

Article

Finite and Infinite Weighted Exchangeable Sequences

Wenpin Tang

Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA; wt2319@columbia.edu

Abstract

Motivated by recent interest in predictive inference under distribution shifts, we study the problem of approximating finite weighted exchangeable sequences by a mixture of finite sequences with independent terms. Two different approaches are developed: a combinatorial approach and a probabilistic approach. Various bounds are derived in terms of weight functions, extending previous results on finite exchangeable sequences. As a byproduct, we obtain a version of de Finetti's theorem for infinite weighted exchangeable sequences, recovering an instance of Barber, Candes, Ramdas and Tibshirani 1 (Bernoulli, 30 (2024), pp. 3004–3028) in the Polish space with bounded and continuous weights.

Keywords: finite/infinite exchangeability; weighted exchangeability; de Finetti theorem

MSC: 60G09; 28A33

1. Introduction

Exchangeability is a fundamental concept in probability theory and statistics, alongside the notion of independent and identically distributed (i.i.d.) random objects. It features a variety of research topics, including exchangeable random graphs and networks [1–3], random walk with exchangeable increments [4,5], exchangeable random partitions and clustering [6,7], and Bayesian inference [8–10], to name just a few. Perhaps the best known result is de Finetti's representation of an infinite exchangeable binary sequence [11]: an exchangeable sequence $X_1, X_2, \dots \in \{0, 1\}$ can be constructed by

$$\text{sample } p \sim \mu, \text{ then draw } X_1, X_2, \dots \stackrel{i.i.d.}{\sim} \text{Bernoulli}(p),$$

for some distribution μ on $[0, 1]$. In other words, any infinite exchangeable binary sequence is a mixture of i.i.d. Bernoulli sequences. The result was extended by Hewitt and Savage [12] to a general space $\mathcal{X}$. See [13,14] for a comprehensive review on the topic of exchangeability.

Recently, refs. [15–17] proposed the notion of *weighted exchangeability* to deal with covariate shifts in predictive inference. Roughly speaking, the sequence $X_1, \dots, X_n$ on $\mathcal{X}$ is weighted exchangeable with respect to the weight functions $\lambda_1, \dots, \lambda_n$ if the probability mass/density function f_n (if $\mathcal{X}$ is discrete/continuous) can be factorized as:

$$f_n(x_1, \dots, x_n) = \left(\prod_{i=1}^n \lambda_i(x_i) \right) g_n(x_1, \dots, x_n), \quad (1)$$

where $g_n : \mathcal{X}^n \rightarrow (0, \infty)$ is symmetric, i.e., $g_n(x_1, \dots, x_n) = g_n(x_{\sigma(1)}, \dots, x_{\sigma(n)})$. For instance, if $\lambda_i \equiv 1$, then we are in the exchangeable setting. The form (1) also appeared



Academic Editor: Steve Drekic

Received: 10 August 2026

Revised: 31 August 2026

Accepted: 3 September 2026

Published: 10 September 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

in the context of extremal families of a statistical model, where $\mathcal{X} = \{0, 1\}$ [18]. See Definition 1 for details, and an (obvious) extension to infinite sequences.

Among many possible questions on weighted exchangeability, a natural first one is whether there is a de Finetti-like theorem for weighted exchangeable sequences. That is, can a weighted exchangeable sequence be represented or approximated by a mixture of weighted i.i.d. (i.e., independent but not necessarily identically distributed) sequences? Intuitively, the weight functions $\lambda_1, \lambda_2, \dots$ should play a role in the answer to this question.

- In [18], it was proved that for $\mathcal{X} = \{0, 1\}$, de Finetti's theorem for infinite weighted exchangeable sequences holds *if and only if*

$$\sum_{i=1}^{\infty} \frac{\min(\lambda_i(0), \lambda_i(1))}{\max(\lambda_i(0), \lambda_i(1))} = \infty. \quad (2)$$

- In [15], it was shown that for a fairly general $\mathcal{X}$, de Finetti's theorem for infinite weighted exchangeable sequences holds *if*

$$\sum_{i=1}^{\infty} \tilde{r}_i = \infty, \quad \text{where } \tilde{r}_i := \frac{\inf_{\mathcal{X}} \lambda_i(x) / \lambda^*(x)}{\sup_{\mathcal{X}} \lambda_i(x) / \lambda^*(x)}, \quad (3)$$

for a reference weight λ^* . It is unknown whether the condition (3) is necessary except when $\mathcal{X} = \{0, 1\}$.

The proofs in [15] rely heavily on measure-theoretic machinery, and that paper also explores the connection between different concepts, including the weighted de Finetti property, the weighted zero-one law (see also [19]), and the weighted law of large numbers.

This paper aims to provide a “quantitative” approach to de Finetti's representation of weighted exchangeable sequences, and the perspective is different from that of [15,18]. To this end, we consider *finite* weighted exchangeable sequences in the spirit of Diaconis and Freedman [20]. See also [21,22] for the theory of finite exchangeability. It is known that a finite (weighted) exchangeable sequence is not a mixture of finite (weighted) i.i.d. sequences, and the goal is to understand how close the former can be approximated by the latter. We will establish various estimates of the difference between sampling with and without replacement, and show how they can be used to derive bounds in terms of the weight functions. Two different approaches will be developed: one is purely combinatorial, involving the oscillation of the likelihood ratio between sampling with and without replacement, and the other is probabilistically based on a thinning argument and stochastic comparisons. As a corollary, we obtain a version of de Finetti's theorem for infinite weighted exchangeable sequences by taking limits. As we will see, the probabilistic bound is sharper than the combinatorial one in most cases. We also present the combinatorial argument for pedagogical purposes, as a direct generalization of the Diaconis–Freedman approach. We point out that the problem of quantifying the approximation of finite weighted exchangeable sequences by a mixture of weighted i.i.d. sequences is of interest in its own right.

The remainder of the paper is organized as follows. In Section 2, we provide background on weighted exchangeability. We develop the combinatorial approach to the weighted de Finetti results in Section 3, and the probabilistic approach is presented in Section 4. Concluding remarks are given in Section 5.

2. Background of Weighted Exchangeability

Throughout this paper, we consider random variables on a Polish space $\mathcal{X}$, equipped with the Borel σ -algebra $\mathcal{B} = \mathcal{B}(\mathcal{X})$. Furthermore, define $\mathcal{X}^n$ and $\mathcal{X}^\infty$ to be the n -fold and countably infinite product spaces, with respective Borel σ -algebras $\mathcal{B}^n = \mathcal{B}(\mathcal{X})^n$ and $\mathcal{B}^\infty = \mathcal{B}(\mathcal{X})^\infty$. Denote by Λ the set of measurable functions $\mathcal{X}$ to $(0, \infty)$, and Λ^n and

Λ^∞ the corresponding n -fold and countably infinite product spaces. Let $\mathfrak{S}_n$ be the set of permutations on $[n] := \{1, \dots, n\}$, and for $\sigma \in \mathfrak{S}_n$, we denote $\sigma[i, j] := (\sigma(i), \dots, \sigma(j))$ for $1 \leq i \leq j \leq n$.

Below, we define the notion of weighted exchangeability for a sequence of random variables.

Definition 1 ([15,17]).

1. (Finite weighted exchangeability). Let n be a positive integer, and P_n be the distribution of $(X_1, \dots, X_n) \in \mathcal{X}^n$. We say that P_n is weighted exchangeable if there is $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda^n$ such that the measure $\bar{P}_n$, defined by

$$\bar{P}_n(B) = \int_B \frac{dP_n(x_1, \dots, x_n)}{\prod_{i=1}^n \lambda_i(x_i)}, \quad \text{for } B \in \mathcal{B}^n, \quad (4)$$

is an exchangeable probability measure. The sequence $(X_1, \dots, X_n)$ or the distribution P_n is called λ -exchangeable.

2. (Infinite weighted exchangeability). Let P be the distribution of $(X_1, X_2, \dots) \in \mathcal{X}^\infty$, and P_n be the marginal measure of P on $(\mathcal{X}^n, \mathcal{B}^n)$ defined by $P_n(B) := P(B \times \mathcal{X} \times \dots)$ for $B \in \mathcal{B}^n$. We say that P is weighted exchangeable if there is $\lambda = (\lambda_1, \lambda_2, \dots) \in \Lambda^\infty$ such that for each $n \geq 1$, the distribution P_n is λ_n -exchangeable, where $\lambda_n := (\lambda_1, \dots, \lambda_n) \in \Lambda^n$. The sequence $(X_1, X_2, \dots)$, or the distribution P , is called λ -exchangeable.

Let us make a few remarks. First, Definition 1 is a slight adaptation of [15] (Definitions 3 and 4), where the notion of weighted exchangeability is defined for general measures in the following obvious way. A measure Q_n on $(\mathcal{X}^n, \mathcal{B}^n)$ is called λ -exchangeable if the measure $\bar{Q}_n$, defined by

$$\bar{Q}_n(B) = \int_B \frac{dQ_n(x_1, \dots, x_n)}{\prod_{i=1}^n \lambda_i(x_i)}, \quad \text{for } B \in \mathcal{B}^n,$$

is exchangeable in the sense that $\bar{Q}_n(B_1 \times \dots \times B_n) = \bar{Q}_n(B_{\sigma(1)} \times \dots \times B_{\sigma(n)})$ for any $B_1, \dots, B_n \in \mathcal{B}$ and any permutation $\sigma \in \mathfrak{S}_n$. Similarly, a measure Q on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$ is λ -exchangeable if for each n , its marginal measure Q_n is λ_n -exchangeable. Here, we focus on weighted exchangeable probability measures, which are useful for practical purposes. Most results can be easily extended to finite measures.

Second, the key ingredient (4) in the definition of weighted exchangeability is basically the factorization (1). In other words, weighted exchangeable sequences can be viewed as a change in measure of exchangeable sequences, with separable Radon–Nikodym derivatives.

Finally, the weights λ are not unique for a weighted exchangeable sequence. For instance, consider (X_1, X_2) with density $f(x_1, x_2) = x_1 x_2$ on $[0, 1]^2$. Note that $x_1 x_2 = x_1 \cdot x_2 \cdot 1 = 1 \cdot 1 \cdot x_1 x_2$, so (X_1, X_2) is λ -exchangeable with $\lambda_1(x) = \lambda_2(x) \equiv 1$ or $\lambda_1(x) = \lambda_2(x) = x$. For $\lambda \in \Lambda^n$ and $\theta = (\theta_1, \dots, \theta_n) \in (0, \infty)^n$, define $\lambda_\theta := \theta \cdot \lambda = (\theta_1 \lambda_1, \dots, \theta_n \lambda_n)$. Obviously, if P_n is λ -exchangeable, then P_n is λ_θ -exchangeable provided $\prod_{i=1}^n \theta_i = 1$. Thus, the notion of λ -exchangeability is defined up to equivalence classes on λ . To illustrate, consider the three-point space $\mathcal{X} = \{a, b, c\}$ and the probability mass function:

$$p(x_1, x_2) \propto \lambda_1(x_1) \lambda_2(x_2) \quad \text{on } \mathcal{X}^2,$$

where $(\lambda_1(a), \lambda_1(b), \lambda_1(c)) = (1, 2, 3)$ and $(\lambda_2(a), \lambda_2(b), \lambda_2(c)) = (3, 2, 1)$. Clearly, this probability distribution is (λ_1, λ_2) -exchangeable. The same distribution is also (λ'_1, λ'_2) -exchangeable, where $(\lambda'_1(a), \lambda'_1(b), \lambda'_1(c)) = (2, 4, 6)$ and $(\lambda'_2(a), \lambda'_2(b), \lambda'_2(c)) = (\frac{3}{2}, 1, \frac{1}{2})$, by taking $\theta = (2, \frac{1}{2})$, so $\lambda'_1(x_1) \lambda'_2(x_2) = \lambda_1(x_1) \lambda_2(x_2)$.

For a probability measure μ on $\mathcal{X}$, let $T_i\mu(dx) \propto \lambda_i(x)\mu(dx)$. We say that a probability measure Q on $\mathcal{X}$ is a (λ) -mixture of weighted i.i.d. if

$$Q = \int_{\mathcal{P}(\mathcal{X})} \bigotimes_{i=1}^n T_i\mu \Pi(d\mu), \quad (5)$$

for some probability measure Π on $\mathcal{P}(\mathcal{X})$. The definition extends to infinite sequences whenever the probability measures $T_i\mu$ are well defined. Here, we use the terminology “weighted i.i.d.” following [15]. It refers to independent coordinates obtained by applying the coordinate-dependent tilts T_i to a common base distribution μ , and does not mean that the resulting coordinates are identically distributed.

De Finetti’s theorem states that any infinite exchangeable sequence is a mixture of i.i.d. sequences (sampling with replacement); see [13] (Section 3) and [14] (Theorem 1.1). A finite exchangeable sequence needs not be a mixture of i.i.d. sequences; instead it is a mixture of urn sequences (sampling without replacement); see [14] (Proposition 1.8) and [20]. Recall that an element in a convex set is called *extreme* if it cannot be expressed as a strict mixture of the elements in the set. The following proposition records some useful results on finite weighted exchangeable sequences.

Proposition 1. Let $\lambda \in \Lambda^n$, and $(X_1, \dots, X_n) \in \mathcal{X}^n$ be λ -exchangeable.

1. The conditional distribution of $(X_1, \dots, X_n)$ given the empirical measure $\hat{P}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ (or its possibly repeated values $\mathcal{U}_X := \{X_1, \dots, X_n\}$) is

$$(X_1, \dots, X_n | \hat{P}_n) \stackrel{d}{=} \sum_{\sigma} \frac{\prod_{i=1}^n \lambda_i(X_{\sigma(i)})}{\sum_{\sigma} \prod_{i=1}^n \lambda_i(X_{\sigma(i)})} \delta_{X_{\sigma(1)}, \dots, X_{\sigma(n)}}, \quad (6)$$

where $\sigma \in \mathfrak{S}_n$ and the sum is over all permutations in $\mathfrak{S}_n$. Denote by $P_{\mathcal{U}_X, n}$ the conditional distribution given by (6).

2. The set of λ -exchangeable sequences is convex, and each extreme λ -exchangeable distribution is of form (6) for some non-random empirical measure $\frac{1}{n} \sum_{i=1}^n \delta_{x_i}$. Consequently, the distribution of $(X_1, \dots, X_n)$ is a unique mixture of extreme points:

$$P_n(\cdot) = \int P_{\mathcal{U}_X, n}(\cdot) \Pi(d\mathcal{U}_X), \quad (7)$$

for a probability measure Π on the space of multisets.

Proof. (1) It was proved in [15] (Proposition 6) that the conditional (marginal) distribution of X_i given $\frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ is

$$(X_i | \hat{P}_n) \stackrel{d}{=} \sum_{j=1}^n \frac{\sum_{\sigma(i)=j} \prod_{i=1}^n \lambda_i(X_{\sigma(i)})}{\sum_{\sigma} \prod_{i=1}^n \lambda_i(X_{\sigma(i)})} \delta_{X_j}, \quad (8)$$

and the Formula (6) is an easy extension that can be read from the proof. Furthermore, note that the joint distribution (6) is consistent with (8) by summing over all permutations $\sigma \in \mathfrak{S}_n$ with $\sigma(i) = j$.

(2) The fact that the set of λ -exchangeable distributions is convex is straightforward from Definition 1; see [15] (Proposition 2). The characterization of extreme λ -exchangeable distributions follows from the general machinery as in [14] (Proposition 1.4 and 1.8). $\square$

Several remarks are necessary.

First, it is clear from Formula (6) that the conditional distribution of $(X_1, \dots, X_k)$ for $k \leq n$ given $\frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ is

$$\begin{aligned}
& (X_1, \dots, X_k | \hat{P}_n) \\
& \stackrel{d}{=} \sum_{j_1, \dots, j_k} \frac{\sum_{\sigma(1)=j_1, \dots, \sigma(k)=j_k} \prod_{i=1}^n \lambda_i(X_{\sigma(i)})}{\sum_{\sigma} \prod_{i=1}^n \lambda_i(X_{\sigma(i)})} \delta_{X_{j_1}, \dots, X_{j_k}}, \\
& \stackrel{d}{=} \sum_{j_1, \dots, j_k} \left(\prod_{i=1}^k \lambda_i(X_{j_i}) \right) \frac{\sum_{\sigma[k+1, n]=[n] \setminus \{j_1, \dots, j_k\}} \prod_{i=k+1}^n \lambda_i(X_{\sigma(i)})}{\sum_{\sigma} \prod_{i=1}^n \lambda_i(X_{\sigma(i)})} \delta_{X_{j_1}, \dots, X_{j_k}},
\end{aligned} \tag{9}$$

where $j_1, \dots, j_k$ take k distinct values from $[n]$, and $\sigma[k+1, n] = [n] \setminus \{j_1, \dots, j_k\}$ as sets. Denote by $\mathcal{P}_{\mathcal{U}_X, k}$ the conditional distribution in (9). There are $(n)_k := n!/(n-k)!$ terms in the sum $\sum_{j_1, \dots, j_k}$, and $(n-k)!$ terms in $\sum_{\sigma[k+1, n]=[n] \setminus \{j_1, \dots, j_k\}}$. The second part of Proposition 1 implies that each extreme point of the k -marginal λ -exchangeable distributions is of form $\mathcal{P}_{\mathcal{U}_X, k}$ for some non-random $\frac{1}{n} \sum_{i=1}^n \delta_{x_i}$. This way, the Formula (9) will be very useful in deriving an error bound between the distribution of $(X_1, \dots, X_k)$ and its closest mixture of i.i.d. sequences.

Second, if $\lambda_i \equiv 1$ for all i (exchangeable), then the Formulas (6) and (9) specialize to

$$(X_1, \dots, X_k | \hat{P}_n) \stackrel{d}{=} \frac{1}{(n)_k} \sum_{j_1, \dots, j_k} \delta_{X_{j_1}, \dots, X_{j_k}}, \quad k \leq n, \tag{10}$$

which is exactly sampling without replacement from the “urn” $\mathcal{U}_X$. So $(X_1, \dots, X_n)$ given the urn $\mathcal{U}_X$ can be sampled sequentially: pick X_{j_1} uniformly at random from $\mathcal{U}_X$, and pick X_{j_2} uniformly at random from $\mathcal{U}_X \setminus \{X_{j_1}\}$, and so on. However, for general weights λ , there does not seem to be an easy way to sample $(X_1, \dots, X_n)$ from $\mathcal{U}_X$.

Finally, the extreme points of k -marginal exchangeable distributions can be made explicit. By [23] (Theorem 4.5), each extreme point is of the form

$$\frac{n^k}{(n)_k} \hat{P}_n^{\otimes} + \varepsilon_{n,k}(\hat{P}_n) = \hat{P}_n^{\otimes} + \left(\left(\frac{n^k}{(n)_k} - 1 \right) \hat{P}_n^{\otimes} + \varepsilon_{n,k}(\hat{P}_n) \right), \tag{11}$$

where $\varepsilon_{n,k}(\hat{P}_n)$ is a “polynomial” involving alternating sums of pushforward and symmetrization of $\hat{P}_n$ (see [23] (4.10), and also [24]). The term $\left(\frac{n^k}{(n)_k} - 1 \right) \hat{P}_n^{\otimes} + \varepsilon_{n,k}(\hat{P}_n)$ gives the difference between the distribution (10) and k i.i.d. samples from the empirical measure $\hat{P}_n$. It is also possible to derive a formula analogous to (11) for weighted exchangeable distributions, but we do not pursue this here, and leave it to the readers.

3. A Combinatorial Approach

In this section, we present a combinatorial approach to approximating finite weighted exchangeable sequences by a mixture of weighted i.i.d. (or independent but not identically distributed) sequences. Recall that the total variation distance between two distributions P and Q is $|P - Q|_{TV} := \sup_A |P(A) - Q(A)|$ (In [20], the total variation is defined by $|P - Q|_{TV} := 2 \sup_A |P(A) - Q(A)|$, so their bounds differ from ours by a factor of 2).

Theorem 1. Let P_n be the distribution of a λ -exchangeable sequence $(X_1, \dots, X_n)$ on $\mathcal{X}$, and P_k be the marginal distribution of $(X_1, \dots, X_k)$ for $k \leq n$. Assume that

$$r_i := \frac{\inf_{\mathcal{X}} \lambda_i(x)}{\sup_{\mathcal{X}} \lambda_i(x)} > 0, \quad \text{for } i \in [n]. \tag{12}$$

Then there exists a mixture of weighted i.i.d. distributions Q_n on $\mathcal{X}^n$, with $Q_{n,k}$ the marginal distribution on its first k coordinates such that

$$|P_k - Q_{n,k}|_{TV} \leq \frac{k(k-1)}{2n} \left(\prod_{i=1}^k r_i \right)^{-1} + \frac{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} - 1}{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} + 1}. \quad (13)$$

Proof. By Proposition 1 (2), P_n is of form $P_n = \int P_{\mathcal{U}_x,n} \Pi(d\mathcal{U}_x)$, and $P_k = P_{\mathcal{U}_x,k} \Pi(d\mathcal{U}_x)$. Define the probability measure $Q_{\mathcal{U}_x,n}$ on $\mathcal{U}_x$ by sampling X_i according to $\frac{\lambda_i(\cdot)}{\sum_{j=1}^n \lambda_i(x_j)}$, independently for each i . So

$$Q_{\mathcal{U}_x,n} = \sum_{s(1), \dots, s(n)} \left(\prod_{i=1}^n \frac{\lambda_i(x_{s(i)})}{\sum_{j=1}^n \lambda_i(x_j)} \right) \delta_{x_{s(1)}, \dots, x_{s(n)}}, \quad (14)$$

where the sum is over all $s(1), \dots, s(n) \in [n]$. Observe that

$$Q_{\mathcal{U}_x,n} = \bigotimes_{i=1}^n T_i \hat{P}_{\mathcal{U}_x}, \quad \text{where } \hat{P}_{\mathcal{U}_x} = \frac{1}{|\mathcal{U}_x|} \sum_{x_i \in \mathcal{U}_x} \delta_{x_i},$$

and hence, $Q_{\mathcal{U}_x,n}$ is a (λ) -weighted i.i.d. distribution. Then

$$Q_n := \int Q_{\mathcal{U}_x,n} \Pi(d\mathcal{U}_x) \quad (15)$$

is a mixture of (λ) -weighted i.i.d. distributions in the sense of (5).

We have

$$|P_k - Q_{n,k}|_{TV} \leq \text{ess sup}_{\mathcal{U}_x \sim \Pi} |P_{\mathcal{U}_x,k} - Q_{\mathcal{U}_x,k}|_{TV}. \quad (16)$$

The goal is to bound $|P_{\mathcal{U}_x,k} - Q_{\mathcal{U}_x,k}|_{TV}$ for a given urn $\mathcal{U}_x$. Note that the support of $Q_{\mathcal{U}_x,k}$ is $\mathcal{U}_x^k$; and the support of $P_{\mathcal{U}_x,k}$ is by picking k elements without replacement from $\mathcal{U}_x$, which we denote by $\tilde{\mathcal{U}}_x^k$. Importantly, $P_{\mathcal{U}_x,k}$ is generally not $Q_{\mathcal{U}_x,k}$ conditioned on $\tilde{\mathcal{U}}_x^k$.

Let $q := Q_{\mathcal{U}_x,k}(\tilde{\mathcal{U}}_x^k)$, and $Q_{\mathcal{U}_x,k}^S := Q_{\mathcal{U}_x,k}(\cdot | S)$ for any $S \subset \mathcal{U}_x^k$. Since $\tilde{\mathcal{U}}_x^k \subset \mathcal{U}_x^k$, we write

$$Q_{\mathcal{U}_x,k} = q Q_{\mathcal{U}_x,k}^{\tilde{\mathcal{U}}_x^k} + (1-q) Q_{\mathcal{U}_x,k}^{(\tilde{\mathcal{U}}_x^k)^c}.$$

By convexity of total variation, we obtain

$$|P_{\mathcal{U}_x,k} - Q_{\mathcal{U}_x,k}|_{TV} \leq (1-q) + q \left| P_{\mathcal{U}_x,k} - Q_{\mathcal{U}_x,k}^{\tilde{\mathcal{U}}_x^k} \right|_{TV} \leq (1-q) + \left| P_{\mathcal{U}_x,k} - Q_{\mathcal{U}_x,k}^{\tilde{\mathcal{U}}_x^k} \right|_{TV}. \quad (17)$$

Bounding $(1-q)$: Note that

$$\begin{aligned} 1-q &= 1 - \sum_{\sigma(1), \dots, \sigma(k)} \prod_{i=1}^k \frac{\lambda_i(x_{\sigma(i)})}{\sum_{j=1}^n \lambda_i(x_j)} \\ &= \frac{\prod_{i=1}^k \left(\sum_{j=1}^n \lambda_i(x_j) \right) - \sum_{\sigma(1), \dots, \sigma(k)} \prod_{i=1}^k \lambda_i(x_{\sigma(i)})}{\prod_{i=1}^k \left(\sum_{j=1}^n \lambda_i(x_j) \right)}, \end{aligned} \quad (18)$$

where the sum $\sum_{\sigma(1), \dots, \sigma(k)}$ is over all k -tuples picking without replacement from $\mathcal{U}_x$ (or equivalently the restriction of $\sigma \in \mathfrak{S}_n$ to $[k]$). There are $n^k - (n)_k$ terms in the numerator of (18), each of which is bounded from above by $\prod_{i=1}^k \sup_{\mathcal{X}} \lambda_i(x)$; and there are n^k terms in the denominator of (18), each of which is bounded from below by $\prod_{i=1}^k \inf_{\mathcal{X}} \lambda_i(x)$. So

$$1 - q \leq \left(\prod_{i=1}^k r_i \right)^{-1} \left(1 - \frac{n_{(k)}}{n^k} \right), \quad (19)$$

where $1 - \frac{n_{(k)}}{n^k}$ is the total variation distance between sampling k elements uniformly from $[n]$ with and without replacement. By [25], we have $1 - \frac{n_{(k)}}{n^k} \leq \frac{k(k-1)}{2n}$, which implies

$$1 - q \leq \frac{k(k-1)}{2n} \left(\prod_{i=1}^k r_i \right)^{-1}. \quad (20)$$

Bounding $\left| P_{\mathcal{U}_{x,k}} - Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} \right|_{TV}$: By (9) and (14),

$$\frac{dP_{\mathcal{U}_{x,k}}}{dQ_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k}} \propto \sum_{\sigma[k+1,n]=[n] \setminus \{j_1, \dots, j_k\}} \prod_{i=k+1}^n \lambda_i(X_{\sigma(i)}) := H(j_1, \dots, j_k).$$

A standard likelihood ratio estimate gives

$$\left| P_{\mathcal{U}_{x,k}} - Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} \right|_{TV} \leq \frac{\sqrt{R_H} - 1}{\sqrt{R_H} + 1}, \quad (21)$$

where $R_H := \max_{\tilde{\mathcal{U}}_x^k} H / \min_{\tilde{\mathcal{U}}_x^k} H$ is the oscillation of the likelihood ratio. Furthermore, each term in $H(j_1, \dots, j_k)$ is between $\prod_{i=k+1}^n \inf_{\mathcal{X}} \lambda_i(x)$ and $\prod_{i=k+1}^n \sup_{\mathcal{X}} \lambda_i(x)$. Thus,

$$R_H \leq \prod_{i=k+1}^n r_i^{-1}. \quad (22)$$

Combining (21) and (22) yields

$$\left| P_{\mathcal{U}_{x,k}} - Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} \right|_{TV} \leq \frac{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} - 1}{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} + 1}. \quad (23)$$

Now, (17), (20) and (23) yields the bound (13). $\square$

If P_n is exchangeable (i.e., λ_i is constant for each i), the bound (13) reduces to the Diaconis–Freedman bound ([14] (Proposition 1.9)) and [20] (Theorem 13), which was shown to be sharp (see [20] (Section 4)). The bound (13) is stated for the first k coordinates of P_n , which is easily extended to any k -subset $\mathcal{S}$:

$$|P_{\mathcal{S}} - Q_{n,\mathcal{S}}|_{TV} \leq \frac{k(k-1)}{2n} \left(\prod_{i \in \mathcal{S}} r_i \right)^{-1} + \frac{(\prod_{i \in [n] \setminus \mathcal{S}} r_i)^{-\frac{1}{2}} - 1}{(\prod_{i \in [n] \setminus \mathcal{S}} r_i)^{-\frac{1}{2}} + 1}, \quad (24)$$

where $P_{\mathcal{S}}$ (resp. $Q_{n,\mathcal{S}}$) is the marginal distribution of P_n (resp. Q_n) on the coordinates $\mathcal{S}$.

The condition (12) is required so that the bound (13) is finite. The bound (13) indicates that the more spread the weights of λ_i are (or the smaller the r_i ratios are), the larger the error of approximation $|P_k - Q_{n,k}|_{TV}$ may be. While we do not know if the factor $(\prod_{i=1}^k r_i)^{-1}$ in the bound is sharp, the order of error k^2/n is optimal for the exchangeable case [20].

Next, we consider the special case where $\mathcal{X}$ is finite. The following result shows that a better bound k/n can be obtained.

Theorem 2. Assume that $\text{card}(\mathcal{X}) = c < \infty$. Let P_n be the distribution of a λ -exchangeable sequence $(X_1, \dots, X_n)$ on $\mathcal{X}$, and P_k be the marginal distribution of $(X_1, \dots, X_k)$ for $k \leq n$. Assume that (12) holds. Then there exists a mixture of weighted i.i.d. distributions Q_n on $\mathcal{X}^n$, with $Q_{n,k}$ the marginal distribution on its first k coordinates such that

$$|P_k - Q_{n,k}|_{TV} \leq \frac{ck}{n} \frac{3 - \prod_{i=1}^k r_i}{2 \prod_{i=1}^k r_i} + \frac{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} - 1}{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} + 1}. \quad (25)$$

Proof. We follow the notations as in the proof of Theorem 1. By (16), it suffices to bound $|P_{\mathcal{U}_{x,k}} - Q_{\mathcal{U}_{x,k}}|_{TV}$. By triangle inequality and (23), we get

$$\begin{aligned} |P_{\mathcal{U}_{x,k}} - Q_{\mathcal{U}_{x,k}}|_{TV} &\leq \left| P_{\mathcal{U}_{x,k}} - Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} \right|_{TV} + \left| Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} - Q_{\mathcal{U}_{x,k}} \right|_{TV} \\ &\leq \frac{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} - 1}{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} + 1} + \left| Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} - Q_{\mathcal{U}_{x,k}} \right|_{TV}. \end{aligned} \quad (26)$$

So it remains to bound $\left| Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} - Q_{\mathcal{U}_{x,k}} \right|_{TV}$. Observe that

$$Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k}(dx) \propto \prod_{i=1}^k \lambda_i(x_i) P_{o,k}(dx), \quad Q_{\mathcal{U}_{x,k}}(dx) \propto \prod_{i=1}^k \lambda_i(x_i) Q_{o,k}(dx),$$

where $P_{o,k}$ (resp. $Q_{o,k}$) denotes the k -marginal distribution of uniform sampling without replacement (resp. with replacement). We need the following lemma to bound $\left| Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} - Q_{\mathcal{U}_{x,k}} \right|_{TV}$. $\square$

Lemma 1. Let μ, ν be two probability measures on $\mathcal{X}$, and $w : \mathcal{X} \rightarrow \mathbb{R}$ be such that $0 < \alpha \leq w(x) \leq 1$ for all x . Define

$$\mu^w(dx) \propto w(x)\mu(dx), \quad \nu^w(dx) \propto w(x)\nu(dx).$$

Then

$$|\mu^w - \nu^w|_{TV} \leq \frac{3 - \alpha}{2\alpha} |\mu - \nu|_{TV}.$$

Proof. Denote by $a := \int w(x)\mu(dx)$ and $b := \int w(x)\nu(dx)$. We have

$$\begin{aligned} 2|\mu^w - \nu^w|_{TV} &\leq \frac{1}{a} \int w(x) |\mu(dx) - \nu(dx)| + \left| \frac{1}{a} - \frac{1}{b} \right| \int w(x) \nu(dx) \\ &= \frac{1}{a} \int w(x) |\mu(dx) - \nu(dx)| + \frac{|a - b|}{a} \\ &\leq \frac{2|\mu - \nu|_{TV}}{a} + \frac{|a - b|}{a}, \end{aligned}$$

where the last inequality follows from $w \leq 1$. Note that $|a - b| = \left| \int (w(x) - \alpha)(\mu(dx) - \nu(dx)) \right| \leq (1 - \alpha) |\mu - \nu|_{TV}$. As a result,

$$|\mu^w - \nu^w|_{TV} \leq \left(\frac{1}{a} + \frac{1 - \alpha}{2a} \right) |\mu - \nu|_{TV} \leq \left(\frac{1}{\alpha} + \frac{1 - \alpha}{2\alpha} \right) |\mu - \nu|_{TV}.$$

$\square$

Now let us go back to the proof of Theorem 2. Applying Lemma 1 with $w(x) = \prod_{i=1}^k \lambda_i(x_i) \geq \prod_{i=1}^k r_i$, $\mu = P_{o,k}$ and $\nu = Q_{o,k}$ yields:

$$\left| Q_{\mathcal{U}_{x,k}}^{\tilde{\mathcal{U}}_x^k} - Q_{\mathcal{U}_{x,k}} \right|_{TV} \leq \frac{3 - \prod_{i=1}^k r_i}{2 \prod_{i=1}^k r_i} |P_{o,k} - Q_{o,k}|_{TV}. \quad (27)$$

Assume without loss of generality that $\mathcal{X} = \{1, \dots, c\}$. It follows from [20] (Theorem 4) that

$$|P_{o,k} - Q_{o,k}|_{TV} \leq \frac{ck}{n}. \quad (28)$$

Combining (26), (27) and (28) yields (25).

Again, if P_n is exchangeable, then the bound (25) reduces to the Diaconis–Freedman bound ([20] (Theorem 3)), which is sharp.

As a consequence of Theorem 1, we derive a version of de Finetti’s theorem for infinite weighted exchangeable sequences, which is an analog to the Hewitt–Savage theorem [12] (Theorem 7.2). A stronger statement is given in Corollary 2. The only difference in the proof is that the bound (13) is replaced with the one in Theorem 3, yielding weaker assumptions.

Corollary 1. *Let P be the distribution of a λ -exchangeable sequence $(X_1, X_2, \dots)$ on $\mathcal{X}^\infty$. Assume that λ_i are bounded continuous functions such that*

$$r_i := \frac{\inf_{\mathcal{X}} \lambda_i(x)}{\sup_{\mathcal{X}} \lambda_i(x)} > 0, \text{ for } i = 1, 2, \dots, \text{ and } \sum_{i=1}^{\infty} (1 - r_i) < \infty. \quad (29)$$

Then P is a mixture of weighted i.i.d. distributions on $\mathcal{X}^\infty$.

Proof. The proof is standard, and we provide it here for completeness. Without loss of generality, assume that $\sup_{\mathcal{X}} \lambda_i(x) = 1$ for each i , so $r_i \leq \lambda_i(x) \leq 1$. The assumption

$$\sum_{i=1}^{\infty} (1 - r_i) < \infty \text{ is equivalent to } \prod_{i=1}^{\infty} r_i > 0.$$

Step 1. Fix $k \leq n$. Applying Theorem 1 to the n -marginal P_n yields

$$|P_k - Q_{n,k}|_{TV} \leq \frac{k(k-1)}{2n} \left(\prod_{i=1}^k r_i \right)^{-1} + \frac{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} - 1}{(\prod_{i=k+1}^n r_i)^{-\frac{1}{2}} + 1}, \quad (30)$$

for $Q_{n,k}$, with some mixture of weighted i.i.d. laws on $\mathcal{X}^k$.

Step 2. Choose $k_j \rightarrow \infty$ such that $\frac{(\prod_{i=k_j+1}^{\infty} r_i)^{-\frac{1}{2}} - 1}{(\prod_{i=k_j+1}^{\infty} r_i)^{-\frac{1}{2}} + 1} \leq \frac{1}{j}$. For each j , choose $n_j > k_j$

sufficiently large that

$$\frac{k_j(k_j-1)}{2n_j} \left(\prod_{i=1}^{k_j} r_i \right)^{-1} \leq \frac{1}{j}.$$

By (30), there exists Q_j some mixture of weighted i.i.d. laws on $\mathcal{X}^{k_j}$ such that

$$|P_{k_j} - Q_j|_{TV} \leq 2/j. \quad (31)$$

Recall from (5) that $T_i\mu(dx) \propto \lambda_i(x)\mu(dx)$, and write

$$Q_j = \int_{\mathcal{P}(\mathcal{X})} \otimes_{i=1}^{k_j} T_i\mu \Pi_j(d\mu).$$

for some probability measure Π_j on $\mathcal{P}(\mathcal{X})$.

Step 3. We claim that $\{\Pi_j\}$ is tight on $\mathcal{P}(\mathcal{X})$. First, let $Q_{j,1}$ be the first marginal of Q_j . We have $|P_1 - Q_{j,1}|_{TV} \leq 2/j$, and hence, $\{Q_{j,1}\}$ is tight. Since $r_1 \leq \lambda_1(x) \leq 1$, we get

$$\mu(A) \leq \frac{1}{r_1} T_1\mu(A) \quad \text{for any Borel set } A. \quad (32)$$

Fix $\varepsilon > 0$. Because $\{Q_{j,1}\}$ is tight, we can choose increasingly compact sets $K_\ell \subset \mathcal{X}$ such that

$$\sup_j Q_{j,1}(K_\ell^c) \leq r_1 \varepsilon 2^{-2\ell}.$$

By (32), $\int \mu(K_\ell^c) \Pi_j(d\mu) \leq \frac{1}{r_1} Q_{j,1}(K_\ell^c) \leq \varepsilon 2^{-2\ell}$. Markov's inequality yields $\Pi_j\{\mu : \mu(K_\ell^c) > 2^{-\ell}\} \leq \varepsilon 2^{-\ell}$. Set

$$\mathcal{K} := \bigcap_{\ell \geq 1} \left\{ \mu : \mu(K_\ell^c) \leq 2^{-\ell} \right\}.$$

We have $\Pi_j(\mathcal{K}) \geq 1 - \varepsilon$. Thus, $\mathcal{K}$ is a uniformly tight set of probability measures on $\mathcal{X}$, and hence is closed and relatively compact in $\mathcal{P}(\mathcal{X})$ by Prokhorov's theorem. This implies the tightness of $\{\Pi_j\}$. Passing to a subsequence, Π_j converges weakly to some probability measure Π on $\mathcal{P}(\mathcal{X})$.

Step 4. Since λ_i are bounded continuous functions, $\mu \rightarrow T_i\mu$ is weakly continuous. So for each fixed m ,

$$\mu \rightarrow \otimes_{i=1}^m T_i\mu \text{ is weakly continuous.}$$

Fix m . We have $k_j \geq m$ for sufficiently large j . Let $Q_{j,m}$ be the first m marginal of Q_j , so

$$Q_{j,m} = \int_{\mathcal{P}(\mathcal{X})} \otimes_{i=1}^m T_i\mu \Pi_j(d\mu).$$

By (31), we get $|P_m - Q_{j,m}|_{TV} \leq 2/j$. Thus, $Q_{j,m}$ converges to P_m as $j \rightarrow \infty$. On the other hand, the convergence of Π_j to Π and the weak continuity of $\mu \rightarrow \otimes_{i=1}^m T_i\mu$ yields

$$Q_{j,m} \text{ converges to } \int_{\mathcal{P}(\mathcal{X})} \otimes_{i=1}^m T_i\mu \Pi(d\mu).$$

Therefore, $P_m = \int_{\mathcal{P}(\mathcal{X})} \otimes_{i=1}^m T_i\mu \Pi(d\mu)$ for each m . (The map $\mu \rightarrow \otimes_{i=1}^m T_i\mu$ is measurable, because all of its finite-dimensional marginals are continuous.) By the Kolmogorov extension theorem, $P = \int_{\mathcal{P}(\mathcal{X})} \otimes_{i=1}^\infty T_i\mu \Pi(d\mu)$. $\square$

As mentioned in the introduction (see (3)), ref. [15] (Theorem 6) showed that if

$$\sum_{i=1}^\infty r_i = \infty, \quad (33)$$

then P is a mixture of weighted i.i.d. distributions. Note that the condition (29) is more restrictive than (33). Essentially, it implies that an infinite “nearly” exchangeable sequence is a mixture of weighted i.i.d. sequences. This calls for a sharper bound away from the exchangeable case (especially, to bypass the second term in (13)), so that de Finetti's representation for infinite weighted i.i.d. sequences can be derived from finite approximations under broader sufficient (and possibly necessary) conditions such as (3). This is the goal of the next section. Also, the assumption that $\mathcal{X}$ is a Polish space may be relaxed. How-

ever, without any topological assumptions on $\mathcal{X}$, de Finetti's theorem for exchangeable sequences fails [26,27].

4. A Probabilistic Approach

In this section, we present a probabilistic approach to the finite weighted de Finetti theory. We rely on a thinning argument, together with stochastic comparisons.

Theorem 3. Let P_n be the distribution of a λ -exchangeable sequence $(X_1, \dots, X_n)$ on $\mathcal{X}$, and P_k be the marginal distribution of $(X_1, \dots, X_k)$ for $k \leq n$. Assume that (12) holds. Then there exists a mixture of weighted i.i.d. distributions Q_n on $\mathcal{X}^n$, with $Q_{n,k}$ the marginal distribution on its first k coordinates such that

$$|P_k - Q_{n,k}|_{TV} \leq \frac{C_k(r)}{k-1 + \sum_{i=k+1}^n r_i}, \quad (34)$$

where $C_k(r) := \sum_{1 \leq i < j \leq k} \max(r_i, r_j)^{-1}$.

Proof. Without loss of generality, assume that $\sup_{x \in \mathcal{X}} \lambda_i(x) = 1$ for each i , so $r_i \leq \lambda_i(x) \leq 1$.

Fix $k < n$, and introduce i.i.d. $U_{k+1}, \dots, U_n \sim \text{Unif}(0, 1)$ independent of $(X_1, \dots, X_n)$. Define

$$B_i := 1 \left(U_i \leq \frac{r_i}{\lambda_i(X_i)} \right) \quad \text{for } i = k+1, \dots, n \quad \text{and} \quad A := \{i \in [k+1, n] : B_i = 1\}.$$

We claim that conditional on A , the “retained” coordinates $J := [k] \cup A$ are weighted exchangeable with weights

$$\lambda_1, \dots, \lambda_k, \underbrace{1, \dots, 1}_{\text{card}(A) \text{ times}}.$$

Recall that $P_n(dx_1 \cdots dx_n) = (\prod_{i=1}^n \lambda_i(x_i)) \bar{P}_n(dx_1 \cdots dx_n)$, where $\bar{P}_n$ is some exchangeable probability measure. Fixing $A \subset [k+1, n]$, the joint distribution of X and the event $\{B_i = 1(i \in A)\}$ is proportional to

$$\left(\prod_{i=1}^k \lambda_i(x_i) \right) \left(\prod_{i \in A} r_i \right) \left(\prod_{i \in J^c} (\lambda_i(x_i) - r_i) \right) \bar{P}_n(dx_1 \cdots dx_n).$$

Integrating out the “rejected” coordinates J^c , the retained coordinates have the distribution of form

$$\left(\prod_{i=1}^k \lambda_i(x_i) \right) P_A(dx_J), \quad \text{where } P_A(dx_J) := \int_{\mathcal{X}^{J^c}} \left(\prod_{i \in J^c} (\lambda_i(x_i) - r_i) \right) \bar{P}_n(dx_J, dx_{J^c}).$$

Clearly, P_A is symmetric in the coordinates indexed by J , and the claim follows.

The next condition is on a (deterministic) retained urn, and by Proposition 1, it suffices to consider the corresponding extremal weighted urn distribution $P_{\mathcal{U}_y}$ on

$$\mathcal{U}_y := \{y_1, \dots, y_{k+\text{card}(A)}\}.$$

Because all weights after the first coordinates are equal to 1, the weights are all equal, and there are $(N-k)!$ completions. Therefore,

$$P_{\mathcal{U}_y, k}(j_1, \dots, j_k) \propto \prod_{i=1}^k \lambda_i(y_{j_i}), \quad j_1, \dots, j_k \text{ distinct}. \quad (35)$$

Define the weighted sampling-with-replacement distribution $Q_{\mathcal{U}_y,k}$ by choosing labels $I_1, \dots, I_k \in [k + \text{card}(A)]$ independently according to

$$\mathbb{P}(I_i = j) = \frac{\lambda_i(y_j)}{\sum_{\ell=1}^{k+\text{card}(A)} \lambda_i(y_\ell)}, \quad j = 1, \dots, k + \text{card}(A).$$

As a result,

$$Q_{\mathcal{U}_y,k}(j_1, \dots, j_k) = \prod_{i=1}^k \frac{\lambda_i(y_{j_i})}{\sum_{\ell=1}^{k+\text{card}(A)} \lambda_i(y_\ell)}. \quad (36)$$

So $Q_{\mathcal{U}_y,k} = \otimes_{i=1}^k T_i \hat{P}_{\mathcal{U}_y}$ is a (λ) -weighted i.i.d. distribution.

From (35) and (36), we see that $P_{\mathcal{U}_y,k}$ has the same distribution as $Q_{\mathcal{U}_y,k}$ conditioned on $\{I_1, \dots, I_k \text{ distinct}\}$. Therefore,

$$\begin{aligned} |P_{\mathcal{U}_y,k} - Q_{\mathcal{U}_y,k}|_{TV} &\leq Q_{\mathcal{U}_y,k}(\{I_1, \dots, I_k \text{ distinct}\}^c) \\ &\leq \sum_{1 \leq i < j \leq k} \mathbb{P}(I_i = I_j), \end{aligned} \quad (37)$$

where the last inequality is obtained by a union bound. Note that for $i < j$,

$$\begin{aligned} \mathbb{P}(I_i = I_j) &= \frac{\sum_{\ell=1}^{k+\text{card}(A)} \lambda_i(y_\ell) \lambda_j(y_\ell)}{\left(\sum_{\ell=1}^{k+\text{card}(A)} \lambda_i(y_\ell)\right) \left(\sum_{\ell=1}^{k+\text{card}(A)} \lambda_j(y_\ell)\right)} \\ &\leq \frac{\min\left\{\sum_{\ell=1}^{k+\text{card}(A)} \lambda_i(y_\ell), \sum_{\ell=1}^{k+\text{card}(A)} \lambda_j(y_\ell)\right\}}{\left(\sum_{\ell=1}^{k+\text{card}(A)} \lambda_i(y_\ell)\right) \left(\sum_{\ell=1}^{k+\text{card}(A)} \lambda_j(y_\ell)\right)} \\ &\leq \frac{1}{(k + \text{card}(A)) \max(r_i, r_j)}. \end{aligned} \quad (38)$$

It follows from (37) and (38) that

$$|P_{\mathcal{U}_y,k} - Q_{\mathcal{U}_y,k}|_{TV} \leq \frac{1}{k + \text{card}(A)} C_k(r).$$

Averaging first over the retained-urn mixture and then over the thinning pattern yields a mixture $Q_{n,k}$ of weighted i.i.d. laws such that

$$|P_k - Q_{n,k}|_{TV} \leq C_k(r) \mathbb{E}\left(\frac{1}{k + \text{card}(A)}\right). \quad (39)$$

Recall that $\frac{r_i}{\lambda_i(X_i)} \geq r_i$. Define the coupling $Z_i := 1(U_i \leq r_i)$ with the same U_i , so Z_i are independent $\text{Bernoulli}(r_i)$ and $Z_i \leq B_i$ almost certainly. Thus, $\text{card}(A) \geq \sum_{i=k+1}^n Z_i$ is almost certain, with the sum $\sum_{i=k+1}^n Z_i$ being Poisson binomial. Consequently,

$$\begin{aligned} \mathbb{E}\left(\frac{1}{k + \text{card}(A)}\right) &\leq \mathbb{E}\left(\frac{1}{k + \sum_{i=k+1}^n Z_i}\right) \\ &= \int_0^1 t^{k-1} \prod_{i=k+1}^n (1 - r_i + r_i t) dt \\ &\leq \int_0^1 t^{k-1} \exp\left(-(1-t) \sum_{i=k+1}^n r_i\right) dt \\ &\leq \int_0^1 \exp\left(-(1-t) \left(k-1 + \sum_{i=k+1}^n r_i\right)\right) dt \\ &= \frac{1 - e^{-(k-1 + \sum_{i=k+1}^n r_i)}}{k-1 + \sum_{i=k+1}^n r_i}, \end{aligned} \quad (40)$$

where the second equality follows from the identity $\frac{1}{k+z} = \int_0^1 t^{k+z-1} dt$ and the probability-generating function of Poisson binomial ([28] (16)), and the fourth inequality is from the fact that $t \leq e^{-(1-t)}$ for $0 \leq t \leq 1$. Combining (39) and (40) yields (34). $\square$

The key to Theorem 3 is the probabilistic bound (39). If P_n is exchangeable, then this bound reduces to the Diaconis–Freedman bound $\frac{k(k-1)}{2n}$. This is slightly stronger than the bound $\frac{k(k-1)}{2(n-1)}$ in (34), and the loss comes from the last two inequalities in (40).

Now let us compare the bounds obtained from the two approaches in Sections 3 and 4. Neither bound uniformly dominates the other.

- The bound (34) is typically sharper away from the exchangeable regime. Indeed,

$$(34) \leq \frac{k(k-1)}{2 \min_{i \leq k} r_i (k-1 + \sum_{i=k+1}^n r_i)},$$

which is of order $\frac{k^2}{n}$ if the r_i are uniformly bounded away from zero. In contrast, the second term in (13) can be close to one when $\prod_{i=k+1}^n r_i$ is small, even though $\sum_{i=k+1}^n r_i$ is large. For instance, if $n = 10$, $k = 2$ and $r_i = 0.9$ for all i , then

$$(13) \approx 0.3311, \quad (34) \approx 0.1355.$$

- The bound (13) is advantageous when the weights are very close to one. For instance, if $n = 10$, $k = 2$, and $r_i = 0.995$ for all i , then

$$(13) \approx 0.1110, \quad (34) \approx 0.1122.$$

In short, the probabilistic bound (34) is sharper over a broad range of non-exchangeable weights, whereas the combinatorial bound (13) retains an advantage in the nearly exchangeable regime.

Finally, following the same argument in Corollary 1 (from Theorem 1), we get the weighted de Finetti theorem under the weaker assumption $\sum_{i=1}^{\infty} r_i = \infty$. Under the additional assumptions imposed here, i.e., $\mathcal{X}$ is Polish and the weights λ_i are bounded and continuous, this recovers an instance of [15] (Theorem 6), corresponding to the choice $\lambda^* \equiv 1$ in the sufficient condition therein (see (3)).

Corollary 2. *Let P be the distribution of a λ -exchangeable sequence $(X_1, X_2, \dots)$ on $\mathcal{X}^{\infty}$. Assume that λ_i are bounded continuous functions such that*

$$r_i := \frac{\inf_{\mathcal{X}} \lambda_i(x)}{\sup_{\mathcal{X}} \lambda_i(x)} > 0, \text{ for } i = 1, 2, \dots, \quad \text{and} \quad \sum_{i=1}^{\infty} r_i = \infty. \quad (41)$$

Then P is a mixture of weighted i.i.d. distributions on $\mathcal{X}^{\infty}$.

5. Conclusions

This paper developed two quantitative approaches to finite weighted exchangeability. The combinatorial approach extends the classical Diaconis–Freedman argument by comparing weighted sampling with and without replacement through a likelihood-ratio estimate. The probabilistic approach uses a thinning construction to reduce the problem to an urn with unweighted tail coordinates. The resulting bounds have different strengths. The combinatorial bound can be sharper in the nearly exchangeable case, and recovers the sharp Diaconis–Freedman bound for exchangeable sequences. Away from this regime, however, the probabilistic bound is typically sharper: it depends on the tail weights through $\sum_{i=k+1}^n r_i$, rather than through the product $\prod_{i=k+1}^n r_i$. In particular, the probabilistic bound

yields, by passing to the limit, a weighted de Finetti representation under the condition $\sum_{i=1}^{\infty} r_i = \infty$, subject to some regularity assumptions.

There are several directions to extend this work. First, the sharp dependence of the finite approximation bounds on the weights is not fully understood. For instance, it would be interesting to know whether the factors $(\prod_{i=1}^k r_i)^{-1}$ in (13) and $C_k(r)$ in (34) are optimal. Second, for infinite weighted exchangeable sequences, the condition $\sum_{i=1}^{\infty} r_i = \infty$ (in Corollary 2) is known to be sufficient [15], but its necessity remains open for general $\mathcal{X}$. Understanding whether this condition, or its more general formulation involving a reference weight, is necessary appears to be a foundational question in the theory of weighted exchangeability. Finally, since weighted exchangeability was motivated in part by distribution shift and conformal prediction, it would be interesting to translate the finite approximation bounds obtained here into quantitative guarantees for predictive inference with finite calibration samples.

Funding: This research was supported by NSF CAREER Award DMS-2538791 and NSF MFAI Award DMS-2602038.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Bickel, P.J.; Chen, A. A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21068–21073. [CrossRef]
2. Diaconis, P.; Janson, S. Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* **2008**, *28*, 33–61.
3. Orbanz, P.; Roy, D.M. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 437–461. [CrossRef]
4. Andersen, E.S. On sums of symmetrically dependent random variables. *Skand. Aktuarietidskr.* **1953**, *36*, 123–138. [CrossRef]
5. Pitman, J.; Tang, W. Extreme order statistics of random walks. *Ann. Inst. Henri Poincaré Probab. Stat.* **2023**, *59*, 97–116. [CrossRef]
6. Campbell, T.; Cai, D.; Broderick, T. Exchangeable trait allocations. *Electron. J. Stat.* **2018**, *12*, 2290–2322. [CrossRef]
7. Pitman, J. Exchangeable and partially exchangeable random partitions. *Probab. Theory Relat. Fields* **1995**, *102*, 145–158. [CrossRef]
8. Holmes, C.C.; Walker, S.G. Statistical inference with exchangeability and martingales. *Philos. Trans. Roy. Soc. A* **2023**, *381*, 20220143. [CrossRef]
9. Korshunova, I.; Degraeve, J.; Huszár, F.; Gal, Y.; Gretton, A.; Dambre, J. Bruno: A deep recurrent model for exchangeable data. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*; NeurIPS: San Diego, CA, USA, 2018; Volume 31.
10. Rubin, D.B. Bayesian inference for causal effects: The role of randomization. *Ann. Stat.* **1978**, *6*, 34–58. [CrossRef]
11. Finetti, B.D. Funzione caratteristica di un fenomeno aleatorio. In *Atti del Congresso Internazionale dei Matematici*; Zanichelli Editore: Bologna, Italy, 1929; pp. 179–190.
12. Hewitt, E.; Savage, L.J. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **1955**, *80*, 470–501. [CrossRef]
13. Aldous, D.J. Exchangeability and related topics. In *École D’été de Probabilités de Saint-Flour, XIII—1983*; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 1985; Volume 1117, pp. 1–198.
14. Kallenberg, O. *Probabilistic Symmetries and Invariance Principles*; Springer: New York, NY, USA, 2005.
15. Barber, R.F.; Candes, E.J.; Ramdas, A.; Tibshirani, R.J. De Finetti’s theorem and related results for infinite weighted exchangeable sequences. *Bernoulli* **2024**, *30*, 3004–3028. [CrossRef]
16. Barber, R.F.; Candès, E.J.; Ramdas, A.; Tibshirani, R.J. Conformal prediction beyond exchangeability. *Ann. Stat.* **2023**, *51*, 816–845. [CrossRef]
17. Tibshirani, R.J.; Barber, R.F.; Candès, E.J.; Ramdas, A. Conformal prediction under covariate shift. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2019; Volume 32.
18. Lauritzen, S.L. *Extremal Families and Systems of Sufficient Statistics*; Lecture Notes in Statistics; Springer: New York, NY, USA, 1988; Volume 49.

19. Aldous, D.; Pitman, J. On the zero-one law for exchangeable events. *Ann. Probab.* **1979**, *7*, 704–723. [[CrossRef](#)]
20. Diaconis, P.; Freedman, D. Finite exchangeable sequences. *Ann. Probab.* **1980**, *8*, 745–764. [[CrossRef](#)]
21. Diaconis, P.; Freedman, D. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Stat.* **1987**, *23*, 397–423.
22. Kerns, G.J.; Székely, G.J. De Finetti's theorem for abstract finite exchangeable sequences. *J. Theor. Probab.* **2006**, *19*, 589–608. [[CrossRef](#)]
23. Carlier, G.; Friesecke, G.; Vögler, D. Convex geometry of finite exchangeable laws and de Finetti style representation with universal correlated corrections. *Probab. Theory Relat. Fields* **2023**, *185*, 311–351. [[CrossRef](#)]
24. Bobkov, S.G. Generalized symmetric polynomials and an approximate de Finetti representation. *J. Theor. Probab.* **2005**, *18*, 399–412. [[CrossRef](#)]
25. Freedman, D. A remark on the difference between sampling with and without replacement. *J. Amer. Stat. Assoc.* **1977**, *72*, 681. [[CrossRef](#)]
26. Dubins, L.E.; Freedman, D.A. Exchangeable processes need not be mixtures of independent, identically distributed random variables. *Z. Wahrsch. Verw. Geb.* **1979**, *48*, 115–132. [[CrossRef](#)]
27. Freedman, D.A. A mixture of independent identically distributed random variables need not admit a regular conditional probability given the exchangeable σ -field. *Z. Wahrsch. Verw. Geb.* **1980**, *51*, 239–248. [[CrossRef](#)]
28. Tang, W.; Tang, F. The Poisson binomial distribution—Old & new. *Stat. Sci.* **2023**, *38*, 108–119. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.