De Finetti's Theorem and Related Results for Infinite Weighted Exchangeable Sequences

Rina Foygel Barber[*], Emmanuel J. Candès[†], Aaditya Ramdas[‡], Ryan J. Tibshirani[§]

Abstract

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.

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 $X_1, X_2, \dots \in \{0,1\}$ are exchangeable if and only if a draw from their joint distribution can be equivalently represented as:

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

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 $X_1, X_2, \dots \in \mathcal{X}$ 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.

[*]Department of Statistics, University of Chicago
[†]Departments of Statistics and Mathematics, Stanford University
[‡]Departments of Statistics and Machine Learning, Carnegie Mellon University
[§]Department of Statistics, University of California, Berkeley

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 $X_1, \dots, X_n$, we say (following Tibshirani et al. [2019]) that their distribution is *weighted exchangeable* with respect to given weight functions $\lambda_1, \dots, \lambda_n$ if the joint density f_n of $X_1, \dots, X_n$ factorizes as

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

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 $\lambda_1, \lambda_2, \dots$ 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 $\mathcal{X}^\infty = \mathcal{X} \times \mathcal{X} \times \dots$ 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

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

space, and $\mathcal{B}(\mathcal{X}^\infty)$ is generated by the product of individual Borel σ-algebras $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}) \times \dots$ (e.g., see Lemma 6.4.2 part (ii) in Bogachev [2007]). This same property is of course also true for finite products.

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

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

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 $\{P \in \mathcal{M}_\mathcal{X} : P(A) \leq t\}$, for $A \in \mathcal{B}(\mathcal{X})$ and $t \geq 0$. 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 $P^n = P \times P \times \dots \times P$ (n fold) and $P^\infty = P \times P \times \dots$ 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 $A_1, \dots, A_n \in \mathcal{B}(\mathcal{X})$,*

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

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 $n \geq 1$ 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 $\lambda = (\lambda_1, \lambda_2, \dots) \in \Lambda^\infty$, 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 $(\lambda_1, \dots, \lambda_n)$-weighted exchangeable.*

2.1 Background on exchangeable distributions

This section provides background on 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 $P^\infty = P \times P \times \dots$ 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*

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

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

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

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

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

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

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

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

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

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 et al. [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*

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

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

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

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 $X_1, X_2, \dots \stackrel{\text{i.i.d.}}{\sim} P$, 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

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

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

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

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

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

[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.

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

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

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^∞,

$$Q \text{ is both a product distribution and } \lambda\text{-weighted exchangeable} \iff Q = P \circ \lambda \text{ for some } P \in \mathcal{M}_\mathcal{X}. \quad (1)$$

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*

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

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

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

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

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

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:

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

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 $(X_1, X_2, \dots) \sim P \circ \lambda$,

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

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

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

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 $X_1, \dots, X_n$. To compare this to the unweighted case, observe that we would just have $(w_{n,i}(X_1, \dots, X_n))_j \equiv \frac{1}{n}$ for each j, so in this case we can view X_i as a uniform random draw from $X_1, \dots, X_n$ (after conditioning on this list). To represent the above condition compactly, we write:

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

An example. 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

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

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

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

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*

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

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

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

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

To summarize, combining our main results, we have:

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

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:*

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

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

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

By (3), then,

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

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*

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

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

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

The full proof is given in Appendix B.1.

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 Appendix B.2.

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

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

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 $X_1, \ldots, X_n$ (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., Vovk et al. [2005], Lei et al. [2018] 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 et al., 2023]. In each of these examples, the data set in hand is finite (and assumed to be weighted exchangeable). 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.

Acknowledgements

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.

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.

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 $1 \leq i < j \leq n$ (or all $1 \leq i < j$), and all measurable functions $f : \mathcal{X}^n \to [0, \infty)$ (or all measurable functions $f : \mathcal{X}^\infty \to [0, \infty)$),

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

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 $1 \leq i < j \leq n$ (or all $1 \leq i < j$), 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

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

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 $\widehat{P}_m, X_{m+1}, \ldots, X_n$ in the finite setting, or by $\widehat{P}_m, X_{m+1}, X_{m+2}, \ldots$ 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 $1 \leq i \leq m \leq n$ (or for any $1 \leq i \leq m$) that*

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

Note that the above conditional law does not depend on Q. In other words, the distribution of $X_i \mid \mathcal{E}_m$ 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 $1 \leq i \leq m \leq n$ (or for any $1 \leq i \leq m$) that*

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

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

In other words, each $X_i \mid \mathcal{E}_m$ can be viewed as a draw from a *weighted* empirical distribution of $X_1, \ldots, X_m$. The weights $(w_{m,i}(X_1, \ldots, X_m))_j$ that define this weighted empirical distribution, as constructed in (2), are determined by the functions $\lambda_1, \ldots, \lambda_m$ 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 $i \geq 1$ and $A \in \mathcal{B}(\mathcal{X})$. Define weights $(w_{n,i}(x_1, \ldots, x_n))_j$, $j \in [n]$ as in (2). First, by Proposition 7, as $P \circ \lambda$ is λ-weighted exchangeable,

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

(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 $\mathcal{E}_1 \supseteq \mathcal{E}_2 \supseteq \ldots$ with $\mathcal{E}_\infty = \cap_{n=1}^\infty \mathcal{E}_n$, by Levy's Downwards Theorem (e.g., Chapter 14.4 of Williams [1991]), we have

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

or in other words,

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

Let Y_A be an $\mathcal{E}_\infty$-measurable random variable such that $Y_A \overset{\text{a.s.}}{=} \mathbb{P}_{P\circ\lambda}\left\{X_i \in A \mid \mathcal{E}_\infty\right\}$; 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 $\lim_{n\to\infty} \widetilde{P}_{n,i}(A) \overset{\text{a.s.}}{=} Y_A$. Now, recalling that we have assumed $\lambda \in \Lambda_{01}$, we have

$$\mathbb{P}_{P\circ\lambda}\left\{Y_A \leq (P \circ \lambda_i)(A)\right\} \in \{0, 1\}.$$

But by construction, $\mathbb{E}_{P\circ\lambda}\left[Y_A\right] = (P \circ \lambda_i)(A)$, so we conclude from the above display that in fact $\mathbb{P}_{P\circ\lambda}\left\{Y_A \leq (P \circ \lambda_i)(A)\right\} = 1$. A similar argument leads to $\mathbb{P}_{P\circ\lambda}\left\{Y_A \geq (P \circ \lambda_i)(A)\right\} = 1$, 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 $X = (X_1, X_2, \dots)$ 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 Appendix A.4.

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 $(P \circ \lambda)(A) \stackrel{\text{a.s.}}{=} (P_* \circ \lambda)(A)$.*

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

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

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

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

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 $\mathsf{d_{TV}}(P_0 \circ \lambda, P_1 \circ \lambda) = 1$, where $\mathsf{d_{TV}}$ 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

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

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 $\mathsf{d_{TV}}(P_0 \circ \lambda_i, P_1 \circ \lambda_i) < 1$, and moreover,

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

We also know that

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

where the second equality holds by properties of the total variation distance for product distributions. In general, for any sequence $a_1, a_2, \dots \in [0, 1)$ 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

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

and so combining this with (7), we get

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

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 $X_1, \ldots, X_n$:

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

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 $X_1, \ldots, X_n$, 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 $U_1, U_2, \ldots \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[0,1]$, independently of X, and define for each $i = 1, 2, \ldots$,

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

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

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

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 Appendix A.5.

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

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

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

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

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

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

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 $(x_0, x_0, \dots)$ (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 Appendix A.6.

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 $\bar{P}_n(A) = \check{P}_{M_n}(A)$, where $\bar{P}_n$ 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 $\check{X}_i$, $i = 1, 2, \dots$ 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 $\mathcal{F}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$, then define $\mathcal{F}_{\text{tail}} = \cap_{n=0}^\infty \mathcal{F}_n$, which is called the *tail σ-algebra* corresponding to the infinite sequence X. The next lemma, proved in Appendix A.7, 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 $\mathcal{F}_{\text{tail}}$-measurable function $g : \mathcal{X}^\infty \to [0, \infty]$ such that*

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

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 Appendix A.8.

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

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

and also

$$\mathbb{P}_Q \left\{X_k \in A \mid X_{-k}\right\} \overset{\text{a.s.}}{=} (\widetilde{P}_* \circ \lambda_k)(A), \tag{18}$$

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 $A_1, \dots, A_n \in \mathcal{B}(\mathcal{X})$,

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

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

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

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

$$\begin{aligned} &\mathbb{E}_Q\left[\prod_{i=1}^{m-1} f_i(X) \cdot \prod_{i=m}^n \mathbb{1}_{X_i\in A_i}\right] \\ &= \mathbb{E}_Q\left[\prod_{i=1}^{m-1} f_i(X) \cdot \mathbb{P}_Q\left\{X_m \in A_m \mid X_{-m}\right\} \cdot \prod_{i=m+1}^n \mathbb{1}_{X_i\in A_i}\right] \quad \text{(as each } f_i(X) \text{ is } \sigma(X_{-m})\text{-measurable)} \end{aligned}$$

$$= \mathbb{E}_Q \left[\prod_{i=1}^{m-1} f_i(X) \cdot (\widetilde{P}_* \circ \lambda_m)(A_m) \cdot \prod_{i=m+1}^{n} \mathbb{1}_{X_i \in A_i} \right] \text{ (by the fact in (18))}$$

$$= \mathbb{E}_Q \left[\prod_{i=1}^{m-1} f_i(X) \cdot f_m(X) \cdot \prod_{i=m+1}^{n} \mathbb{1}_{X_i \in A_i} \right] \text{ (by construction of } f_m\text{)}$$

$$= \mathbb{E}_Q \left[\prod_{i=1}^{m} f_i(X) \cdot \prod_{i=m+1}^{n} \mathbb{1}_{X_i \in A_i} \right]$$

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

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

where the last step holds by construction of the functions f_i, $i = 1, \ldots, n$. 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.

References

Irfan Alam. Generalizing the de Finetti–Hewitt–Savage theorem. *arXiv preprint arXiv:2008.08754*, 2020.

David J. Aldous. Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII*, pages 1–198, 1985.

Vladimir I. Bogachev. *Measure Theory*. Springer, 2007.

Emmanuel J. Candès, Lihua Lei, and Zhimei Ren. Conformalized survival analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(1):24–45, 2023.

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

Bruno de Finetti. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7(1):1–68, 1937.

Persi Diaconis. Finite forms of de Finetti's theorem on exchangeability. *Synthese*, 36(2):271–281, 1977.

Persi Diaconis and David Freedman. Finite exchangeable sequences. *The Annals of Probability*, 8 (4):745–764, 1980.

Persi Diaconis and David Freedman. A dozen de Finetti-style results in search of a theory. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 23(S2):397–423, 1987.

Lester E. Dubins and David A. Freedman. Exchangeable processes need not be mixtures of independent, identically distributed random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 48(2):115–132, 1979.

Evgenii B. Dynkin. Classes of equivalent random quantities. *Uspekhi Matematicheskikh Nauk*, 8(2): 125–130, 1953.

Arnold M. Faden. The existence of regular conditional probabilities: Necessary and sufficient conditions. *Annals of Probability*, 13(1):288–298, 1985.

Clara Fannjiang, Stephen Bates, Anastasios N. Angelopoulos, Jennifer Listgarten, and Michael I. Jordan. Conformal prediction under feedback covariate shift for biomolecular design. *Proceedings of the National Academy of Sciences*, 119(43):e2204569119, 2022.

Roger H. Farrell. Representation of invariant measures. *Illinois Journal of Mathematics*, 6(3): 447–467, 1962.

Tobias Fritz, Tomáš Gonda, and Paolo Perrone. De Finetti's theorem in categorical probability. *arXiv preprint arXiv:2105.02639*, 2021.

Edwin Hewitt and Leonard J. Savage. Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501, 1955.

John F. C. Kingman. Uses of exchangeability. *Annals of Probability*, 6(2):183–197, 1978.

Konrad Knopp. *Theory and Application of Infinite Series*. Dover Publications, 1990.

Andrey N. Kolmogorov. Sur la loi forte des grands nombres. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 191:910–912, 1930.

Steffen L. Lauritzen. *Extremal families and systems of sufficient statistics*, volume 49. Springer Science & Business Media, 1988.

Jing Lei, Max G?Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523): 1094–1111, 2018.

Lihua Lei and Emmanuel J. Candès. Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83 (5):911–938, 2021.

Ashok Maitra. Integral representations of invariant measures. *Transactions of the American Mathematical Society*, 229:209–225, 1977.

Valentin V. Petrov. *Sums of Independent Random Variables*. De Gruyter, 1975.

Aleksandr Podkopaev and Aaditya Ramdas. Distribution-free uncertainty quantification for classification under label shift. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2021.

Mark J. Schervish. *Theory of Statistics*. Springer Science, 2012.

Wenpin Tang. Finite and infinite weighted exchangeable sequences. *arXiv preprint arXiv:2306.11584*, 2023.

Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel J. Candès, and Aaditya Ramdas. Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*, 2019.

Veeravalli S. Varadarajan. Groups of automorphisms of Borel spaces. *Transactions of the American Mathematical Society*, 109(2):191–220, 1963.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

David Williams. *Probability with Martingales.* Cambridge University Press, 1991.

A Proofs supporting main results

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, $X \mid \mathcal{E}_\infty$, or $(X_{n+1}, X_{n+2}, \dots) \mid (X_1, \dots, X_n)$, etc. For example, if we consider the conditional distribution of $X \mid \mathcal{E}_\infty$, 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 $X \mid \mathcal{E}_\infty$, i.e., $\kappa(X, A) \stackrel{\text{a.s.}}{=} \mathbb{P}\{X \in A \mid \mathcal{E}_\infty\}$. In what follows, 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) $\Longrightarrow$ (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 $A = A_1 \times \cdots \times A_n \times \mathcal{X} \times \mathcal{X} \times \dots$ where $n \geq 1$ and $A_1, \dots, A_n \in \mathcal{B}(\mathcal{X})$, 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

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

We then have

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

and similarly,

$$\begin{aligned}
\mathbb{E}_Q\left[\frac{\mathbb{1}_{X^{ij} \in A}}{\lambda_i(X_i)\lambda_j(X_j)}\right] &= \mathbb{E}_Q\left[\frac{\mathbb{1}_{((X^{ij})_1,\dots,(X^{ij})_n) \in A_1 \times \cdots \times A_n}}{\lambda_i(X_i)\lambda_j(X_j)}\right] \\
&= \mathbb{E}_{Q_n}\left[\frac{\mathbb{1}_{((X^{ij})_1,\dots,(X^{ij})_n) \in A_1 \times \cdots \times A_n} \cdot \prod_{k \in [n] \setminus \{i,j\}} \lambda_k(X_k)}{\prod_{k=1}^n \lambda_k(X_k)}\right]
\end{aligned}$$

$$= \mathbb{E}_{Q_n}\left[\frac{h((X^{ij})_1,\ldots,(X^{ij})_n)}{\prod_{k=1}^n \lambda_k(X_k)}\right]$$
$$= \int_{\mathcal{X}^n} h((x^{ij})_1,\ldots,(x^{ij})_n)\,\frac{\mathsf{d}Q_n(x_1,\ldots,x_n)}{\prod_{k=1}^n \lambda_k(x_k)}.$$

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

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

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

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

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

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

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

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

Define

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

Then since we have assumed (b) holds, we have $\mathbb{E}_Q\left[\frac{f(X)}{\lambda_i(X_i)\lambda_j(X_j)}\right] = \mathbb{E}_Q\left[\frac{f(X^{ij})}{\lambda_i(X_i)\lambda_j(X_j)}\right]$, 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 $A \in \mathcal{B}(\mathcal{X})$,

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

Equivalently, it is sufficient to check that

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

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., $g((x_{\sigma(1)},\ldots,x_{\sigma(m)},x_{m+1},\ldots,x_n)) = g(x)$ for all $x \in \mathcal{X}^n$ and all $\sigma \in \mathbb{R}^m$. We have

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

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

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 $\cup_{n\geq m}\mathcal{E}_{m,n}$. By Levy's Upwards Theorem (e.g., Chapter 14.2 of Williams [1991]),

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

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

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

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

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

A.4 Proof of Lemma 1

Fix any $i\neq j$ and define $\lambda_{ij}(x)=\min\{\lambda_i(x),\lambda_j(x)\}$. For any $P\in\mathcal{M}_{\mathcal{X}}(\lambda)$, a direct calculation shows that

$$
(P\circ\lambda_i)(A)=\frac{\mathbb{E}_{P\circ\lambda_{ij}}\left[\mathbb{1}_{X\in A}\cdot\frac{\lambda_i(X)}{\lambda_{ij}(X)}\right]}{\mathbb{E}_{P\circ\lambda_{ij}}\left[\frac{\lambda_i(X)}{\lambda_{ij}(X)}\right]}
$$

for all $A \in \mathcal{B}(\mathcal{X})$. Therefore, if $(P \circ \lambda_{ij})(A) \overset{\text{a.s.}}{=} (P_* \circ \lambda_{ij})(A)$ for any $i \neq j$ and for all $A \in \mathcal{B}(\mathcal{X})$, then this implies that $(P \circ \lambda_i)(A) \overset{\text{a.s.}}{=} (P_* \circ \lambda_i)(A)$ for any i and for all $A \in \mathcal{B}(\mathcal{X})$. Therefore, it suffices to show that $(P \circ \lambda_{ij})(A) \overset{\text{a.s.}}{=} (P_* \circ \lambda_{ij})(A)$ for any $A \in \mathcal{B}(\mathcal{X})$.

Next define a new distribution $\tilde{\mu}$ on $\mathcal{M}_{\mathcal{X}}(\lambda)$ as

$$\tilde{\mu}(B) = \frac{\int_B \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P(x)} \cdot \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_j(x)\,\mathsf{d}P(x)}\,\mathsf{d}\mu(P)}{\int_{\mathcal{M}_{\mathcal{X}}(\lambda)} \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P'(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P'(x)} \cdot \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P'(x)}{\int_{\mathcal{X}} \lambda_j(x)\,\mathsf{d}P'(x)}\,\mathsf{d}\mu(P')}$$

for all measurable $B \subseteq \mathcal{M}_{\mathcal{X}}(\lambda)$. (This is well-defined, because the integrand in the denominator is positive and is bounded by 1, surely, by construction.) Since μ and $\tilde{\mu}$ are absolutely continuous with respect to each other, with Radon–Nikodym derivative

$$\frac{\mathsf{d}\tilde{\mu}(P)}{\mathsf{d}\mu(P)} = \frac{\frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P(x)} \cdot \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_j(x)\,\mathsf{d}P(x)}}{\int_{\mathcal{M}_{\mathcal{X}}(\lambda)} \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P'(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P'(x)} \cdot \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P'(x)}{\int_{\mathcal{X}} \lambda_j(x)\,\mathsf{d}P'(x)}\,\mathsf{d}\mu(P')} > 0,$$

we see that $(P \circ \lambda_{ij})(A) = (P_* \circ \lambda_{ij})(A)$ holds almost surely under $P \sim \mu$ if and only if it holds almost surely under $P \sim \tilde{\mu}$, so now we will verify the statement for $P \sim \tilde{\mu}$. This is the same as verifying that

$$\mathbb{E}_{P\sim\tilde{\mu}}\left[(P \circ \lambda_{ij})(A)\right] = (P_* \circ \lambda_{ij})(A) \quad \text{and} \quad \mathrm{Var}_{P\sim\tilde{\mu}}((P \circ \lambda_{ij})(A)) = 0,$$

or equivalently,

$$\mathbb{E}_{P\sim\tilde{\mu}}\left[(P \circ \lambda_{ij})(A)\right] = (P_* \circ \lambda_{ij})(A) \quad \text{and} \quad \mathbb{E}_{P\sim\tilde{\mu}}\left[\big((P \circ \lambda_{ij})(A)\big)^2\right] = \big((P_* \circ \lambda_{ij})(A)\big)^2.$$

For any $B_1, B_2 \in \mathcal{B}(\mathcal{X})$,

$$\begin{aligned}
&\mathbb{E}_{P\sim\tilde{\mu}}\left[(P \circ \lambda_{ij})^2(B_1 \times B_2)\right] \\
&= \mathbb{E}_{P\sim\mu}\left[\int (P \circ \lambda_{ij})^2(B_1 \times B_2) \frac{\frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P(x)} \cdot \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_j(x)\,\mathsf{d}P(x)}}{\int_{\mathcal{M}_{\mathcal{X}}(\lambda)} \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P'(x)} \cdot \frac{\int_{\mathcal{X}} \lambda_{ij}(x)\,\mathsf{d}P'(x)}{\int_{\mathcal{X}} \lambda_j(x)\,\mathsf{d}P'(x)}\,\mathsf{d}\mu(P')}\right] \quad \text{(by definition of } \tilde{\mu})\\
&= \frac{\mathbb{E}_{P\sim\mu}\left[\mathbb{E}_{P\circ\lambda}\left[\mathbb{1}_{(X_i,X_j)\in B_1\times B_2} \cdot \frac{\lambda_{ij}(X_i)}{\lambda_i(X_i)} \cdot \frac{\lambda_{ij}(X_j)}{\lambda_j(X_j)}\right]\right]}{\mathbb{E}_{P\sim\mu}\left[\mathbb{E}_{P\circ\lambda}\left[\frac{\lambda_{ij}(X_i)}{\lambda_i(X_i)} \cdot \frac{\lambda_{ij}(X_j)}{\lambda_j(X_j)}\right]\right]} \quad \text{(by definition of } P\circ\lambda \text{ and of } P\circ\lambda_{ij})\\
&= \frac{\mathbb{E}_{P_*\circ\lambda}\left[\mathbb{1}_{(X_i,X_j)\in B_1\times B_2} \cdot \frac{\lambda_{ij}(X_i)}{\lambda_i(X_i)} \cdot \frac{\lambda_{ij}(X_j)}{\lambda_j(X_j)}\right]}{\mathbb{E}_{P_*\circ\lambda}\left[\frac{\lambda_{ij}(X_i)}{\lambda_i(X_i)} \cdot \frac{\lambda_{ij}(X_j)}{\lambda_j(X_j)}\right]} \quad \text{(since } P_*\circ\lambda = Q = (P\circ\lambda)_\mu)\\
&= \frac{\mathbb{E}_{P_*\circ\lambda_i}\left[\mathbb{1}_{X\in B_1} \cdot \frac{\lambda_{ij}(X)}{\lambda_i(X)}\right]}{\mathbb{E}_{P_*\circ\lambda_i}\left[\frac{\lambda_{ij}(X)}{\lambda_i(X)}\right]} \cdot \frac{\mathbb{E}_{P_*\circ\lambda_j}\left[\mathbb{1}_{X\in B_2} \cdot \frac{\lambda_{ij}(X)}{\lambda_j(X)}\right]}{\mathbb{E}_{P_*\circ\lambda_j}\left[\frac{\lambda_{ij}(X)}{\lambda_j(X)}\right]} \quad \text{(by definition of } P_*\circ\lambda)\\
&= (P_* \circ \lambda_{ij})(B_1) \cdot (P_* \circ \lambda_{ij})(B_2) \quad \text{(by definition of } P_*\circ\lambda_i,\ P_*\circ\lambda_j,\ P_*\circ\lambda_{ij}).
\end{aligned}$$

Applying this calculation with $B_1 = A$ and $B_2 = \mathcal{X}$ we obtain

$$\mathbb{E}_{P\sim\tilde{\mu}}\left[(P \circ \lambda_{ij})(A)\right] = \mathbb{E}_{P\sim\tilde{\mu}}\left[(P \circ \lambda_{ij})(A) \cdot (P \circ \lambda_{ij})(\mathcal{X})\right] = (P_*\circ\lambda_{ij})(A)\cdot(P_*\circ\lambda_{ij})(\mathcal{X}) = (P_*\circ\lambda_{ij})(A).$$

Applying the calculation again with $B_1 = B_2 = A$, we obtain

$$\mathbb{E}_{P\sim\tilde{\mu}}\left[\big((P \circ \lambda_{ij})(A)\big)^2\right] = \big((P_* \circ \lambda_{ij})(A)\big)^2,$$

which completes the proof.

A.5 Proof of Lemma 2

First consider the case that $\sum_{i=1}^{\infty} p_i(X_i) \cdot \mathbb{1}_{X_i \in A} = \infty$. For n sufficiently large so that $\sum_{i=1}^{n} B_i > 0$ and $\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A} > 0$, we calculate

$$\begin{aligned}\bar{P}_n(A) &= \frac{\sum_{i=1}^{n} B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} B_i} \\ &= \frac{\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} p_i(X_i)} \cdot \frac{\sum_{i=1}^{n} p_i(X_i)}{\sum_{i=1}^{n} B_i} \cdot \frac{\sum_{i=1}^{n} B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A}} \\ &= \widetilde{P}_n(A) \cdot \frac{\sum_{i=1}^{n} p_i(X_i)}{\sum_{i=1}^{n} B_i} \cdot \frac{\sum_{i=1}^{n} B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A}},\end{aligned}$$

where the last step holds by the definition of $\widetilde{P}_n$ and of $p_i(X_i)$ given in (8) and (10). Thus,

$$\left|\bar{P}_n(A) - \widetilde{P}_n(A)\right| = \widetilde{P}_n(A) \cdot \left|\frac{\sum_{i=1}^{n} p_i(X_i)}{\sum_{i=1}^{n} B_i} \cdot \frac{\sum_{i=1}^{n} B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A}} - 1\right|,$$

and so, since $\widetilde{P}_n(A) \leq 1$ because $\widetilde{P}_n$ is a distribution,

$$\limsup_{n\to\infty} \left|\bar{P}_n(A) - \widetilde{P}_n(A)\right| \leq \limsup_{n\to\infty} \left|\frac{\sum_{i=1}^{n} p_i(X_i)}{\sum_{i=1}^{n} B_i} \cdot \frac{\sum_{i=1}^{n} B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A}} - 1\right|.$$

Next, since $\sum_{i=1}^{\infty} p_i(X_i) = \infty$ and $\sum_{i=1}^{\infty} p_i(X_i)\mathbb{1}_{X_i \in A} = \infty$ in this case, we have

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} B_i}{\sum_{i=1}^{n} p_i(X_i)} = 1, \quad \text{almost surely (conditional on } X),$$

by Chapter IX, Section §3, Theorem 12 of Petrov [1975] (applied by checking the variance condition $\sum_{i \geq i_*} \frac{\mathrm{Var}(B_i|X_i)}{(\sum_{j\leq i} p_j(X_j))^2} \leq \sum_{i \geq i_*} \frac{p_i(X_i)}{(\sum_{j\leq i} p_j(X_j))^2} < \infty$, for $i_* = \min\{i : p_i(X_i) > 0\}$). Thus, similarly,

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A}} = 1, \quad \text{almost surely (conditional on } X).$$

Thus, on the event $\sum_{i=1}^{\infty} p_i(X_i) \cdot \mathbb{1}_{X_i \in A} = \infty$, we have

$$\limsup_{n\to\infty} \left|\frac{\sum_{i=1}^{n} p_i(X_i)}{\sum_{i=1}^{n} B_i} \cdot \frac{\sum_{i=1}^{n} B_i \cdot \mathbb{1}_{X_i \in A}}{\sum_{i=1}^{n} p_i(X_i) \cdot \mathbb{1}_{X_i \in A}} - 1\right| = 0, \quad \text{almost surely (conditional on } X),$$

and so

$$\limsup_{n\to\infty} \left|\bar{P}_n(A) - \widetilde{P}_n(A)\right| = 0, \quad \text{almost surely (conditional on } X). \tag{20}$$

Next, consider the case that $\sum_{i=1}^{\infty} p_i(X_i) \cdot \mathbb{1}_{X_i \in A} < \infty$ while $\sum_{i=1}^{\infty} p_i(X_i) = \infty$. Then in this case, conditional on X, it holds almost surely that $\sum_{i=1}^{\infty} B_i \cdot \mathbb{1}_{X_i \in A} < \infty$ while $\sum_{i=1}^{\infty} B_i = \infty$ (by the first and second Borel–Cantelli Lemmas, respectively); on these events, we have

$$\lim_{n\to\infty} \widetilde{P}_n(A) = \lim_{n\to\infty} \bar{P}_n(A) = 0.$$

This implies (20) again holds conditional on X.

Finally, we have shown that $\sum_{i=1}^{\infty} p_i(X_i) = \infty$ almost surely. Combining the cases considered above, we see that the claim (20) holds marginally, i.e., with respect to the distribution of X as well as $B_1, B_2, \ldots$, which completes the proof.

A.6 Proof of Lemma 3

By Proposition 4, it suffices to show that, for any $j \neq k$ and any measurable $f : \mathcal{X}^\infty \to [0, \infty)$,

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

It is equivalent to verify this for functions of the form $f(x) = \mathbb{1}_{x \in A}$, where $A \in \mathcal{B}(\mathcal{X}^\infty)$, i.e., to show that for any $j \neq k$ and for any $A \in \mathcal{B}(\mathcal{X}^\infty)$,

$$\mathbb{P}\left\{ \check{X} \in A \right\} = \mathbb{P}\left\{ \check{X}^{jk} \in A \right\}.$$

Since $\mathcal{B}(\mathcal{X}^\infty)$ is the product σ-algebra, it suffices to show that, for all $n \geq \max\{j, k\}$ and all $A \in \mathcal{B}(\mathcal{X}^n)$,

$$\mathbb{P}\left\{ (\check{X}_1, \ldots, \check{X}_n) \in A \right\} = \mathbb{P}\left\{ ((\check{X}^{jk})_1, \ldots, (\check{X}^{jk})_n) \in A \right\}.$$

We will prove the stronger statement

$$\mathbb{P}\left\{ (\check{X}_1, \ldots, \check{X}_n) \in A \;\middle|\; I_1, \ldots, I_n \right\} \stackrel{\text{a.s.}}{=} \mathbb{P}\left\{ ((\check{X}^{jk})_1, \ldots, (\check{X}^{jk})_n) \in A \;\middle|\; I_1, \ldots, I_n \right\}.$$

Equivalently, for any indices $1 \leq i_1 < \cdots < i_n < \infty$ such that the event $E = \{I_1 = i_1, \ldots, I_n = i_n\}$ has positive probability, we need to verify that

$$\mathbb{P}\left\{ (\check{X}_1, \ldots, \check{X}_n) \in A \;\middle|\; I_1 = i_1, \ldots, I_n = i_n \right\} = \mathbb{P}\left\{ ((\check{X}^{jk})_1, \ldots, (\check{X}^{jk})_n) \in A \;\middle|\; I_1 = i_1, \ldots, I_n = i_n \right\},$$

or equivalently,

$$\mathbb{E}\left[\mathbb{1}_{(\check{X}_1, \ldots, \check{X}_n) \in A} \mathbb{1}_E \right] = \mathbb{E}\left[\mathbb{1}_{((\check{X}^{jk})_1, \ldots, (\check{X}^{jk})_n) \in A} \mathbb{1}_E \right].$$

Now let us rewrite this to be more interpretable. Define

$$A' = \{x \in \mathcal{X}^\infty : (x_{i_1}, \ldots, x_{i_n}) \in A\},$$

thus $\mathbb{1}_{(X_{i_1}, \ldots, X_{i_n}) \in A} = \mathbb{1}_{X \in A'}$ holds surely, meaning that on E, it holds that $\mathbb{1}_{(\check{X}_1, \ldots, \check{X}_n) \in A} = \mathbb{1}_{X \in A'}$. Moreover, by construction,

$$(X_{i_1}, \ldots, X_{i_n})^{jk} = ((X^{i_j i_k})_{i_1}, \ldots, (X^{i_j i_k})_{i_n})$$

and thus $\mathbb{1}_{(X_{i_1}, \ldots, X_{i_n})^{jk} \in A} = \mathbb{1}_{X^{i_j i_k} \in A'}$, so that on E, it holds that $\mathbb{1}_{(\check{X}_1, \ldots, \check{X}_n)^{jk} \in A} = \mathbb{1}_{X^{i_j i_k} \in A'}$. Therefore, we now just need to show

$$\mathbb{E}\left[\mathbb{1}_{X \in A'} \mathbb{1}_E \right] = \mathbb{E}\left[\mathbb{1}_{X^{i_j i_k} \in A} \mathbb{1}_E \right].$$

Next, we calculate

$$\begin{aligned}
\mathbb{P}\left\{ E \mid X \right\} &= \mathbb{P}\left\{ B_\ell = \mathbb{1}_{\ell \in \{i_1, \ldots, i_n\}}, \ell = 1, \ldots, i_n \;\middle|\; X \right\} \\
&= \prod_{r=1}^{n} \mathbb{P}\left\{ B_{i_r} = 1 \mid X \right\} \cdot \prod_{\ell \in [i_n] \setminus \{i_1, \ldots, i_n\}} \mathbb{P}\left\{ B_{i_r} = 0 \mid X \right\} \\
&= \prod_{r=1}^{n} p_{i_r}(X_{i_r}) \cdot \prod_{\ell \in [i_n] \setminus \{i_1, \ldots, i_n\}} (1 - p_\ell(X_\ell))
\end{aligned}$$

$$= \prod_{r=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_{i_r}(x)/\lambda_*(x)}{\lambda_{i_r}(X_{i_r})/\lambda_*(X_{i_r})} \cdot \prod_{\ell\in[i_n]\setminus\{i_1,\dots,i_n\}} \left(1 - \frac{\inf_{x\in\mathcal{X}} \lambda_\ell(x)/\lambda_*(x)}{\lambda_\ell(X_\ell)/\lambda_*(X_\ell)}\right)$$
$$= \frac{h(X)}{\lambda_{i_j}(X_{i_j})\lambda_{i_k}(X_{i_k})},$$

where

$$h(x) = \frac{\inf_{x'\in\mathcal{X}} \lambda_{i_j}(x')/\lambda_*(x')}{1/\lambda_*(x_{i_j})} \cdot \frac{\inf_{x'\in\mathcal{X}} \lambda_{i_k}(x')/\lambda_*(x')}{1/\lambda_*(x_{i_k})} \cdot \prod_{r\in[n]\setminus\{j,k\}} \frac{\inf_{x'\in\mathcal{X}} \lambda_{i_r}(x')/\lambda_*(x')}{\lambda_{i_r}(x_{i_r})/\lambda_*(x_{i_r})} \cdot \prod_{\ell\in[i_n]\setminus\{i_1,\dots,i_n\}} \left(1 - \frac{\inf_{x'\in\mathcal{X}} \lambda_\ell(x')/\lambda_*(x')}{\lambda_\ell(x_\ell)/\lambda_*(x_\ell)}\right).$$

We can therefore write

$$\mathbb{E}\left[\mathbb{1}_{X\in A'}\mathbb{1}_E\right] = \mathbb{E}\left[\mathbb{1}_{X\in A'} \cdot \mathbb{P}\{E \mid X\}\right] = \mathbb{E}\left[\mathbb{1}_{X\in A'} \cdot \frac{h(X)}{\lambda_{i_j}(X_{i_j})\lambda_{i_k}(X_{i_k})}\right],$$

and similarly

$$\mathbb{E}\left[\mathbb{1}_{X^{i_j i_k}\in A'}\mathbb{1}_E\right] = \mathbb{E}\left[\mathbb{1}_{X^{i_j i_k}\in A'} \cdot \frac{h(X)}{\lambda_{i_j}(X_{i_j})\lambda_{i_k}(X_{i_k})}\right] = \mathbb{E}\left[\mathbb{1}_{X^{i_j i_k}\in A'} \cdot \frac{h(X^{i_j i_k})}{\lambda_{i_j}(X_{i_j})\lambda_{i_k}(X_{i_k})}\right],$$

where the last step holds because $h(x) = h(x^{i_j i_k})$ for all x, by definition. By Proposition 5, we have

$$\mathbb{E}\left[\mathbb{1}_{X^{i_j i_k}\in A'} \cdot \frac{h(X^{i_j i_k})}{\lambda_{i_j}(X_{i_j})\lambda_{i_k}(X_{i_k})}\right] = \mathbb{E}\left[\mathbb{1}_{X\in A'} \cdot \frac{h(X)}{\lambda_{i_j}(X_{i_j})\lambda_{i_k}(X_{i_k})}\right],$$

which proves that

$$\mathbb{E}\left[\mathbb{1}_{X\in A'}\mathbb{1}_E\right] = \mathbb{E}\left[\mathbb{1}_{X^{i_j i_k}\in A'}\mathbb{1}_E\right],$$

and thus completes the proof.

A.7 Proof of Lemma 4

First, we calculate

$$\begin{aligned}
\mathbb{E}_{X'\sim\widetilde{P}}\left[f(X')\right] &= \int_{\mathcal{X}} f(x)\,\mathsf{d}\widetilde{P}(x) \\
&= \int_{\mathcal{X}}\int_{t\geq 0} \mathbb{1}_{f(x)\geq t}\,\mathsf{d}t\,\mathsf{d}\widetilde{P}(x) \\
&= \int_{t\geq 0}\int_{\mathcal{X}} \mathbb{1}_{f(x)\geq t}\,\mathsf{d}\widetilde{P}(x)\,\mathsf{d}t \quad \text{(by Tonelli's theorem)} \\
&= \int_{t\geq 0} \widetilde{P}(L_t)\,\mathsf{d}t,
\end{aligned}$$

where L_t is the nested family of sets given by $L_t = \{x \in \mathcal{X} : f(x) \geq t\} \in \mathcal{B}(\mathcal{X})$.

Simplifying further, it is therefore sufficient to show that, for any $C \in \mathcal{B}(\mathcal{X})$, there exists an $\mathcal{F}_{\text{tail}}$-measurable function $h_C : \mathcal{X}^\infty \to [0, 1]$ such that $h_C(X) \overset{\text{a.s.}}{=} \widetilde{P}(C)$. Taking

$$h_C(X) = \limsup_{n\to\infty} \widetilde{P}_n(C),$$

by (16), we see that $h_C(X) \stackrel{\text{a.s.}}{=} \widetilde{P}(C)$ must hold. It remains to show that the random variable $h_C(X)$ is $\mathcal{F}_{\text{tail}}$-measurable.

For any $n > m \geq 0$ define

$$T_{m,n} = \frac{\sum_{i=m+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)} \cdot \mathbb{1}_{X_i\in C}}{\sum_{i=m+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)}},$$

where, for any fixed m, $T_{m,n}$ is well-defined (i.e., the denominator is positive) for sufficiently large n by (4). Note that $T_{m,n}$ is $\mathcal{F}_m$-measurable, and so $\limsup_{n\to\infty} T_{m,n}$ is also $\mathcal{F}_m$-measurable. To complete the proof, we will verify that

$$\limsup_{n\to\infty} \widetilde{P}_n(C) = \limsup_{n\to\infty} T_{m,n}$$

holds surely, for any m; if this is the case, then $h_C(X)$ is $\mathcal{F}_m$-measurable for all $m \geq 0$, and thus is $\mathcal{F}_{\text{tail}}$-measurable. It is sufficient to verify that

$$\limsup_{n\to\infty} \left| T_{m,n} - \widetilde{P}_n(C) \right| = 0 \tag{21}$$

holds surely. We calculate (for any $n > m$ sufficiently large, so that $T_{m,n}$ is well-defined),

$$\begin{aligned}\left| T_{m,n} - \widetilde{P}_n(C) \right| &= \left| \frac{\sum_{i=m+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)} \cdot \mathbb{1}_{X_i\in C}}{\sum_{i=m+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)}} - \frac{\sum_{i=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)} \cdot \mathbb{1}_{X_i\in C}}{\sum_{i=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)}} \right| \\ &\leq \frac{\sum_{i=1}^{m} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)}}{\sum_{i=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_*(X_i)}} \leq \frac{m}{\sum_{i=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}},\end{aligned}$$

and note that the limit of the denominator is ∞ by (4). Therefore, (21) holds surely, and thus we have shown that $h_C(X)$ is $\mathcal{F}_{\text{tail}}$-measurable, which completes the proof.

A.8 Proof of Lemma 5

We split the proof into two parts. First, we verify that the two claims hold for a particular choice of k. Then we extend to all $k \geq 1$.

A.8.1 Special case: proof for a specific k

First we fix any $k \geq 1$ satisfying

$$\frac{\inf_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x)} > 0.$$

Note that there must be infinitely many such k, by (4). We will first show that the two claims hold for this particular choice of k.

First, we check (17): since $\widetilde{P}$ is a distribution, we have

$$\int_{\mathcal{X}} \lambda_k(x)\, \mathsf{d}\widetilde{P}_*(x) = \int_{\mathcal{X}} \frac{\lambda_k(x)}{\lambda_*(x)}\, \mathsf{d}\widetilde{P}(x) \in \left[\inf_{x\in\mathcal{X}} \frac{\lambda_k(x)}{\lambda_*(x)}, \sup_{x\in\mathcal{X}} \frac{\lambda_k(x)}{\lambda_*(x)}\right] \subseteq (0,\infty),$$

where the last step holds by our choice of k. Therefore, (17) holds surely for our choice of k.

Next, we need to prove that (18) holds for our choice of k. It is equivalent to show that, for $X \sim Q$,

$$\mathbb{P}\left\{X_k \in A, X_{-k} \in B\right\} = \mathbb{E}\left[(\widetilde{P}_* \circ \lambda_k)(A) \cdot \mathbb{1}_{X_{-k} \in B}\right], \quad \text{for all } A \in \mathcal{B}(\mathcal{X}),\ B \in \mathcal{B}(\mathcal{X}^\infty).$$

Note that we have shown that $(\widetilde{P}_* \circ \lambda_k)$ is well-defined surely, in our proof of (17), above, and thus this expected value is well-defined. As $\mathcal{B}(\mathcal{X}^\infty)$ is a product σ-algebra, it is sufficient to show that, for all $\ell \geq k$,

$$\mathbb{P}\left\{X_k \in A, X_{[\ell]\setminus k} \in B\right\} = \mathbb{E}\left[(\widetilde{P}_* \circ \lambda_k)(A) \cdot \mathbb{1}_{X_{[\ell]\setminus k} \in B}\right], \quad \text{for all } A \in \mathcal{B}(\mathcal{X}),\ B \in \mathcal{B}(\mathcal{X}^{\ell-1}), \tag{22}$$

where $X_{[\ell]\setminus k} = (X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_\ell) \in \mathcal{X}^{\ell-1}$.

The next lemma, proved in Appendix A.9, reduces the above condition to one involving a limit of the weighted empirical distribution.

Lemma 6. *Under the notation and assumptions above, it holds that*

$$\lim_{n\to\infty} \left|\mathbb{P}\left\{X_k \in A, X_{[\ell]\setminus k} \in B\right\} - \mathbb{E}\left[(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) \cdot \mathbb{1}_{X_{[\ell]\setminus k} \in B}\right]\right| = 0,$$

where $(\lambda_k/\lambda_*) \in \Lambda$ *should be interpreted elementwise, i.e.,* $(\lambda_k/\lambda_*)(x) = \lambda_k(x)/\lambda_*(x)$.

Thus to show (22), we only need to show that

$$\lim_{n\to\infty} \left|\mathbb{E}\left[(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) \cdot \mathbb{1}_{X_{[\ell]\setminus k} \in B}\right] - \mathbb{E}\left[(\widetilde{P}_* \circ \lambda_k)(A) \cdot \mathbb{1}_{X_{[\ell]\setminus k} \in B}\right]\right| = 0,$$

for which it is sufficient to verify

$$\lim_{n\to\infty} \mathbb{E}\left[\left|(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) - (\widetilde{P}_* \circ \lambda_k)(A)\right|\right] = 0.$$

Since the term inside the expectation above is at most 1 in absolute value, by the dominated convergence theorem, it suffices to verify that

$$\left|(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) - (\widetilde{P}_* \circ \lambda_k)(A)\right| \overset{\text{a.s.}}{\to} 0. \tag{23}$$

By construction, we have $\widetilde{P}_* \circ \lambda_k = \widetilde{P} \circ (\lambda_k/\lambda_*)$, so we calculate

$$\left|(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) - (\widetilde{P}_* \circ \lambda_k)(A)\right| = \left|\frac{\int_A \frac{\lambda_k(x)}{\lambda_*(x)}\,\mathsf{d}\widetilde{P}_n(x)}{\int_{\mathcal{X}} \frac{\lambda_k(x)}{\lambda_*(x)}\,\mathsf{d}\widetilde{P}_n(x)} - \frac{\int_A \frac{\lambda_k(x)}{\lambda_*(x)}\,\mathsf{d}\widetilde{P}(x)}{\int_{\mathcal{X}} \frac{\lambda_k(x)}{\lambda_*(x)}\,\mathsf{d}\widetilde{P}(x)}\right|.$$

Since $\lambda_k(x)/\lambda_*(x)$ is positive and bounded by our choice of k, it suffices to show that

$$\int_{A'} \frac{\lambda_k(x)}{\lambda_*(x)}\,\mathsf{d}\widetilde{P}_n(x) \overset{\text{a.s.}}{\to} \int_{A'} \frac{\lambda_k(x)}{\lambda_*(x)}\,\mathsf{d}\widetilde{P}(x).$$

holds for any $A' \in \mathcal{B}(\mathcal{X})$, which is true because we have constructed $\widetilde{P}$ to be the almost sure weak limit of $\widetilde{P}_n$, as in (16). We have therefore established that (23) holds, and so we have shown that (18) holds for our choice of k.

A.8.2 General case

Now that the two claims are established for our particular choice of k, we will next show that both claims hold with ℓ in place of k, for any $\ell \geq 1$ with $\ell \neq k$.

First, we need to show, for any $\ell \neq k$,

$$0 < \int_{\mathcal{X}} \lambda_\ell(x) \, \mathsf{d}\widetilde{P}_*(x) < \infty$$

holds almost surely, to verify (17) (with ℓ in place of k). By definition of $\widetilde{P}_*$, we have

$$\int_{\mathcal{X}} \lambda_\ell(x) \, \mathsf{d}\widetilde{P}_*(x) = \int_{\mathcal{X}} \frac{\lambda_\ell(x)}{\lambda_*(x)} \, \mathsf{d}\widetilde{P}(x) > 0,$$

surely, as $\frac{\lambda_\ell(x)}{\lambda_*(x)}$ is always positive, and $\widetilde{P} \in \mathcal{P}_{\mathcal{X}}$. Therefore the lower bound holds surely. Next we address the upper bound. We calculate

$$\begin{aligned}
1 &= \mathbb{E}_Q\left[\frac{\lambda_k(X_k)\lambda_\ell(X_\ell)}{\lambda_k(X_k)\lambda_\ell(X_\ell)}\right] \\
&= \mathbb{E}_Q\left[\frac{\lambda_k(X_\ell)\lambda_\ell(X_k)}{\lambda_k(X_k)\lambda_\ell(X_\ell)}\right] \quad \text{(by Proposition 5)} \\
&= \mathbb{E}_Q\left[\mathbb{E}_Q\left[\frac{\lambda_\ell(X_k)}{\lambda_k(X_k)} \;\middle|\; X_{-k}\right] \frac{\lambda_k(X_\ell)}{\lambda_\ell(X_\ell)}\right] \\
&= \mathbb{E}_Q\left[\int_{\mathcal{X}} \frac{\lambda_\ell(x)}{\lambda_k(x)} \, \mathsf{d}(\widetilde{P}_* \circ \lambda_k)(x) \cdot \frac{\lambda_k(X_\ell)}{\lambda_\ell(X_\ell)}\right],
\end{aligned}$$

where the last step holds since we have proved (18) for k (and recall that we have shown that $\widetilde{P}_* \circ \lambda_k$ is well-defined, surely). Next we have

$$\int_{\mathcal{X}} \frac{\lambda_\ell(x)}{\lambda_k(x)} \, \mathsf{d}(\widetilde{P}_* \circ \lambda_k) = \frac{\int_{\mathcal{X}} \lambda_\ell(x) \, \mathsf{d}\widetilde{P}_*(x)}{\int_{\mathcal{X}} \lambda_k(x) \, \mathsf{d}\widetilde{P}_*(x)},$$

and therefore,

$$1 = \mathbb{E}_Q\left[\int_{\mathcal{X}} \lambda_\ell(x) \, \mathsf{d}\widetilde{P}_*(x) \cdot \frac{\lambda_k(X_\ell)/\lambda_\ell(X_\ell)}{\int_{\mathcal{X}} \lambda_k(x) \, \mathsf{d}\widetilde{P}_*(x)}\right].$$

Note that $\lambda_k(X_\ell)/\lambda_\ell(X_\ell)/\int_{\mathcal{X}} \lambda_k(x) \, \mathsf{d}\widetilde{P}_*(x) > 0$ almost surely—in the numerator, λ_k, λ_ℓ both take positive finite values, while in the denominator, we apply (17) with our choice of k. Therefore, $\int_{\mathcal{X}} \lambda_\ell(x) \, \mathsf{d}\widetilde{P}_*(x) < \infty$ almost surely, which verifies that (17) holds with ℓ in place of k.

Finally, we need to show, for any $\ell \neq k$ and for $X \sim Q$,

$$\mathbb{P}\left\{X_\ell \in A \mid X_{-\ell}\right\} \overset{\text{a.s.}}{=} (\widetilde{P}_* \circ \lambda_\ell)(A), \quad \text{for all } A \in \mathcal{B}(\mathcal{X}),$$

to verify that (18) holds with ℓ in place of k. Equivalently, we need to show that, for any $A \in \mathcal{B}(\mathcal{X})$ and $B \in \mathcal{B}(\mathcal{X}^\infty)$,

$$\mathbb{P}\left\{X_\ell \in A, X_{-\ell} \in B\right\} = \mathbb{E}\left[(\widetilde{P}_* \circ \lambda_\ell)(A) \cdot \mathbb{1}_{X_{-\ell} \in B}\right],$$

since $\widetilde{P}_* \circ \lambda_\ell$ is well-defined almost surely, but not necessarily surely, we should interpret the expected value as being computed only over the almost sure event that $\widetilde{P}_* \circ \lambda_\ell$ is well-defined. Since $\mathcal{B}(\mathcal{X}^\infty)$ is the product σ-algebra, it suffices to show that, for all $n \geq \ell$ and all $B \in \mathcal{B}(\mathcal{X}^{n-1})$,

$$\mathbb{P}\left\{X_\ell \in A, X_{[n]\setminus\ell} \in B\right\} = \mathbb{E}\left[(\widetilde{P}_* \circ \lambda_\ell)(A) \cdot \mathbb{1}_{X_{[n]\setminus\ell} \in B}\right].$$

We have

$$\begin{aligned}
\mathbb{P}\left\{X_\ell \in A, X_{[n]\setminus\ell} \in B\right\} &= \mathbb{E}\left[\mathbb{1}_{X_\ell\in A}\cdot \mathbb{1}_{X_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_k)\lambda_\ell(X_\ell)}{\lambda_k(X_k)\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[\mathbb{1}_{(X^{k\ell})_\ell\in A}\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k((X^{k\ell})_k)\lambda_\ell((X^{k\ell})_\ell)}{\lambda_k(X_k)\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[\mathbb{1}_{X_k\in A}\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_\ell)\lambda_\ell(X_k)}{\lambda_k(X_k)\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{X_k\in A}\cdot \frac{\lambda_\ell(X_k)}{\lambda_k(X_k)}\;\middle|\; X_{-k}\right]\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_\ell)}{\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[\mathbb{E}_{X'\sim\widetilde{P}_*\circ\lambda_k}\left[\mathbb{1}_{X'\in A}\cdot \frac{\lambda_\ell(X')}{\lambda_k(X')}\right]\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_\ell)}{\lambda_\ell(X_\ell)}\right],
\end{aligned}$$

where the second step holds by Proposition 5 since $X \sim Q$ where Q is λ-weighted exchangeable, while the fourth step holds since $(X^{k\ell})_{[n]\setminus\ell}$ does not depend on X_k by construction, and the final step holds by (18) with our choice of k. Next,

$$\begin{aligned}
\mathbb{E}_{X'\sim\widetilde{P}_*\circ\lambda_k}\left[\mathbb{1}_{X'\in A}\cdot \frac{\lambda_\ell(X')}{\lambda_k(X')}\right] &= \frac{\int_{\mathcal{X}} \mathbb{1}_{x\in A}\cdot \lambda_\ell(x)\,\mathsf{d}\widetilde{P}_*(x)}{\int_{\mathcal{X}} \lambda_k(x)\,\mathsf{d}\widetilde{P}_*(x)}\\
&= \frac{\int_{\mathcal{X}} \mathbb{1}_{x\in A}\cdot \lambda_\ell(x)\,\mathsf{d}\widetilde{P}_*(x)}{\int_{\mathcal{X}} \lambda_\ell(x)\,\mathsf{d}\widetilde{P}_*(x)}\cdot \frac{\int_{\mathcal{X}} \frac{\lambda_\ell(x)}{\lambda_k(x)}\cdot \lambda_k(x)\,\mathsf{d}\widetilde{P}_*(x)}{\int_{\mathcal{X}} \lambda_k(x)\,\mathsf{d}\widetilde{P}_*(x)}\\
&= (\widetilde{P}_*\circ\lambda_\ell)(A)\cdot \mathbb{E}_{X'\sim\widetilde{P}_*\circ\lambda_k}\left[\frac{\lambda_\ell(X')}{\lambda_k(X')}\right]
\end{aligned}$$

almost surely, where we use the fact that $\int_{\mathcal{X}} \lambda_\ell(x)\,\mathsf{d}\widetilde{P}_*(x) \in (0,\infty)$ and thus $\widetilde{P}_*\circ\lambda_\ell$ is well-defined, almost surely, by (17) applied with ℓ in place of k. Returning to the calculations above, then,

$$\begin{aligned}
\mathbb{P}\left\{X_\ell \in A, X_{[n]\setminus\ell} \in B\right\} &= \mathbb{E}\left[(\widetilde{P}_*\circ\lambda_\ell)(A)\cdot \mathbb{E}_{X'\sim\widetilde{P}_*\circ\lambda_k}\left[\frac{\lambda_\ell(X')}{\lambda_k(X')}\right]\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_\ell)}{\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[(\widetilde{P}_*\circ\lambda_\ell)(A)\cdot \mathbb{E}\left[\frac{\lambda_\ell(X_k)}{\lambda_k(X_k)}\;\middle|\; X_{-k}\right]\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_\ell)}{\lambda_\ell(X_\ell)}\right],
\end{aligned}$$

where the second step again applies (18) with our choice of k. Next, by Lemma 4, as in (19) we can construct an $\mathcal{F}_{\mathsf{tail}}$-measurable function $f : \mathcal{X}^\infty \to [0,1]$, such that $f(X) \stackrel{\text{a.s.}}{=} (\widetilde{P}_*\circ\lambda_\ell)(A)$. We then have

$$\begin{aligned}
\mathbb{P}\left\{X_\ell \in A, X_{[n]\setminus\ell} \in B\right\} &= \mathbb{E}\left[f(X)\cdot \mathbb{E}\left[\frac{\lambda_\ell(X_k)}{\lambda_k(X_k)}\;\middle|\; X_{-k}\right]\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_\ell)}{\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[f(X)\cdot \mathbb{1}_{(X^{k\ell})_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_\ell)\lambda_\ell(X_k)}{\lambda_k(X_k)\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[f(X^{k\ell})\cdot \mathbb{1}_{X_{[n]\setminus\ell}\in B}\cdot \frac{\lambda_k(X_k)\lambda_\ell(X_\ell)}{\lambda_k(X_k)\lambda_\ell(X_\ell)}\right]\\
&= \mathbb{E}\left[f(X)\cdot \mathbb{1}_{X_{[n]\setminus\ell}\in B}\right]\\
&= \mathbb{E}\left[(\widetilde{P}_*\circ\lambda_\ell)(A)\cdot \mathbb{1}_{X_{[n]\setminus\ell}\in B}\right],
\end{aligned}$$

where the third step holds by Proposition 5, and the second and fourth steps hold since f is $\mathcal{F}_{\mathsf{tail}}$-measurable (and thus also measurable with respect to $\sigma(X_{-k}) \supseteq \mathcal{F}_{\mathsf{tail}}$, for the second step where we apply the tower law, and with respect to $\mathcal{E}_\infty \supseteq \mathcal{F}_{\mathsf{tail}}$ so that $f(X) = f(X^{k\ell})$, for the fourth step). This completes the proof that (18) holds with ℓ in place of k.

A.9 Proof of Lemma 6

First, we verify that, for sufficiently large n, it must hold that $\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(x')/\lambda_k(x')} > 0$ for all $x' \in \mathcal{X}$. Indeed, we have

$$\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(x')/\lambda_k(x')} \geq \sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_j(x)/\lambda_k(x)}$$

$$\geq \inf_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x) \cdot \sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}.$$

The first term is positive by our choice of k, while the second term is positive for sufficiently large n by our assumption that the sufficient condition (4) holds.

Next, fixing any sufficiently large $n > \ell$, we calculate

$$\begin{aligned}
&\mathbb{P}\left\{X_k \in A, X_{[\ell]\setminus k} \in B\right\} = \mathbb{E}\left[\mathbb{1}_{X_k\in A} \cdot \mathbb{1}_{X_{[\ell]\setminus k}\in B}\right]\\
&= \mathbb{E}\left[\mathbb{1}_{X_k\in A} \cdot \mathbb{1}_{X_{[\ell]\setminus k}\in B} \cdot \sum_{i=\ell+1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)}}{\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}}\right]\\
&= \sum_{i=\ell+1}^{n} \mathbb{E}\left[\mathbb{1}_{X_k\in A} \cdot \mathbb{1}_{X_{[\ell]\setminus k}\in B} \cdot \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)}}{\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}} \cdot \frac{\lambda_i(X_i)\lambda_k(X_k)}{\lambda_i(X_i)\lambda_k(X_k)}\right]\\
&= \sum_{i=\ell+1}^{n} \mathbb{E}\left[\mathbb{1}_{(X^{ik})_k\in A} \cdot \mathbb{1}_{(X^{ik})_{[\ell]\setminus k}\in B} \cdot \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i((X^{ik})_i)/\lambda_k((X^{ik})_i)}}{\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j((X^{ik})_j)/\lambda_k((X^{ik})_j)}} \cdot \frac{\lambda_i((X^{ik})_i)\lambda_k((X^{ik})_k)}{\lambda_i(X_i)\lambda_k(X_k)}\right],
\end{aligned}$$

where in the last step we apply Proposition 5. As $(X^{ik})_{[\ell]\setminus k} = X_{[\ell]\setminus k}$ for any $i > \ell$ by definition, we can simplify the above to

$$\begin{aligned}
&\mathbb{P}\left\{X_k \in A, X_{[\ell]\setminus k} \in B\right\}\\
&= \mathbb{E}\left[\mathbb{1}_{X_{[\ell]\setminus k}\in B} \cdot \sum_{i=\ell+1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_k)/\lambda_k(X_k)} \cdot \mathbb{1}_{X_i\in A}}{\left(\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}\right) + \Delta_i} \cdot \frac{\lambda_i(X_k)\lambda_k(X_i)}{\lambda_i(X_i)\lambda_k(X_k)}\right]\\
&= \mathbb{E}\left[\mathbb{1}_{X_{[\ell]\setminus k}\in B} \cdot \sum_{i=\ell+1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)} \cdot \mathbb{1}_{X_i\in A}}{\left(\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}\right) + \Delta_i}\right],
\end{aligned}$$

where

$$\Delta_i = \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_k)/\lambda_k(X_k)} - \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)}.$$

Recall the weighted empirical distribution $\widetilde{P}_n$ defined in (8). We can calculate

$$(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) = \frac{\sum_{i=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)} \cdot \mathbb{1}_{X_i\in A}}{\sum_{j=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}},$$

and so

$$\left|(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) - \sum_{i=\ell+1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)} \cdot \mathbb{1}_{X_i\in A}}{\left(\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}\right) + \Delta_i}\right|$$

$$= \left|\sum_{i=1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)} \cdot \mathbb{1}_{X_i\in A}}{\sum_{j=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}} - \sum_{i=\ell+1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)} \cdot \mathbb{1}_{X_i\in A}}{\left(\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)}\right) + \Delta_i}\right|$$

$$\leq \frac{\ell+1}{\left(\frac{\inf_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x)} \cdot \sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)} - 1\right)_+},$$

where the last step follows from some direct calculations, making use of the fact that

$$|\Delta_i| \leq \sup_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x)$$

holds surely for every i. Therefore,

$$\left|\mathbb{P}\left\{X_k \in A, X_{[\ell]\setminus k} \in B\right\} - \mathbb{E}\left[(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) \cdot \mathbb{1}_{X_{[\ell]\setminus k}\in B}\right]\right|$$

$$= \left|\mathbb{E}\left[\mathbb{1}_{X_{[\ell]\setminus k}\in B} \cdot \left(\sum_{i=\ell+1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)} \cdot \mathbb{1}_{X_i\in A}}{\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)} + \Delta_i} - (\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A)\right)\right]\right|$$

$$\leq \mathbb{E}\left[\mathbb{1}_{X_{[\ell]\setminus k}\in B} \cdot \left|\sum_{i=\ell+1}^{n} \frac{\frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\lambda_i(X_i)/\lambda_k(X_i)} \cdot \mathbb{1}_{X_i\in A}}{\sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_j(x)/\lambda_*(x)}{\lambda_j(X_j)/\lambda_k(X_j)} + \Delta_i} - (\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A)\right|\right]$$

$$\leq \frac{\ell+1}{\left(\frac{\inf_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_k(x)/\lambda_*(x)} \cdot \sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_*(x)/\lambda_j(x)}{\sup_{x\in\mathcal{X}} \lambda_*(x)/\lambda_j(x)} - 1\right)_+}.$$

By our choice of k, we know that the first factor in the denominator above is positive, and by (4),

$$\lim_{n\to\infty} \sum_{j=\ell+1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_*(x)/\lambda_j(x)}{\sup_{x\in\mathcal{X}} \lambda_*(x)/\lambda_j(x)} = \infty.$$

Therefore,

$$\lim_{n\to\infty} \left|\mathbb{P}\left\{X_k \in A, X_{[\ell]\setminus k} \in B\right\} - \mathbb{E}\left[(\widetilde{P}_n \circ (\lambda_k/\lambda_*))(A) \cdot \mathbb{1}_{X_{[\ell]\setminus k}\in B}\right]\right| = 0,$$

which completes the proof.

B Proofs supporting auxiliary results

B.1 Proof of Proposition 3

Since $|\mathcal{X}| \geq 3$, we can choose a partition $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1 \cup \mathcal{X}_2$ for some nonempty $\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2 \in \mathcal{B}(\mathcal{X})$. Now define

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

First we examine the sufficient condition (4). Let $\lambda_* : \mathcal{X} \to \mathbb{R}_+$ be any measurable function. Fix any $x_\ell \in \mathcal{X}_\ell$ for each $\ell = 0, 1, 2$. Then we have

$$\sum_{i=1}^{\infty} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)}{\sup_{x\in\mathcal{X}} \lambda_i(x)/\lambda_*(x)} \leq \sum_{i=1}^{\infty} \frac{\min_{\ell=0,1,2} \lambda_i(x_\ell)/\lambda_*(x_\ell)}{\max_{\ell=0,1,2} \lambda_i(x_\ell)/\lambda_*(x_\ell)}$$

$$\leq \frac{\max_{\ell=0,1,2} \lambda_*(x_\ell)}{\min_{\ell=0,1,2} \lambda_*(x_\ell)} \cdot \sum_{i=1}^{\infty} \frac{\min_{\ell=0,1,2} \lambda_i(x_\ell)}{\max_{\ell=0,1,2} \lambda_i(x_\ell)} = \frac{\max_{\ell=0,1,2} \lambda_*(x_\ell)}{\min_{\ell=0,1,2} \lambda_*(x_\ell)} \cdot \sum_{i=1}^{\infty} e^{-i} < \infty,$$

and therefore the sufficient condition (4) is not satisfied. Now we check the necessary condition (3). By construction, we can see that $\mathcal{M}_{\mathcal{X}}(\lambda) = \{P \in \mathcal{M}(\mathcal{X}) : 0 < P(\mathcal{X}) < \infty\}$. Fix any $A, A^c \in \mathcal{B}(\mathcal{X})$ with $P(A), P(A^c) > 0$. Since $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1 \cup \mathcal{X}_2$ is a partition, we can choose some $\ell_A \in \{0, 1, 2\}$ such that $P(A \cap \mathcal{X}_{\ell_A}) > 0$, and similarly some $\ell_{A^c} \in \{0, 1, 2\}$ such that $P(A^c \cap \mathcal{X}_{\ell_{A^c}}) > 0$.

Now let $\ell \in \{0, 1, 2\}\backslash\{\ell_A, \ell_{A^c}\}$, so that for any i with $\mathrm{mod}(i, 3) = \ell$, we have $\lambda_i(x) = 1$ for $x \in \mathcal{X}_{\ell_A}$ and also for $x \in \mathcal{X}_{\ell_{A^c}}$. Then for any such i, we calculate

$$(P \circ \lambda_i)(A) = \frac{\int_A \lambda_i(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P(x)} \geq \frac{\int_{A\cap\mathcal{X}_{\ell_A}} \lambda_i(x)\,\mathsf{d}P(x)}{\int_{\mathcal{X}} \lambda_i(x)\,\mathsf{d}P(x)} \geq \frac{\int_{A\cap\mathcal{X}_{\ell_A}} \mathsf{d}P(x)}{\int_{\mathcal{X}} \mathsf{d}P(x)} = \frac{P(A \cap \mathcal{X}_{\ell_A})}{P(\mathcal{X})} > 0,$$

where the third step holds since $\lambda_i(x) = 1$ for $x \in \mathcal{X}_{\ell_A}$ by the choice of i, and $\lambda_i(x) \leq 1$ for all x. Similarly, for any i with $\mathrm{mod}(i, 3) = \ell$,

$$(P \circ \lambda_i)(A^c) \geq \frac{P(A^c \cap \mathcal{X}_{\ell_{A^c}})}{P(\mathcal{X})} > 0.$$

Therefore, we have

$$\sum_{i=1}^{\infty} \min\{(P \circ \lambda_i)(A), (P \circ \lambda_i)(A^c\} \geq \sum_{i=1}^{\infty} \mathbb{1}_{\mathrm{mod}(i,3)=\ell} \cdot \min\left\{\frac{P(A \cap \mathcal{X}_{\ell_A)}}{P(\mathcal{X})}, \frac{P(A^c \cap \mathcal{X}_{\ell_{A^c})}}{P(\mathcal{X})}\right\} = \infty,$$

proving that the condition (3) is satisfied.

B.2 Proof of Theorem 8

Once we apply the results of Theorems 4 and 5, we see that we only need to verify that

$$\Lambda_{\mathsf{dF}} \supseteq \{\lambda \text{ satisfying (3)}\},$$

i.e., that the necessary condition (3) is in fact sufficient for proving $\lambda \in \Lambda_{\mathsf{dF}}$, for the finite case.

B.2.1 An equivalent characterization via a graph

First, we give an equivalent characterization of the necessary condition for this finite case. For any nonempty subset $S \subseteq \mathcal{X}$, define an undirected graph on vertices S, denoted by $G_S = (S, E_S)$, where the set of edges is given by

$$E_S = \left\{(x_0, x_1) \in S \times S : \sum_{i=1}^{\infty} \frac{\min\{\lambda_i(x_0), \lambda_i(x_1)\}}{\max_{x\in S} \lambda_i(x)} = \infty\right\}. \tag{24}$$

Lemma 7. *If $|\mathcal{X}| < \infty$ then, for any $\lambda \in \Lambda^\infty$, λ satisfies the condition (3) if and only if, for every nonempty $S \subseteq \mathcal{X}$, G_S is a connected graph.*

From this point onward, we will choose a *directed* and *rooted* spanning tree $E^*_S \subseteq E_S$, and will work with the rooted directed acyclic graph (DAG) $G^*_S = (S, E^*_S)$.

B.2.2 Writing Q via a mixture of supports

Next for any $x \in \mathcal{X}^\infty$ define its support

$$S(x) = \left\{ x \in \mathcal{X} : \sum_{i=1}^{\infty} \mathbb{1}_{x_i = x} > 0 \right\},$$

and for each nonempty $S \subseteq \mathcal{X}$ define

$$p_S = \mathbb{P}_Q \left\{ S(X) = S \right\},$$

the probability that $S(X) = S$ when we sample $X \sim Q$. We can write Q as a mixture,

$$Q = \sum_{S : p_S > 0} p_S \cdot Q_S,$$

where Q_S is the distribution Q conditional on the event $S(X) = S$.

For any S with $p_S > 0$, as the event $\{S(X) = S\}$ is in $\mathcal{E}_\infty$, Q_S is also λ-weighted exchangeable (this can be verified exactly as in the proof of Theorem 4—specifically, the proof that $\Lambda_{\mathsf{dF}} \subseteq \Lambda_{\mathsf{01}}$). In what follows, we will show that, for each S, Q_S has a (weighted) de Finetti representation. As Q is ultimately a mixture of distributions Q_S over subsets S, this will be sufficient: if each Q_S can be written as a mixture of λ-weighted i.i.d. distributions, then the same is also true for Q.

B.2.3 Defining the probability measure

From this point on, we will work with Q_S for a specific subset $S \subseteq \mathcal{X}$, and all probabilities and expectations should be interpreted as taken with respect to Q_S, unless otherwise specified. Next we need another lemma.

Lemma 8. *Under the notation and assumptions above, for any $(x, x') \in E^*_S$,*

$$\sum_{i \geq 1} \frac{\min\{\lambda_i(x), \lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i \in \{x, x'\}} = \infty$$

almost surely under Q_S.

Next, for each $(x, x') \in E^*_S$, define

$$\widetilde{P}_n(x; x') = \frac{\sum_{i=1}^n \frac{\min\{\lambda_i(x), \lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i = x}}{\sum_{i=1}^n \frac{\min\{\lambda_i(x), \lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i \in \{x, x'\}}}.$$

By Lemma 8, the denominator tends to infinity almost surely as $n \to \infty$. In particular, it holds almost surely that the denominator is positive for sufficiently large n. In other words, for sufficiently large n, $\widetilde{P}_n(x; x')$ is well-defined.

Next we need to verify that this quantity converges. As in the proof of Theorem 6, the following lemma will be proved by extracting an exchangeable subsequence.

Lemma 9. *Under the notation and assumptions above, for any $(x, x') \in E^*_S$, there is a $\mathcal{F}_{\mathrm{tail}}$-measurable random variable $\widetilde{P}(x; x') \in (0, 1)$ such that*

$$\widetilde{P}(x; x') = \lim_{n \to \infty} \widetilde{P}_n(x; x')$$

almost surely under Q_S.

Next, we will convert these values to a probability distribution $\{\widetilde{P}(x)\}_{x\in\mathcal{X}}$, which again is $\mathcal{F}_{\text{tail}}$-measurable. Let x_* be the root of the rooted DAG G^*_S. For $x \in S$, consider the unique path in G^*_S, from the root x_* to x: this path can be expressed as a finite sequence $y_0(x) = x_*, y_1(x), \ldots, y_{k(x)}(x) = x$ such that $(y_{i-1}(x), y_i(x)) \in E^*_S$ for each i. Then define

$$\widetilde{P}'(x) = \prod_{i=1}^{k(x)} \frac{1 - \widetilde{P}(y_{i-1}(x); y_i(x))}{\widetilde{P}(y_{i-1}(x); y_i(x))},$$

where for the case $x = x_*$ (and so $k(x) = k(x_*) = 0$), the empty product should be interpreted as 1, i.e., $\widetilde{P}'(x_*) = 1$. Note that, since $\widetilde{P}(x; x') \in (0, 1)$ for all $(x, x') \in E^*_S$, therefore the above product is positive and finite for all $x \in S$. Then define

$$\widetilde{P}(x) = \begin{cases} \frac{\widetilde{P}'(x)}{\sum_{x'\in S} \widetilde{P}'(x')} & x \in S, \\ 0 & x \notin S. \end{cases}$$

Now we will see the motivation for this construction: for any $(x, x') \in E^*_S$, we now verify that

$$\frac{\widetilde{P}(x)}{\widetilde{P}(x) + \widetilde{P}(x')} = \widetilde{P}(x; x'). \tag{25}$$

In other words, $\widetilde{P}(x; x')$ defines the conditional probability $\mathbb{P}_{\widetilde{P}}\{X = x \mid X \in \{x, x'\}\}$, for each edge $(x, x') \in E^*_S$ in the rooted DAG. To see why this holds, first note that since G^*_S is a rooted DAG, the path from the root to x, and the path from the root to x', are each unique. Recall that

$$y_0(x) = x_*, y_1(x), \ldots, y_{k(x)}(x) = x$$

is the path from x_* to x. Then the unique path from the root to x' must therefore be equal to

$$y_0(x) = x_*, y_1(x), \ldots, y_{k(x)}(x) = x, y_{k(x)+1}(x) = x'.$$

In other words, we have $k(x') = k(x) + 1$, and

$$y_i(x') = \begin{cases} y_i(x) & i = 1, \ldots, k(x), \\ x' & i = k(x) + 1 = k(x'). \end{cases}$$

We can therefore calculate

$$\begin{aligned} \widetilde{P}'(x') = \prod_{i=1}^{k(x')} \frac{1 - \widetilde{P}(y_{i-1}(x'); y_i(x'))}{\widetilde{P}(y_{i-1}(x'); y_i(x'))} &= \left[\prod_{i=1}^{k(x')-1} \frac{1 - \widetilde{P}(y_{i-1}(x'); y_i(x'))}{\widetilde{P}(y_{i-1}(x'); y_i(x'))}\right] \cdot \frac{1 - \widetilde{P}(y_{k(x')-1}(x'); y_{k(x')}(x'))}{\widetilde{P}(y_{k(x')-1}(x'); y_{k(x')}(x'))} \\ &= \left[\prod_{i=1}^{k(x)} \frac{1 - \widetilde{P}(y_{i-1}(x); y_i(x))}{\widetilde{P}(y_{i-1}(x); y_i(x))}\right] \cdot \frac{1 - \widetilde{P}(x; x')}{\widetilde{P}(x; x')} = \widetilde{P}'(x) \cdot \frac{1 - \widetilde{P}(x; x')}{\widetilde{P}(x; x')}, \end{aligned}$$

which proves the desired equality as we then have

$$\frac{\widetilde{P}(x)}{\widetilde{P}(x) + \widetilde{P}(x')} = \frac{\widetilde{P}'(x)}{\widetilde{P}'(x) + \widetilde{P}'(x')} = \frac{\widetilde{P}'(x)}{\widetilde{P}'(x) + \widetilde{P}'(x) \cdot \frac{1-\widetilde{P}(x;x')}{\widetilde{P}(x;x')}} = \widetilde{P}(x; x').$$

B.2.4 Representing Q_S as a mixture distribution

Finally, we need to check that it holds that Q_S is equal to the mixture defined by $\widetilde{P} \circ \lambda$, for the random measure $\widetilde{P}$. In fact, as in the proof of Theorem 6, it is sufficient to verify that

$$\mathbb{P}\{X_k = x \mid X_{-k}\} \stackrel{\text{a.s.}}{=} (\widetilde{P} \circ \lambda_k)(x)$$

for all k and all $x \in S$ (this claim is analogous to the claim (18) in Lemma 5; the argument that appears after Lemma 5, in the proof of Theorem 6, explains why verifying this claim is sufficient). Since G_S is a connected graph, it is sufficient to check that

$$\frac{\mathbb{P}\{X_k = x \mid X_{-k}\}}{\mathbb{P}\{X_k = x' \mid X_{-k}\}} \stackrel{\text{a.s.}}{=} \frac{\widetilde{P}(x)\lambda_k(x)}{\widetilde{P}(x')\lambda_k(x')},$$

for all k and for all $(x, x') \in E_S^*$, or equivalently,

$$\frac{\mathbb{P}\{X_k = x \mid X_{-k}\}}{\mathbb{P}\{X_k = x' \mid X_{-k}\}} \stackrel{\text{a.s.}}{\geq} \frac{\widetilde{P}(x)\lambda_k(x)}{\widetilde{P}(x')\lambda_k(x')}, \quad \text{and} \quad \frac{\mathbb{P}\{X_k = x \mid X_{-k}\}}{\mathbb{P}\{X_k = x' \mid X_{-k}\}} \stackrel{\text{a.s.}}{\leq} \frac{\widetilde{P}(x)\lambda_k(x)}{\widetilde{P}(x')\lambda_k(x')}.$$

Since the proofs of these two inequalities are essentially identical, we will prove only the first. By the property (25) calculated above, equivalently, we need to check that

$$\mathbb{P}\{X_k = x \mid X_{-k}\} \stackrel{\text{a.s.}}{\geq} \frac{\widetilde{P}(x;x')}{1-\widetilde{P}(x;x')} \cdot \frac{\lambda_k(x)}{\lambda_k(x')} \cdot \mathbb{P}\left\{X_k = x' \mid X_{-k}\right\},$$

for all k and for all $(x, x') \in E_S^*$. (Note that $\widetilde{P}(x;x')$ is $\mathcal{F}_{\text{tail}}$-measurable, and is therefore measurable with respect to $\sigma(X_{-k})$.) By definition of conditional probability, this is equivalent to checking that

$$\mathbb{P}\{X_k = x, X_{-k} \in A\} \geq \frac{\lambda_k(x)}{\lambda_k(x')} \cdot \mathbb{E}\left[\frac{\widetilde{P}(x;x')}{1-\widetilde{P}(x;x')} \cdot \mathbb{1}_{X_k=x'} \mathbb{1}_{X_{-k}\in A}\right]$$

for all $A \in \mathcal{B}(\mathcal{X}^\infty)$. By construction of the product σ-algebra, it is sufficient to check that, for every $\ell \geq k$ and any $A \in \mathcal{B}(\mathcal{X}^{\ell-1})$,

$$\mathbb{P}\left\{X_k = x, X_{[\ell]\setminus k} \in A\right\} \geq \frac{\lambda_k(x)}{\lambda_k(x')} \cdot \mathbb{E}\left[\frac{\widetilde{P}(x;x')}{1-\widetilde{P}(x;x')} \cdot \mathbb{1}_{X_k=x'} \mathbb{1}_{X_{[\ell]\setminus k}\in A}\right]. \tag{26}$$

We can derive that

$$\frac{\widetilde{P}(x;x')}{1-\widetilde{P}(x;x')} = \lim_{n\to\infty} \frac{\sum_{i=1}^n \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x}}{\sum_{i=1}^n \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}} = \lim_{n\to\infty} \frac{\sum_{i=\ell+1}^n \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x}}{1+\sum_{i=\ell+1}^n \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}},$$

almost surely, where the first step holds from Lemma 9, while the second step is true since, by Lemmas 8 and 9, the sums in the numerator and denominator both diverge. By Fatou's lemma,

$$\mathbb{E}\left[\frac{\widetilde{P}(x;x')}{1-\widetilde{P}(x;x')} \cdot \mathbb{1}_{X_k=x'} \mathbb{1}_{X_{[\ell]\setminus k}\in A}\right]$$
$$\leq \liminf_{n\to\infty} \mathbb{E}\left[\frac{\sum_{i=\ell+1}^n \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x}}{1+\sum_{i=\ell+1}^n \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}} \cdot \mathbb{1}_{X_k=x'} \mathbb{1}_{X_{[\ell]\setminus k}\in A}\right].$$

For each n, we calculate

$$\mathbb{E}\left[\frac{\sum_{i=\ell+1}^{n}\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x}}{1+\sum_{i=\ell+1}^{n}\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}\cdot\mathbb{1}_{X_k=x'}\mathbb{1}_{X_{[\ell]\setminus k}\in A}\right]$$
$$=\sum_{i=\ell+1}^{n}\mathbb{E}\left[\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}}\cdot\mathbb{1}_{X_k=x'}\mathbb{1}_{X_{[\ell]\setminus k}\in A}\cdot\frac{\lambda_i(X_i)\lambda_k(X_k)}{\lambda_i(X_i)\lambda_k(X_k)}\right]$$
$$=\sum_{i=\ell+1}^{n}\mathbb{E}\left[\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_k)}\cdot\mathbb{1}_{X_k=x}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}((X^{ik})_{i'})}\cdot\mathbb{1}_{(X^{ik})_{i'}=x'}}\cdot\mathbb{1}_{X_i=x'}\mathbb{1}_{(X^{ik})_{[\ell]\setminus k}\in A}\frac{\lambda_i(X_k)\lambda_k(X_i)}{\lambda_i(X_i)\lambda_k(X_k)}\right]$$
$$=\sum_{i=\ell+1}^{n}\mathbb{E}\left[\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_k=x}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}((X^{ik})_{i'})}\cdot\mathbb{1}_{(X^{ik})_{i'}=x'}}\cdot\mathbb{1}_{X_i=x'}\mathbb{1}_{X_{[\ell]\setminus k}\in A}\right]\cdot\frac{\lambda_k(x')}{\lambda_k(x)}$$
$$=\sum_{i=\ell+1}^{n}\mathbb{E}\left[\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}+\Delta_i}\cdot\mathbb{1}_{X_k=x}\mathbb{1}_{X_{[\ell]\setminus k}\in A}\right]\cdot\frac{\lambda_k(x')}{\lambda_k(x)},$$

where the second step holds by Proposition 5, and where

$$\Delta_i=\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(x')}\cdot(\mathbb{1}_{X_k=x'}-\mathbb{1}_{X_i=x'})\in[-1,1].$$

Therefore,

$$\left|\frac{\lambda_k(x)}{\lambda_k(x')}\mathbb{E}\left[\frac{\sum_{i=\ell+1}^{n}\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x}}{1+\sum_{i=\ell+1}^{n}\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}\cdot\mathbb{1}_{X_k=x'}\mathbb{1}_{X_{[\ell]\setminus k}\in A}\right]-\mathbb{P}\left\{X_k=x,X_{[\ell]\setminus k}\in A\right\}\right|$$
$$=\left|\sum_{i=\ell+1}^{n}\mathbb{E}\left[\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}+\Delta_i}\cdot\mathbb{1}_{X_k=x}\mathbb{1}_{X_{[\ell]\setminus k}\in A}\right]-\mathbb{P}\left\{X_k=x,X_{[\ell]\setminus k}\in A\right\}\right|$$
$$\le\mathbb{E}\left[\left|\sum_{i=\ell+1}^{n}\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}+\Delta_i}\cdot\mathbb{1}_{X_k=x}\mathbb{1}_{X_{[\ell]\setminus k}\in A}-\mathbb{1}_{X_k=x,X_{[\ell]\setminus k}\in A}\right|\right]$$
$$\le\mathbb{E}\left[\left|\sum_{i=\ell+1}^{n}\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}+\Delta_i}-1\right|\right].$$

Now we bound the sum inside the expectation. If $\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}=0$ then the sum is zero. If not, then since $\Delta_i\in[-1,1]$ by construction, the sum is upper bounded as

$$\sum_{i=\ell+1}^{n}\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}+\Delta_i}$$
$$\le\sum_{i=\ell+1}^{n}\frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)}\cdot\mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n}\frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})}\cdot\mathbb{1}_{X_{i'}=x'}+(-1)}=1,$$

and is lower bounded as

$$\sum_{i=\ell+1}^{n} \frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n} \frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})} \cdot \mathbb{1}_{X_{i'}=x'} + \Delta_i} \geq \sum_{i=\ell+1}^{n} \frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n} \frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})} \cdot \mathbb{1}_{X_{i'}=x'} + 1}$$
$$= \frac{\sum_{i=\ell+1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}}{2+\sum_{i'=\ell+1}^{n} \frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})} \cdot \mathbb{1}_{X_{i'}=x'}} = 1 - \frac{2}{2+\sum_{i'=\ell+1}^{n} \frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})} \cdot \mathbb{1}_{X_{i'}=x'}}.$$

In all cases, then, we have

$$\left| \sum_{i=\ell+1}^{n} \frac{\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}}{1+\sum_{i'=\ell+1}^{n} \frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})} \cdot \mathbb{1}_{X_{i'}=x'} + \Delta_i} - 1 \right| \leq \frac{2}{2+\sum_{i'=\ell+1}^{n} \frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})} \cdot \mathbb{1}_{X_{i'}=x'}}.$$

Therefore,

$$\left| \frac{\lambda_k(x)}{\lambda_k(x')} \lim_{n\to\infty} \mathbb{E}\left[\frac{\sum_{i=\ell+1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x}}{1+\sum_{i=\ell+1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}} \right] - \mathbb{P}\left\{X_k = x, X_{[\ell]\backslash k} \in A\right\} \right|$$
$$= \lim_{n\to\infty} \left| \frac{\lambda_k(x)}{\lambda_k(x')} \mathbb{E}\left[\frac{\sum_{i=\ell+1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x}}{1+\sum_{i=\ell+1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x'}} \cdot \mathbb{1}_{X_k=x'} \mathbb{1}_{X_{[\ell]\backslash k}\in A} \right] - \mathbb{P}\left\{X_k = x, X_{[\ell]\backslash k} \in A\right\} \right|$$
$$\leq \lim_{n\to\infty} \mathbb{E}\left[\frac{2}{2+\sum_{i'=\ell+1}^{n} \frac{\min\{\lambda_{i'}(x),\lambda_{i'}(x')\}}{\lambda_{i'}(X_{i'})} \cdot \mathbb{1}_{X_{i'}=x'}} \right]$$
$$= 0,$$

where the last step holds since the denominator tends to infinity almost surely, by Lemmas 8 and 9. This verifies (26), as desired.

B.3 Proof of Lemma 7

First suppose the graph G_S is disconnected for some nonempty $S \subseteq \mathcal{X}$. Then we can find a subset $A \subseteq S$ with $A, S\backslash A$ both nonempty, such that $(x_0, x_1) \notin E_S$ for all $x_0 \in A, x_1 \in S\backslash A$. Let P be the uniform distribution on S. We will now show that the necessary condition (3) fails for this P and this A. For any i we have

$$\begin{aligned}
\min\{(P\circ\lambda_i)(A), (P\circ\lambda_i)(S\backslash A)\} &= \min\left\{ \sum_{x_0\in A} (P\circ\lambda_i)(\{x_0\}), \sum_{x_1\in S\backslash A} (P\circ\lambda_i)(\{x_1\}) \right\} \\
&= \min\left\{ \sum_{x_0\in A} \frac{\lambda_i(x_0)}{\sum_{x\in S}\lambda_i(x)}, \sum_{x_1\in S\backslash A} \frac{\lambda_i(x_1)}{\sum_{x\in S}\lambda_i(x)} \right\} \\
&\leq \min\left\{ \frac{|A|\cdot\max_{x_0\in A}\lambda_i(x_0)}{\max_{x\in S}\lambda_i(x)}, \frac{|S\backslash A|\cdot\max_{x_1\in S\backslash A}\lambda_i(x_1)}{\max_{x\in S}\lambda_i(x)} \right\} \\
&\leq |S| \min\left\{ \max_{x_0\in A} \frac{\lambda_i(x_0)}{\max_{x\in S}\lambda_i(x)}, \max_{x_1\in S\backslash A} \frac{\lambda_i(x_1)}{\max_{x\in S}\lambda_i(x)} \right\} \\
&= |S| \max_{x_0\in A, x_1\in S\backslash A} \min\left\{ \frac{\lambda_i(x_0)}{\max_{x\in S}\lambda_i(x)}, \frac{\lambda_i(x_1)}{\max_{x\in S}\lambda_i(x)} \right\}
\end{aligned}$$

$$\le |S| \sum_{x_0\in A, x_1\in S\setminus A} \min\left\{\frac{\lambda_i(x_0)}{\max_{x\in S}\lambda_i(x)}, \frac{\lambda_i(x_1)}{\max_{x\in S}\lambda_i(x)}\right\}.$$

Therefore,

$$\sum_{i=1}^{\infty} \min\{(P\circ\lambda_i)(A), (P\circ\lambda_i)(S\setminus A)\} \le |S| \sum_{x_0\in A, x_1\in S\setminus A} \sum_{i\ge 1} \min\left\{\frac{\lambda_i(x_0)}{\max_{x\in S}\lambda_i(x)}, \frac{\lambda_i(x_1)}{\max_{x\in S}\lambda_i(x)}\right\} < \infty,$$

where the last step holds since, for all $x_0 \in A$ and $x_1 \in S\setminus A$, we have $(x_0, x_1) \notin E_S$ and therefore $\sum_{i=1}^{\infty} \frac{\min\{\lambda_i(x_0),\lambda_i(x_1)\}}{\max_{x\in S}\lambda_i(x)} < \infty$.

Next suppose that G_S is connected for all S. We will show that the necessary condition (3) must hold. Fix any $A \subseteq \mathcal{X}$ and any $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$ with $P(A), P(A^c) > 0$. Note that, since $\lambda(x) > 0$ for all x, we must have $P(\mathcal{X}) < \infty$ in order to have $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$. Now let $S = \{x \in \mathcal{X} : P(x) > 0\}$ be the support of P. Since G_S is a connected graph we must have some $x_0^* \in A$, $x_1^* \in S\setminus A$ with $\sum_{i=1}^{\infty} \frac{\min\{\lambda_i(x_0^*),\lambda_i(x_1^*)\}}{\max_{x\in S}\lambda_i(x)} = \infty$. We now calculate

$$\begin{aligned}
&\min\{(P\circ\lambda_i)(A), (P\circ\lambda_i)(S\setminus A)\}\\
&= \min\left\{\sum_{x_0\in A}(P\circ\lambda_i)(\{x_0\}), \sum_{x_1\in S\setminus A}(P\circ\lambda_i)(\{x_1\})\right\}\\
&= \min\left\{\sum_{x_0\in A}\frac{P(\{x_0\})\cdot\lambda_i(x_0)}{\sum_{x\in S}P(\{x\})\cdot\lambda_i(x)}, \sum_{x_1\in S\setminus A}\frac{P(\{x_1\})\cdot\lambda_i(x_1)}{\sum_{x\in S}P(\{x_1\})\cdot\lambda_i(x)}\right\}\\
&\ge \min\left\{\sum_{x_0\in A}\frac{P(\{x_0\})\cdot\lambda_i(x_0)}{\sum_{x\in S}P(\{x\})\cdot\max_{x\in S}\lambda_i(x)}, \sum_{x_1\in S\setminus A}\frac{P(\{x_1\})\cdot\lambda_i(x_1)}{\sum_{x\in S}P(\{x\})\cdot\max_{x\in S}\lambda_i(x)}\right\}\\
&\ge \min\left\{\frac{P(\{x_0^*\})\cdot\lambda_i(x_0^*)}{\sum_{x\in S}P(\{x\})\cdot\max_{x\in S}\lambda_i(x)}, \frac{P(\{x_1^*\})\cdot\lambda_i(x_1^*)}{\sum_{x\in S}P(\{x\})\cdot\max_{x\in S}\lambda_i(x)}\right\}\\
&\ge \min\left\{\frac{P(\{x_0^*\})}{P(\mathcal{X})}, \frac{P(\{x_1^*\})}{P(\mathcal{X})},\right\}\cdot\min\left\{\frac{\lambda_i(x_0^*)}{\max_{x\in S}\lambda_i(x)}, \frac{\lambda_i(x_1^*)}{\max_{x\in S}\lambda_i(x)}\right\},
\end{aligned}$$

and therefore,

$$\begin{aligned}
&\sum_{i=1}^{\infty}\min\{(P\circ\lambda_i)(A), (P\circ\lambda_i)(S\setminus A)\}\\
&\qquad\ge \min\left\{\frac{P(\{x_0^*\})}{P(\mathcal{X})}, \frac{P(\{x_1^*\})}{P(\mathcal{X})},\right\}\cdot\sum_{i=1}^{\infty}\min\left\{\frac{\lambda_i(x_0^*)}{\max_{x\in S}\lambda_i(x)}, \frac{\lambda_i(x_1^*)}{\max_{x\in S}\lambda_i(x)}\right\} = \infty,
\end{aligned}$$

where the last step holds because the first term is positive (as $x_0^*, x_1^* \in S$, and S is the support of P), while the second term is infinite by our choice of x_0^*, x_1^*. This completes the proof.

B.4 Proof of Lemma 8

Let $\mathcal{I} = \{i \ge 1 : \lambda_i(x) < \lambda_i(x')\}$. Then, since $(x, x') \in E_S^*$,

$$\infty = \sum_{i=1}^{\infty}\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\max_{x''\in S}\lambda_i(x'')} = \sum_{i\in\mathcal{I}}\frac{\lambda_i(x)}{\max_{x''\in S}\lambda_i(x'')} + \sum_{i\in\mathcal{I}^c}\frac{\lambda_i(x')}{\max_{x''\in S}\lambda_i(x'')},$$

where the last step holds by definition of $\mathcal{I}$. Then we must have either

$$\sum_{i\in\mathcal{I}} \frac{\lambda_i(x)}{\max_{x''\in S}\lambda_i(x'')} = \infty \quad \text{or} \quad \sum_{i\in\mathcal{I}^c} \frac{\lambda_i(x')}{\max_{x''\in S}\lambda_i(x'')} = \infty.$$

Without loss of generality, suppose the first sum above is infinite. In this case, we will show that $\sum_{i\in\mathcal{I}} \mathbb{1}_{X_i=x} = \infty$. Note that, since $\frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x} = \mathbb{1}_{X_i=x}$ for all $i \in \mathcal{I}$, this will be sufficient to establish the lemma.

Now consider the subsequence $X_\mathcal{I}$. By definition of weighted exchangeability, this subsequence is $\lambda_\mathcal{I}$-weighted exchangeable, and the weight sequence $\lambda_\mathcal{I}$ satisfies

$$\sum_{j=1}^\infty \frac{(\lambda_\mathcal{I})_j(x)}{\max_{x''\in\mathcal{X}}(\lambda_\mathcal{I})_j(x'')} = \sum_{i\in\mathcal{I}} \frac{\lambda_i(x)}{\max_{x''\in S}\lambda_i(x'')} = \infty.$$

Now we need another lemma.

Lemma 10. *Suppose that Q is a λ-weighted exchangeable distribution on a finite set $\mathcal{X}$. Fix any $x \in \mathcal{X}$, and suppose also that*

$$\sum_{i=1}^\infty \frac{\lambda_i(x)}{\max_{x'\in\mathcal{X}}\lambda_i(x)} = \infty.$$

Then

$$\mathbb{P}_Q\left\{0 < \sum_{i=1}^\infty \mathbb{1}_{X_i=x} < \infty\right\} = 0.$$

That is, it holds almost surely that $X_i = x$ occurs either zero times or infinitely many times.

This result, applied to the $\lambda_\mathcal{I}$-weighted exchangeable sequence $X_\mathcal{I}$, implies that

$$\mathbb{P}\left\{0 < \sum_{j=1}^\infty \mathbb{1}_{(X_\mathcal{I})_j=x} < \infty\right\} = 0,$$

or equivalently,

$$\mathbb{P}\left\{0 < \sum_{i\in\mathcal{I}} \mathbb{1}_{X_i=x} < \infty\right\} = 0.$$

To complete the proof, defining

$$p = \mathbb{P}\left\{\sum_{i\in\mathcal{I}} \mathbb{1}_{X_i=x} = 0\right\},$$

we need to prove that $p = 0$. Suppose instead that $p > 0$. Recall that by definition of Q_S, we know that $\mathbb{P}_{Q_S}\left\{\sum_{i\geq 1} \mathbb{1}_{X_i=x} = 0\right\} = 0$, and so for sufficiently large $n \geq 1$ we have

$$\mathbb{P}\left\{\sum_{i=1}^n \mathbb{1}_{X_i=x} = 0\right\} < p.$$

Fix any such n, and define $\mathcal{I}' = \mathcal{I} \cup \{1, \ldots, n\}$. Then applying Lemma 10 to the subsequence $X_{\mathcal{I}'}$, which is $\lambda_{\mathcal{I}'}$-weighted exchangeable, we have

$$\mathbb{P}\left\{0 < \sum_{i\in\mathcal{I}'} \mathbb{1}_{X_i=x} < \infty\right\} = 0.$$

We can also calculate

$$\mathbb{P}\left\{0 < \sum_{i\in\mathcal{I}'} \mathbb{1}_{X_i=x} < \infty\right\} \geq \mathbb{P}\left\{\sum_{i=1}^{n} \mathbb{1}_{X_i=x} > 0, \sum_{i\in\mathcal{I}} \mathbb{1}_{X_i=x} = 0\right\}$$

$$\geq \mathbb{P}\left\{\sum_{i\in\mathcal{I}} \mathbb{1}_{X_i=x} = 0\right\} - \mathbb{P}\left\{\sum_{i=1}^{n} \mathbb{1}_{X_i=x} = 0\right\} = p - \mathbb{P}\left\{\sum_{i=1}^{n} \mathbb{1}_{X_i=x} = 0\right\} > 0.$$

We have reached a contradiction, as desired.

B.5 Proof of Lemma 9

Fix $(x, x') \in E_S^*$. For any $n > \ell \geq 0$, define

$$\widetilde{P}_{\ell,n}(x; x') = \frac{\sum_{i=\ell+1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i=x}}{\sum_{i=\ell+1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i\in\{x,x'\}}}.$$

Note that

$$\left|\widetilde{P}_{\ell,n}(x; x') - \widetilde{P}_n(x; x')\right| \leq \frac{\ell}{\sum_{i=1}^{n} \frac{\min\{\lambda_i(x),\lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i\in\{x,x'\}}}$$

by construction, and this denominator tends to infinity almost surely by Lemma 8, so we therefore have

$$\limsup_{n\to\infty} \widetilde{P}_n(x; x') = \limsup_{n\to\infty} \widetilde{P}_{\ell,n}(x; x')$$

holding almost surely, for all ℓ. Therefore,

$$\limsup_{n\to\infty} \widetilde{P}_n(x; x') = \limsup_{\ell\to\infty} \left(\limsup_{n\to\infty} \widetilde{P}_{\ell,n}(x; x')\right)$$

holds almost surely. Since $\widetilde{P}_{\ell,n}(x; x')$ is $\mathcal{F}_\ell$-measurable by construction, $\limsup_{n\to\infty} \widetilde{P}_{\ell,n}(x; x')$ is also $\mathcal{F}_\ell$-measurable. Therefore, $\limsup_{\ell\to\infty}(\limsup_{n\to\infty} \widetilde{P}_{\ell,n}(x; x'))$ is $\mathcal{F}_\ell$-measurable for every ℓ, and is therefore $\mathcal{F}_{\text{tail}}$-measurable.

Now let

$$\widetilde{P}'(x; x') = \limsup_{\ell\to\infty} \left(\limsup_{n\to\infty} \widetilde{P}_{\ell,n}(x; x')\right),$$

and let

$$\widetilde{P}(x; x') = \begin{cases} \widetilde{P}'(x; x') & \text{if } \widetilde{P}'(x; x') \in (0, 1), \\ 0.5 & \text{otherwise.} \end{cases}$$

By construction, $\widetilde{P}(x; x')$ is $\mathcal{F}_{\text{tail}}$-measurable.

From the work above, we see that $\widetilde{P}'(x; x') = \limsup_{n\to\infty} \widetilde{P}_n(x; x')$ holds almost surely. We next need to show that $\lim_{n\to\infty} \widetilde{P}_n(x; x')$ exists almost surely, to verify that $\widetilde{P}'(x; x') = \lim_{n\to\infty} \widetilde{P}_n(x; x')$ holds almost surely. Define

$$p_i = p_i(X_i) = \frac{\min\{\lambda_i(x), \lambda_i(x')\}}{\lambda_i(X_i)} \mathbb{1}_{X_i\in\{x,x'\}} \in [0, 1].$$

Similar to our earlier construction, draw $U_1, U_2, \ldots \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 1]$, independently of X, and define for each $i = 1, 2, \ldots,$

$$B_i = \mathbb{1}_{U_i \leq p_i(X_i)}.$$

Define, as before, $M = \sum_{i=1}^{\infty} B_i$, and

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

i.e., $I_1 < I_2 < \dots$ enumerates all indices i for which $B_i = 1$. Then define

$$\check{X} = \begin{cases} (X_{I_1}, X_{I_2}, X_{I_3}, \dots) & \text{if } M = \infty, \\ (X_{I_1}, \dots, X_{I_M}, x, x, \dots) & \text{otherwise.} \end{cases}$$

Note that $\check{X} \in \{x, x'\}^\infty$ by construction. By Lemma 8, together with the second Borel–Cantelli Lemma, we have $M = \infty$ almost surely. Next, we can verify that $\check{X}$ is exchangeable—the proof of this claim is essentially identical to the proof of Lemma 3 (where we establish the analogous result for establishing Theorem 6). Therefore, by de Finetti's theorem,

$$\lim_{m\to\infty} \frac{\mathbb{1}_{\check{X}_1 = x} + \dots + \mathbb{1}_{\check{X}_m = x}}{m}$$

converges almost surely; as in the proof of Theorem 6 we can see that this is equal (almost surely) to

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n B_i \cdot \mathbb{1}_{X_i = x}}{\sum_{i=1}^n B_i},$$

and again following the same steps as in the proof of Theorem 6, this is equal (almost surely) to

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n p_i \cdot \mathbb{1}_{X_i = x}}{\sum_{i=1}^n p_i} = \lim_{n\to\infty} \frac{\sum_{i=1}^n \frac{\min\{\lambda_i(x), \lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i = x}}{\sum_{i=1}^n \frac{\min\{\lambda_i(x), \lambda_i(x')\}}{\lambda_i(X_i)} \cdot \mathbb{1}_{X_i \in \{x, x'\}}} = \lim_{n\to\infty} \widetilde{P}_n(x; x').$$

In particular this verifies that $\lim_{n\to\infty} \widetilde{P}_n(x; x')$ exists almost surely, and returning to de Finetti's theorem for the exchangeable subsequence, conditional on the random value $\widetilde{P}'(x; x')$ to which the above limit converges, $\check{X}_{I_i}$, $i = 1, 2, \dots$ are i.i.d. draws from $\widetilde{P}'(x; x') \cdot \delta_x + (1 - \widetilde{P}'(x; x')) \cdot \delta_{x'}$.

To complete the proof we need to check that $\widetilde{P}'(x; x') \in (0, 1)$ almost surely, so that we have $\widetilde{P}(x; x') \overset{\text{a.s.}}{=} \lim_{n\to\infty} \widetilde{P}_n(x; x')$. Suppose instead that $\widetilde{P}'(x; x') = 0$ with positive probability. Then, on this event, we have

$$\mathbb{P}\left\{ \sum_{m=1}^{\infty} \mathbb{1}_{\check{X}_{I_m} = x} = 0 \;\middle|\; X \right\} = 1.$$

In order to have $\sum_{m=1}^{\infty} \mathbb{1}_{\check{X}_{I_m} = x} = 0$, we need to have $B_i = 0$ for all i with $X_i = x$, that is,

$$1 = \mathbb{P}\left\{ \sum_{m=1}^{\infty} \mathbb{1}_{\check{X}_{I_m} = x} = 0 \;\middle|\; X \right\} = \mathbb{P}\left\{ B_i = 0 \text{ for all } i \text{ with } X_i = x \mid X \right\} = \prod_{i : X_i = x} (1 - p_i).$$

Since $p_i > 0$ for any i with $X_i = x$ (because λ_i is positive-valued for each i), this implies that we must have $\sum_{i=1}^{\infty} \mathbb{1}_{X_i = x} = 0$ in order for this equality to hold; in other words, if $\widetilde{P}'(x; x') = 0$ with positive probability, then $\sum_{i=1}^{\infty} \mathbb{1}_{X_i = x} = 0$ with positive probability. But by definition of Q_S we have $\mathbb{P}\{\sum_{i=1}^{\infty} \mathbb{1}_{X_i = x} > 0\} = 1$. This is a contradiction, so $\widetilde{P}'(x; x') = 0$ can only hold with probability zero. We can similarly show that $\widetilde{P}'(x; x') = 1$ can only hold with probability zero, which completes the proof.

B.6 Proof of Lemma 10

Assume for the sake of contradiction that

$$\mathbb{P}_Q\left\{\sum_{i=1}^{\infty} \mathbb{1}_{X_i=x} = k\right\} > 0,$$

for some finite and positive k. The event that $\sum_{i=1}^{\infty} \mathbb{1}_{X_i=x} = k$, i.e., that x is observed exactly k many times, is $\mathcal{E}_\infty$-measurable. Let Q' be the distribution of X conditional on this event. Then Q' is also λ-weighted exchangeable (this can be verified exactly as in the proof of Theorem 4—specifically, the proof that $\Lambda_{\mathsf{dF}} \subseteq \Lambda_{\mathsf{01}}$). Now define $p_i = \mathbb{P}_{Q'}\left\{X_i = x\right\}$. We must have

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \mathbb{P}_{Q'}\left\{X_i = x\right\} = \mathbb{E}_{Q'}\left[\sum_{i=1}^{\infty} \mathbb{1}_{X_i=x}\right] = \mathbb{E}_{Q'}\left[k\right] = k.$$

We also have, for any $i \geq j$,

$$\begin{aligned}
p_i &= \mathbb{E}_{Q'}\left[\mathbb{1}_{X_i=x}\right] \\
&= \mathbb{E}_{Q'}\left[\mathbb{1}_{X_i=x} \cdot \frac{\lambda_i(X_i)\lambda_j(X_j)}{\lambda_i(X_i)\lambda_j(X_j)}\right] \\
&= \mathbb{E}_{Q'}\left[\mathbb{1}_{X_j=x} \cdot \frac{\lambda_i(X_j)\lambda_j(X_i)}{\lambda_i(X_i)\lambda_j(X_j)}\right] \\
&\leq \mathbb{E}_{Q'}\left[\mathbb{1}_{X_j=x}\right] \cdot \frac{\lambda_i(x)/\min_{x'\in\mathcal{X}} \lambda_i(x')}{\lambda_j(x)/\max_{x'\in\mathcal{X}} \lambda_j(x')} \\
&= p_j \cdot \frac{\lambda_i(x)/\min_{x'\in\mathcal{X}} \lambda_i(x')}{\lambda_j(x)/\max_{x'\in\mathcal{X}} \lambda_j(x')},
\end{aligned}$$

where the third equality applies Proposition 5. In particular, this implies that each p_i is positive (since these inequalities can be satisfied either if each p_i is positive, or each p_i is zero—and they cannot all be zero when we choose $k > 0$).

Therefore, for any i,

$$\sum_{j=1}^{\infty} p_j \geq p_i \cdot \left(\frac{\lambda_i(x)/\min_{x'\in\mathcal{X}} \lambda_i(x')}{\lambda_j(x)/\max_{x'\in\mathcal{X}} \lambda_j(x')}\right)^{-1} = p_i \cdot \frac{\min_{x'\in\mathcal{X}} \lambda_i(x')}{\lambda_i(x)} \cdot \sum_{j=1}^{\infty} \frac{\lambda_j(x)}{\max_{x'\in\mathcal{X}} \lambda_j(x')}.$$

But this is a contradiction, because on the left-hand side we have $\sum_{j=1}^{\infty} p_j = k$ by construction, while on the right-hand side, $p_i \cdot \frac{\min_{x'\in\mathcal{X}} \lambda_i(x')}{\lambda_i(x)} > 0$, and $\sum_{j=1}^{\infty} \frac{\lambda_j(x)}{\max_{x'\in\mathcal{X}} \lambda_j(x')} = \infty$ by the assumption in the lemma.