Bernoulli **30**(4), 2024, 3004–3028
https://doi.org/10.3150/23-BEJ1704

De Finetti's theorem and related results for infinite weighted exchangeable sequences

RINA FOYGEL BARBER[1,a], EMMANUEL J. CANDÈS[2,b], AADITYA RAMDAS[3,c] and RYAN J. TIBSHIRANI[4,d]

[1]*Department of Statistics, University of Chicago, Chicago, IL, USA,* [a]*rina@uchicago.edu*
[2]*Departments of Statistics and Mathematics, Stanford University, Stanford, CA, USA,* [b]*candes@stanford.edu*
[3]*Departments of Statistics and Machine Learning, Carnegie Mellon University, Pittsburgh, PA, USA,* [c]*aramdas@cmu.edu*
[4]*Department of Statistics, University of California, Berkeley, CA, USA,* [d]*ryantibs@berkeley.edu*

De Finetti's theorem, also called the de Finetti–Hewitt–Savage theorem, is a foundational result in probability and statistics. Roughly, it says that an infinite sequence of exchangeable random variables can always be written as a mixture of independent and identically distributed (i.i.d.) sequences of random variables. In this paper, we consider a weighted generalization of exchangeability that allows for weight functions to modify the individual distributions of the random variables along the sequence, provided that – modulo these weight functions – there is still some common exchangeable base measure. We study conditions under which a de Finetti-type representation exists for weighted exchangeable sequences, as a mixture of distributions which satisfy a weighted form of the i.i.d. property. Our approach establishes a nested family of conditions that lead to weighted extensions of other well-known related results as well, in particular, extensions of the zero-one law and the law of large numbers.

Keywords: De Finetti; exchangeability

1. Introduction

Nearly 100 years ago, de Finetti (1929) established a result that connects an infinite exchangeable sequence of binary random variables to a mixture of i.i.d. sequences of binary random variables. De Finetti's result says that $X_1, X_2, \ldots \in \{0,1\}$ are exchangeable if and only if a draw from their joint distribution can be equivalently represented as:

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

for some distribution μ on $[0,1]$. In other words, any infinite exchangeable binary sequence can be represented as a mixture of i.i.d. Bernoulli sequences.

This result has been extended well beyond the binary case, to a general space $\mathcal{X}$, due to work by de Finetti and others, most notably Hewitt and Savage (1955). The more general result is known as the de Finetti–Hewitt–Savage theorem, but is also often simply called de Finetti's theorem. It says that, under fairly mild assumptions on $\mathcal{X}$, any infinite exchangeable sequence $X_1, X_2, \ldots \in \mathcal{X}$ can be represented as a mixture of i.i.d. sequences. See Theorem 1 below for a formal statement, and the paragraphs that follow for more references and discussion of the history of contributions in this area. De Finetti's theorem is widely considered to be of foundational importance in probability and statistics. Many authors also view it as a central point of motivation for Bayesian inference; e.g., see Schervish (2012) for additional background on the role de Finetti's theorem plays in statistical theory and methodology.

Some authors have also studied a *weighted* notion of exchangeability. This was considered by Lauritzen (1988) in the binary case $\mathcal{X} = \{0,1\}$, and by Tibshirani et al. (2019) for general $\mathcal{X}$ in the context of predictive inference under distribution shift. For example, in the case of continuously-distributed and real-valued random variables $X_1, \ldots, X_n$, we say (following Tibshirani et al. (2019)) that their distribution is *weighted exchangeable* with respect to given weight functions $\lambda_1, \ldots, \lambda_n$ if the joint density f_n of $X_1, \ldots, X_n$ factorizes as

$$f_n(x_1, \ldots, x_n) = \left(\prod_{i=1}^{n} \lambda_i(x_i)\right) \cdot g_n(x_1, \ldots, x_n), \quad \text{for } x_1, \ldots, x_n \in \mathbb{R},$$

for a function g_n that is symmetric, i.e., invariant to any permutation of its arguments. Note that if instead f_n itself is symmetric (equivalently, if the above holds with all constant weight functions $\lambda_i \equiv 1$), then this reduces to the ordinary (unweighted) notion of exchangeability. The definition of weighted exchangeability can be extended to infinite sequences in the usual way: by requiring that the joint distribution of each finite subsequence of random variables be weighted exchangeable as per the above. It also extends beyond continuously-distributed, real-valued random variables. See Definitions 3 and 4 for the formal details.

In light of what de Finetti's theorem teaches us about exchangeable distributions, it is natural to ask whether infinite weighted exchangeable distributions have an analogous property: to put it informally, can an infinite weighted exchangeable sequence be represented as a mixture of infinite weighted i.i.d. (i.e., weighted exchangeable and mutually independent) sequences?

It turns out that the weight functions $\lambda_1, \lambda_2, \ldots$ play a critical role in determining whether or not this is true, and a central question in this paper is:[1]

Q1. Which sequences of weight functions lead to a generalized de Finetti representation?

Rather than trying to answer the above question on its own, we find it interesting to embed it into a larger problem of comparing the answers to *three* questions – Q1, and the following two questions, on weighted extensions of two other well-known results in the i.i.d. case, namely, the zero-one law (Theorem 2) and the law of large numbers (Theorem 3):

Q2. Which sequences of weight functions lead to a generalized zero-one law?
Q3. Which sequences of weight functions lead to a generalized law of large numbers?

Our main result, in Theorem 4 below, relates questions Q1–Q3. In short we show that "answers to Q1" (weight functions leading to a de Finetti-type representation) are a subset of "answers to Q2", which are themselves a subset of "answers to Q3". We also provide complementary necessary and sufficient conditions.

General assumptions and notation.

Throughout this paper, we will work in a space $\mathcal{X}$ that we assume is a separable complete metric space (this certainly includes, for example, any finite-dimensional Euclidean space, $\mathcal{X} = \mathbb{R}^d$). We denote the Borel σ-algebra of $\mathcal{X}$ by $\mathcal{B}(\mathcal{X})$. Under these assumptions, we call $\mathcal{X}$, equipped with its Borel σ-algebra $\mathcal{B}(\mathcal{X})$, a standard Borel space.

We write $\mathcal{X}^n = \mathcal{X} \times \cdots \times \mathcal{X}$ (n fold) and $\mathcal{X}^\infty = \mathcal{X} \times \mathcal{X} \times \cdots$ for the finite and countably infinite product spaces, respectively, and $\mathcal{B}(\mathcal{X}^n)$ and $\mathcal{B}(\mathcal{X}^\infty)$ for their respective Borel σ-algebras. We note that since $\mathcal{X}$ is a standard Borel space, it is second countable, which means it has a countable base. This implies (among other important facts) that the countable product $\mathcal{X}^\infty$ is also a standard Borel space, and $\mathcal{B}(\mathcal{X}^\infty)$ is generated by the product of individual Borel σ-algebras $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}) \times \cdots$ (e.g., see Lemma 6.4.2 part (ii) in Bogachev (2007)). This same property is of course also true for finite products.

[1] As we will describe later, this question was already answered by Lauritzen (1988) for the binary case $\mathcal{X} = \{0, 1\}$.

For a measure Q on $\mathcal{X}^\infty$ and any $n \ge 1$, we use Q_n for the associated marginal measure on $\mathcal{X}^n$,

$$Q_n(A) = Q(A \times \mathcal{X} \times \mathcal{X} \times \cdots), \quad \text{for } A \in \mathcal{B}(\mathcal{X}^n).$$

We use $\mathcal{M}_\mathcal{X}$ to denote the set of measures on $\mathcal{X}$.[2] This is itself a measure space, with a σ-algebra generated by sets of the form $\{P \in \mathcal{M}_\mathcal{X} : P(A) \le t\}$, for $A \in \mathcal{B}(\mathcal{X})$ and $t \ge 0$. We use $\mathcal{P}_\mathcal{X} \subseteq \mathcal{M}_\mathcal{X}$ to denote the set of distributions on $\mathcal{X}$, that is, $\mathcal{P}_\mathcal{X} = \{P \in \mathcal{M}_\mathcal{X} : P(\mathcal{X}) = 1\}$. For $P \in \mathcal{P}_\mathcal{X}$, we write $P^n = P \times P \times \cdots \times P$ (n fold) and $P^\infty = P \times P \times \cdots$ for the finite and countably infinite product distributions on $\mathcal{X}^n$ and $\mathcal{X}^\infty$, respectively; so that writing $(X_1, X_2, \dots) \sim P^\infty$ is the same as writing $X_1, X_2, \dots \overset{\text{i.i.d.}}{\sim} P$, and similarly for P^n. We use δ_x to denote the distribution defined by a point mass at any given $x \in \mathcal{X}$. Finally, we use $\overset{\text{a.s.}}{=}$ and $\overset{\text{a.s.}}{\to}$ to denote almost sure equality and almost sure convergence, respectively.

2. Exchangeability and weighted exchangeability

Exchangeability is a property of a sequence of random variables that expresses the notion, roughly speaking, that the sequence is "equally likely to appear in any order". For a distribution on a finite sequence $(X_1, \dots, X_n) \in \mathcal{X}^n$, or more generally, for a measure on $\mathcal{X}^n$, we can define this property in terms of invariance to permutations, as follows.[3]

Definition 1 (Finite exchangeability). A measure Q on $\mathcal{X}^n$ is called *exchangeable* provided that, for all $A_1, \dots, A_n \in \mathcal{B}(\mathcal{X})$,

$$Q(A_1 \times \cdots \times A_n) = Q(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}), \quad \text{for all } \sigma \in \mathcal{S}_n,$$

where $\mathcal{S}_n$ is the set of permutations on $[n] = \{1, \dots, n\}$.

This definition can be extended to a distribution on an infinite sequence $(X_1, X_2, \dots) \in \mathcal{X}^\infty$, or more generally, to a measure on $\mathcal{X}^\infty$, as follows.

Definition 2 (Infinite exchangeability). A measure Q on $\mathcal{X}^\infty$ is called *exchangeable* if for all $n \ge 1$ the corresponding marginal measure Q_n is exchangeable.

In this paper, we will study a weighted generalization of exchangeability. We denote by $\Lambda = \Lambda_\mathcal{X}$ the set of measurable functions from $\mathcal{X}$ to $(0, \infty)$, with Λ^n or Λ^∞ denoting the finite or countably infinite product, respectively, of this space. First, we define the notion of weighted exchangeability for finite sequences.

Definition 3 (Finite weighted exchangeability). Given $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda^n$, a measure Q on $\mathcal{X}^n$ is called *λ-weighted exchangeable* if the measure $\bar{Q}$ defined as

$$\bar{Q}(A) = \int_A \frac{\mathsf{d}Q(x_1, \dots, x_n)}{\lambda_1(x_1) \cdots \lambda_n(x_n)}, \quad \text{for } A \in \mathcal{B}(\mathcal{X}^n)$$

is an exchangeable measure.

[2]We allow measures on $\mathcal{X}$ to be nonfinite, i.e., a measure $Q \in \mathcal{M}_\mathcal{X}$ may have $Q(\mathcal{X}) = \infty$.

[3]Throughout this paper, we will use the terms "exchangeable" and "i.i.d." (and later on, "weighted exchangeable" and "weighted i.i.d.") to refer either to a measure Q itself, or to a random sequence X drawn from Q (when Q is a distribution), depending on the context.

Similar to the unweighted case, we can extend this definition to infinite sequences.

Definition 4 (Infinite weighted exchangeability). Given $\lambda = (\lambda_1, \lambda_2, \ldots) \in \Lambda^\infty$, a measure Q on $\mathcal{X}^\infty$ is called *λ-weighted exchangeable* if for all $n \geq 1$ the corresponding marginal measure Q_n is $(\lambda_1, \ldots, \lambda_n)$-weighted exchangeable.

2.1. Background on exchangeable distributions

This section provides background on (unweighted) exchangeability, and reviews key properties. Readers familiar with these topics may wish to skip ahead to Section 2.2.

2.1.1. Mixtures and i.i.d. sequences

An important special case of exchangeability is given by an independent and identically distributed (i.i.d.) process. For any distribution P on $\mathcal{X}$, the product P^n is a finitely exchangeable distribution on $\mathcal{X}^n$, whereas the countable product $P^\infty = P \times P \times \cdots$ is an infinitely exchangeable distribution on $\mathcal{X}^\infty$. More generally, exchangeability is always preserved under mixtures, as the following result recalls, which we state without proof.

Proposition 1. *Any mixture of exchangeable measures on $\mathcal{X}^n$ (or $\mathcal{X}^\infty$) is itself an exchangeable measure on $\mathcal{X}^n$ (or $\mathcal{X}^\infty$).*

In particular, this means that any mixture of i.i.d. sequences has an exchangeable distribution.

Corollary 1. *For any distribution μ on $\mathcal{P}_\mathcal{X}$, the distribution Q on $\mathcal{X}^n$ defined by*

$$\textit{sample } P \sim \mu, \textit{ then draw } X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} P$$

is exchangeable. Similarly, the distribution Q on $\mathcal{X}^\infty$ defined by

$$\textit{sample } P \sim \mu, \textit{ then draw } X_1, X_2, \ldots \stackrel{i.i.d.}{\sim} P$$

is exchangeable.

For the infinite setting, we use $Q = (P^\infty)_\mu$ to denote this mixture. That is, for a distribution μ on $\mathcal{P}_\mathcal{X}$, the distribution $Q = (P^\infty)_\mu$ is the infinitely exchangeable distribution defined by

$$Q(A) = \mathbb{E}_{P \sim \mu}\left[\left(P^\infty\right)(A)\right], \quad \text{for } A \in \mathcal{B}(\mathcal{X}^\infty).$$

To unpack this further, note that we can also equivalently define this mixture distribution via

$$Q(A_1 \times A_2 \times \cdots) = \int \prod_{i=1}^\infty P(A_i) \mathsf{d}\mu(P), \quad \text{for } A_1, A_2, \ldots \in \mathcal{B}(\mathcal{X}),$$

as $\mathcal{B}(\mathcal{X}^\infty)$ is generated by $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}) \times \cdots$ (which holds because $\mathcal{X}$ is standard Borel).

2.1.2. De Finetti's theorem

The well-known de Finetti theorem, sometimes called the de Finetti–Hewitt–Savage theorem, gives a converse to Corollary 1, for the infinite setting.

Theorem 1 (De Finetti–Hewitt–Savage). *Let Q be an exchangeable distribution on $\mathcal{X}^\infty$. Then there exists a distribution μ on $\mathcal{P}_\mathcal{X}$ such that $Q = (P^\infty)_\mu$. Moreover, the distribution μ satisfying this equality is unique.*

This result was initially proved by de Finetti (1929) for the special case of binary sequences, where $\mathcal{X} = \{0,1\}$. In this case the mixing distribution μ can simply be viewed as a distribution on $p \in [0,1]$, where p gives the parameter for the Bernoulli distribution P on $\mathcal{X}$, and then Theorem 1 has a particularly simple interpretation: given any exchangeable distribution Q on infinite binary sequences, there exists a distribution μ on $[0,1]$ such that draws from Q can be expressed as

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

as described earlier in Section 1. After his pioneering 1929 work, de Finetti (1937) extended this theorem to real-valued sequences, where $\mathcal{X} = \mathbb{R}$, which was also later established independently by Dynkin (1953). Hewitt and Savage (1955) generalized de Finetti's result to a much more abstract setting, which covers what are called Baire measurable random variables taking values in a compact Hausdorff space. Further generalizations to broader classes of spaces can be found in Farrell (1962), Maitra (1977), among others. Quite recently, Alam (2020) generalized this to cover any Hausdorff space, under the assumption that the common marginal distribution Q_1 on $\mathcal{X}$ is a Radon measure. In particular, as stated in Theorem 1, we emphasize that de Finetti's theorem holds when $\mathcal{X}$ is a standard Borel space – this can be seen a consequence of the general topological result in Hewitt and Savage (1955), as pointed out by Varadarajan (1963). See also the discussion after Theorem 1.4 in Alam (2020), or Theorem 2.1 in Fritz, Gonda and Perrone (2021).

Notably, completeness – inherent to standard Borel spaces – is critical here because without it de Finetti's theorem can fail. This was shown by Dubins and Freedman (1979), who constructed an infinite sequence of Borel measurable exchangeable random variables in a separable metric space that cannot be expressed as a mixture of i.i.d. processes. Furthermore, the assumption of infinite exchangeability is also critical, because the result can fail in the finite setting. As a simple example, let $\mathcal{X} = \{0,1\}$, $n = 2$, and let Q place equal mass on $(1,0)$ and on $(0,1)$ (and no mass on $(0,0)$ or $(1,1)$). Then Q is exchangeable, but clearly cannot be realized by mixing the distributions of i.i.d. Bernoulli sequences.

Lastly, we remark that the above discussion is by no means a comprehensive treatment of the work on de Finetti's theorem, its extensions and applications, or alternative proofs. For a broader perspective, see, for example, Aldous (1985), Diaconis and Freedman (1987), Lauritzen (1988).

2.1.3. Special properties of i.i.d. sequences

Before moving on to discuss our main results, we mention some additional well-known results that apply to infinite i.i.d. sequences (but not to infinitely exchangeable sequences more generally), so that we can compare to the weighted case later on.

We begin by recalling the Hewitt–Savage zero-one law, which is also due to Hewitt and Savage (1955). See also Kingman (1978).

Theorem 2 (Hewitt–Savage zero-one law). *For any distribution $P \in \mathcal{P}_{\mathcal{X}}$, it holds that*

$$P^{\infty}(A) \in \{0,1\}, \quad \textit{for all } A \in \mathcal{E}_{\infty},$$

where $\mathcal{E}_{\infty} \subseteq \mathcal{B}(\mathcal{X}^{\infty})$ is the sub-σ-algebra of exchangeable events.

The sub-σ-algebra $\mathcal{E}_{\infty} \subseteq \mathcal{B}(\mathcal{X}^{\infty})$ referenced in the theorem is defined as

$$\mathcal{E}_{\infty} = \Big\{A \in \mathcal{B}\big(\mathcal{X}^{\infty}\big) : \text{for all } x \in A,\, n \geq 1, \text{ and } \sigma \in \mathcal{S}_n,$$
$$\text{it holds that } (x_{\sigma(1)}, \ldots, x_{\sigma(n)}, x_{n+1}, x_{n+2}, \ldots) \in A\Big\}.$$

This is often called the *exchangeable σ-algebra*. To give a concrete example, the event "we never observe the value zero" is in $\mathcal{E}_{\infty}$, since we can express it as $A = \{x \in \mathcal{X}^{\infty} : \sum_{i=1}^{\infty} \mathbb{1}_{x_i = 0} = 0\}$, and this satisfies the permutation invariance condition in the last display.

Next, we recall the strong law of large numbers,[4] due to Kolmogorov (1930) (whereas weaker versions date back earlier to Bernoulli, Chebyshev, Markov, Borel, and others). For simplicity, we will drop the specifier "strong" henceforth, and simply refer to this as the "law of large numbers".

Theorem 3 (Law of large numbers). *For any distribution $P \in \mathcal{P}_{\mathcal{X}}$, given $X_1, X_2, \ldots \stackrel{i.i.d.}{\sim} P$, write*

$$\widehat{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

to denote the empirical distribution of the first n random variables. Then it holds that

$$\widehat{P}_n(A) \stackrel{a.s.}{\to} P(A), \quad \textit{as } n \to \infty, \textit{ for all } A \in \mathcal{B}(\mathcal{X}).$$

This is particularly interesting when combined with de Finetti's theorem (recall Theorem 1). For any exchangeable distribution Q on $\mathcal{X}^{\infty}$, by de Finetti's theorem, we can express $Q = (P^{\infty})_{\mu}$ for a distribution μ on $\mathcal{P}_{\mathcal{X}}$; then by the law of large numbers (which we apply after conditioning on the draw $P \sim \mu$), the unknown (random) distribution P can be recovered from the observed sequence $X \in \mathcal{X}^{\infty}$ by taking the limit of its empirical distribution.

2.2. Mixtures in the weighted case, and weighted i.i.d. sequences

The next result extends Proposition 1 to weighted exchangeable distributions. We omit its proof since the result follows immediately from the definition of weighted exchangeability.

Proposition 2. *For any $\lambda \in \Lambda^n$ (or Λ^{∞}), any mixture of λ-weighted exchangeable distributions on $\mathcal{X}^n$ (or $\mathcal{X}^{\infty}$) is itself a λ-weighted exchangeable distribution on $\mathcal{X}^n$ (or $\mathcal{X}^{\infty}$).*

We next introduce some notation that will help us concisely represent certain product distributions. For $P \in \mathcal{M}_{\mathcal{X}}$ and $\lambda \in \Lambda$, we define the distribution $P \circ \lambda \in \mathcal{P}_{\mathcal{X}}$ by

$$(P \circ \lambda)(A) = \frac{\int_A \lambda(x) \, \mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda(x) \, \mathsf{d}P(x)}, \quad \text{for } A \in \mathcal{B}(\mathcal{X}).$$

[4] To be precise, what we state in Theorem 3 is actually simply the application of the law of large numbers to indicator variables, as this special case is most pertinent to our study.

Here, we are effectively "reweighting" the measure P according to the weight function λ. Note that $P \circ \lambda$ is well-defined for all P in the set $\mathcal{M}_{\mathcal{X}}(\lambda) = \{P \in \mathcal{M}_{\mathcal{X}} : 0 < \int_{\mathcal{X}} \lambda(x)\,\mathsf{d}P(x) < \infty\}$. For finite or infinite sequences of weight functions, we will overload this notation as follows: when $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda^n$, we use $P \circ \lambda$ to denote the distribution on $\mathcal{X}^n$ defined by

$$P \circ \lambda = (P \circ \lambda_1) \times \cdots \times (P \circ \lambda_n),$$

and when $\lambda = (\lambda_1, \lambda_2 \dots) \in \Lambda^\infty$, we use $P \circ \lambda$ to denote the distribution on $\mathcal{X}^\infty$ defined by

$$P \circ \lambda = (P \circ \lambda_1) \times (P \circ \lambda_2) \times \cdots,$$

and either of these mixtures are well-defined for all P in the set (overloading notation once more) $\mathcal{M}_{\mathcal{X}}(\lambda) = \{P \in \mathcal{M}_{\mathcal{X}} : 0 < \int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P(x) < \infty, \text{ for all } i\}$.

With this notation in hand, it is worth noting that for a distribution Q on $\mathcal{X}^n$ or $\mathcal{X}^\infty$ and any λ in Λ^n or Λ^∞,

$$\begin{aligned} &Q \text{ is both a product distribution and } \lambda\text{-weighted exchangeable} \quad \Longleftrightarrow \\ &\quad Q = P \circ \lambda \text{ for some } P \in \mathcal{M}_{\mathcal{X}}. \end{aligned} \tag{1}$$

Each direction is simple to check using Definition 3 or 4. We will use the term *λ-weighted i.i.d.* to refer to any distribution Q satisfying (1).

Finally, as in the original unweighted case, Proposition 2 has a nice interpretation for weighted i.i.d. sequences.

Corollary 2. *For any $\lambda \in \Lambda^n$ and distribution μ on $\mathcal{M}_{\mathcal{X}}(\lambda)$, the distribution Q on $\mathcal{X}^n$ defined by*

$$\textit{sample } P \sim \mu, \textit{ then draw } X_1, \dots, X_n \sim P \circ \lambda$$

is λ-weighted exchangeable. Similarly, for any $\lambda \in \Lambda^\infty$ and distribution μ on $\mathcal{M}_{\mathcal{X}}(\lambda)$, the distribution Q on $\mathcal{X}^\infty$ defined by

$$\textit{sample } P \sim \mu, \textit{ then draw } X_1, X_2, \dots \sim P \circ \lambda$$

is λ-weighted exchangeable.

For the infinite setting, we will use $Q = (P \circ \lambda)_\mu$ to denote this mixture, that is, for any $\lambda \in \Lambda^\infty$ and any distribution μ on $\mathcal{M}_{\mathcal{X}}(\lambda)$, the distribution $Q = (P \circ \lambda)_\mu$ is the λ-weighted exchangeable distribution given by

$$Q(A) = \mathbb{E}_{P \sim \mu}\big[(P \circ \lambda)(A)\big], \quad \text{for } A \in \mathcal{B}(\mathcal{X}^\infty),$$

or equivalently, in less compact notation,

$$Q(A_1 \times A_2 \times \cdots) = \int \prod_{i=1}^{\infty} (P \circ \lambda_i)(A_i)\,\mathsf{d}\mu(P), \quad \text{for } A_1, A_2, \dots \in \mathcal{B}(\mathcal{X}).$$

3. Main results

We now present the central questions and main findings of this work. In short, we are interested in examining whether de Finetti's theorem for infinitely exchangeable sequences can be generalized to the setting of infinite weighted exchangeability. Along the way, we will also consider whether results for i.i.d. sequences – namely, the Hewitt–Savage zero-one law and the law of large numbers – can be extended to the weighted case, as well.

3.1. Framework

We now define the properties we wish to examine. These are all weighted generalizations of the properties in the ordinary unweighted setting, described above.

The weighted de Finetti property.

We say that a sequence $\lambda \in \Lambda^\infty$ satisfies the *weighted de Finetti property* if, for any λ-weighted exchangeable distribution Q on $\mathcal{X}^\infty$, there is a distribution μ on $\mathcal{M}_\mathcal{X}(\lambda)$ such that $Q = (P \circ \lambda)_\mu$, or equivalently,

$$Q(A) = \mathbb{E}_{P\sim\mu}\left[(P \circ \lambda)(A)\right], \quad \text{for all } A \in \mathcal{B}(\mathcal{X}^\infty).$$

In other words, this says that for any λ-weighted exchangeable distribution Q, we can represent it by mixing the distributions of λ-weighted i.i.d. sequences. For our main results that follow, it will be useful to denote this condition compactly. To this end, we write:

$$\Lambda_{\mathsf{dF}} = \big\{\lambda \in \Lambda^\infty : \text{ for all } \lambda\text{-weighted exchangeable distributions } Q \text{ on } \mathcal{X}^\infty, \\ \text{there exists a distribution } \mu \text{ on } \mathcal{M}_\mathcal{X}(\lambda) \text{ such that } Q = (P \circ \lambda)_\mu\big\}.$$

The weighted zero-one law.

We say that a sequence $\lambda \in \Lambda^\infty$ satisfies the *weighted zero-one law* if, for any $P \in \mathcal{M}_\mathcal{X}(\lambda)$, each event $A \in \mathcal{E}_\infty$ is assigned either probability zero or one by $P \circ \lambda$. To represent this condition compactly, we write:

$$\Lambda_{01} = \big\{\lambda \in \Lambda^\infty : \text{for all } P \in \mathcal{M}_\mathcal{X}(\lambda) \text{ and } A \in \mathcal{E}_\infty, \text{ it holds that } (P \circ \lambda)(A) \in \{0,1\}\big\}.$$

The weighted law of large numbers.

We say that a sequence $\lambda \in \Lambda^\infty$ satisfies the *weighted law of large numbers* if, for any $P \in \mathcal{M}_\mathcal{X}(\lambda)$, the following holds: for $(X_1, X_2, \ldots) \sim P \circ \lambda$,

$$\widetilde{P}_{n,i}(A) \stackrel{\text{a.s.}}{\to} (P \circ \lambda_i)(A), \quad \text{as } n \to \infty, \text{ for all } i \geq 1 \text{ and } A \in \mathcal{B}(\mathcal{X}),$$

where $\widetilde{P}_{n,i}$ is a certain weighted empirical distribution of $X_1, \ldots, X_n$, defined by

$$\widetilde{P}_{n,i} = \sum_{j=1}^{n} \big(w_{n,i}(X_1, \ldots, X_n)\big)_j \cdot \delta_{X_j},$$
$$\text{where} \big(w_{n,i}(x_1, \ldots, x_n)\big)_j = \frac{\sum_{\sigma \in \mathcal{S}_n : \sigma(i)=j} \prod_{k=1}^{n} \lambda_k(x_{\sigma(k)})}{\sum_{\sigma \in \mathcal{S}_n} \prod_{k=1}^{n} \lambda_k(x_{\sigma(k)})}, \; j = 1, \ldots, n. \tag{2}$$

(Later on, in Proposition 7, we will see that $\widetilde{P}_{n,i}$ determines the distribution of the random variable X_i, if we condition on observing the *unordered* collection of values $X_1, \ldots, X_n$. To compare this to the unweighted case, observe that we would just have $(w_{n,i}(X_1, \ldots, X_n))_j \equiv \frac{1}{n}$ for each j, so in this case we can view X_i as a uniform random draw from $X_1, \ldots, X_n$ after conditioning on this list.) To represent the above condition compactly, we write:

$$\Lambda_{\mathsf{LLN}} = \Big\{\lambda \in \Lambda^\infty : \text{ for all } P \in \mathcal{M}_{\mathcal{X}}(\lambda), i \ge 1, \text{ and } A \in \mathcal{B}(\mathcal{X}), \text{ if } X \sim P \circ \lambda,$$
$$\text{then } \widetilde{P}_{n,i}(A) \overset{\text{a.s.}}{\to} (P \circ \lambda_i)(A), \text{ for } \widetilde{P}_{n,i} \text{ as defined in (2)}\Big\}.$$

Example 1. To build intuition for these various sets, here we pause to give a simple example. Consider the binary setting, $\mathcal{X} = \{0,1\}$, and define $\lambda \in \Lambda^\infty$ as the sequence of functions

$$\lambda_i(0) = 1, \qquad \lambda_i(1) = 2^{-i}, \quad i \ge 1.$$

A straightforward calculation verifies that this sequence does not belong to any of the three sets $\Lambda_{\mathsf{dF}}, \Lambda_{\mathsf{01}}, \Lambda_{\mathsf{LLN}}$ defined above. For instance, to see that $\lambda \notin \Lambda_{\mathsf{dF}}$, consider the following distribution Q on $\mathcal{X}^\infty$: defining $e_i = (0,\dots,0,1,0,0,\dots) \in \mathcal{X}^\infty$ as the sequence with a 1 in position i and 0s elsewhere, let

$$Q\big(\{e_i\}\big) = 2^{-i}, \quad i \ge 1,$$

and $Q(\{x\}) = 0$ for all other $x \in \mathcal{X}^\infty$. We can verify that Q is λ-weighted exchangeable, but since the sequence $X = (X_1, X_2, \dots)$ must contain a single 1 almost surely under Q, we can see that Q cannot be written as a mixture of λ-weighted i.i.d. distributions.

3.2. Main theorems

Our first main result in this paper establishes a connection between these three weighted properties. Its proof will be covered in Section 5.

Theorem 4 (Embedding of conditions). *It holds that*

$$\Lambda_{\mathsf{dF}} \subseteq \Lambda_{\mathsf{01}} \subseteq \Lambda_{\mathsf{LLN}}.$$

In other words, the above theorem says that for any $\lambda \in \Lambda^\infty$, the weighted de Finetti property implies the weighted zero-one law, which in turn implies the weighted law of large numbers.

Our next main results pertain to necessary and sufficient conditions for $\lambda \in \Lambda^\infty$ to lie in these sets. Their proofs are also covered in Section 5.

Theorem 5 (Necessary condition). *If $\lambda \in \Lambda_{\mathsf{LLN}}$, then λ satisfies*

$$\sum_{i=1}^\infty \min\big\{(P \circ \lambda_i)(A), (P \circ \lambda_i)\big(A^c\big)\big\} = \infty,$$
$$\textit{for all } P \in \mathcal{M}_{\mathcal{X}}(\lambda), A \in \mathcal{B}(\mathcal{X}) \textit{ with } P(A), P(A^c) > 0. \tag{3}$$

Theorem 6 (Sufficient condition). *If $\lambda \in \Lambda^\infty$ satisfies*

$$\sum_{i=1}^\infty \frac{\inf_{x \in \mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\sup_{x \in \mathcal{X}} \lambda_i(x)/\lambda_*(x)} = \infty \quad \textit{for some } \lambda_* \in \Lambda, \tag{4}$$

then $\lambda \in \Lambda_{\mathsf{dF}}$.

To summarize, combining our main results, we have:

$$\{\lambda \text{ satisfying (4)}\} \subseteq \Lambda_{\mathsf{dF}} \subseteq \Lambda_{\mathsf{01}} \subseteq \Lambda_{\mathsf{LLN}} \subseteq \{\lambda \text{ satisfying (3)}\}. \tag{5}$$

In what follows, we explore whether these set inclusions are strict, or whether they are equalities. In particular, we will derive precise answers to these questions for the special case where $\mathcal{X}$ has finite cardinality; in the infinite case, we will see that open questions remain.

3.3. Special case: Binary random variables

Before treating the finite case in full generality, it is useful to consider the case of binary random variables, where $\mathcal{X}$ contains two elements, and we can take $\mathcal{X} = \{0, 1\}$ without loss of generality. In this case, it turns out that all inclusions above, in (5), are equalities. This was already established by Lauritzen (1988). Below we show that this can be derived as a consequence of Theorem 4.

Theorem 7 (Adapted from Sections II.9.1, II.9.2 and Chapter II Theorem 4.4 of Lauritzen (1988)). *For the binary case where $\mathcal{X} = \{0, 1\}$, the five sets defined above are equal:*

$$\{\lambda \textit{ satisfying } (4)\} = \Lambda_{\mathsf{dF}} = \Lambda_{\mathsf{01}} = \Lambda_{\mathsf{LLN}} = \{\lambda \textit{ satisfying } (3)\}.$$

Moreover, the common necessary (3) and sufficient (4) conditions can equivalently be expressed as:

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

Proof. Lauritzen (1988) establishes this result for $\mathcal{X} = \{0, 1\}$ through the lens of extremal families of distributions. Here we instead give a simple proof by applying Theorems 4, 5, and 6, which hold for a general standard Borel space $\mathcal{X}$. It suffices to show that the necessary condition (3) implies the sufficient condition (4) when $\mathcal{X} = \{0, 1\}$.

Towards this end, let λ satisfy (3), thus $\sum_{i=1}^{\infty} \min\{(P \circ \lambda_i)(A), (P \circ \lambda_i)(A^c)\} = \infty$ holds for all measures $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$ and $A \in \mathcal{B}(\mathcal{X})$ with $P(A), P(A^c) > 0$. Fixing $P = \text{Bernoulli}(0.5) \in \mathcal{M}_{\mathcal{X}}(\lambda)$, and $A = \{0\}$, we have $P(A), P(A^c) > 0$, so we can apply (3). We calculate

$$(P \circ \lambda_i)(\{0\}) = \frac{\lambda_i(0)}{\lambda_i(1) + \lambda_i(0)}, \qquad (P \circ \lambda_i)(\{1\}) = \frac{\lambda_i(1)}{\lambda_i(1) + \lambda_i(0)}.$$

By (3), then,

$$\begin{aligned}
\infty &= \sum_{i=1}^{\infty} \min\{(P \circ \lambda_i)(A), (P \circ \lambda_i)(A^c)\} = \sum_{i=1}^{\infty} \min\{(P \circ \lambda_i)(\{0\}), (P \circ \lambda_i)(\{1\})\} \\
&= \sum_{i=1}^{\infty} \frac{\min\{\lambda_i(0), \lambda_i(1)\}}{\lambda_i(1) + \lambda_i(0)} \le \sum_{i=1}^{\infty} \frac{\min\{\lambda_i(0), \lambda_i(1)\}}{\max\{\lambda_i(0), \lambda_i(1)\}} = \sum_{i=1}^{\infty} \frac{\inf_{x \in \mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\sup_{x \in \mathcal{X}} \lambda_i(x)/\lambda_*(x)},
\end{aligned}$$

where in the last step we define the function $\lambda_* : \mathcal{X} \to (0, \infty)$ by $\lambda_*(x) \equiv 1$. This proves that, if the necessary condition (3) holds, then the sufficient condition (4) also holds. Finally, the fact that the condition: $\sum_{i=1}^{\infty} \min\{\lambda_i(0), \lambda_i(1)\}/\max\{\lambda_i(0), \lambda_i(1)\} = \infty$ is equivalent to the common necessary and sufficient condition is a consequence of the last display. □

3.4. Beyond the binary case

Now we move beyond the binary case. First we will see that when $|\mathcal{X}| \geq 3$, it will no longer be the case that the five sets in (5) are all equal: specifically, there is always a gap between the necessary condition (3) and the sufficient condition (4).

Proposition 3. *If* $|\mathcal{X}| \geq 3$, *then*

$$\{\lambda \textit{ satisfying } (4)\} \subsetneq \{\lambda \textit{ satisfying } (3)\},$$

that is, the necessary condition (3) *is strictly weaker than the sufficient condition* (4).

This result is proved with a simple example: we choose a partition $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1 \cup \mathcal{X}_2$ for some nonempty $\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2 \in \mathcal{B}(\mathcal{X})$, and define

$$\lambda_i(x) = \begin{cases} e^{-i} & x \in \mathcal{X}_{\mathrm{mod}(i,3)}, \\ 1 & x \notin \mathcal{X}_{\mathrm{mod}(i,3)}, \end{cases} \quad i \geq 1.$$

The full proof is given in the supplementary material (Barber et al., 2024).

Proposition 3 implies that, in the sequence of four set inclusions in (5), at least one of these set inclusions must be strict whenever $|\mathcal{X}| \geq 3$ – but the result does not specify exactly where this gap might occur. The following theorem resolves this question for the finite case, where $\mathcal{X}$ has finite cardinality. Its proof is given in the supplementary material (Barber et al., 2024).

Theorem 8. *If* $|\mathcal{X}| < \infty$, *then*

$$\Lambda_{\mathsf{dF}} = \Lambda_{\mathsf{01}} = \Lambda_{\mathsf{LLN}} = \{\lambda \textit{ satisfying } (3)\},$$

that is, the condition (3) *is in fact both necessary and sufficient for the sets* $\Lambda_{\mathsf{dF}}, \Lambda_{\mathsf{01}}, \Lambda_{\mathsf{LLN}}$.

For the infinite case, however, no such result is known at present. Combining all of our results so far, we can now summarize these different regimes as follows.

- **Binary case.** If $|\mathcal{X}| = 2$, then by Theorem 7,

  $$\{\lambda \text{ satisfying } (4)\} = \Lambda_{\mathsf{dF}} = \Lambda_{\mathsf{01}} = \Lambda_{\mathsf{LLN}} = \{\lambda \text{ satisfying } (3)\}.$$

  (The same holds trivially for the singleton case, $|\mathcal{X}| = 1$.)
- **Finite case.** If $2 < |\mathcal{X}| < \infty$, then by Proposition 3 and Theorem 8,

  $$\{\lambda \text{ satisfying } (4)\} \subsetneq \Lambda_{\mathsf{dF}} = \Lambda_{\mathsf{01}} = \Lambda_{\mathsf{LLN}} = \{\lambda \text{ satisfying } (3)\}.$$

  That is, condition (3) is both necessary and sufficient, and condition (4) is strictly stronger than needed.
- **Infinite case.** If $|\mathcal{X}| = \infty$, then by Proposition 3,

  $$\{\lambda \text{ satisfying } (4)\} \subsetneq \{\lambda \text{ satisfying } (3)\}.$$

  However, for the infinite case, it is currently an open question to determine which of the four set inclusions in (5) are strict.

4. Discussion

This work studies and generalizes de Finetti's theorem through the lens of what we call weighted exchangeability. Our main result shows that if a sequence of weight functions λ satisfies a weighted de Finetti representation, then this implies λ also satisfies a weighted zero-one law, which in turn implies λ satisfies a weighted law of large numbers. We also present more explicit sufficient (for the weighted de Finetti theorem representation) and necessary (for the weighted law of large numbers) conditions. After the initial version of our paper appeared online, Tang (2023) built on our work to derive interesting, approximate de Finetti representations for *finite* weighted exchangeable sequences $X_1, \dots, X_n$ (analogous to well-known results by Diaconis (1977), Diaconis and Freedman (1980) for finite unweighted exchangeable sequences).

A potentially important contribution of this work, which we have not yet discussed at this point in the paper, lies in the proof of the sufficient condition – in Section 5.4 below. There we establish that, if the sufficient condition holds, one can start with an infinite weighted exchangeable sequence X and construct an *exchangeable* infinite subsequence $\check{X}$ by a careful rejection sampling scheme (described in Section 5.4.2), whose limiting *unweighted* empirical distribution matches a weighted empirical distribution of the original sequence. This is a key to our proof of the sufficient condition, and may be useful in other problems in which weighted exchangeability arises.

We finish by discussing the connection between weighted exchangeability and the literature on distribution-free statistical inference. Conformal prediction is a general framework for quantifying uncertainty in the predictions made by arbitrary prediction algorithms. It does so by acting as a wrapper method, converting the predictions made by an algorithm into prediction sets which have distribution-free, finite-sample coverage properties; see, e.g., Lei et al. (2018), Vovk, Gammerman and Shafer (2005) for background. Importantly, the coverage guarantees for conformal prediction rely on the assumption that all of the data – the training samples (fed into the algorithm, to fit the predictive model) and test sample (at which we want to a form prediction set) – are exchangeable.

In previous work (Tibshirani et al., 2019), we extended the conformal prediction framework to a setting where the data is not exchangeable, but instead weighted exchangeable (precisely as defined in the current paper). This framework allows conformal prediction to be applied in problems with covariate shift (where the training and test covariate distributions differ); moreover, it can be used as a basis for developing new conformal methods in various settings, such as label shift (Podkopaev and Ramdas, 2021), causal inference (Lei and Candès, 2021), experimental design (Fannjiang et al., 2022), and survival analysis (Candès, Lei and Ren, 2023). In each of these examples, the data set in hand is finite (and assumed to be weighted exchangeable). In contrast, the current paper considers an infinite weighted exchangeable sequence; the characterizations that we developed for such sequences (as mixtures of weighted i.i.d. sequences) may be useful for understanding online (streaming data) problems in nonexchangeable conformal prediction. Developing such connections, as well as broadly pursuing applications of our theorems to other problems in statistical modeling and inference, will be topics for future work.

5. Proofs of main results

5.1. Important properties

Before proving our main results, we state a number of basic but important properties of exchangeability and weighted exchangeability. These results will not only be helpful later on in our proofs; they will also help build intuition about (weighted) exchangeability, and some could be of potential interest in their own right.

5.1.1. Equivalent characterizations

In the unweighted case, checking exchangeability of a distribution can be reduced to checking that the joint distribution of $X_1, X_2, \dots$ is unchanged by swapping any pair of random variables. For concreteness, we record this fact in the next result. For a finite sequence or infinite sequence x, we will write x^{ij} to denote x but with i^{th} and j^{th} entries swapped.

Proposition 4. *For any distribution Q on $\mathcal{X}^n$ (or on $\mathcal{X}^\infty$), the following statements are equivalent:*

(a) *Q is exchangeable.*
(b) *For all $1 \le i < j \le n$ (or all $1 \le i < j$), and all measurable functions $f : \mathcal{X}^n \to [0,\infty)$ (or all measurable functions $f : \mathcal{X}^\infty \to [0,\infty)$),*

$$\mathbb{E}_Q\left[f(X)\right] = \mathbb{E}_Q\left[f\left(X^{ij}\right)\right].$$

An analogous result holds in the weighted case: checking λ-weighted exchangeability of $X \sim Q$ is equivalent to considering *weighted* swaps of the entries of X.

Proposition 5. *Fix $\lambda \in \Lambda^n$ (or $\lambda \in \Lambda^\infty$). For any distribution Q on $\mathcal{X}^n$ (or on $\mathcal{X}^\infty$), the following statements are equivalent:*

(a) *Q is λ-weighted exchangeable.*
(b) *For all $1 \le i < j \le n$ (or all $1 \le i < j$), and all measurable functions $f : \mathcal{X}^n \to [0,\infty)$ (or all measurable functions $f : \mathcal{X}^\infty \to [0,\infty)$),*

$$\mathbb{E}_Q\left[\frac{f(X)}{\lambda_i(X_i)\lambda_j(X_j)}\right] = \mathbb{E}_Q\left[\frac{f(X^{ij})}{\lambda_i(X_i)\lambda_j(X_j)}\right].$$

We remark that an analogous result holds more generally when Q is a λ-weighted exchangeable measure, but for simplicity we state the result in Proposition 5 only for distributions. The proof of Proposition 5 (which generalizes Proposition 4) is deferred to Appendix A.2.

5.1.2. Conditional distributions

For a sequence X, we denote by $\widehat{P}_m$ the empirical distribution of the first m terms in the sequence,

$$\widehat{P}_m(A) = \frac{1}{m}\sum_{i=1}^m \mathbb{1}_{X_i \in A}, \quad \text{for } A \in \mathcal{B}(\mathcal{X}).$$

Define $\mathcal{E}_m$ to be the sub-σ-algebra containing events that are symmetric in the first m coordinates, that is, if $B \in \mathcal{E}_m$ then it must hold that

$$x \in B \quad \Longleftrightarrow \quad x^{ij} \in B \quad \text{for all } i \ne j \in [m].$$

We think of $\mathcal{E}_m$ as a sub-σ-algebra of either $\mathcal{B}(\mathcal{X}^n)$ or $\mathcal{B}(\mathcal{X}^\infty)$, depending on if we are working in the finite or infinite setting; this will be clear from context. Equivalently, we can define $\mathcal{E}_m$ as the σ-algebra generated by $\widehat{P}_m, X_{m+1}, \dots, X_n$ in the finite setting, or by $\widehat{P}_m, X_{m+1}, X_{m+2}, \dots$ in the infinite setting.

The following result on conditional distributions holds for the exchangeable case.

Proposition 6. *For any exchangeable distribution Q on $\mathcal{X}^n$ (or on $\mathcal{X}^\infty$), and $X \sim Q$, it holds for any $1 \le i \le m \le n$ (or for any $1 \le i \le m$) that*

$$X_i | \mathcal{E}_m \sim \widehat{P}_m.$$

Note that the above conditional law does not depend on Q. In other words, the distribution of $X_i|\mathcal{E}_m$ is the same – it is simply the empirical distribution on the first m coordinates – for *any* exchangeable distribution Q.

The next result characterizes the analogous conditional distribution for the terms in a weighted exchangeable sequence.

Proposition 7. *For any $\lambda \in \Lambda^n$ (or $\lambda \in \Lambda^\infty$), any λ-exchangeable distribution Q on $\mathcal{X}^n$ (or on $\mathcal{X}^\infty$), and $X \sim Q$, it holds for any $1 \le i \le m \le n$ (or for any $1 \le i \le m$) that*

$$X_i|\mathcal{E}_m \sim \widetilde{P}_{m,i},$$

where $\widetilde{P}_{m,i} = \sum_{j=1}^m (w_{m,i}(X_1,\dots,X_m))_j \cdot \delta_{X_j}$ is the weighted empirical distribution defined in (2).

In other words, each $X_i|\mathcal{E}_m$ can be viewed as a draw from a *weighted* empirical distribution of $X_1,\dots,X_m$. The weights $(w_{m,i}(X_1,\dots,X_m))_j$ that define this weighted empirical distribution, as constructed in (2), are determined by the functions $\lambda_1,\dots,\lambda_m$ but do not otherwise depend on the original distribution Q. This is crucial, and is analogous to the unweighted case in Proposition 6. The proof of Proposition 7 (which generalizes Proposition 6) is deferred to Appendix A.3.

5.2. Proof of Theorem 4

We turn to the proof of our first main result, Theorem 4.

5.2.1. *Proof of* $\Lambda_{01} \subseteq \Lambda_{\mathsf{LLN}}$

Let $\lambda \in \Lambda_{01}$, and draw $X \sim P \circ \lambda$ for some $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$. Fix $i \ge 1$ and $A \in \mathcal{B}(\mathcal{X})$. Define weights $(w_{n,i}(x_1,\dots,x_n))_j$, $j \in [n]$ as in (2). First, by Proposition 7, as $P \circ \lambda$ is λ-weighted exchangeable,

$$\mathbb{P}_{P\circ\lambda}\{X_i \in A|\mathcal{E}_n\} \stackrel{\text{a.s.}}{=} \sum_{j=1}^n \big(w_{n,i}(X_1,\dots,X_n)\big)_j \cdot \mathbb{1}_{X_j \in A} = \widetilde{P}_{n,i}(A).$$

(Note that $\widetilde{P}_{n,i}(A)$ is $\mathcal{E}_n$-measurable by definition.) Next, fixing an arbitrary set $A \in \mathcal{B}(\mathcal{X})$, since $\mathcal{E}_1 \supseteq \mathcal{E}_2 \supseteq \dots$ with $\mathcal{E}_\infty = \bigcap_{n=1}^\infty \mathcal{E}_n$, by Levy's Downwards Theorem (e.g., Chapter 14.4 of Williams (1991)), we have

$$\lim_{n\to\infty} \mathbb{P}_{P\circ\lambda}\{X_i \in A|\mathcal{E}_n\} \stackrel{\text{a.s.}}{=} \mathbb{P}_{P\circ\lambda}\{X_i \in A|\mathcal{E}_\infty\},$$

or in other words,

$$\lim_{n\to\infty} \widetilde{P}_{n,i}(A) \stackrel{\text{a.s.}}{=} \mathbb{P}_{P\circ\lambda}\{X_i \in A|\mathcal{E}_\infty\}. \tag{6}$$

Let Y_A be an $\mathcal{E}_\infty$-measurable random variable such that $Y_A \stackrel{\text{a.s.}}{=} \mathbb{P}_{P\circ\lambda}\{X_i \in A|\mathcal{E}_\infty\}$; the existence of such a random variable is ensured (Faden, 1985) since $\mathcal{X}$, and hence $\mathcal{X}^\infty$, is standard Borel. We thus have $\lim_{n\to\infty} \widetilde{P}_{n,i}(A) \stackrel{\text{a.s.}}{=} Y_A$. Now, recalling that we have assumed $\lambda \in \Lambda_{01}$, we have

$$\mathbb{P}_{P\circ\lambda}\big\{Y_A \le (P \circ \lambda_i)(A)\big\} \in \{0,1\}.$$

But by construction, $\mathbb{E}_{P\circ\lambda}[Y_A] = (P \circ \lambda_i)(A)$, so we conclude from the above display that in fact $\mathbb{P}_{P\circ\lambda}\{Y_A \le (P \circ \lambda_i)(A)\} = 1$. A similar argument leads to $\mathbb{P}_{P\circ\lambda}\{Y_A \ge (P \circ \lambda_i)(A)\} = 1$, and thus combining these conclusions, we have shown that $Y_A \overset{\text{a.s.}}{=} (P \circ \lambda_i)(A)$ and

$$\lim_{n\to\infty} \widetilde{P}_{n,i}(A) \overset{\text{a.s.}}{=} (P \circ \lambda_i)(A),$$

as desired.

5.2.2. Proof of $\Lambda_{\mathsf{dF}} \subseteq \Lambda_{\mathsf{01}}$

Let $\lambda \in \Lambda_{\mathsf{dF}}$ and suppose for the sake of contradiction that $\lambda \notin \Lambda_{\mathsf{01}}$, so there is some $P_* \in \mathcal{M}_{\mathcal{X}}(\lambda)$ and some $B_0 \in \mathcal{E}_\infty$ such that $p = (P_* \circ \lambda)(B_0) \in (0, 1)$. Now let Q_0 and Q_1 denote the distribution of $X = (X_1, X_2, \ldots)$ conditional on the event B_0 and on the event $B_1 = B_0^c$, respectively, for $X \sim P_* \circ \lambda$. Note that $P_* \circ \lambda$ can be written as a mixture,

$$P_* \circ \lambda = p \cdot Q_0 + (1 - p) \cdot Q_1.$$

Next we verify that Q_ℓ is λ-weighted exchangeable for each $\ell = 0, 1$. Proposition 5 tells us that to verify Q_ℓ is λ-weighted exchangeable, we must only verify that

$$\mathbb{E}_{Q_\ell}\left[\frac{f(X)}{\lambda_i(X_i)\lambda_j(X_j)}\right] = \mathbb{E}_{Q_\ell}\left[\frac{f(X^{ij})}{\lambda_i(X_i)\lambda_j(X_j)}\right],$$

for all $i \neq j$ and all measurable $f : \mathcal{X}^\infty \to [0, \infty)$. But since Q_ℓ is equal to $P_* \circ \lambda$ conditional on the event B_ℓ, and $(P_* \circ \lambda)(B_\ell) > 0$, it is equivalent to verify that

$$\mathbb{E}_{P_*\circ\lambda}\left[\frac{f(X)}{\lambda_i(X_i)\lambda_j(X_j)} \cdot \mathbb{1}_{X\in B_\ell}\right] = \mathbb{E}_{P_*\circ\lambda}\left[\frac{f(X^{ij})}{\lambda_i(X_i)\lambda_j(X_j)} \cdot \mathbb{1}_{X\in B_\ell}\right].$$

Note that $X \in B_\ell$ if and only if $X^{ij} \in B_\ell$, because $B_\ell \in \mathcal{E}_\infty$, so it is equivalent to check that

$$\mathbb{E}_{P_*\circ\lambda}\left[\frac{f(X) \cdot \mathbb{1}_{X\in B_\ell}}{\lambda_i(X_i)\lambda_j(X_j)}\right] = \mathbb{E}_{P_*\circ\lambda}\left[\frac{f(X^{ij}) \cdot \mathbb{1}_{X^{ij}\in B_\ell}}{\lambda_i(X_i)\lambda_j(X_j)}\right].$$

Lastly, another application of Proposition 5 (with $f(X) \cdot \mathbb{1}_{X\in B_\ell}$ in place of $f(X)$) tells us that the above display holds, since $P_* \circ \lambda$ itself is λ-weighted exchangeable.

Recalling that we have assumed $\lambda \in \Lambda_{\mathsf{dF}}$, and having just established that each Q_ℓ is λ-weighted exchangeable, we know that there is a distribution μ_ℓ on $\mathcal{M}_{\mathcal{X}}(\lambda)$ such that $Q_\ell = (P \circ \lambda)_{\mu_\ell}$, for each $\ell = 0, 1$. In particular, writing the mixture of μ_0 and μ_1 as

$$\mu = p \cdot \mu_0 + (1 - p) \cdot \mu_1,$$

we see that $P_* \circ \lambda = (P \circ \lambda)_\mu$. We now need an additional lemma, whose proof is in the supplementary material (Barber et al., 2024).

Lemma 1. *Fix $\lambda \in \Lambda^\infty$ and let μ be a distribution on $\mathcal{M}_{\mathcal{X}}(\lambda)$. Suppose that $(P \circ \lambda)_\mu = P_* \circ \lambda$ for some $P_* \in \mathcal{M}_{\mathcal{X}}(\lambda)$. Then, under $P \sim \mu$, for any $A \in \mathcal{B}(\mathcal{X}^\infty)$ it holds that $(P \circ \lambda)(A) \overset{a.s.}{=} (P_* \circ \lambda)(A)$.*

Therefore we know $(P_* \circ \lambda)(B_0) = (P \circ \lambda)(B_0)$ holds almost surely under $P \sim \mu$. We have now reached our desired contradiction, because $(P_* \circ \lambda)(B_0) = p$ is nonrandom, whereas $(P \circ \lambda)(B_0)$ is distributed as Bernoulli(p) under $P \sim \mu$, by construction of $\mu = p \cdot \mu_0 + (1 - p) \cdot \mu_1$.

5.3. Proof of Theorem 5

Let $\lambda \in \Lambda_{\mathsf{LLN}}$, and fix any $P_0 \in \mathcal{M}_{\mathcal{X}}(\lambda)$ and any $A \in \mathcal{B}(\mathcal{X})$ with $P_0(A) \in (0,1)$. We will prove that the necessary condition (3) must hold for $P = P_0$. Fix any $c \in (0,1)$ and define a measure P_1 via

$$P_1(B) = c \cdot P_0(A \cap B) + P_0\big(A^c \cap B\big), \quad \text{for } B \in \mathcal{B}(\mathcal{X}).$$

Observe that $P_1 \in \mathcal{M}_{\mathcal{X}}(\lambda)$ by construction. As $\lambda \in \Lambda_{\mathsf{LLN}}$, by assumption we have

$$\mathbb{P}_{P_\ell \circ \lambda}\left\{\lim_{n\to\infty} \sum_{j=1}^n \big(w_{n,1}(X_1,\dots,X_n)\big)_j \cdot \mathbb{1}_{X_j \in A} = (P_\ell \circ \lambda_1)(A)\right\} = 1,$$

for each $\ell = 0,1$, where recall $(w_{n,1}(X_1,\dots,X_n))_j$ is defined as in (2). Next we let E be the event that $\lim_{n\to\infty} \sum_{j=1}^n (w_{n,1}(X_1,\dots,X_n))_j \cdot \mathbb{1}_{X_j \in A} = (P_0 \circ \lambda_1)(A)$. As $(P_0 \circ \lambda_1)(A) \neq (P_1 \circ \lambda_1)(A)$ by construction, this implies

$$(P_0 \circ \lambda)(E) = 1, \quad \text{and} \quad (P_1 \circ \lambda)(E) = 0,$$

and thus $\mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda, P_1 \circ \lambda) = 1$, where d_{TV} denotes total variation distance.

Next, for any $i \geq 1$ and any $B \in \mathcal{B}(\mathcal{X})$, a straightforward calculation shows that

$$(P_1 \circ \lambda_i)(B) = \frac{c \cdot (P_0 \circ \lambda_i)(A \cap B) + (P_0 \circ \lambda_i)(A^c \cap B)}{c \cdot (P_0 \circ \lambda_i)(A) + (P_0 \circ \lambda_i)(A^c)},$$

and thus

$$\begin{aligned}\mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda_i, P_1 \circ \lambda_i) &= \sup_{B \in \mathcal{B}(\mathcal{X})} \big|(P_0 \circ \lambda_i)(B) - (P_1 \circ \lambda_i)(B)\big| \\ &= (1-c) \cdot \frac{(P_0 \circ \lambda_i)(A) \cdot (P_0 \circ \lambda_i)(A^c)}{c \cdot (P_0 \circ \lambda_i)(A) + (P_0 \circ \lambda_i)(A^c)},\end{aligned}$$

where the last equality holds because the supremum is attained at $B = A$ (or equivalently, at $B = A^c$). From this we can verify that $\mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda_i, P_1 \circ \lambda_i) < 1$, and moreover,

$$\mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda_i, P_1 \circ \lambda_i) \leq \big(c^{-1} - 1\big) \cdot \min\big\{(P_0 \circ \lambda_i)(A), (P_0 \circ \lambda_i)\big(A^c\big)\big\}. \tag{7}$$

We also know that

$$0 = 1 - \mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda, P_1 \circ \lambda) = \prod_{i=1}^\infty \big(1 - \mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda_i, P_1 \circ \lambda_i)\big),$$

where the second equality holds by properties of the total variation distance for product distributions. In general, for any sequence $a_1, a_2, \ldots \in [0,1)$ with finite sum $\sum_{i=1}^\infty a_i < \infty$, it holds that $\prod_{i=1}^\infty (1 - a_i) > 0$ (see, e.g., Theorem 4 in Section 28 of Knopp (1990)). Therefore, we must have

$$\sum_{i=1}^\infty \mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda_i, P_1 \circ \lambda_i) = \infty,$$

and so combining this with (7), we get

$$\infty = \sum_{i=1}^\infty \mathsf{d}_{\mathsf{TV}}(P_0 \circ \lambda_i, P_1 \circ \lambda_i) \leq \big(c^{-1} - 1\big) \cdot \sum_{i=1}^\infty \min\big\{(P_0 \circ \lambda_i)(A), (P_0 \circ \lambda_i)\big(A^c\big)\big\}.$$

Since $(c^{-1} - 1)$ is finite, the sum on the right-hand side must be infinite, which completes the proof.

5.4. Proof of Theorem 6

5.4.1. Outline of proof

We begin by sketching the main ideas of the proof in order to build intuition. In the exchangeable case, where $X \sim Q$ for an exchangeable distribution Q, de Finetti's theorem tells us that we can view $X_1, X_2, \dots$ as i.i.d. draws from some random distribution P (that is, for some $P \sim \mu$) and we can recover this underlying distribution P almost surely via the limit $\lim_{n\to\infty} \widehat{P}_n$, where $\widehat{P}_n$ is the empirical distribution of $X_1, \dots, X_n$:

$$P(A) \overset{\text{a.s.}}{=} \lim_{n\to\infty} \widehat{P}_n(A), \quad \text{for all } A \in \mathcal{B}(\mathcal{X}).$$

In the weighted exchangeable case, where $X \sim Q$ for a λ-weighted exchangeable Q, we will work towards a similar conclusion. To recover the underlying random distribution, we will take the limit of a *weighted* empirical distribution of $X_1, \dots, X_n$, denoted by $\widetilde{P}_n$, and defined by

$$\widetilde{P}_n(A) = \frac{\sum_{i=1}^n \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)} \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^n \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)}}, \quad \text{for } A \in \mathcal{B}(\mathcal{X}), \tag{8}$$

with $\lambda_* \in \Lambda$ chosen such that the sufficient condition (4) is satisfied. Note that by (4), for n large enough the denominator in (8) is positive, hence $\widetilde{P}_n$ is well-defined. (Also, to avoid confusion, we note that $\widetilde{P}_n$ is different from the weighted empirical distribution $\widetilde{P}_{n,i}$ in (2) that is used to define the weighted law of large numbers.)

In the rest of the proof of Theorem 6, we will first construct a random distribution $\widetilde{P}$ that is an almost sure limit of the weighted empirical distribution $\widetilde{P}_n$ in (8). Then we will show that the given weighted exchangeable distribution Q is equal to the mixture distribution $(P \circ \lambda)_\mu$, where μ is the distribution of the random measure $\widetilde{P}_*$ defined as

$$\widetilde{P}_*(A) = \int_A \frac{\mathsf{d}\widetilde{P}(x)}{\lambda_*(x)}, \quad \text{for } A \in \mathcal{B}(\mathcal{X}). \tag{9}$$

5.4.2. Constructing the limit $\widetilde{P}$ via an exchangeable subsequence

Our first task is to construct a distribution $\widetilde{P}$ that is an almost sure limit of the weighted empirical distribution $\widetilde{P}_n$ defined in (8). We begin by approximating this weighted empirical distribution. Let $U_1, U_2, \dots \overset{\text{i.i.d.}}{\sim} \text{Unif}[0,1]$, independently of X, and define for each $i = 1, 2, \dots$,

$$B_i = \mathbb{1}_{U_i \le p_i(X_i)}, \quad \text{where } p_i(x) = \frac{\inf_{x'\in\mathcal{X}} \lambda_i(x')/\lambda_*(x')}{\lambda_i(x)/\lambda_*(x)} \in [0,1]. \tag{10}$$

We have

$$\sum_{i=1}^\infty p_i(X_i) = \sum_{i=1}^\infty \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)} \ge \sum_{i=1}^\infty \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)} = \infty,$$

where the last step holds by (4). Thus, defining

$$M = \sum_{i=1}^\infty B_i,$$

by the second Borel–Cantelli Lemma we see that $M = \infty$ almost surely. Now define

$$P_n(A) = \frac{\sum_{i=1}^n B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^n B_i}, \quad \text{for } A \in \mathcal{B}(\mathcal{X}), \tag{11}$$

which, almost surely, is well-defined for sufficiently large n, because $M \stackrel{\text{a.s.}}{=} \infty$. This turns out to be a helpful approximation for the weighted empirical distribution in (8), with almost sure equality in the limit, as we state next. The proof of this result is deferred to the supplementary material (Barber et al., 2024).

Lemma 2. *Under the notation and assumptions above, it holds that*

$$\limsup_{n\to\infty} \left| P_n(A) - \widetilde{P}_n(A) \right| \stackrel{a.s.}{=} 0, \quad \textit{for all } A \in \mathcal{B}(\mathcal{X}).$$

Next we define an infinite subsequence $\check{X}$ of the original sequence X. Define $I_0 = 0$, and

$$I_m = \min\{i > I_{m-1} : B_i = 1\}, \quad \text{for } m = 1, \ldots, M. \tag{12}$$

In other words, $I_1 < I_2 < \ldots$ enumerates all indices i for which $B_i = 1$. Then define

$$\check{X} = \begin{cases} (X_{I_1}, X_{I_2}, X_{I_3}, \ldots) & \text{if } M = \infty, \\ (X_{I_1}, \ldots, X_{I_M}, x_0, x_0 \ldots) & \text{otherwise,} \end{cases} \tag{13}$$

where $x_0 \in \mathcal{X}$ is some fixed value. In other words, in the case $M < \infty$, we augment the subsequence $(X_{I_m})_{m=1}^M$ with an infinite string $(x_0, x_0, \ldots)$ (but of course, since $M = \infty$ almost surely, this occurs with probability zero). Critically, this construction results in $\check{X}$ being exchangeable. The proof of the lemma below is deferred to the supplementary material (Barber et al., 2024).

Lemma 3. *Under the notation and assumptions above, the sequence $\check{X}$ that is defined in* (13) *has an exchangeable distribution.*

Why are we interested in $\check{X}$? This will become more apparent once we inspect the empirical distribution of its first m elements: for $m \geq 1$, define

$$\check{P}_m(A) = \frac{\sum_{i=1}^m \mathbb{1}_{\check{X}_i \in A}}{m}, \quad \text{for } A \in \mathcal{B}(\mathcal{X}). \tag{14}$$

Observe that, for any $n \geq 1$, denoting $M_n = \sum_{i=1}^n B_i$, if $M_n > 0$ then $P_n(A) = \check{P}_{M_n}(A)$, where P_n is defined in (11) above. Combining this with Lemma 2, together with the fact that $M = \lim_{n\to\infty} M_n$, we see that

$$\text{if } M = \infty \text{ and } \lim_{m\to\infty} \check{P}_m(A) \text{ exists, then } \lim_{n\to\infty} \widetilde{P}_n(A) \text{ exists and is equal to the same value} \tag{15}$$

holds almost surely, for all $A \in \mathcal{B}(\mathcal{X})$.

Exchangeability of the distribution of $\check{X}$ now allows us to apply the original de Finetti–Hewitt–Savage theorem (Theorem 1) and the law of large numbers (Theorem 3) to assert the existence of a random distribution $\widetilde{P} \in \mathcal{P}_{\mathcal{X}}$ such that $\check{P}_m$ (and thus also $\widetilde{P}_n$) converges to $\widetilde{P}$. To be more specific, the de Finetti–Hewitt–Savage theorem tells us that we can represent a draw from the distribution of $\check{X}$ as follows: we first sample a random distribution $\widetilde{P}$, then we sample components $\check{X}_i$, $i = 1, 2, \ldots$ independently from $\widetilde{P}$. The law of large numbers tells us that $\widetilde{P}$ can be recovered by taking the limit of the empirical distribution $\check{P}_m$. Thus,

$$\text{there exists a random distribution } \widetilde{P} \in \mathcal{P}_{\mathcal{X}} \text{ such that } \lim_{m\to\infty} \check{P}_m(A) \stackrel{\text{a.s.}}{=} \widetilde{P}(A), \text{ for all } A \in \mathcal{B}(\mathcal{X}),$$

and consequently, combining this with (15) along with the fact that $M \stackrel{\text{a.s.}}{=} \infty$,

$$\text{there exists a random distribution } \widetilde{P} \in \mathcal{P}_{\mathcal{X}} \text{ such that } \lim_{n\to\infty} \widetilde{P}_n(A) \stackrel{\text{a.s.}}{=} \widetilde{P}(A), \text{ for all } A \in \mathcal{B}(\mathcal{X}). \tag{16}$$

Next, for $n \geq 1$, we define $\mathcal{F}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$, then define $\mathcal{F}_{\text{tail}} = \bigcap_{n=0}^{\infty} \mathcal{F}_n$, which is called the *tail σ-algebra* corresponding to the infinite sequence X. The next lemma, proved in the supplementary material (Barber et al., 2024), establishes that $\widetilde{P}$ can be recovered via $\mathcal{F}_{\text{tail}}$-measurable random variables. This will be important later on.

Lemma 4. *Under the notation and assumptions above, for any measurable function $f : \mathcal{X} \to [0, \infty)$, there exists an $\mathcal{F}_{tail}$-measurable function $g : \mathcal{X}^\infty \to [0, \infty]$ such that*

$$g(X) \stackrel{a.s.}{=} \mathbb{E}_{X' \sim \widetilde{P}}\left[f(X')\right].$$

5.4.3. Representing Q as a mixture distribution

In the last part of the proof, we construct a random measure $\widetilde{P}_* \in \mathcal{M}_{\mathcal{X}}$ as specified in (9), with $\widetilde{P}$ the distribution constructed in the last subsection, from (16). We will show that the distribution Q of interest is equal to the mixture $\widetilde{P}_* \circ \lambda$. However, we have not established that $\widetilde{P}_* \circ \lambda$ is always well-defined, so we will need to treat this carefully.

We will need an additional lemma, which we prove in the supplementary material (Barber et al., 2024).

Lemma 5. *Under the notation and assumptions above, for all $k \geq 1$ and all $A \in \mathcal{B}(\mathcal{X})$,*

$$\int_{\mathcal{X}} \lambda_k(x)\, \mathsf{d}\widetilde{P}_*(x) \in (0, \infty) \quad \textit{almost surely}, \tag{17}$$

and also

$$\mathbb{P}_Q\{X_k \in A | X_{-k}\} \stackrel{a.s.}{=} (\widetilde{P}_* \circ \lambda_k)(A), \tag{18}$$

where we note that $(\widetilde{P}_* \circ \lambda_k)$ is well-defined almost surely, by the first claim (17).

By (17), we see that $\widetilde{P}_* \in \mathcal{M}_{\mathcal{X}}(\lambda)$ almost surely. Therefore, we can define μ as the distribution of this random measure $\widetilde{P}_*$ conditional on the event $\widetilde{P}_* \in \mathcal{M}_{\mathcal{X}}(\lambda)$, so that μ is a distribution over $\mathcal{M}_{\mathcal{X}}(\lambda)$. Next we need to verify that $Q = (P \circ \lambda)_\mu$. As $\mathcal{B}(\mathcal{X}^\infty)$ is the product σ-algebra, it suffices to prove that, for all $n \geq 1$ and all $A_1, \ldots, A_n \in \mathcal{B}(\mathcal{X})$,

$$\mathbb{P}_Q\{X_1 \in A_1, \ldots, X_n \in A_n\} = \mathbb{E}_{P\sim\mu}\left[\prod_{i=1}^n (P \circ \lambda_i)(A_i)\right].$$

By Lemma 4, for each $i \geq 1$, and any $A \in \mathcal{B}(\mathcal{X})$, we can construct an $\mathcal{F}_{\text{tail}}$-measurable function $f_{A,i} : \mathcal{X}^\infty \to [0, \infty]$ such that

$$f_{A,i}(X) \stackrel{\text{a.s.}}{=} \mathbb{E}_{X'\sim\widetilde{P}}\left[\mathbb{1}_{X'\in A} \cdot \frac{\lambda_i(X')}{\lambda_*(X')}\right] = \int_A \lambda_i(x)\, \mathsf{d}\widetilde{P}_*.$$

Applying this with $A = A_i$ and again with $A = \mathcal{X}$, we can define an $\mathcal{F}_{\text{tail}}$-measurable function

$$f_i(x) = \begin{cases} f_{A_i,i}(x)/f_{\mathcal{X},i}(x) & \text{if } f_{A_i,i}(x) < \infty \text{ and } f_{\mathcal{X},i}(x) \in (0,\infty), \\ 0 & \text{otherwise.} \end{cases}$$

Applying (17), this satisfies

$$f_i(X) \overset{\text{a.s.}}{=} \frac{\int_{A_i} \lambda_i(x)\,\mathsf{d}\widetilde{P}_*}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}\widetilde{P}_*} = (\widetilde{P}_* \circ \lambda_i)(A_i). \tag{19}$$

By the tower law, noting that each $f_i(X)$ is $\mathcal{F}_{\text{tail}}$-measurable and thus, is measurable with respect to $\sigma(X_{-j}) \supseteq \mathcal{F}_{\text{tail}}$ for any j, we can calculate

$$\begin{aligned}
&\mathbb{E}_Q\left[\prod_{i=1}^{m-1} f_i(X) \cdot \prod_{i=m}^{n} \mathbb{1}_{X_i \in A_i}\right] \\
&= \mathbb{E}_Q\left[\prod_{i=1}^{m-1} f_i(X) \cdot \mathbb{P}_Q\{X_m \in A_m | X_{-m}\} \cdot \prod_{i=m+1}^{n} \mathbb{1}_{X_i \in A_i}\right] \quad \text{(as each } f_i(X) \text{ is } \sigma(X_{-m})\text{-measurable)} \\
&= \mathbb{E}_Q\left[\prod_{i=1}^{m-1} f_i(X) \cdot (\widetilde{P}_* \circ \lambda_m)(A_m) \cdot \prod_{i=m+1}^{n} \mathbb{1}_{X_i \in A_i}\right] \quad \text{(by the fact in (18))} \\
&= \mathbb{E}_Q\left[\prod_{i=1}^{m-1} f_i(X) \cdot f_m(X) \cdot \prod_{i=m+1}^{n} \mathbb{1}_{X_i \in A_i}\right] \quad \text{(by construction of } f_m) \\
&= \mathbb{E}_Q\left[\prod_{i=1}^{m} f_i(X) \cdot \prod_{i=m+1}^{n} \mathbb{1}_{X_i \in A_i}\right]
\end{aligned}$$

for each $m = 1,\dots,n$. Combining these calculations over all $i = 1,\dots,n$, we obtain

$$\mathbb{P}_Q\{X_1 \in A_1,\dots,X_n \in A_n\} = \mathbb{E}_Q\left[\prod_{i=1}^{n} \mathbb{1}_{X_i \in A_i}\right] = \mathbb{E}_Q\left[\prod_{i=1}^{n} f_i(X)\right] = \mathbb{E}\left[\prod_{i=1}^{n} (\widetilde{P}_* \circ \lambda_i)(A_i)\right],$$

where the last step holds by construction of the functions f_i, $i = 1,\dots,n$. Since μ is equal to the distribution of $\widetilde{P}_*$ conditional on the almost sure event $\widetilde{P}_* \in \mathcal{M}_{\mathcal{X}}(\lambda)$, we therefore see that

$$\mathbb{E}\left[\prod_{i=1}^{n} (\widetilde{P}_* \circ \lambda_i)(A_i)\right] = \mathbb{E}_{P\sim\mu}\left[\prod_{i=1}^{n} (P \circ \lambda_i)(A_i)\right],$$

which completes the proof.

Appendix A: Proofs of properties of weighted exchangeability

In this section, we prove Propositions 5 and 7, which establish some key properties of weighted exchangeable distributions. All other proofs are deferred to the supplementary materials (Barber et al., 2024).

A.1. A note on regularity conditions

Before proceeding, we pause to comment on a technical point about conditional distributions. As $\mathcal{X}$ (and thus also $\mathcal{X}^n$ and $\mathcal{X}^\infty$) is a standard Borel space, this suffices (see, e.g., Faden (1985)) to ensure the existence of a regular conditional distribution (also called a product regular conditional probability kernel) for, say, $X|\mathcal{E}_\infty$, or $(X_{n+1}, X_{n+2}, \ldots)|(X_1, \ldots, X_n)$, etc. For example, if we consider the conditional distribution of $X|\mathcal{E}_\infty$, the standard Borel property guarantees the existence of a kernel $\kappa : \mathcal{X}^\infty \times \mathcal{B}(\mathcal{X}^\infty) \to [0,1]$ such that $x \mapsto \kappa(x, A)$ is measurable for all $A \in \mathcal{B}(\mathcal{X}^\infty)$, and such that $\kappa(X, \cdot)$ gives the conditional distribution of $X|\mathcal{E}_\infty$, i.e., $\kappa(X, A) \stackrel{\text{a.s.}}{=} \mathbb{P}\{X \in A|\mathcal{E}_\infty\}$. In what follows (and in the proofs in the supplementary material (Barber et al., 2024) as well), in any part of the proof where a conditional distribution is used, it should be understood that we have constructed a regular conditional distribution, which is guaranteed to exist by our standard Borel assumption.

A.2. Proof of Proposition 5

We will present proofs for the infinite case, where Q is a distribution on $\mathcal{X}^\infty$; the calculations for the finite case, where Q is a distribution on $\mathcal{X}^n$, are similar and are omitted for brevity.

First we prove (a) $\implies$ (b). Suppose that Q is λ-weighted exchangeable. Note that it suffices to consider functions of the form $f(x) = \mathbb{1}_{x \in A}$, because any measurable function can be generated by such functions. That is, we want to prove

$$\mathbb{E}_Q\left[\frac{\mathbb{1}_{X \in A}}{\lambda_i(X_i)\lambda_j(X_j)}\right] = \mathbb{E}_Q\left[\frac{\mathbb{1}_{X^{ij} \in A}}{\lambda_i(X_i)\lambda_j(X_j)}\right],$$

for any $A \in \mathcal{B}(\mathcal{X}^\infty)$. As the σ-algebra on $\mathcal{X}^\infty$ is generated by sets of the form $A = A_1 \times \cdots \times A_n \times \mathcal{X} \times \mathcal{X} \times \cdots$ where $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{B}(\mathcal{X})$, it suffices to check that the expected values are equal for sets of this form. Without loss of generality it suffices to consider $n \geq \max\{i, j\}$. Define the function

$$h(x_1, \ldots, x_n) = \mathbb{1}_{(x_1, \ldots, x_n) \in A_1 \times \cdots \times A_n} \cdot \prod_{k \in [n] \setminus \{i,j\}} \lambda_k(x_k).$$

We then have

$$\begin{aligned}
\mathbb{E}_Q\left[\frac{\mathbb{1}_{X \in A}}{\lambda_i(X_i)\lambda_j(X_j)}\right] &= \mathbb{E}_Q\left[\frac{\mathbb{1}_{(X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n}}{\lambda_i(X_i)\lambda_j(X_j)}\right] \\
&= \mathbb{E}_{Q_n}\left[\frac{\mathbb{1}_{(X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n} \cdot \prod_{k \in [n] \setminus \{i,j\}} \lambda_k(X_k)}{\prod_{k=1}^n \lambda_k(X_k)}\right] \\
&= \mathbb{E}_{Q_n}\left[\frac{h(X_1, \ldots, X_n)}{\prod_{k=1}^n \lambda_k(X_k)}\right] \\
&= \int_{\mathcal{X}^n} h(x_1, \ldots, x_n) \frac{\mathsf{d}Q_n(x_1, \ldots, x_n)}{\prod_{k=1}^n \lambda_k(x_k)},
\end{aligned}$$

and similarly,

$$\mathbb{E}_Q\left[\frac{\mathbb{1}_{X^{ij} \in A}}{\lambda_i(X_i)\lambda_j(X_j)}\right] = \mathbb{E}_Q\left[\frac{\mathbb{1}_{((X^{ij})_1, \ldots, (X^{ij})_n) \in A_1 \times \cdots \times A_n}}{\lambda_i(X_i)\lambda_j(X_j)}\right]$$

$$= \mathbb{E}_{Q_n}\left[\frac{\mathbb{1}_{((X^{ij})_1,\dots,(X^{ij})_n)\in A_1\times\cdots\times A_n}\cdot\prod_{k\in[n]\setminus\{i,j\}}\lambda_k(X_k)}{\prod_{k=1}^n \lambda_k(X_k)}\right]$$

$$= \mathbb{E}_{Q_n}\left[\frac{h((X^{ij})_1,\dots,(X^{ij})_n)}{\prod_{k=1}^n \lambda_k(X_k)}\right]$$

$$= \int_{\mathcal{X}^n} h\big((x^{ij})_1,\dots,(x^{ij})_n\big)\,\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)}.$$

Finally, since Q_n is $(\lambda_1,\dots,\lambda_n)$-weighted exchangeable,

$$\int_{\mathcal{X}^n} h(x_1,\dots,x_n)\,\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)} = \int_{\mathcal{X}^n} h\big((x^{ij})_1,\dots,(x^{ij})_n\big)\,\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)}.$$

Second we prove (b) $\Longrightarrow$ (a). For each $n\ge 1$, we need to verify that the measure induced by $\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)}$ is finite exchangeable, i.e., for any $A\in\mathcal{B}(\mathcal{X}^n)$ and any permutation σ on $[n]$,

$$\int_{\mathcal{X}^n} \mathbb{1}_{(x_1,\dots,x_n)\in A}\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)} = \int_{\mathcal{X}^n} \mathbb{1}_{(x_{\sigma(1)},\dots,x_{\sigma(n)})\in A}\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)}.$$

Since the permutations on $[n]$ can be generated by pairwise swaps, it is sufficient to show that

$$\int_{\mathcal{X}^n} \mathbb{1}_{(x_1,\dots,x_n)\in A}\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)} = \int_{\mathcal{X}^n} \mathbb{1}_{((x^{ij})_1,\dots,(x^{ij})_n)\in A}\frac{\mathsf{d}Q_n(x_1,\dots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)}$$

for all $1\le i<j\le n$. Equivalently we need to show that

$$\mathbb{E}_Q\left[\frac{\mathbb{1}_{(X_1,\dots,X_n)\in A}}{\prod_{k=1}^n \lambda_k(X_k)}\right] = \mathbb{E}_Q\left[\frac{\mathbb{1}_{((X^{ij})_1,\dots,(X^{ij})_n)\in A}}{\prod_{k=1}^n \lambda_k(X_k)}\right].$$

Define

$$f(x) = \frac{\mathbb{1}_{(x_1,\dots,x_n)\in A}}{\prod_{k\in[n]\setminus\{i,j\}}\lambda_k(x_k)}.$$

Then since we have assumed (b) holds, we have $\mathbb{E}_Q[\frac{f(X)}{\lambda_i(X_i)\lambda_j(X_j)}] = \mathbb{E}_Q[\frac{f(X^{ij})}{\lambda_i(X_i)\lambda_j(X_j)}]$, which completes the proof.

A.3. Proof of Proposition 7

First consider the case that Q is a λ-weighted exchangeable distribution on $\mathcal{X}^n$. We need to show that, for any $A\in\mathcal{B}(\mathcal{X})$,

$$\mathbb{P}_Q\{X_i\in A|\mathcal{E}_m\} \stackrel{\text{a.s.}}{=} \sum_{j=1}^m \frac{\sum_{\sigma\in\mathcal{S}_m:\sigma(i)=j}\prod_{k=1}^m \lambda_k(X_{\sigma(k)})}{\sum_{\sigma\in\mathcal{S}_m}\prod_{k=1}^m \lambda_k(X_{\sigma(k)})}\cdot\mathbb{1}_{X_j\in A}.$$

Equivalently, it is sufficient to check that

$$\mathbb{E}_Q\left[\sum_{j=1}^m \frac{\sum_{\sigma\in\mathcal{S}_m:\sigma(i)=j}\prod_{k=1}^m \lambda_k(X_{\sigma(k)})}{\sum_{\sigma\in\mathcal{S}_m}\prod_{k=1}^m \lambda_k(X_{\sigma(k)})}\cdot\mathbb{1}_{X_j\in A}\cdot\mathbb{1}_{X\in C}\right] = \mathbb{E}_Q[\mathbb{1}_{X_i\in A}\cdot\mathbb{1}_{X\in C}]$$

holds for any $C \in \mathcal{E}_m$. Define $g(x) = \frac{\mathbb{1}_{x \in C}}{\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(x_{\sigma(k)})}$, and note that as $C \in \mathcal{E}_m$, we see that g is $\mathcal{E}_m$-measurable, i.e., $g((x_{\sigma(1)}, \dots, x_{\sigma(m)}, x_{m+1}, \dots, x_n)) = g(x)$ for all $x \in \mathcal{X}^n$ and all $\sigma \in \mathbb{R}^m$. We have

$$
\begin{aligned}
&\mathbb{E}_Q\left[\sum_{j=1}^m \frac{\sum_{\sigma \in \mathcal{S}_m : \sigma(i)=j} \prod_{k=1}^m \lambda_k(X_{\sigma(k)})}{\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(X_{\sigma(k)})} \cdot \mathbb{1}_{X_j \in A} \cdot \mathbb{1}_{X \in C}\right] \\
&= \mathbb{E}_Q\left[\frac{\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(X_{\sigma(k)}) \cdot \mathbb{1}_{X_{\sigma(i)} \in A}}{\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(X_{\sigma(k)})} \cdot \mathbb{1}_{X \in C}\right] \\
&= \mathbb{E}_Q\left[\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(X_{\sigma(k)}) \cdot \mathbb{1}_{X_{\sigma(i)} \in A} \cdot g(X)\right] \\
&= \int_{\mathcal{X}^n} \sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(x_{\sigma(k)}) \cdot \mathbb{1}_{x_{\sigma(i)} \in A} \cdot g(x) \, \mathsf{d}Q(x) \\
&= \sum_{\sigma \in \mathcal{S}_m} \int_{\mathcal{X}^n} \prod_{k=1}^m \lambda_k(x_{\sigma(k)}) \cdot \mathbb{1}_{x_{\sigma(i)} \in A} \cdot g(x) \cdot \prod_{k=1}^m \lambda_k(x_k) \cdot \prod_{k=m+1}^n \lambda_k(x_k) \frac{\mathsf{d}Q(x)}{\prod_{k=1}^n \lambda_k(x_k)} \\
&= \sum_{\sigma \in \mathcal{S}_m} \int_{\mathcal{X}^n} \prod_{k=1}^m \lambda_k(x_k) \cdot \mathbb{1}_{x_i \in A} \cdot g(x) \cdot \prod_{k=1}^m \lambda_k(x_{\sigma^{-1}(k)}) \cdot \prod_{k=m+1}^n \lambda_k(x_k) \frac{\mathsf{d}Q(x)}{\prod_{k=1}^n \lambda_k(x_k)} \\
&= \sum_{\sigma \in \mathcal{S}_m} \int_{\mathcal{X}^n} \mathbb{1}_{x_i \in A} \cdot g(x) \cdot \prod_{k=1}^m \lambda_k(x_{\sigma^{-1}(k)}) \, \mathsf{d}Q(x) \\
&= \int_{\mathcal{X}^n} \mathbb{1}_{x_i \in A} \cdot g(x) \cdot \left(\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(x_{\sigma^{-1}(k)})\right) \mathsf{d}Q(x) \\
&= \int_{\mathcal{X}^n} \mathbb{1}_{x_i \in A} \cdot \mathbb{1}_{x \in C} \, \mathsf{d}Q(x) \\
&= \mathbb{E}_Q[\mathbb{1}_{X_i \in A} \cdot \mathbb{1}_{X \in C}],
\end{aligned}
$$

where the fifth equality holds by replacing x with $(x_{\sigma^{-1}(1)}, \dots, x_{\sigma^{-1}(m)}, x_{m+1}, \dots, x_n)$, and recalling that Q is λ-weighted exchangeable while g is $\mathcal{E}_m$-measurable. This verifies the desired equality.

Now suppose Q is a λ-weighted exchangeable distribution on $\mathcal{X}^\infty$. For any $n \geq m$, define

$$
\mathcal{E}_{m,n} = \left\{A \in \mathcal{B}(\mathcal{X}^n) : \mathbb{1}_{(x_1, \dots, x_m) \in A} = \mathbb{1}_{(x_{\sigma(1)}, \dots, x_{\sigma(m)}, x_{m+1}, \dots, x_n) \in A}, \text{ for all } x \in \mathcal{X}^n \text{ and } \sigma \in \mathcal{S}_m\right\}.
$$

We can then verify that $\mathcal{E}_m \subseteq \mathcal{B}(\mathcal{X}^\infty)$ is the minimal σ-algebra generated by $\bigcup_{n \geq m} \mathcal{E}_{m,n}$. By Levy's Upwards Theorem (e.g., Chapter 14.2 of Williams (1991)),

$$
\mathbb{P}_Q\{X_i \in A | \mathcal{E}_m\} \stackrel{\text{a.s.}}{=} \lim_{n \to \infty} \mathbb{P}_Q\{X_i \in A | \mathcal{E}_{m,n}\}.
$$

Applying our work above to the finite $(\lambda_1, \dots, \lambda_n)$-weighted exchangeable distribution Q_n, we have

$$
\mathbb{P}_Q\{X_i \in A | \mathcal{E}_{m,n}\} \stackrel{\text{a.s.}}{=} \sum_{j=1}^m \frac{\sum_{\sigma \in \mathcal{S}_m : \sigma(i)=j} \prod_{k=1}^m \lambda_k(X_{\sigma(k)})}{\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^m \lambda_k(X_{\sigma(k)})} \cdot \mathbb{1}_{X_j \in A}
$$

for each $n \geq m$, and therefore, as desired, we conclude that

$$\mathbb{P}_Q\{X_i \in A | \mathcal{E}_m\} \overset{\text{a.s.}}{=} \sum_{j=1}^{m} \frac{\sum_{\sigma \in \mathcal{S}_m : \sigma(i)=j} \prod_{k=1}^{m} \lambda_k(X_{\sigma(k)})}{\sum_{\sigma \in \mathcal{S}_m} \prod_{k=1}^{m} \lambda_k(X_{\sigma(k)})} \cdot \mathbb{1}_{X_j \in A}.$$

Acknowledgments

We are grateful to Vladimir Vovk for suggesting that we study a de Finetti-type representation in the weighted setting, and for introducing us to the results of Lauritzen (1988). We are also grateful to Isaac Gibbs for helpful feedback on an earlier draft of this manuscript.

Funding

We thank the American Institute of Mathematics for supporting and hosting our collaboration. R.F.B. was supported by the National Science Foundation via grants DMS-1654076 and DMS-2023109, and by the Office of Naval Research via grant N00014-20-1-2337. E.J.C. was supported by the Office of Naval Research grant N00014-20-1-2157, the National Science Foundation grant DMS-2032014, the Simons Foundation under award 814641, and the ARO grant 2003514594.

Supplementary Material

Supplement to "De Finetti's theorem and related results for infinite weighted exchangeable sequences" (DOI: 10.3150/23-BEJ1704SUPP; .pdf). Additional proofs and technical lemmas.

References

Alam, I. (2020). Generalizing the de Finetti–Hewitt–Savage theorem. arXiv preprint. Available at arXiv:2008.08754.

Aldous, D.J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII – 1983*. *Lecture Notes in Math.* **1117** 1–198. Berlin: Springer. MR0883646 https://doi.org/10.1007/BFb0099421

Barber, R.F., Candès, E.J., Ramdas, A. and Tibshirani, R.J. (2024). Supplement to "De Finetti's theorem and related results for infinite weighted exchangeable sequences." https://doi.org/10.3150/23-BEJ1704SUPP

Bogachev, V.I. (2007). *Measure Theory. Vol. I, II*. Berlin: Springer. MR2267655 https://doi.org/10.1007/978-3-540-34514-5

Candès, E., Lei, L. and Ren, Z. (2023). Conformalized survival analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 24–45.

de Finetti, B. (1937). La prévision : Ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincaré* **7** 1–68. MR1508036

de Finetti, B. (1929). Funzione caratteristica di un fenomeno aleatorio. In *Atti del Congresso Internazionale dei Matematici* 179–190.

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese* **36** 271–281. Foundations of probability and statistics, II. MR0517222 https://doi.org/10.1007/BF00486116

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. *Ann. Probab.* **8** 745–764. MR0577313

Diaconis, P. and Freedman, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 397–423. MR0898502

Dubins, L.E. and Freedman, D.A. (1979). Exchangeable processes need not be mixtures of independent, identically distributed random variables. *Z. Wahrsch. Verw. Gebiete* **48** 115–132. MR0534840 https://doi.org/10.1007/BF01886868

Dynkin, E.B. (1953). Classes of equivalent random quantities. *Uspekhi Mat. Nauk* **8** 125–130. MR0055601

Faden, A.M. (1985). The existence of regular conditional probabilities: Necessary and sufficient conditions. *Ann. Probab.* **13** 288–298. MR0770643

Fannjiang, C., Bates, S., Angelopoulos, A.N., Listgarten, J. and Jordan, M.I. (2022). Conformal prediction under feedback covariate shift for biomolecular design. *Proc. Natl. Acad. Sci. USA* **119** Paper No. e2204569119, 12. MR4542869 https://doi.org/10.1073/pnas.2204569119

Farrell, R.H. (1962). Representation of invariant measures. *Illinois J. Math.* **6** 447–467. MR0150264

Fritz, T., Gonda, T. and Perrone, P. (2021). De Finetti's theorem in categorical probability. *J. Stoch. Anal.* **2** 4. MR4344737

Hewitt, E. and Savage, L.J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501. MR0076206 https://doi.org/10.2307/1992999

Kingman, J.F.C. (1978). Uses of exchangeability. *Ann. Probab.* **6** 183–197. MR0494344 https://doi.org/10.1214/aop/1176995566

Knopp, K. (1990). *Theory and Application of Infinite Series*. Dover Publications. MR0077665

Kolmogorov, A.N. (1930). Sur la loi forte des grands nombres. *C. R. Acad. Sci., Sér. 1 Math.* **191** 910–912.

Lauritzen, S.L. (1988). *Extremal Families and Systems of Sufficient Statistics*. *Lecture Notes in Statistics* **49**. New York: Springer. MR0971253 https://doi.org/10.1007/978-1-4612-1023-8

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J. and Wasserman, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342 https://doi.org/10.1080/01621459.2017.1307116

Lei, L. and Candès, E.J. (2021). Conformal inference of counterfactuals and individual treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 911–938. MR4349122

Maitra, A. (1977). Integral representations of invariant measures. *Trans. Amer. Math. Soc.* **229** 209–225. MR0442197 https://doi.org/10.2307/1998506

Podkopaev, A. and Ramdas, A. (2021). Distribution-free uncertainty quantification for classification under label shift. In *Uncertainty in Artificial Intelligence* 844–853. PMLR.

Schervish, M.J. (1995). *Theory of Statistics*. *Springer Series in Statistics*. New York: Springer. MR1354146 https://doi.org/10.1007/978-1-4612-4250-5

Tang, W. (2023). Finite and infinite weighted exchangeable sequences. arXiv preprint. Available at arXiv:2306.11584.

Tibshirani, R.J., Barber, R.F., Candès, E.J. and Ramdas, A. (2019). Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*.

Varadarajan, V.S. (1963). Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109** 191–220. MR0159923 https://doi.org/10.2307/1993903

Vovk, V., Gammerman, A. and Shafer, G. (2005). *Algorithmic Learning in a Random World*. New York: Springer. MR2161220

Williams, D. (1991). *Probability with Martingales*. *Cambridge Mathematical Textbooks*. Cambridge: Cambridge Univ. Press. MR1155402 https://doi.org/10.1017/CBO9780511813658

Received April 2023 and revised November 2023