta_k \to \mathbb{R}$, $c_{\mathbf{m}}(\theta) = \prod_i^k \theta_i^{\lambda_{\mathbf{m}}(i)}$ are continuous, linearly independent and span a dense subset in Δ_k. Furthermore, for each $i \in [k]$ the sequence $(A_n^i)_n$ given by $\frac{1}{n}\sum_{j=1}^n \mathbb{I}(Y_j = i)$ forms a reverse martingale. These together allow us to apply our proof technique to prove **Fact 4.1** and the corresponding signed measure representation for finite exchangeable sequences.

A similar proof of the integral representation is available if the state space S has a finitely generated σ-algebra, which is true if e.g. S is compact and assumes the Borel σ-algebra. This is also incorporated in the general setting of the next section.

4.2 Partially exchangeable binary sequences

Partial exchangeability is often referred to [Leo18], [Dia23] the joint law of random binary sequences $(Y_n^1), \dots (Y_n^k)$ that are invariant under separate permutations of their indices. That is, for permutations $\pi_1, \dots \pi_k$ and π_2 and finite integers $N_1, \dots N_k$

$$\{\{Y^i_{\pi_i(n)}\}_{n=1}^{N_i}\}_{i=1}^k \overset{d}{=} \{\{Y^i_n\}_{n=1}^{N_i}\}_{i=1}^k. \tag{40}$$

This is a stronger assumption than the vectors $(Y_i^1, , \dots, Y_i+)$ being exchangeable, but weaker than that the zipped sequence of the $\mathbf{Y}^i$-s is exchangeable (e.g. $\mathbf{Y}^0$ and $\mathbf{Y}^1$ are not necessarily identically distributed). For such, there is a de Finetti-style representation theorem as follows.

Fact 4.2. A the joint law of k infinite random binary sequences $\mathbf{Y}^1, \dots \mathbf{Y}^k$ and is partially exchangeable if any only if there is a measure μ on $[0,1]^k$ such that

$$\mathbb{P}(\{\{Y_n^i = y_j^i\}_{j=1}^{N_i}\}_{i=1}^k) = \int_{[0,1]^k} \prod_{i=1}^k \prod_j^{N_i} \theta_i^{y_i^j} (1-\theta_i)^{(1-y_i^j)} \mu(d(\theta_1, \dots d\theta_k)). \tag{41}$$

In this case, μ is unique and the empirical averages $\frac{1}{n}(\sum_{j=1}^n Y_j^1, \cdots \sum_{j=1}^n Y_j^k)$ converge to μ in distribution.

The corresponding signed measure representation for finite partially exchangeable sequences $\mathbf{X}^1, \dots, X^k$ is also available and was first proven in [Leo18]. The argument using our proof recipe can be directly applied with the parameter space $[0,1]^k$ and polynomials $\prod_{i=1}^{k} \prod_{j}^{N_i} \theta_i^{y_i^j} (1-\theta_i)^{(1-y_i^j)}$.

4.3 Bipartite graphs

By a bipartite graph G we mean a graph with two disjoint vertex sets V_G^1 and V_G^2 with edges connecting nodes in separate vertex sets. In [DJ08] the authors sketch an analogous exchangeability and limit theory of bipartite graphs to the simple graphs case. A graphon in this case need not be symmetric, it is a measurable function $W : [0,1] \times [0,1] \to [0,1]$. The functions $t(F, \cdot)$, $t_{\text{inj}}(F, \cdot)$, $t_{\text{ind}}(F, \cdot)$ for a bipartite F can be defined in a similar manner and have analogous properties [DJ08]. We can again define the cut norm

$$\|W\|_\square = \sup_{S,T \subset [0,1]} \left| \int_{S \times T} W(x,y) dx\, dy \right|, \tag{42}$$

and in this case, the cut distance is obtained by

$$\delta_\square(U, W) = \inf_{\phi \in \Psi} \|U - W^{\phi,\psi}\|_\square \tag{43}$$

where Ψ is the set of measure preserving maps $[0,1] \to [0,1]$ and $W^{\phi,\psi}(x,y) = W(\phi(x), \psi(y))$. If we quotient out by the relation $\delta_\square(W, U) = 0$, we again obtain a compact space $\overline{\mathcal{W}}_{\text{as}}$, by close analogy to the proof in [LS07]. Checking all the required properties, we can obtain the analogous integral representation results to **Fact 3.7** and **Theorem 3.8**.

5 More general invariant structures

As we have seen, for an infinite exchangeable sequence $\mathbf{X}$ the statistic $(\frac{1}{n}(\sum_i X_i))_{n \in \mathbb{N}}$ forms a backwards martingale and the limit specifies the law of the de Finetti mixing measure μ in distribution. Similarly, for an infinite exchangeable graph $\mathbb{H}$, the sequence $(t_{\text{inj}}(F, \mathbb{H}|_n))_{n > |V(F)|}$ forms a backwards martingale for each finite simple graph F and the collection of the limits of these martingales specifies the law of the mixing measure on graphons. These backwards martingales are specific examples of the more general construction given in **Lemma 5.3** which enables us to derive an analogous exchangeability and limit theory for more general group actions. The starting point is the abstract ergodic decomposition of invariant probability measures discussed in various flavours in e.g. [Far62], [Var63][Mai77], and [Dyn78], see also [Kal05].

By measurable action of a group $\mathbb{G}$ on a Polish space Ω we mean measurable functions $\phi \colon \Omega \to \Omega$ for each $\phi \in \mathbb{G}$, which respect the product structure of $\mathbb{G}$ under composition. A measurable function $f \colon \Omega \to \mathbb{R}$ is $\mathbb{G}$-invariant if $f \circ \phi = f$ for each $\phi \in \mathbb{G}$. A measurable set $A \subset \Omega$ is invariant, if $\mathbb{I}_A$ is an invariant function. The set of $\mathbb{G}$-invariant sets form a σ-algebra, which we call $\Sigma_\mathbb{G}$. A measure μ on Ω is invariant if it equals its pushforward $\phi_\# \mu$ for any $\phi \in \mathbb{G}$, that is $\int f d\mu = \int f \circ \phi d\mu$ for any integrable function f and $\phi \in \mathbb{G}$. We write $\mathcal{P}_\mathbb{G}$ for the set of $\mathbb{G}$-invariant probability measures. An invariant probability measure μ is ergodic, if it is trivial on $\Sigma_\mathbb{G}$, that is $\mu(A) \in \{0,1\}$ for each $A \in \Sigma_\mathbb{G}$. Ergodic decomposition holds under the technical assumption that $\mathbb{G}$ is amenable. All the groups that we discuss here are amenable, we refer to [AO22] for the definition in a similar treatment to ours.

Fact 5.1 (Varadarajan)**.** Let $\mathbb{G}$ be an amenable group that acts measurably on a Polish space Ω. Then:

- The ergodic probability measures are precisely the extreme points of $\mathcal{P}_{\mathbb{G}}$. The set $\mathrm{ex}(\mathcal{P}_{\mathbb{G}})$ of ergodic measures is measurable in $\mathcal{P}$.
- A probability measure P on Ω is G-invariant if and only if

$$P(A) = \int_{\mathrm{ex}\mathcal{P}_{\mathbb{G}}} \eta(A)\mu_P(d\eta) \quad \text{for each measurable } A \subset \Omega$$

for a probability measure μ_P on $\mathrm{ex}\mathcal{P}_{\mathbb{G}}$, which is then uniquely determined by P.

Note that all the exchangeable structures that we discussed are invariant distributions under some action of the symmetric group $\mathbb{S}$ on the space of infinite sequences, graphs, etc by some index permutations. These are all compact, metrizable spaces in their respective product topologies. E.g. for infinite graphs, if we write $\mathcal{N}$ for the set of possible edges on the vertex set $\mathbb{N}$, then the action of the symmetric group has the base space $\mathbf{Z}_{\mathrm{graph}} = \{0,1\}^{\mathcal{N}}$, which is a compact metric space in the product topology, and the action is by permuting vertices. In the case of exchangeable binary sequences, we have the standard action of the symmetric group on $\mathbf{Z}_{\mathrm{seq}} = \{0,1\}^{\mathbb{N}}$. The corresponding de Finetti theorems thus fit in the framework of this ergodic decomposition theorem, but the additional structure gives further three important properties.

1. The characterisation of ergodic measures by some independence criterion usually referred to as dissociatedness. For the sequence case this is the fact that the ergodic distributions are precisely the i.i.d. ones, often called the Hewitt-Savage theorem.
2. The de Finetti integrals can be stated over a parameter space where the law of large numbers gives rise to the limit theory of infinite sequences, graphs, etc, in the sense that their finite components (initial segments, restrictions) converge almost surely to a random element of the parameter space: $[0,1]$ for sequences and $\overline{\mathcal{W}}$ for graphs.
3. The set of ergodic measures are weakly compact, which is of independent interest and also renders the parameter space compact via a homeomprhism.

We extract structural, algebraic assumptions on the action of $\mathbb{S}$ on graphs and sequences (and the other combinatorial structures that we leave implicit here) that allow for these properties. We then state these properties as abstract assumptions on the action of a group $\mathbb{G}$ on a space $\mathbf{Z}$. We assume that $\mathbf{Z}$ is a compact metric space, which is the case for $\mathbf{Z}_{\mathrm{seq}} = \{0,1\}^{\mathbb{N}}$ and $\mathbf{Z}_{\mathrm{graph}} = \{0,1\}^{\mathcal{N}}$. We draw analogies from the treatment of graph limits and exchangeable graphs in [DJ08]. In particular, our parameter space is inspired by the product space $\overline{\mathcal{U}^*} \subset [0,1]^{\mathcal{U}}$ discussed there, which is homeomorhpic to the graphon space ([Lov12] Remark 11.4). We assume that the group $\mathbb{G}$ is a direct limit of compact groups.

Definition 5.2. The group $\mathbb{G}$ is called the direct limit of the groups $\mathbb{G}_n$, if each $\mathbb{G}_n$ is compact, $\mathbb{G}_n \subset \mathbb{G}_{n+1}$ for each n and $\mathbb{G} = \cup_n \mathbb{G}_n$.

We call $\mathbb{G}$ a direct limit in short, if it is a direct limit of compact groups. Note the symmetric group $\mathbb{S}$ is the direct limit of the finite symmetric gropus $\mathbb{S}_n$. For direct limits, the following pointwise ergodic theorem is available.

Lemma 5.3. Let a direct limit $\mathbb{G}$ of compact groups $\mathbb{G}_n$ act measurably on a standard Borel space $(\mathbf{Z}, \Sigma)$. Let Z be a $\mathbb{G}$-invariant random element of $\mathbf{Z}$ and let $f \in L_1(Z)$. Then, $(\mathbb{E}_Z(f(Z|\Sigma_{\mathbb{G}_n}))_{n\in\mathbb{N}}$ is a backwards martingale and

$$\mathbb{E}_Z(f(Z)|\Sigma_{\mathbb{G}_n}) \to \mathbb{E}_Z(f(Z)|\Sigma_{\mathbb{G}}) \tag{44}$$

almost surely. Moreover,

$$\mathbb{E}_Z(f(Z)|\Sigma_{\mathbb{G}_n}) = \frac{1}{|\mathbb{G}_n|}\int_{\mathbb{G}_n} f(\phi Z)|d\phi| \tag{45}$$

almost surely.

Proof. Abbreviate $f_n(x) := |\mathbb{G}_n|^{-1}\int_{\mathbb{G}_n} f(\phi x)|d\phi|$. The direct limit structure implies that $\cap_n \Sigma_{\mathbb{G}_n} = \Sigma_{\mathbb{G}}$, so we have to show $f_n \to \mathbb{E}_Z[f(Z)| \cap_n \Sigma_{\mathbb{G}_n}]$ a.s.

Let P be the law of Z. Observe f_n is $\mathbb{G}_n$-invariant, since $\mathbb{G}_n$ is a group and hence

$$\int_{\mathbb{G}_n} f(\phi\psi x)|d\phi| = \int_{\mathbb{G}_n\psi} f(\phi x)|d\phi| = \int_{\mathbb{G}_n} f(\phi x)|d\phi| \qquad \text{for } \psi \in \mathbb{G}_n \,.$$

It follows that $f_n \in \Sigma_{\mathbb{G}_n}$, which implies $f_n = P(f_n|\Sigma_{\mathbb{G}_n})$ almost surely. Consider any $A \in \Sigma_{\mathbb{G}_n}$ and $\phi \in \mathbb{G}_n$. Since P is ϕ-invariant,

$$\mathbb{I}_A \circ \phi = \mathbb{I}_A \;\; P\text{-a.s.} \quad \text{and hence} \quad P(\mathbb{I}_A f \circ \phi) = P(\mathbb{I}_A \circ \phi^{-1} f) = P(\mathbb{I}_A f) \,.$$

It follows that $P(f_n|\Sigma_{\mathbb{G}_n}) = P(f|\Sigma_{\mathbb{G}_n})$ almost surely, since

$$P(\mathbb{I}_A f_n) = \frac{1}{|\mathbb{G}_n|}\int_{\mathbb{G}_n} P(\mathbb{I}_A f \circ \phi)|d\phi| = \frac{1}{|\mathbb{G}_n|}\int_{\mathbb{G}_n} P(\mathbb{I}_A f)|d\phi| = P(\mathbb{I}_A f)$$

for all $A \in \Sigma_{\mathbb{G}_n}$. In summary, we have shown $f_n = P(f_n|\Sigma_{\mathbb{G}_n}) = P(f|\Sigma_{\mathbb{G}_n})$ almost surely.

Note $\mathbb{G}_n \subset \mathbb{G}_{n+1}$ implies $\Sigma_{\mathbb{G}_n} \supset \Sigma_{\mathbb{G}_{n+1}}$, so by the law of total probability,

$$P(P(f|\Sigma_{\mathbb{G}_n})|\Sigma_{\mathbb{G}_{n+1}}) \;=\; P(f|\Sigma_{\mathbb{G}_{n+1}}) \qquad \text{almost surely.}$$

That shows $(P(f|\Sigma_{\mathbb{G}_n}), \Sigma_{\mathbb{G}_n})_{n\in\mathbb{N}}$ is a reverse martingale, so

$$f_n \;=\; P(f|\Sigma_{\mathbb{G}_n}) \xrightarrow{n\to\infty} P(f|\cap_n \Sigma_{\mathbb{G}_n}) = P(f|\Sigma_{\mathbb{G}}) \qquad \text{almost surely}$$

by the reverse martingale convergence theorem. □

For a finite, labelled graph F, consider the bounded, indeed continuous function f_F on $\mathbf{Z}_{\text{graph}} = \{0,1\}^{\mathcal{N}}$ given by $f_F(H) = \mathbb{I}(F \subset H)$. Here, F is formally considered to have vertex set $\mathbb{N}$ with only finitely many edges and inclusion is understood on the edge set. Then, by rewriting the definition of $t_{\text{inj}}(F,\cdot)$ it follows that for an infinite graph H, finite graph F and $n \geq |V(F)|$,

$$\frac{1}{|\Sigma_{\mathbb{S}_n}|}\sum_{\sigma\in\mathbb{S}_n} f_F(\sigma H) = t_{\text{inj}}(F, H|_n). \tag{46}$$

Thus, by (45), the backwards martingale structure of the injective homomorphism densities is a special case of **Lemma 5.3**, when choosing the function f as f_F.

For the sequence case, similar backwards martingales can be identified by looking at the functions $f_{\mathbf{x}}(X) = \mathbb{I}(\mathbf{x} = (X_1, \dots X_n))$ for each finite assignment $\mathbf{x} = (x_1, \dots, x_n)$

as the initial segment of an infinite sequence. The empirical average corresponds to the backwards martingale obtained from the assignment $\mathbf{x} = (1)$ and has the random limit θ in $[0,1]$. If $|\mathbf{x}| = n$ and $\sum_i x_i = k$, then the limit is $c_k^n(\theta) = \theta^k(1-\theta)^{n-k}$.

Note that **Lemma 5.3** provides a similar backwards martingale for each function f that is integrable under each invariant law on $\mathbf{Z}_{\text{graph}}$ or $\mathbf{Z}_{\text{seq}}$. Our aim is to understand what are the properties of the set of functions $\mathcal{F}_{graph} = \{f_F\colon F \in \mathcal{L}\}$ and $\mathcal{F}_{seq} = \{f_{\mathbf{x}}\colon \mathbf{x} \in \cup_n\{0,1\}^n\}$ that make them suitable for the development of the parallel limit and exchangeability theories that satisfy the properties 1-3. above. Specifically, we aim to construct a parameter space $\mathcal{C}$ and a set of real valued functions $\mathcal{F}$ on $\mathbf{Z}$ and for each $f \in \mathcal{F}$ a function $\pi_f : \mathcal{C} \to \mathbb{R}$, such that a distribution $\mathbb{P}$ on $\mathbf{Z}$ is $\mathbb{G}-$invariant, if and only if there is a necessarily unique measure $\mu_{\mathbb{P}}$ on $\mathcal{C}$, such that

$$\mathbb{P}(f) = \int_{\mathcal{C}} \pi_f(c)\, \mu_{\mathbb{P}}(dc) \text{ for each } f \in \mathcal{F}, \tag{47}$$

and exhibit a statistic Z_n of a random element $Z \sim \mathbb{P}$ such that $Z_n \to Z \in \mathcal{C}$ almost surely with $Z \sim \mu_{\mathbb{P}}$. Then Z_n will correspond to 'random invariant finite objects' like the restrictions $\mathbb{H}|_n$ of infinite exchangeable graphs to the vertex set $[n]$, and Z to their limit.

Firstly, all elements of $\mathcal{F}_{\text{graph}}$ (the analogy with $\mathcal{F}_{\text{seq}}$ will be left implicit in the rest) are continuous and in particular in $L_1(\mathbf{Z}_{\text{graph}}, \mu)$ for any measure μ on $\mathbf{Z}_{\text{graph}}$, as $\mathbf{Z}_{\text{graph}}$ is compact. Also, it can be checked by the Stone-Weierstrass theorem that the linear span of $\mathcal{F}_{\text{graph}}$ is dense in $C(\mathbf{Z}_{\text{graph}})$. Consequently, the values $(\mathbb{P}(f), f \in \mathcal{F}_{\text{graph}})$ determine probability measures on $\mathbf{Z}_{\text{graph}}$. In the general case, since $\mathbf{Z}$ is assumed metrizable, $C(\mathbf{Z})$ is separable, so we can always extract such a countable dense set. This is the first property of $\mathcal{F}$ we require.

Definition 5.4 (Fullness)**.** A countable set $\mathcal{F}$ of continuous functions $\mathbf{Z} \to \mathbb{R}$ is full if its linear span is dense in $C(\mathbf{Z})$.

We now identify the parameter space. Let $F : \mathbf{Z} \to \mathbb{R}^{\mathcal{F}}$ be the Cartesian product of the functions $f \in \mathcal{F}$. For an ergodic random element Z, note that $\mathbb{E}_Z(f(Z)|\Sigma_{\mathbb{G}})$ is almost surely constant for each $f \in \mathcal{F}$ and takes value $\mathbb{E}_Z(f(Z))$. Because $\mathcal{F}$ is countable, this holds jointly for each $f \in \mathcal{F}$, almost surely. Consider the set

$$\mathcal{C}_0 = \{\mathbb{E}_{\mathbb{Q}}(F(Z)) : \mathbb{Q} \text{ is an ergodic measure on } \mathbf{Z}\} \subset \overline{\text{conv}}(F(\mathbf{Z})). \tag{48}$$

We then have the following.

Lemma 5.5. For an $\mathbb{G}$-invariant measure $\mathbb{P}$ on $\mathbf{Z}$, we have that $\mathbb{P}(F(Z)|\Sigma_{\mathbb{G}}) \in \mathcal{C}_0$, almost surely under $\mathbb{P}$.

Proof. Let us write $F(Z_n) = \frac{1}{|\mathbb{G}_n|}\int_{\mathbb{G}_n} F(\phi Z)|d\phi|$ for $Z \in \mathbf{Z}$. By **Lemma 5.3** we have that $F(Z_n) = \mathbb{E}_{\mathbb{P}}(F(Z)|\Sigma_{\mathbb{G}_n})$ and converges to $\mathbb{E}_{\mathbb{P}}(F(Z)|\Sigma_{\mathbb{G}})$ $\mathbb{P}$-almost surely. By the abstract ergodic decomposition theorem **Fact 5.1**,

$$\mathbb{E}_{\mathbb{P}}(F(Z_n) \in \mathcal{C}_0) = \int_{\text{ext}(\mathcal{P}_{\mathbb{G}})} Q(F(Z_n) \in \mathcal{C}_0)\, \alpha(dQ) \tag{49}$$

for a mixing measure α concentrated on the ergodic measures $\text{ex}(\mathcal{P}_{\mathbb{G}})$. Also note that for ergodic Q, the sequence $F(Z_n) \to \mathbb{E}_Q(F(Z))$ almost surely, and so it follows that $Q(Z_n \in \mathcal{C}) \nearrow 1$ for each such Q. Then it follows by the monotone convergence theorem that the left hand side of (49) tends to 1. $\square$

This motivates the choice of $\mathcal{C}_0$ as our parameter space. Then $\mu_{\mathbb{P}}$ is taken as the law of $\mathbb{E}_{\mathbb{P}}(F(Z)|\Sigma_{\mathbb{G}})$ and $\pi_f : \mathcal{C} \to \mathbb{R}$ as the projection on the coordinate given by f. A further property that we require is that any distribution $\mathbb{P}$ on $\mathbf{Z}$ that satisfies (47) is $\mathbb{G}-$invariant. We note the following.

Lemma 5.6. Let $\mathcal{F}$ be full. Suppose for $f \in \mathcal{F}$ we also have $f \circ \phi \in \mathcal{F}$ for each $\phi \in \mathbb{G}$. Then any distribution $\mathbb{P}$ satisfying (47) for each $f \in \mathcal{F}$ is $\mathbb{G}-$invariant.

Proof. Fix $\phi \in \mathbb{G}$. Note that for each ergodic distribution Q, we have $\mathbb{E}_Q(f) = \mathbb{E}_Q(f \circ \phi)$. In particular, the coordinate projections π_f and π_g on the components f and $g = f \circ \phi$ coincide. It thus follows that if $\mathbb{P}$ satisfies (47), then $\mathbb{P}(f) = \mathbb{P}(f \circ \phi)$ for each $f \in \mathcal{F}$. Since $\text{span}(\mathcal{F})$ is dense in $C(\mathbf{Z})$, this holds for each $f \in C(\mathbf{Z})$ and so $\mathbb{P} = \mathbb{P} \circ \phi$. □

Note that for graphs, we have that $\mathbb{I}(F \subset \sigma H) = \mathbb{I}(\sigma^{-1} F \subset H)$ and so in particular, $\mathcal{F}_{\text{graph}}$ has this closure property that we state as an assumption.

Definition 5.7 (Closure)**.** The set $\mathcal{F}$ is closed under the action of $\mathbb{G}$, if for $f \in \mathcal{F}$ and $\phi \in \mathbb{G}$, we also have $f \circ \phi \in \mathcal{F}$.

For graphs, the argument for the uniqueness of the mixing measure $\mu_{\mathbb{P}}$ via the Stone-Weierstrass theorem relies on the compactness of the parameter space and the factorisation property $t(F, W) \cdot t(G, W) = t(FG, W)$ of homomorphism densities. It turns out that both of these properties are a consequence of the dissociatedness of ergodic measures that we define in the general setting in two steps. The first are two structural assumptions motivated by the graph case.

Definition 5.8 (Shift property)**.** The set $\mathcal{F}$ has the shift property, if for each $n \in \mathbb{N}$ and $f \in \mathcal{F}$ there is some $\phi \in \mathbb{G}$ such that $f_n = f \circ \phi$ is $\mathbb{G}_n$-invariant.

Definition 5.9 (Adaptedness)**.** The set $\mathcal{F}$ with the shift property is adapted to the action of $\mathbb{G}$ if for each $g \in \mathcal{F}$, there is an N such that for any $f \in \mathcal{F}$ and $m, n \geq N$, we have that $g \cdot f_n \in \mathcal{F}$ and $g \cdot f_n = (g \cdot f_m) \circ \phi$ for some $\phi \in \mathbb{G}$. We call N the adaptation degree of g.

For a finite graph F, its shifted version F_n is the one whose vertex set is shifted by n. Then for graphs, $\mathcal{F}_{\text{graph}}$ satisfies the shift property with the choice $\mathbb{I}(F \subset \cdot)_n = \mathbb{I}(F_n \subset \cdot)$. Indeed any permutation σ on $\mathbb{N}$ suffices that shifts the support of F and maps 0-s to the first n coordinates. Given another finite graph G, for large enough n, G and F_n are disjoint. So $G \cup F_n$ can be permuted into $G \cup F_m$ for $m > n$ by a permutation that leaves the vertex set of G intact and shifts the vertex set of F_n. This is the content of the adaptedness property. The adaptation degree of $\mathbb{I}(G \subset \cdot)$ is then $|V(G)|$.

Definition 5.10 (Dissociatedness)**.** A distribution $\mathbb{P}$ on $\mathbf{Z}$ is dissociated with respect to an adapted set $\mathcal{F}$, if for each $g \in \mathcal{F}$ with adaptation degree N and another $f \in \mathcal{F}$ and $n \geq N$, we have that

$$\mathbb{P}(gf_n) = \mathbb{P}(g)\mathbb{P}(f_n). \tag{50}$$

For graphs, this corresponds to the property $\mathbb{P}(F \cup G \subset H) = \mathbb{P}(F \subset H) \cdot \mathbb{P}(G \subset H)$ for disjoint finite graphs F and G, which in this case characterises ergodic distributions. It is apparent that if all ergodic distributions on $\mathbf{Z}$ are dissociated with respect to $\mathcal{F}$, then the projections $\{\pi_f : f \in \mathcal{F}\}$ have the factorisation property $\pi_f \cdot \pi_g = \pi_{gf_n}$. The link between dissociatedness and compactness is summarised as follows.

Lemma 5.11. Suppose the set $\mathcal{F}$ of functions on the (compact metric) space $\mathbf{Z}$ is adapted to the action of $\mathbb{G}$. Then the set $\mathcal{P}_{\text{diss}}$ of dissociated distributions on $\mathbf{Z}$ is compact in the weak topology.

Proof. It is well-known that the set $\mathcal{P}$ of all probability measures on a compact metric space $\mathbf{Z}$ is compact and metrizable, see e.g. [Kle20]. So we need to show that $\mathcal{P}_{\text{diss}}$ is weakly closed in $\mathcal{P}$. Because $\mathcal{P}$ is metrizable, it is sufficient to check sequential closure.

So take a sequence $P_m \in \mathcal{P}_{\text{diss}}$ and suppose $P_m \rightharpoonup P$. Take $g \in \mathcal{F}$ with adaptatioin degree N and $n \geq N$. Then for $f \in \mathcal{F}$, we have that $\mathbb{P}_m(gf_n) \to \mathbb{P}(gf_n)$ as $m \to \infty$. On the other hand, $\mathbb{P}_m(gf_n) = \mathbb{P}_m(g)\mathbb{P}_m(f_n) \to \mathbb{P}(g)\mathbb{P}(f_n)$. It follows that $\mathbb{P}$ is dissociated and so $\mathcal{P}_{\text{diss}}$ is weakly sequentially closed as required. □

This leads to the following.

Lemma 5.12. Suppose that the set $\text{ex}(\mathcal{P}_{\mathbb{G}})$ of ergodic measures coincides with $\mathcal{P}_{\text{diss}} \cap \mathcal{P}_{\mathbb{G}}$. Then $\text{ex}(\mathcal{P}_{\mathbb{G}})$ and $\mathcal{C}_0$ are compact and homeomorphic. Furthermore, the integral representation (47) uniquely specifies the mixing measure $\mu_{\mathbb{P}}$.

Proof. We have that $\text{ex}(\mathcal{P}_{\mathbb{G}})$ is weakly compact by **Lemma 5.11**. A similar argument shows that $\mathcal{P}_{\mathbb{G}}$ is compact as well and thus so is their intersection. The map $\tau : \text{ex}(\mathcal{P}_{\mathbb{G}}) \to \mathcal{C}_0$ with $\tau(Q) = \mathbb{E}_Q(F)$ is a continuous bijection from a compact space, so is a homeomorphism, and in particular $\mathcal{C}_0$ is compact. Moreover, by the factorisation property, $\text{span}(\{\pi_f \colon f \in \mathcal{F}\})$ is an algebra that separates points and contains constant functions and so is dense. In particular, the equations in (47) uniquely specify $\mu_{\mathbb{P}}$. □

It is then now clear that our final aim is to structurally characterise when $\text{ex}(\mathcal{P}_{\mathbb{G}}) = \mathcal{P}_{\mathbb{G}} \cap \mathcal{P}_{\text{diss}}$. One direction does not require further assumptions.

Lemma 5.13. If the set $\mathcal{F}$ is adapted to the action of $\mathbb{G}$, then all ergodic measures are dissociated.

Proof. We follow along the proof of **Theorem 5.5** in [DJ08]. Take an ergodic distribution Q on $\mathbf{Z}$. Take $g \in \mathcal{F}$ with adaptation degree N. For any $f \in \mathcal{F}$ and $n \geq N$, we have that

$$\mathbb{E}(gf_n) = \mathbb{E}(\mathbb{E}(g|\Sigma_{\mathbb{G}_n}) \cdot f_n),$$

since f_n is $\Sigma_{\mathbb{G}_n}$-measurable. Note $\mathbb{E}(g|\Sigma_{\mathbb{G}_n}) \to \mathbb{E}(g)$, almost surely, by **Lemma 5.3** since Q is ergodic. Hence, by the dominated convergence theorem,

$$\mathbb{E}((\mathbb{E}(g|\Sigma_{\mathbb{G}_n}) - \mathbb{E}(g)f_n) \to 0$$

as $n \to \infty$. It thus follows that $\mathbb{E}(gf_n) - \mathbb{E}(g)\mathbb{E}(f_n) \to 0$. But by adaptedness, both $\mathbb{E}(f_n)$ $\mathbb{E}(gf_n)$ are unchanged for $n \geq N$ and so the result follows. □

For the converse, we make our last structural assumption on $\mathcal{F}$.

Definition 5.14 (Shift-completeness)**.** The set $\mathcal{F}$ is said to be shift-complete, if

$$\sigma(\{f_n : f \in \mathcal{F}\}) = \Sigma_{\mathbb{G}_n}.$$

For graphs, shift-completeness follows from the fact that finite graphs shifted by n generate infinite graphs on vertex set $[n, \infty)$, which generate $\mathbb{S}_n$-invariant sets.

Lemma 5.15. Suppose that $\mathcal{F}$ is full, adapted to the action of $\mathbb{G}$ and is shift complete. Then any $\mathbb{G}$-invariant dissociated distribution is ergodic.

Proof. Take a dissociated distribution R on $\mathbf{Z}$. For $N \in \mathbb{N}$, let $\mathcal{F}_N \subset \mathcal{F}$ contain elements with dissociation degree N. Since R is dissociated, it follows that $\sigma(\mathcal{F}_N)$ is independent of $\sigma(\{f_n : f \in \mathcal{F}\})$, for any $n > N$, which is $\Sigma_{\mathbb{G}_n}$ by shift completeness. It thus follows, that $\Sigma_{\mathbb{G}}$ is independent of $\sigma(\mathcal{F}_N)$ for each N. But $\cup_n \mathcal{F}_n = \mathcal{F}$ and $\sigma(\mathcal{F}) = \mathcal{B}(\mathbf{Z})$, as $\mathcal{F}$ is full. Hence $\Sigma_{\mathbb{G}} \subset \mathcal{B}(\mathbf{Z})$ is independent of itself and so is trivial. If further R is $\mathbb{G}$-invariant, it follows that it is ergodic. □

We are now ready to put everything together into a final statement.

Theorem 5.16. Let $\mathbb{G}$ be the direct limit of the compact groups $\mathbb{G}_n$ that act measurably on the compact metric space $\mathbf{Z}$. Suppose that there is a countable set $\mathcal{F}$ of continuous functions on $\mathbf{Z}$ that is full, closed under and adapted to the action of $\mathbb{G}$ and is shift-complete.

Then a $\mathbb{G}$-invariant measure is ergodic, if and only if it is dissociated with respect to $\mathcal{F}$. The set $\mathrm{ex}(\mathcal{P}_{\mathbb{G}})$ of ergodic measures is compact and has an embedding $\tau : \mathrm{ex}(\mathcal{P}_{\mathbb{G}}) \hookrightarrow \mathbb{R}^{\mathcal{F}}$ with image $\mathcal{C}$. A distribution $\mathbb{P}$ on $\mathbf{Z}$ is $\mathbb{G}$-invariant if and only if there is a measure $\mu_{\mathbb{P}}$ on $\mathcal{C}$ such that

$$\mathbb{P}(f) = \int_{\mathcal{C}} \pi_f(c)\mu_{\mathbb{P}}(dc) \text{ for each } f \in \mathcal{F}, \tag{51}$$

for some continuous functions π_f on $\mathcal{C}$. The measure $\mu_{\mathbb{P}}$ is then unique. Moreover, $f(Z_n) = \frac{1}{|\mathbb{G}_n|}\int_{\mathbb{G}_n} f(\phi Z)|d\phi| \to \pi_f(C)$ $\mathbb{P}$-almost surely, where $C \in \mathcal{C}$ is a random element with distribution $\mu_{\mathbb{P}}$.

Proof. Take a probability measure $\mathbb{P}$ on $\mathbf{Z}$. Since $\mathcal{F}$ is full, the integral equations (51) uniquely specify $\mathbb{P}$. As above, let $F : \mathbf{Z} \to \mathbb{R}^{\mathcal{F}}$ be the Cartesian product of the functions $f \in \mathcal{F}$ and let $\mathcal{C} = \{\mathbb{E}_Q(F) : Q \in \mathrm{ex}(\mathcal{P}_{\mathbb{G}})\}$. First, if (51) holds, then by the closedness of $\mathcal{F}$ under $\mathbb{G}$, **Lemma 5.6** implies that $\mathbb{P}$ is $\mathbb{G}$-invariant.

Suppose now that $\mathbb{P}$ is $\mathbb{G}$-invariant. By **Lemma 5.5**, we have that $\mathbb{E}_{\mathbb{P}}(F|\Sigma_{\mathbb{G}_n}) = F(Z_n) \to \mathbb{E}_{\mathbb{P}}(F|\Sigma_{\mathbb{G}}) \in \mathcal{C}$ almost surely. Let $\mu_{\mathbb{P}}$ be the law of $\mathbb{E}_{\mathbb{P}}(F|\Sigma_{\mathbb{G}})$. Take the coordinate projection function $c_f : \mathcal{C} \to \mathbb{R}$ that projects $c \in \mathcal{C}$ to its compomponent indexed by f. Then

$$\mu_{\mathbb{P}}(c_f) = \mathbb{E}\left(\mathbb{E}_{\mathbb{P}}(f|\Sigma_{\mathbb{G}})\right) = \mathbb{P}(f)$$

by the law of total probability, so $\mu_{\mathbb{P}}$ satisfies (51) with the choice of π_f as the coordinate projections c_f. Now the adaptedness and shift-completeness of $\mathcal{F}$ imply by **Lemma 5.13** and **Lemma 5.15** that $\mathrm{ex}(\mathcal{P}_{\mathbb{G}})$ coincides with the set $\mathcal{P}_{\mathbb{G}} \cap \mathcal{P}_{\mathrm{diss}}$ of $\mathcal{F}$-dissociated measures. Then, by **Lemma 5.12**, $\mathrm{ex}(\mathcal{P}_{\mathbb{G}})$ is compact and homeomorphic to $\mathcal{C}$ and the equations (51) uniquely specify $\mu_{\mathbb{P}}$. □

5.1 Invariance under $\mathbb{G}_n$

In this subsection, we discuss an analog of finite exchangeability in the framework presented.

6 Information projections and exchangeable binary sequences

In this section, we introduce the connection between exchangeability and the theory of maximum entropy distributions developed in, among others, [Csi75], [Csi84][CM03], [CM01] [CM08]. In this literature, the authors study probability measures Q satisfying

$$Q = \mathrm{argmin}_{P \in \mathcal{C}} \, \mathrm{kl}(P \| \mu), \tag{52}$$

where $\mathcal{C}$ is a convex subset of probability measures on a measurable space $(S, \mathcal{S})$ and $\mu \notin \mathcal{C}$ is another probability measure. Such Q is called the information projection of μ on $\mathcal{C}$. Of particular interest is the case when $\mathcal{C}$ is taken as a *linear set* $L_{\mathbf{a},\mathbf{f}}$, where

$$L_{\mathbf{a},\mathbf{f}} = \{P\colon \int_S f_i(s)P(ds) = a_i\} \tag{53}$$

for vectors of measurable functions $\mathbf{f} = (f_1, \dots f_n)$ and scalars $\mathbf{a} = (a_1, \dots a_n)$. In [Csi75] and [CM03] the authors propose precise conditions for the existence of an the information projection Q and study its density with respect to μ. They show that at least approximately, Q is in the exponential family

$$\mathcal{E}_{\mu,f} = \big\{Q_\theta\colon \frac{dQ_\theta}{d\mu}(x) = e^{\langle \theta, f(x)\rangle - \Lambda_f(\theta)},\ \theta \in \mathrm{dom}(\Lambda_f)\big\}, \tag{54}$$

where $\Lambda_f(\theta) = \int_X e^{\langle \theta, f(x)\rangle} d\mu$ and $\mathrm{dom}(\Lambda_f) = \{\theta\colon \Lambda_f(\theta) < \infty\}$.

Note that for a an infinite exchangeable random sequence $\mathbf{Y}$, for each n, the mixing measure μ in de Finetti's theorem can be viewed as an element $\mu \in L_{\mathbf{a},\mathbf{f}}$, where $a_k = p_k^n$ is the law vector of $(Y_1, \dots Y_n)$ and $f_k(s) = c_k^n(s)$ and $S = [0, 1]$. Further, note that any $\nu' \in L_{\mathbf{a},\mathbf{f}}$ is the mixing measure of an infinite exchangeable sequence $\mathbf{Z}$ with $(Z_1, \dots Z_n) \overset{d}{=} (Y_1, \dots Y_n)$.

This connection enables us to study the following questions. Given a finitely exchangeable random sequence $\mathbf{X} \in \{0, 1\}^n$, how large is the set $\mathcal{R}_{\mathbf{X}}^{n,+}$ of possible positive mixing measures for $\mathbf{X}$? Also, given a Borel measure μ, which $\mu_0 \in \mathcal{R}_{\mathbf{X}}^{n,+}$ is the closest to μ in the kl-sense? Note that by de Finetti's theorem, we need that $\mathbf{X}$ is infinitely extendible for $\mathcal{R}_{\mathbf{X}}^{n,+}$ to be nonempty. Recall from **Section 1.1** that this is equivalent to $p_{\mathbf{X}} \in \mathrm{conv}(\mathbf{c}([0, 1])$, where $\mathbf{c}^n(\theta) = (c_1^n(\theta), \dots, c_n^n(\theta))$ is the i.i.d. curve.

Theorem 6.1. Let $\mathbf{X}$ be a finite exchangeable sequence and μ a measure on $[0, 1]$ with infinite support s_μ Then $\mathrm{int}(\mathrm{conv}(\mathbf{c}(s_\mu)))$ is not empty and there is some $\mu_1 \in \mathcal{E}_{\mu,\mathbf{c}}$ that is also in the set of possible mixing measures $\mathcal{R}_{\mathbf{X}}^{n,+}$, if and only if

$$p_{\mathbf{X}} \in \mathrm{int}(\mathrm{conv}(\mathbf{c}(s_\mu))). \tag{55}$$

Recall the discussion in **Section 1.1** with the geometric interpretation of this result and the possible interpretation of this theorem that the set $\mathcal{R}_{\mathbf{X}}^{n,+}$ is as large as possible.

The proof of **Theorem 6.1** is delayed until the end of this section. The boundary case $p_{\mathbf{X}} \in \partial(\mathrm{conv}(\mathbf{c}(\mathrm{supp}(\mu))))$ is more delicate and requires a notion of extension of the exponential family discussed below. This is the subject of **Theorem 6.12**.

We now develop the background on information projections to the extent neccessary for the proofs of **Theorem 6.1 and 6.12**. For further details, we refer to [CM03], [Csi75] and [CM01]. We start with the naive existence result of [Csi75]. Recall the total variation norm of (signed) defined in (7). We use the shorthand $\mathrm{kl}(\mathcal{C}\|\mu) = \inf_{\eta \in \mathcal{C}} \mathrm{kl}(\eta\|\mu)$.

Fact 6.2 (Csiszár)**.** Let $\mathcal{C}$ be a convex set of probability measures that is total variation closed and let $\mu \notin \mathcal{C}$ be another probability measure. Suppose

$$\mathrm{kl}(\mathcal{C}\|\mu) < \infty. \tag{56}$$

Then $\mathcal{C}$ admits a unique information projection Q under μ.

Recall that in the case $S = [0, 1]$ the Riesz Representation Theorem implies that the total variation norm is the dual space norm of linear functionals on $C([0, 1])$. Since the functions c_i^n are continuous, it follows that the linear sets $L_{\mathbf{a},\mathbf{f}}$ as in (53) with $f_i = c_n^i$ are weak-* closed. So they are certainly norm (and thus total-variation) closed. Thus **Fact 6.2** implies that a measure μ on $[0, 1]$ has a unique information projection on $L_{\mathbf{p_X},\mathbf{c}}$, if $L_{\mathbf{p_X},\mathbf{c}}$ and μ satisfy (56). In [CM01] and [CM03] the authors came up with a device, the *convex core* $\mathrm{cc}(\mu)$ of a measure to characterise when this is the case in terms of the vector of values $\mathbf{a} = \mathbf{p}_X$ and the function $\mathbf{f} = \mathbf{c}$. The definitions are stated for $S = \mathbb{R}^d$, but as we will see, we can transfer them to more general spaces.

Definition 6.3 (Csiszár, Matus)**.** For a finite Borel measure ν on $\mathbb{R}^d$, its *convex support* is the intersection of all closed and full-measure sets, denoted $\mathrm{cs}(\nu)$. The intersection of all full-measure, convex sets is called the *convex core*, written as $\mathrm{cc}(\nu)$.

In [CM01], convex cores of measures are characterised as those convex sets that have at most countably many faces. Thus they are in particular Borel. The convex core and the well-understood convex support only differ on their boundaries.

Fact 6.4 (Csiszár, Matus)**.** We have $\mathrm{cl}(\mathrm{cc}(\mu)) = \mathrm{cs}(\mu)$, where cl stands for the topological closure. Also, their relative interiors coincide, i.e $\mathrm{ri}(\mathrm{cc}(\mu)) = \mathrm{ri}(\mathrm{cs}(\mu))$.

The crucial characterisation for the convex core is the following result in [CM01]. It says that the convex core contains the right amount of boundary to contain means of absolutely continuous measures.

Fact 6.5 (Csiszár, Matus)**.** We have

$$\mathrm{cc}(\nu) = \left\{ \int_{\mathbb{R}^d} x\, dP \colon\ P \ll \nu,\ P \text{ a probability measure with a mean} \right\}. \tag{57}$$

Moreover, for each $a \in \mathrm{cc}(\nu)$, there exists such P with $\frac{dP}{d\nu}$ bounded.

Let us write $L_{\mathbf{a}} = L_{\mathbf{a},\mathbf{f}}$ with $\mathbf{f}$ taken as the identity on $\mathbb{R}^d$. In [CM03] they then point out the following corollary. We recite the proof because of the significance of this result for our discussion.

Fact 6.6 (Csiszár, Matus)**.** . For the linear set $L_{\mathbf{a}}$ and a finite measure μ, we have

$$\mathrm{kl}(L_{\mathbf{a}} \| \mu) < \infty,$$

if and only if $\mathbf{a} \in \mathrm{cc}(\mu)$.

Proof. Suppose $\mathbf{a} \in \mathrm{cc}(\mu) \subset \mathbb{R}^d$. Then by**Fact 6.5** there is $P \in L_{\mathbf{a}}$ that has a bounded Radon-Nikodym derivative f. Since μ is finite, then

$$\mathrm{kl}(P\|\mu) = \int_{\mathbb{R}^d} f \log(f) d\mu < \infty.$$

Conversely, if $\mathbf{a} \notin \mathrm{cc}(\mu)$, then again by **Fact 6.5**, there is no $P \in L_{\mathbf{a}}$ with $P << \mu$ and so by definition, for all $P \in L_{\mathbf{a}}$ we have $\mathrm{kl}(P\|\mu) = \infty$. □

Our aim is now to transfer **Fact 6.6** to the more general linear sets $L_{\mathbf{a},\mathbf{f}}$. This can be done by pushforward and pullback constructions as shown in [Csi84]. We follow [CM03] in our outline and notation. Let $g \colon S \to \mathbb{R}^n$ be measurable. For a probability measure μ on S, we write μ_g for its image measure on $\mathbb{R}^n$. For a measure $\tilde{\nu} << \mu_g$ on $\mathbb{R}^n$, we write $\tilde{\nu}_{g^{-1},\mu}$ for the measure on S with μ-density $\frac{d\tilde{\nu}}{d\mu_g}(g(x))$, where For a convex set $\mathcal{C}$ of probability measures on $\mathbb{R}^n$, let

$$\mathcal{C}_{g^{-1}} = \{P \in \mathcal{P}(S) \colon P_g \in \mathcal{C}\}. \tag{58}$$

Fact 6.7 (Csiszár)**.** For μ and $\mathcal{C}$ and g as above, we have

$$\mathrm{kl}(\mathcal{C}_{g^{-1}}\|\mu) = \mathrm{kl}(\mathcal{C}, \|\mu_g). \tag{59}$$

Moreover, μ_g has information projection $\tilde{\eta}$ on $\mathcal{C}$ if and only if $\tilde{\eta}_{g^{-1},\mu}$ is the information projection of μ on $\mathcal{C}_{g^{-1}}$.

Now notice that the linear set $L_{\mathbf{a},\mathbf{f}}$ we have that $L_{\mathbf{a},\mathbf{f}} = (L_{\mathbf{a}})_{\mathbf{f}^{-1}}$. We can thus conclude the following.

Fact 6.8 (Csiszár, Matus)**.** For the linear set $L_{\mathbf{a},\mathbf{f}}$ on the measurable space $(S, \mathcal{S})$ and finite measure μ on S, we have $\mathrm{kl}(L_{\mathbf{a},\mathbf{f}}|\mu) < \infty$ if and only of $\mathbf{a} \in \mathrm{cc}(\mu_{\mathbf{f}})$.

Using the connection with exchangeability and **Fact 6.2**, we conclude the following lemma.

Lemma 6.9. Given a finitely exchangeable binary sequence $\mathbf{X} = (X_1, \dots X_n)$ with law-vector $\mathbf{p}_X$ and a measure μ on $[0,1]$. Let $\mathbf{c}(p) \to (p^n, p^{n-1}p, \dots (1-p)^n)$ be as before. Then μ has an information projection on the set $\mathcal{R}^{n,+}_{\mathbf{X}}$ of possible positive mixing measures for $\mathbf{X}$, if and only if

$$\mathbf{p}_{\mathbf{X}} \in \mathrm{cc}(\mu_{\mathbf{c}}). \tag{60}$$

Then the information projection Q is unique.

Proof. Because $[0,1]$ is compact and Hausdorff, we have seen that the unique information projection on the convex set $\mathcal{R}^{n,+}_{\mathbf{X}} = L_{\mathbf{p}_{\mathbf{X}},\mathbf{c}}$, exists, if and only if $\mathrm{kl}(\mathcal{R}^{n,+}_{\mathbf{X}}\|\mu) < \infty$. Thus, by **Fact 6.8**, we conclude. □

If exists, then $Q << \mu$ for the information projection Q of μ on $L_{\mathbf{a},\mathbf{f}}$. Our aim now is to characterise its μ-density. **Fact 6.11** will give us that it comes from an extension of the exponential family $\mathcal{E}_{\mu,\mathbf{f}}$ to measures supported on the faces of $\mathrm{cc}(\mu_{\mathbf{f}})$ defined in [CM03].

Let μ be a measure on S and $\mathbf{f}\colon S \to \mathbb{R}^n$ be measurable. Let $F \subset \mathrm{cc}(\mu_{\mathbf{f}})$ be a face of the convex set $\mathrm{cc}(\mu_{\mathbf{f}}) \subset \mathbb{R}^n$. Let μ_F be the restriction of μ to the set $\mathbf{f}^{-1}(\mathrm{cl}(F))$. The following Lemma from [CM01] is a key technical tool.

Fact 6.10 (Csiszár, Matus)**.** Let F be a face of $\mathrm{cc}(\mu_{\mathbf{f}})$. The measure μ_F has convex core F. In particular, it is a nontrivial measure.

Thus exponential families with base measure μ_F are well-defined. A member of $\mathcal{E}_{\mu_F,\mathbf{f}}$ has a μ-density $e^{\langle\theta,\mathbf{f}(s)\rangle - \Lambda^F_{\mathbf{f}}(\theta)}$ for $s \in \mathbf{f}^{-1}(F)$ and 0 otherwise. Here $\Lambda^F_{\mathbf{f}}(\theta) = \int_S e^{\langle\theta,\mathbf{f}(s)\rangle}\mu(ds)$. Then

$$\mathrm{ext}(\mathcal{E}_{\mu,\mathbf{f}}) = \bigcup_{F \text{ is a face of } \mathrm{cc}(\mu_{\mathbf{f}})} \mathcal{E}_{\mu_F,\mathbf{f}} \tag{61}$$

is called the *extended exponential family* with base measure μ and sufficient statistic $\mathbf{f}$.

Observe that **Fact 6.10** implies that the extreme points of $\mathrm{cc}(\mu_{\mathbf{f}})$ are atoms of $\mu_{\mathbf{f}}$. In particular, if $\mu_{\mathbf{f}}$ is non-atomic, there are no non-trivial faces F and we have $\mathrm{ext}(\mathcal{E}_{\mu,\mathbf{f}}) = \mathcal{E}_{\mu,\mathbf{f}}$. In general, the extension is a way of forming a closure of the exponential families by including densities supported on convex hulls of atoms. See [CM05] for details.

Fact 6.11 (Csiszár, Matus)**.** Let the linear set $L_{\mathbf{a},\mathbf{f}}$ and the measure μ be such that μ has an information projection Q on $L_{\mathbf{a},\mathbf{f}}$. Then $Q \in \mathrm{ext}(\mathcal{E}_{\mu,\mathbf{f}})$. Furthermore, Q belongs to the component based on μ_F, where F is the face with $\mathbf{a} \in \mathrm{ri}(F)$. Moreover, Q satisfies the Pythagorean identity

$$\mathrm{kl}(P\|\mu) = \mathrm{kl}(P\|Q) + \mathrm{kl}(Q\|\mu) \ \text{ for } P \in L_{\mathbf{a},\mathbf{f}}. \tag{62}$$

Note in particular that if $\mathbf{a} \in \mathrm{ri}(\mathrm{cc}(\mu))$, then Q is in the actual exponential family $\mathcal{E}_{\mu,\mathbf{f}}$. We are now ready to state and prove our general result.

Theorem 6.12. Let $\mathbf{X}$ be a finite exchangeable binary sequence and μ a Borel measure on $[0,1]$, such that $\mathbf{p_X} \in \mathrm{cc}(\mu_\mathbf{c})$. Then there is some μ_1 in the extended exponential family $\mathrm{ext}(\mathcal{E}_{\mu,\mathbf{c}})$ that is a possible mixing measure for $\mathbf{X}$. Furthermore, μ_1 belongs to the component based on μ_F, where F is the face with $\mathbf{p_X} \in \mathrm{ri}(F)$. In particular, if $\mathbf{p_X} \in \mathrm{ri}(\mathrm{cc}(\mu_\mathbf{c})$, then $\mu_1 \in \mathcal{E}_{\mu,\mathbf{c}}$.

Proof. By **Lemma 6.9**, the information projection μ_1 of μ exists on the set of possible mixing measures $\mathcal{R}^{n,+}_\mathbf{X} = L_{\mathbf{p_X},\mathbf{c}}$, if and only if $\mathbf{p_X} \in \mathrm{cc}(\mu_\mathbf{c})$. We can conclude by **Fact 6.11**. Note that this result also provides the Pythagorean identity (62). ☐

To obtain the more tangible results like **Theorem 6.1** stated above, our aim now is to evaluate the convex core $\mathrm{cc}(\mu_\mathbf{c})$. This is the content of the following results. They are adapted from [CM03], where the function $\theta \to (\theta, \theta^2, \ldots, \theta^n)$ is used as an example, instead of our $\mathbf{c}(\theta) = (\theta^n, (1-\theta)\theta^{n-1}, \ldots, (1-\theta)^n)$.

Corollary 6.13. Suppose that μ has finite support s_μ. Then $\mathrm{cc}(\mu_\mathbf{c}) = \mathrm{cs}(\mu_\mathbf{c})$ and is a polytope. The extended exponential family $\mathrm{ext}(\mathcal{E})$ coincides with the closure of $\mathcal{E}_{\mu,\mathbf{c}}$, when viewed as a subset of $\mathbb{R}^T$.

Corollary 6.14. Suppose that μ has infinite support s_μ. Let $Y \subset s_\mu$ be the set of atoms of μ.

i) The convex set $\mathrm{cc}(\mu_\mathbf{c})$ has a non-empty interior which equals to that of $\mathrm{conv}(\mathbf{c}(s_\mu))$.

ii) Each proper face F of $\mathrm{cc}(\mu_F)$ equals a simplex $\mathrm{conv}(K)$ with $K \subset \mathbf{c}(Y)$ of size at most n. In addition $\mu_{\mathbf{c}F} = \mu_{\mathbf{c}K}$.

iii) Each set $K \subset \mathbf{c}(Y)$ of size $\leq \frac{n}{2}$ spans a face of $\mathrm{cc}(\mu_\mathbf{c})$.

Proof. We only prove part i), which we directly need for the proof of **Theorem 6.1**, for the rest we refer to [CM03]. First of all, we observe that

$$\mathrm{cc}(\mu_\mathbf{c}) \subset \mathrm{conv}(\mathbf{c}(s_\mu)) = \overline{\mathrm{conv}}(\mathbf{c}(s_\mu)) = \mathrm{cs}(\mu_\mathbf{c}).$$

Indeed, $\mathbf{c}(s_\mu)$ is a compact set of full measure and its convex hull is again a compact subset of $\mathbb{R}$, which is also full and convex. Thus $\mathrm{cc}(\mu_\mathbf{c}) \subset \mathrm{cs}(\mu_\mathbf{c}) \subset \mathrm{conv}(\mathbf{c}(s_\mu))$. But clearly $\mathrm{conv}(\mathbf{c}(\mathrm{s}_\mu)$ is a subset of any closed convex set of full measure, so it follows that $\mathrm{conv}(\mathbf{c}(\mathrm{s}_\mu)) \subset \mathrm{cs}(\mu_\mathbf{c})$. By **Fact 6.4**, the sets $\mathrm{cc}(\mu_\mathbf{c})$ and $\mathrm{conv}(\mathbf{c}(s_\mu))$ share the same relative interior. We can thus conclude if we can show that the latter has nonempty interior. We closely follow the a argument in [CM03].

The i.i.d curve $\mathbf{c}$ intersects any hyperplane $H = \{\tilde{\theta} \colon \langle \mathbf{d}, \tilde{\theta} \rangle = r\} \subset \mathbb{R}^{n+1}$ in at most n points, because $\mathbf{c}(\theta) \in H$ implies that θ is a real root of the polynomial $\sum_i d_i \theta^i (1-\theta)^{n-i} - r$ of degree at most n. This implies that any $1 \leq k \leq n+1$ points on the i.i.d. curve $\mathbf{c}(\theta)$ are affinely independent, i.e. span a simplex of dimension $k-1$. In particular, $\mathrm{conv}(\mathbf{c}(s_\mu))$ has nonempty interior, because s_μ is infinite. ☐

We can finally conclude **Theorem 6.1** as a consequence of the first part of **Corollary 6.14** and **Theorem 6.12**.

Proof of Theorem 6.1. By **Corollary 6.14 i)**, the interior of $\mathrm{cc}(\mu_\mathbf{c})$ is not empty and equals $\mathrm{conv}(\mathbf{c}(s_\mu))$. If $\mathbf{X}$ and μ are such that $\mathbf{p_X} \in \mathrm{int}(\mathrm{cc}(\mu_\mathbf{c}))$, then the information projection μ_1 of μ on the set $L_{\mathbf{p_X},\mathbf{c}}$ of possible mixing measures for $\mathbf{X}$ is in the (non-extended) exponential family $\mathcal{E}_{\mu,\mathbf{c}}$ by **Theorem 6.12**. This μ_1 then satisfies the requirements. ☐

7 Information projection and exchangeable graphs

In this section, we adapt the arguments in **Section 6** to obtain analogous information projection results for exchangeable graphs. In this case the density of the mixing measure will come from an exonential random graph distribution. Since the the applications of the information projection results are completely analogous, we only state the main results and refer to **Section 6** for more details.

We start by connecting an exchangeable graphs to the geometric view presented in Section 1.1. Recall the graphon distributions, which are the infinite exchangeable graph distributions that correspond to a deterministic graphon, i.e. which satisfy

$$P(\mathbb{H}|_{[n]} = G) = t_{\mathrm{ind}}(G, W)$$

for a fixed $W \in \overline{\mathcal{W}}$. By the de Finetti theorem for graphs **Fact 3.7**, these are the ergodic exchangeable graph distributions. For graphs on n vertices, the exchangeable distributions form a simplex, of dimension $m = |\mathcal{U}_n| - 1$ and with extreme points as uniform distributions on isomorphism classes of graphs in $\mathcal{L}_n$, see [LRS18a]. The ergodic distributions again form a parametric curve $\mathcal{C}$ in the simplex, given by

$$\begin{aligned} \mathbf{ind}^n \colon \overline{\mathcal{W}} &\to [0,1]^m \\ \mathbf{ind}^n(W) &= (t_{\mathrm{ind}}(G_1, W), \dots t_{\mathrm{ind}}(G_m, W)),\ G_i \in \mathcal{U}_n. \end{aligned}$$

By **Fact 3.7**, the distributions in the convex hull of the curve $\mathcal{C}$ are exactly the exchangeable graph distributions that are infinitely extendible. The curves $\mathbf{ind}^n$ (and coordinate projections thereof) are of great importance in extremal graph theory and are quite delicate objects, see [Lov12] Chapter 16. Let $\mathbb{F} \in \mathcal{L}_n$ be a finitely exchangeable random graph. We write $\mathbf{p}_{\mathbb{F}}$ for the law-vector of $\mathbb{F}$ indexed by graphs $G \in \mathcal{L}_n$, i.e.

$$\mathbf{p}_{\mathbb{F}}[G] = P(\mathbb{F} = G).$$

Fact 3.7 then gives us that if $\mathbf{p}_{\mathbb{F}} \in \mathrm{conv}(\mathbf{ind}^n(\overline{\mathcal{W}}))$, then there is a probability measure μ in the linear set $L_{\mathbf{a},\mathbf{f}}$, with $\mathbf{a} = \mathbf{p}_{\mathbb{F}}$ and $\mathbf{f} = \mathbf{ind}^n$ of possible mixing measures for $\mathbb{F}$. We also writhe $\mathcal{R}^{n,+}_{\mathbb{F}}$ for this set. Since $\overline{\mathcal{W}}$ is compact, we can again conclude, using **Fact 6.8**, a Lemma about existence of the information projection.

Lemma 7.1. Let $\mathbb{F} \in \mathcal{L}_n$ be a finite exchangeable graph with law vector $\mathbf{p}_{\mathbb{F}}$ and a measure μ on $\overline{\mathcal{W}}$.. Let $\mathbf{ind}^n(W) \to (t_{\mathrm{ind}}(G_1, W), \dots, t_{\mathrm{ind}}(G_m, W))$ be the ergodic curve as before. Then μ has an information projection on the set $\mathcal{R}^{n,+}_{\mathbb{F}}$ of possible positive mixing measures for $\mathbb{F}$, if and only if $\mathbf{p}_{\mathbb{F}} \in \mathrm{cc}(\mu_{\mathbf{ind}^n})$.

Following the same steps, we also conclude the abstract extended exponential density result for the projection.

Theorem 7.2. Let $\mathbb{F} \in \mathcal{L}_n$ be a finite exchangeable graph and μ a Borel measure on $\overline{\mathcal{W}}$, such that $\mathbf{p}_{\mathbb{F}} \in \mathrm{cc}(\mu_{\mathbf{ind}^n})$. Then there is some μ_1 in the extended exponential family $\mathrm{ext}(\mathcal{E}_{\mu,\mathbf{ind}^n})$, which is a possible mixing measure for $\mathbb{F}$. Furthermore, μ_1 belongs to the component based on μ_F, where F is the face with $\mathbf{p}_{\mathbb{F}} \in \mathrm{ri}(F)$.

The real challenge is to evaluate the convex core of $\mu_{\mathbf{ind}^n}$, which is a hard problem, given the complexity of the curve $\mathbf{ind}^n(\overline{\mathcal{W}})$. Partial steps can be made in this direction along the lines of **Corollary 6.14**. Using an old result from graph theory in [ELS79], we can conclude a similar result to **Corollary 6.14** *i*). The result in [ELS79] is stated in terms of the standard homomorphism densities of graphs, we restate it in the graph limit language. See also [Lov12], [DGKR15] and [LRS18a].

Fact 7.3 (Erdős, Lovász, Spencer)**.** The set $\mathrm{ind}^n(\overline{\mathcal{W}})$ contains an open ball B in $\mathbb{R}^m$.

Proposition 7.4. If μ is a measure on $\overline{\mathcal{W}}$ of full support, then $\mathrm{cc}(\mu^n_{\mathbf{ind}})$ has a nonempty interior that equals that of $\mathrm{conv}(\mathbf{ind}^n(\overline{\mathcal{W}}))$.

Proof. We can similarly argue as in the proof of **Corollary 6.14 i)** that the two interiors coincide. **Fact 7.3** then gives us that $\mathrm{int}(\mathrm{conv}(\mathbf{ind}^n(\overline{\mathcal{W}})))$ is not empty. □

Note that unlike in the proof of **Corollary 6.14 i)** we cannot resort to the argument that polynomials have finite number of roots. We thus needed to assume a full support to be able to use **Fact 7.3**. Putting these results all together, we can finally conclude a similar result to **Theorem 6.1**.

Theorem 7.5. Let $\mathbb{F}$ be a finite exchangeable graph in $\mathcal{L}_n$ and μ a measure on $\overleftarrow{\mathcal{W}}$ of full support. Then $\mathrm{int}(\mathrm{conv}(\mathbf{ind}^n(\overline{\mathcal{W}})))$ is not empty and there is some $\mu_1 \in \mathcal{E}_{\mu,\mathbf{ind}^n}$ that is also in the set of possible mixing measures $\mathcal{R}^{n,+}_{\mathbb{F}}$, if and only if

$$p_{\mathbb{F}} \in \mathrm{int}(\mathrm{conv}(\mathbf{ind}^n(\overline{\mathcal{W}}))). \tag{63}$$

The densities in the exponential family $\mathcal{E}_{\mu,\mathbf{ind}^n}$ are the so called exponential random graph distributions. They are of great interest for statistical inference in network models, see e.g. [CD13] and references therein and [LRS18b] for a connection to exchangeability. Our result states that a mixing measure μ_1 for an infinitely finite extendible exchangeable graph $\mathbb{F} \in \mathcal{L}_n$ can be chosen from the exponential family random graphs with any base measure μ of full measure. That is, there is a parameter $\theta \in \mathbb{R}^m$, such that

$$P(\mathbb{F}|_{[n]} = G) = \int_{\overline{\mathcal{W}}} t_{\mathrm{ind}}(G, W) e^{\sum_j^m \theta_j t_{\mathrm{ind}}(F_i) - \psi(\theta)}\, \mu(dW). \tag{64}$$

Here $m = |\mathcal{U}_n|$ as usual, and $F_i \in \mathcal{U}_n$.

References

[Ald81] David J. Aldous, *Representations for partially exchangeable arrays of random variables*, Journal of Multivariate Analysis **11** (1981), no. 4, 581–598.

[AO22] Morgane Austern and Peter Orbanz, *Limit theorems for distributions invariant under groups of transformations*, The Annals of Statistics **50** (2022), no. 4, 1960–1991, Publisher: Institute of Mathematical Statistics.

[Aus08] Tim Austin, *On exchangeable random variables and the statistics of large graphs and hypergraphs*, Probability Surveys **5** (2008), no. 1, 80–145, arXiv:0801.1698 [math].

[Aus12] ______, *Exchangeable random arrays*, 2012.

[BC89] Ranko Bojanic and Fuhua Cheng, *Rate of convergence of Bernstein polynomials for functions with derivatives of bounded variation*, Journal of Mathematical Analysis and Applications **141** (1989), no. 1, 136–151.

[BCL+06] Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi, *Counting Graph Homomorphisms*, Topics in Discrete Mathematics (Berlin, Heidelberg) (Martin Klazar, Jan Kratochvíl, Martin Loebl, Jiří Matoušek, Pavel Valtr, and Robin Thomas, eds.), Springer, 2006, pp. 315–371 (en).

[BCL+08] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Advances in Mathematics **219** (2008), no. 6, 1801–1851.

[BRW14] A. Barron, T. Roos, and K. Watanabe, *Bayesian properties of normalized maximum likelihood and its fast computation*, IEEE International Symposium on Information Theory - Proceedings, IEEE Press, 2014, pp. 1667–1671.

[BS98] Bruno Bassan and Marco Scarsini, *Finite row-column exchangeable arrays*, Commentationes Mathematicae Universitatis Carolinae **39** (1998), no. 1, 137–145 (en).

[CD13] Sourav Chatterjee and Persi Diaconis, *Estimating and understanding exponential random graph models*, The Annals of Statistics **41** (2013), no. 5, 2428–2461, Publisher: Institute of Mathematical Statistics.

[CM01] Imre Csiszár and František Matúš, *Convex cores of measures on R^d*, Studia Scientiarum Mathematicarum Hungarica **38** (2001), no. 1-4, 177–190.

[CM03] ______, *Information projections revisited*, IEEE Transactions on Information Theory **49** (2003), 1474–1490.

[CM05] ______, *Closures of exponential families*, The Annals of Probability **33** (2005), no. 2, 582–600.

[CM08] ______, *On minimization of entropy functionals under moment constraints*, 2008 IEEE International Symposium on Information Theory, July 2008, ISSN: 2157-8117, pp. 2101–2105.

[Csi75] Imre Csiszár, *I-Divergence Geometry of Probability Distributions and Minimization Problems*, The Annals of Probability **3** (1975), no. 1, 146–158, Publisher: Institute of Mathematical Statistics.

[Csi84] ______, *Sanov Property, Generalized I-Projection and a Conditional Limit Theorem*, The Annals of Probability **12** (1984), no. 3, 768–793, Publisher: Institute of Mathematical Statistics.

[DF80] Persi Diaconis and David Freedman, *Finite Exchangeable Sequences*, The Annals of Probability **8** (1980), no. 4, 745–764, Publisher: Institute of Mathematical Statistics.

[DF20] Emanuele Dolera and Stefano Favaro, *Rates of convergence in de Finetti's representation theorem, and Hausdorff moment problem*, Bernoulli **26** (2020), no. 2, 1294–1322, Publisher: Bernoulli Society for Mathematical Statistics and Probability.

[DGKR15] Peter Diao, Dominique Guillot, Apoorva Khare, and Bala Rajaratnam, *Differential calculus on graphon space*, Journal of Combinatorial Theory, Series A **133** (2015), 183–227.

[Dia77] Persi Diaconis, *Finite Forms of de Finetti's Theorem on Exchangeability*, Synthese **36** (1977), no. 2, 271–281, Publisher: Springer.

[Dia23] ______, *Approximate exchangeability and de Finetti priors in 2022*, Scandinavian Journal of Statistics **50** (2023), no. 1, 38–53 (en).

[DJ08] Persi Diaconis and Svante Janson, *Graph limits and exchangeable random graphs*, Rendiconti di Matematica e delle sue Applicazioni. Serie VII **28** (2008), 33–61, arXiv:0712.2749.

[Dyn78] Eugene B. Dynkin, *Sufficient Statistics and Extreme Points*, The Annals of Probability **6** (1978), no. 5, 705–730, Publisher: Institute of Mathematical Statistics.

[Dö15] Christian Döbler, *Stein's method of exchangeable pairs for the Beta distribution and generalizations*, Electronic Journal of Probability **20** (2015), no. none, 1–34, Publisher: Institute of Mathematical Statistics and Bernoulli Society.

[ELS79] Pál Erdős, László Lovász, and Joel Spencer, *Strong Independence of Graphcopy Functions*, Graph Theory and Related Topics, Academic Press, 1979.

[ES12] Gábor Elek and Balázs Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Advances in Mathematics **231** (2012), no. 3, 1731–1772.

[ET22] Gábor Elek and Gábor Tardos, *Convergence and Limits of Finite Trees*, Combinatorica **42** (2022), no. 6, 821–852 (en).

[Far62] Rogers H. Farrell, *Representation of invariant measures*, Illinois Journal of Mathematics **6** (1962), 447–467.

[Fel71] William Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971.

[GK21] Lampros Gavalakis and Ioannis Kontoyiannis, *An information-theoretic proof of a finite de Finetti theorem*, Electronic Communications in Probability **26** (2021), 1–5, Publisher: The Institute of Mathematical Statistics and the Bernoulli Society.

[HKM+13] Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, and Rudini Menezes Sampaio, *Limits of permutation sequences*, Journal of Combinatorial Theory, Series B **103** (2013), no. 1, 93–113.

[Hoo79] Douglas N. Hoover, *Relations on probability spaces and arrays of random variables.*, 1979.

[Jan11a] Svante Janson, *Graphons, cut norm and distance, couplings and rearrangements*, June 2011, arXiv:1009.2376 [math].

[Jan11b] ______, *Poset limits and exchangeable random posets*, Combinatorica **31** (2011), no. 5, 529–563 (en).

[Jay86] Edwin Jaynes, Thompson, *Some Applications and Extensions of the de Finetti Representation Theorem*, Bayesian Inference and Decision Techniques: Essays in Honor of Brune de Finetti, Elsevier, Amsterdam, 1986.

[JKY16] Svante Janson, Takis Konstantopoulos, and Linglong Yuan, *On a representation theorem for finitely exchangeable random vectors*, Journal of Mathematical Analysis and Applications **442** (2016), no. 2, 703–714.

[Kal89] Olav Kallenberg, *On the representation theorem for exchangeable arrays*, Journal of Multivariate Analysis **30** (1989), no. 1, 137–154.

[Kal05] ______, *Probabilistic Symmetries and Invariance Principles*, Probability and Its Applications, Springer-Verlag, New York, 2005 (en).

[Kin78] J. F. C. Kingman, *Uses of Exchangeability*, The Annals of Probability **6** (1978), no. 2, 183–197, Publisher: Institute of Mathematical Statistics.

[Kle20] Achim Klenke, *Probability theory*, 3rd ed., Springer, Switzerland, 2020.

[KS06] G. Jay. Kerns and Gábor J. Székely, *Definetti's Theorem for Abstract Finite Exchangeable Sequences*, Journal of Theoretical Probability **19** (2006), no. 3, 589–608.

[KY18] Takis Konstantopoulos and Linglong Yuan, *On the extendibility of finitely exchangeable probability measures*, Transactions of the American Mathematical Society **371** (2018), no. 10, 7067–7092.

[Leo18] Paolo Leonetti, *Finite Partially Exchangeable Laws Are Signed Mixtures of Product Laws*, Sankhya A **80** (2018), no. 2, 195–214.

[Lov12] László Lovász, *Large Networks and Graph Limits*, Colloquium Publications, vol. 60, American Mathematical Society, Providence, Rhode Island, December 2012 (en).

[LRS18a] Steffen Lauritzen, Alessandro Rinaldo, and Kayvan Sadeghi, *On Exchangeability in Network Models*, September 2018, arXiv:1709.03885 [math].

[LRS18b] ———, *Random Networks, Graphical Models and Exchangeability*, Journal of the Royal Statistical Society Series B: Statistical Methodology **80** (2018), no. 3, 481–508.

[LS06] László Lovász and Balázs Szegedy, *Limits of dense graph sequences*, Journal of Combinatorial Theory, Series B **96** (2006), no. 6, 933–957 (en).

[LS07] ———, *Szemerédi's Lemma for the Analyst*, GAFA Geometric And Functional Analysis **17** (2007), no. 1, 252–270 (en).

[LS10] ———, *Limits of compact decorated graphs*, October 2010, arXiv:1010.5155 [math].

[Mai77] A. Maitra, *Integral representations of invariant measures*, Transactions of the American Mathematical Society **229** (1977), 209–225.

[Mat95] František Matúš, *Finite Partially Exchangeable Arrays*, Research Report 1856, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic, November 1995.

[MPS16] Guillaume Mijoule, Giovanni Peccati, and Yvik Swan, *On the rate of convergence in de Finetti's representation theorem*, Alea: Latin American Journal of Probability and Mathematical Statistics **13** (2016), no. 2, 1167–1186.

[OR15] Peter Orbanz and Daniel M. Roy, *Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures*, IEEE Transactions on Pattern Analysis and Machine Intelligence **37** (2015), no. 2, 437–461, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[OS16] Peter Orbanz and Balazs Szegedy, *Borel liftings of graph limits*, Electronic Communications in Probability **21** (2016), no. none, 1–4, Publisher: Institute of Mathematical Statistics and Bernoulli Society.

[Res85] Paul Ressel, *De Finetti-type Theorems: An Analytical Approach*, The Annals of Probability **13** (1985), no. 3, 898–922, Publisher: Institute of Mathematical Statistics.

[Sad20] Kayvan Sadeghi, *On finite exchangeability and conditional independence*, Electronic Journal of Statistics **14** (2020), no. 2, 2773–2797, Publisher: Institute of Mathematical Statistics and Bernoulli Society.

[SBC00] M. Scarsini, Bruno Bassan, and D. Capello, *An improved version of a theorem concerning finite row-column exchangeable arrays*, Commentationes Mathematicae Universitatis Carolinae **41** (2000), 197–198.

[Stu21] Benedikt Stufler, *Graphon convergence of random cographs*, Random Structures & Algorithms **59** (2021), 464–491.

[Sze75] Endre Szemerédi, *On sets of integers containing k elements in arithmetic progression*, Acta Arithmetica **27** (1975), no. 1, 199–245 (eng).

[Var63] V. S. Varadarajan, *Groups of automorphisms of Borel spaces*, Transactions of the American Mathematical Society **109** (1963), no. 2, 191–220.