

Characterization of exchangeable measure-valued Pólya urn sequences*

Hristo Sariev^{†‡} Mladen Savov^{†‡}

Abstract

Measure-valued Pólya urn sequences (MVPS) are a generalization of the observation processes generated by k -color Pólya urn models, where the space of colors \mathbb{X} is a complete separable metric space and the urn composition is a finite measure on \mathbb{X} , in which case reinforcement reduces to a summation of measures. In this paper, we prove a representation theorem for the reinforcement measures R of all exchangeable MVPSs, which leads to a characterization result for their directing random measures \tilde{P} . In particular, when \mathbb{X} is countable or R is dominated by the initial distribution ν , then any exchangeable MVPS is a Dirichlet process mixture model over a family of probability distributions with disjoint supports. Furthermore, for all exchangeable MVPSs, the predictive distributions converge on a set of probability one in total variation to \tilde{P} . Importantly, we do not restrict our analysis to balanced MVPSs, in the terminology of k -color urns, but rather show that the only non-balanced exchangeable MVPSs are sequences of i.i.d. random variables.

Keywords: exchangeability; reinforced processes; Pólya sequences; directing random measures; urn models.

MSC2020 subject classifications: 60G09; 60G25; 60G57; 62G99.

Submitted to EJP on October 21, 2023, final version accepted on April 23, 2024.

Supersedes arXiv:2305.10083v2.

1 Introduction

The classical two-color Pólya urn model, which describes the evolution of an urn reinforced with one ball of the observed color, has a fundamental role in the predictive construction of prior distributions for Bayesian inference [7]. Suppose we have an urn that initially contains $w_0, w_1 > 0$ balls of colors 0 and 1, respectively. Let us denote by

*This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project No. BG-RRP-2.004-0008.

[†]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 8 Acad. Georgi Bonchev Str., Sofia 1113, Bulgaria. E-mail: h.sariev@math.bas.bg

[‡]Faculty of Mathematics and Informatics, Sofia University “St. Kliment Ohridski”, 5 James Bourchier Blvd, Sofia 1164, Bulgaria. E-mail: msavov@fmi.uni-sofia.bg

$X_n \in \{0, 1\}$ the color of the sampled ball at step n . Then $X_1 = 1$ with probability $\frac{w_1}{w_0+w_1}$ and, for every $n = 1, 2, \dots$,

$$\mathbb{P}(X_{n+1} = 1 | X_1, \dots, X_n) = \frac{w_1 + \sum_{i=1}^n X_i}{w_0 + w_1 + n} \quad \text{a.s.}, \quad (1.1)$$

which is the proportion of 1-balls in the urn at time n . It is well-known that the process $(X_n)_{n \geq 1}$ is exchangeable, that is, its law is invariant under finite permutations of the indices, and satisfies, as $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} \tilde{p},$$

where \tilde{p} has a Beta distribution with parameters (w_0, w_1) , see, e.g., [16, p.8]; therefore, by de Finetti representation theorem for exchangeable sequences [1, Theorem 3.1], given \tilde{p} , the X_n are conditionally independent and identically distributed with probability of “success” \tilde{p} . The converse is also true – any exchangeable sequence of Bernoulli random variables with a Beta prior distribution has the predictive structure (1.1).

There has been much interest in generalizing the construction (1.1) to obtain greater model flexibility and, at the same time, retain tractability of the process dynamics, see [7, 13, 16] and references therein. Various extensions of the Pólya urn scheme consider time-dependent, randomized, or generalized reinforcement mechanisms where colors other than those observed are added to the urn. More recently, [2, 10, 14] have proposed an extended class of *measure-valued Pólya urn processes* that are very general in nature, yet retain characteristic features of urn processes and contain, as a special case, most k -color urn models. The idea is to consider the urn composition as a finite measure μ on the space of colors \mathbb{X} , in the sense that, for any measurable set $B \subseteq \mathbb{X}$, the quantity $\mu(B)$ records the total *mass* of balls in the urn whose colors lie in B . Reinforcement is then reduced to a summation of measures, so that the updated urn composition, given that a ball with color x has been observed, becomes $\mu + R_x$, where R is a (random) transition kernel on \mathbb{X} , i.e. a map $x \mapsto R_x$ from \mathbb{X} to the space of (random) measures on \mathbb{X} . Assuming $\mu + R_x$ as the new urn composition, the next draw proceeds in the same way as before, independent of all previous draws. Thus, a sequence of finite measures $(\mu_n)_{n \geq 0}$ is a *measure-valued Pólya urn process* (MVPP) if it is a Markov process with the aforementioned additive structure for a given finite reinforcement kernel R . In this case, Theorem 1 in [9] implies the existence of a companion observation process $(X_n)_{n \geq 1}$ such that $X_1 \sim \frac{\mu_0}{\mu_0(\mathbb{X})} =: \mu'_0$ and, for each $n = 1, 2, \dots$,

$$\mathbb{P}(X_{n+1} \in \cdot | \mu_0, X_1, \mu_1, \dots, X_n, \mu_n) = \frac{\mu_n(\cdot)}{\mu_n(\mathbb{X})} =: \mu'_n(\cdot) \quad \text{a.s.}$$

In particular, when the reinforcement R is non-random, the above becomes

$$\mathbb{P}(X_{n+1} \in \cdot | X_1, \dots, X_n) = \frac{\mu_0(\cdot) + \sum_{i=1}^n R_{X_i}(\cdot)}{\mu_0(\mathbb{X}) + \sum_{i=1}^n R_{X_i}(\mathbb{X})} \quad \text{a.s.}, \quad (1.2)$$

which corresponds to the normalized urn composition at time n in the urn analogy. We call such a sequence $(X_n)_{n \geq 1}$ a *measure-valued Pólya urn sequence* (MVPS) to distinguish it from the process $(\mu_n)_{n \geq 0}$.

Analysis of MVPPs has been centered around two very distinct cases. In [9, 17], the authors study the so-called “diagonal” model, where only the observed color is reinforced, i.e. $R_x = w(x) \cdot \delta_x$, with $w(x) > 0$, for $x \in \mathbb{X}$, and δ_x is the unit mass at x . It follows under certain conditions that there exists a random probability measure \tilde{P} on \mathbb{X} such that, as $n \rightarrow \infty$, (i) the normalized urn composition μ'_n converges almost surely (a.s.) in total

variation to \tilde{P} ; and (ii) \tilde{P} is concentrated on a subset of so-called “dominant” colors. On the other hand, [2, 14, 15] consider MVPPs for which there exists an underlying Markov chain with kernel R that satisfies some “irreducibility”-type conditions (see [15, Section 1.2] for more details). Then, under additional assumptions, they prove that μ'_n converges a.s. weakly to a (deterministic) probability measure on \mathbb{X} .

In this work, we consider MVPPs that generate an exchangeable observation sequence $(X_n)_{n \geq 1}$ via (1.2). Standard results in exchangeability theory then guarantee that the predictive distributions of $(X_n)_{n \geq 1}$ converge a.s. weakly to a random probability measure \tilde{P} on \mathbb{X} , called the *directing random measure*, whose distribution determines the law of the exchangeable process, see, e.g., [1, Section 3]. Therefore, the only exchangeable “irreducible” MVPPs are sequences of i.i.d. random variables. Our goal is to characterize all R for which $(X_n)_{n \geq 1}$ is exchangeable, and in the process prove some new facts about \tilde{P} and hence about the distribution of $(X_n)_{n \geq 1}$. This will help us assess the limitations of the whole class of exchangeable MVPPs when used as models for Bayesian analysis. Important to our understanding are the *kernel based Dirichlet sequences* studied by Berti et al. [4], which are exchangeable MVPPs whose reinforcement R is a regular version of the conditional distribution for μ'_0 given some sub- σ -algebra \mathcal{G} ,

$$R_x(\cdot) = \mu'_0(\cdot \mid \mathcal{G})(x) \quad \text{for } \mu'_0\text{-a.e. } x, \tag{1.3}$$

see Section 2 for a rigorous definition. This assumption greatly simplifies the subsequent analysis and allows [4] to obtain a complete characterization of the directing random measure of any kernel based Dirichlet sequence. In [4, p.18], the authors tentatively raise the conjecture that (1.3) holds true for every exchangeable MVPP such that $R_x(\mathbb{X}) = 1$, not excluding the possibility of counterexamples. Our main theorem states that a normalized version of condition (1.3) is indeed true for every member of the class of exchangeable MVPPs, regardless of whether $R_x(\mathbb{X}) = 1$. That such a representation exists is not obvious, and its proof requires the use of techniques that go beyond those typical of the area and involve intensive use of abstract measure-valued objects. We also give additional results that do not make use of (1.3). In particular, we prove that when $(X_n)_{n \geq 1}$ is not i.i.d., then $R_x(\mathbb{X})$ is constant for almost every x , which implies that a “diagonal” MVPP will be exchangeable only if $w(x)$ is constant, see also Example 2.2. In addition, we show that, for fixed x , the reinforcement R_x is either absolutely continuous or mutually singular with respect to ν . On the other hand, using a refined version of (1.3), we provide a complete description of all possible exchangeable k -color urn models with positive time-homogeneous reinforcement, which to our knowledge has not been done before.

The rest of the paper is structured as follows. In Section 2, we provide notation, state some general facts about exchangeable sequences, and formally define our model. Results and examples are given in Section 3, with proofs postponed to Section 4. We state our representation theorem for general R_x in Section 3.1, while in Section 3.2 we study the case when R_x is dominated by μ'_0 , which includes k -color urns, and in Section 3.3 the case when R_x and μ'_0 are mutually singular. A final section concludes the paper.

2 The model

2.1 Preliminaries

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space, \mathbb{X} a complete separable metric space, and \mathcal{X} the associated Borel σ -algebra on \mathbb{X} . Standard results imply that \mathcal{X} is countably generated. A *transition kernel* on \mathbb{X} is a function $R : \mathbb{X} \times \mathcal{X} \rightarrow \bar{\mathbb{R}}_+$ that satisfies (i) the map $x \mapsto R(x, B) \equiv R_x(B)$ is \mathcal{X} -measurable, for all $B \in \mathcal{X}$; and (ii) $B \mapsto R_x(B)$ is a measure

on \mathbb{X} , for all $x \in \mathbb{X}$. Equivalently, R can be represented as a measurable function from \mathbb{X} to the space of measures on \mathbb{X} . In addition, a transition kernel R is said to be *finite* if $R_x(\mathbb{X}) < \infty$, for all $x \in \mathbb{X}$, *non null* if $R_x(\mathbb{X}) > 0$, for all $x \in \mathbb{X}$, and is called a *probability kernel* if $R_x(\mathbb{X}) = 1$, for all $x \in \mathbb{X}$. A *random (probability) measure* is a transition (probability) kernel $\tilde{P} : \Omega \times \mathcal{X} \rightarrow \bar{\mathbb{R}}_+$ from Ω to \mathbb{X} .

Let ν be a probability measure on \mathcal{X} , R a transition probability kernel on \mathbb{X} , and $\mathcal{G} \subseteq \mathcal{X}$ a sub- σ -algebra on \mathbb{X} . Then R is said to be a *regular version of the conditional distribution (r.c.d.) for ν given \mathcal{G}* , which for emphasis we will denote by

$$R_x(\cdot) = \nu(\cdot | \mathcal{G})(x) \quad \text{for } \nu\text{-a.e. } x,$$

if (i) $x \mapsto R_x(B)$ is \mathcal{G} -measurable, for all $B \in \mathcal{X}$; and (ii) $\int_A R_x(B)\nu(dx) = \nu(A \cap B)$, for all $A \in \mathcal{G}$ and $B \in \mathcal{X}$. It follows from the assumptions on $(\mathbb{X}, \mathcal{X})$ that a r.c.d. for ν under \mathcal{G} exists and is unique up to a ν -null set.

Let X be an \mathbb{X} -valued random variable with marginal distribution $X \sim \mathbb{P}_X$. Depending on the context, we would work with conditional distributions of the type $\mathbb{P}(\cdot | X)$ or $\mathbb{P}(\cdot | X = x)$, which are related by

$$\int_{\{X \in A\}} \mathbb{P}(B|X)(\omega)\mathbb{P}(d\omega) = \int_A \mathbb{P}(B|X = x)\mathbb{P}_X(dx),$$

for all $A \in \mathcal{X}$ and $B \in \mathcal{H}$.

Recall that an \mathbb{X} -valued sequence of random variables $(X_n)_{n \geq 1}$ is exchangeable if

$$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \dots, X_{\sigma(n)}),$$

for every $n \geq 2$ and all permutations σ of $\{1, \dots, n\}$. In that case (see [1, Section 3]), there exists a random probability measure \tilde{P} on \mathbb{X} , called the *directing random measure* of the sequence, such that, given \tilde{P} , the random variables X_1, X_2, \dots are conditionally independent and identically distributed (i.i.d.) with marginal distribution \tilde{P} ,

$$\begin{array}{ccc} X_n | \tilde{P} & \stackrel{i.i.d.}{\sim} & \tilde{P} \\ \tilde{P} & \sim & Q \end{array}$$

where Q is the (nonparametric) *prior* distribution of \tilde{P} . Moreover, for every $A \in \mathcal{X}$,

$$\mathbb{P}(X_{n+1} \in A | X_1, \dots, X_n) \xrightarrow{a.s.} \tilde{P}(A). \tag{2.1}$$

Since we take a predictive approach to model building, the following result from [8], which provides necessary and sufficient conditions for the system of predictive distributions of any stochastic process to be consistent with exchangeability, becomes our starting point of analysis.

Theorem 2.1 (Theorem 3.1, Proposition 3.2 in [8]). *A sequence of random variables $(X_n)_{n \geq 1}$ is exchangeable if and only if, for each $n = 0, 1, 2, \dots$ and every $A, B \in \mathcal{X}$,*

$$\mathbb{P}(X_{n+1} \in A, X_{n+2} \in B | X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in B, X_{n+2} \in A | X_1, \dots, X_n) \quad \text{a.s.}, \tag{2.2}$$

and

$$\mathbb{P}(X_{n+1} \in A | X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} \in A | X_1 = x_{\sigma(1)}, \dots, X_n = x_{\sigma(n)}), \tag{2.3}$$

for all permutations σ of $\{1, \dots, n\}$ and almost every $(x_1, \dots, x_n) \in \mathbb{X}^n$ with respect to the marginal distribution $\mathbb{P}_{(X_1, \dots, X_n)}$ of (X_1, \dots, X_n) on \mathcal{X}^n , where the case $n = 0$ should be understood as an unconditional statement.

2.2 Exchangeable MVPS

We will call any sequence $(X_n)_{n \geq 1}$ of \mathbb{X} -valued random variables on $(\Omega, \mathcal{H}, \mathbb{P})$ a *measure-valued Pólya urn sequence* with parameters θ, ν and R , denoted MVPS (θ, ν, R) , if $X_1 \sim \nu$ and, for each $n = 1, 2, \dots$,

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \dots, X_n) = \frac{\theta \nu(\cdot) + \sum_{i=1}^n R_{X_i}(\cdot)}{\theta + \sum_{i=1}^n R_{X_i}(\mathbb{X})} \quad \text{a.s.}, \quad (2.4)$$

where R is a non null¹ finite transition kernel on \mathbb{X} , called the *replacement/reinforcement kernel* of the process, ν is a probability measure on \mathbb{X} , known as the *base measure*, and $\theta > 0$ is a positive constant. We will say that the MVPS is *balanced* if there is, in addition, some $m > 0$ such that $R_x(\mathbb{X}) = m$ for ν -a.e. x , which in the urn analogy means that we add the same total number of balls each time. Moreover, unlike (1.2), we decompose the initial urn composition into two parameters, θ and ν , each one having a separate effect on the process (see, e.g., the construction in (2.7)).

Consider an exchangeable MVPS (θ, ν, R) with directing random measure \tilde{P} . It follows from (2.1) and (2.4) that the reinforcement R directly influences the form of \tilde{P} . On the other hand, by Theorem 2.1, R itself must satisfy certain conditions that make it admissible under exchangeability. Note, however, that equation (2.3) is always true for MVPSs. Hence, information about R can only be retrieved from the invariance property of the two-step-ahead predictive distributions in (2.2). In fact, [4, Theorem 7] show in their study of kernel based Dirichlet sequences that, for a balanced MVPS to be exchangeable, it is sufficient that equation (2.2) holds for $n = 0, 1$. Remarkably, when the MVPS is unbalanced, then (2.2) for $n = 2$ is also needed (see the proof of Theorem 3.2). Using this information, we show that, when normalized, R must be a r.c.d. for ν given some sub- σ -algebra \mathcal{G} of \mathcal{X} ,

$$\frac{R_x(\cdot)}{R_x(\mathbb{X})} = \nu(\cdot \mid \mathcal{G})(x) \quad \text{for } \nu\text{-a.e. } x, \quad (2.5)$$

and in the process we prove that exchangeable MVPSs are necessarily balanced unless i.i.d. It turns out that the representation (2.5) has major consequences for the shape and distribution of \tilde{P} , including the fact that the convergence in (2.1) is in total variation. We give one important example of an exchangeable MVPS next.

Example 2.2 (Pólya sequence). Let $(X_n)_{n \geq 1}$ be an MVPS with reinforcement kernel

$$R_x = \delta_x \quad \text{for } x \in \mathbb{X}, \quad (2.6)$$

i.e., we only reinforce the observed color with one additional ball. This process, also known as a *Pólya sequence*, was first studied by Blackwell and MacQueen [6] as an extension of the Pólya urn scheme (1.1) to general Polish spaces \mathbb{X} . In [6], the authors prove that $(X_n)_{n \geq 1}$ is exchangeable and its directing random measure \tilde{P} has a *Dirichlet process* distribution with parameters (θ, ν) , denoted $\text{DP}(\theta, \nu)$. We recall that $\tilde{P} \sim \text{DP}(\theta, \nu)$ if, for every finite partition $B_1, \dots, B_n \in \mathcal{X}$ of \mathbb{X} , the random vector $(\tilde{P}(B_1), \dots, \tilde{P}(B_n))$ has a Dirichlet distribution with parameters $(\theta \nu(B_1), \dots, \theta \nu(B_n))$,

$$(\tilde{P}(B_1), \dots, \tilde{P}(B_n)) \sim \text{Dir}(\theta \nu(B_1), \dots, \theta \nu(B_n)).$$

An equivalent statement (see, e.g., [12, p.112]) is that \tilde{P} is equal in law to

$$\tilde{P}(\cdot) \stackrel{w}{=} \sum_{j=1}^{\infty} V_j \delta_{Z_j}(\cdot), \quad (2.7)$$

¹From an application point of view, it may be interesting to see whether exchangeability supports $R_x(\mathbb{X}) = 0$ for all $x \in Z$ in some $Z \in \mathcal{X}$ such that $0 < \nu(Z) < 1$, and to characterize R_x for $x \in Z^c$. Importantly, by slightly modifying the proof of Theorem 3.2, we can show in the non-i.i.d. case that $R_x(\mathbb{X})$ is constant for ν -a.e. $x \in Z^c$. See Remark 4.1 for more information.

where $V_j = W_j \prod_{i=1}^{j-1} (1 - W_i)$ and $(W_j)_{j \geq 1}$ are i.i.d. with $W_1 \sim \text{Beta}(1, \theta)$, and $(Z_j)_{j \geq 1}$ are i.i.d. (ν) , independent of $(V_j)_{j \geq 1}$.

In this case, the reinforcement (2.6) has the anticipated structure (2.5), since

$$R_x(\cdot) = \nu(\cdot | \mathcal{X})(x) \quad \text{for } x \in \mathbb{X}.$$

3 Main results

3.1 General case

We begin this section with a simple characterization result for i.i.d. MVPSs.

Proposition 3.1. *An MVPS (θ, ν, R) is i.i.d. if and only if, for ν -a.e. x ,*

$$\frac{R_x(\cdot)}{R_x(\mathbb{X})} = \nu(\cdot). \tag{3.1}$$

The next theorem states that MVPSs that are exchangeable, but not i.i.d. are necessarily balanced. This fact is essential since it implies that the predictive distributions (2.4) are a linear combination of reinforcement kernels and leads via (2.2) to certain identities for R .

Theorem 3.2. *An unbalanced exchangeable MVPS is necessarily i.i.d.*

Remark 3.3. If an MVPS (θ, ν, R) is balanced, then $R_x(\mathbb{X}) = m$ for ν -a.e. x and some constant $m > 0$. It then follows from the particular form of the predictive distributions (2.4) that we can equivalently reformulate the process as an MVPS with parameters $(\tilde{\theta}, \nu, \tilde{R})$, where $\tilde{\theta} = \frac{\theta}{m}$ and $\tilde{R}_x = \frac{R_x}{m}$ with $\tilde{R}_x(\mathbb{X}) = 1$, for ν -a.e. x . On the other hand, it is not hard to see that any i.i.d. MVPS (θ, ν, R) with unbalanced R is also an i.i.d. MVPS with parameters (θ, ν, ν) . Thus, when we consider i.i.d. MVPSs below, we will implicitly assume that they are balanced. The above remarks are combined in the next corollary.

Corollary 3.4. *An exchangeable MVPS (θ, ν, R) is also an exchangeable MVPS $(\tilde{\theta}, \nu, \tilde{R})$ for some probability kernel \tilde{R} and constant $\tilde{\theta}$.*

We now consider the issue of representing R as in (2.5). For completeness, we first state the converse result, which is found in [4].

Theorem 3.5 (Theorem 7 in [4]). *Any balanced MVPS (θ, ν, R) whose reinforcement kernel R is, up to normalization, a r.c.d. for ν given some sub- σ -algebra \mathcal{G} of \mathcal{X} , is exchangeable.*

Remark 3.6. Although Berti et al. [4] focus only on the balanced case, it follows from Theorem 3.2 that no unbalanced MVPS, having replacement kernel as in Theorem 3.5, can be exchangeable unless $\mathcal{G} = \{\emptyset, \mathbb{X}\}$, in which case the observation sequence is i.i.d.

Our main result studies the necessity of such a representation under exchangeability. In addition, it states that we may take \mathcal{G} to be *countably generated (c.g.)* under ν , in the sense that there exists $C \in \mathcal{G}$ such that $\nu(C) = 1$ and $\mathcal{G} \cap C$ is c.g. In fact (see [3, p.649] and [5]), \mathcal{G} is c.g. under ν if and only if there exists a regular version μ of $\nu(\cdot | \mathcal{G})$ which is a.e. *proper*, that is, μ satisfies, for some $F \in \mathcal{G}$ such that $\nu(F) = 1$,

$$\mu_x(A) = \delta_x(A) \quad \text{for all } A \in \mathcal{G} \text{ and } x \in F.$$

Theorem 3.7. *Let $(X_n)_{n \geq 1}$ be an exchangeable MVPS (θ, ν, R) . Then there exists a sub- σ -algebra \mathcal{G} of \mathcal{X} such that R is a r.c.d. for ν under \mathcal{G} ,*

$$\frac{R_x(\cdot)}{R_x(\mathbb{X})} = \nu(\cdot | \mathcal{G})(x) \quad \text{for } \nu\text{-a.e. } x. \tag{3.2}$$

Moreover, we can find a sub- σ -algebra \mathcal{G} of \mathcal{X} such that (3.2) holds and \mathcal{G} is c.g. under ν .

Remark 3.8. If $(X_n)_{n \geq 1}$ is an MVPS with reinforcement $R(\cdot) = \nu(\cdot | \mathcal{G})$ and thus exchangeable by Theorem 3.5, then there exists by Theorem 3.7 a sub- σ -algebra \mathcal{G}' which is c.g. under ν and for which $\nu(\cdot | \mathcal{G}) = \nu(\cdot | \mathcal{G}')$.

A major consequence of Theorems 3.2 and 3.7 is that one can now characterize the directing random measure \tilde{P} of any exchangeable MVPS using results developed in [4] under the stronger assumptions that R is balanced and of the form (3.2). In that case, Berti et al. [4, Theorem 10] show that the predictive distributions (2.4) converge almost surely in total variation to \tilde{P} . By virtue of Theorems 3.2 and 3.7, the following result, which is based on Theorems 9 and 10 in [4], is now true for the entire class of exchangeable MVPS.

Theorem 3.9. Let $(X_n)_{n \geq 1}$ be an exchangeable MVPS (θ, ν, R) with directing random measure \tilde{P} . Then, as $n \rightarrow \infty$,

$$\sup_{A \in \mathcal{X}} |\mathbb{P}(X_{n+1} \in A | X_1, \dots, X_n) - \tilde{P}(A)| \xrightarrow{a.s.} 0.$$

Moreover, \tilde{P} is equal in law to

$$\tilde{P}(\cdot) \stackrel{w}{=} \sum_{j=1}^{\infty} V_j \frac{R_{Z_j}(\cdot)}{R_{Z_j}(\mathbb{X})},$$

where $(V_j)_{j \geq 1}$ and $(Z_j)_{j \geq 1}$ are as in (2.7), using as parameters $(\frac{\theta}{m}, \nu)$, and $m > 0$ is some constant such that $m = R_x(\mathbb{X})$ for ν -a.e. x .

3.2 Absolutely continuous case

In this section, we consider MVPSs such that, for ν -a.e. x , the replacement R_x is absolutely continuous with respect to the base measure ν , denoted $R_x \ll \nu$. Special cases include k -color urn models and MVPSs with discrete ν . The next theorem states that, under the assumption $R_x \ll \nu$, the σ -algebra \mathcal{G} in (3.2) is c.g. under ν w.r.t. a countable partition.

Theorem 3.10. Let $(X_n)_{n \geq 1}$ be an exchangeable MVPS (θ, ν, R) . Then $R_x \ll \nu$ for ν -a.e. x if and only if there exists a countable partition D_1, D_2, \dots of \mathbb{X} in \mathcal{X} such that

$$\frac{R_x(\cdot)}{R_x(\mathbb{X})} = \sum_k \nu(\cdot | D_k) \cdot \mathbb{1}_{D_k}(x) \quad \text{for } \nu\text{-a.e. } x. \tag{3.3}$$

Note that the existence of the countable partition in (3.3) does not depend on the enumerability of the color space \mathbb{X} or on the form of the base measure ν . The implications of Theorem 3.10 are illustrated in the next Example 3.11, which describes k -color urn models with positive time-homogeneous reinforcement.

Example 3.11 (k -color urn). When the space of colors $\mathbb{X} = \{x_1, \dots, x_k\}$ is finite, then $\nu(\cdot) = \sum_{i=1}^k p_i \delta_{x_i}(\cdot)$, where $p_i \geq 0$ represents the initial fraction of balls with color x_i in the urn. Furthermore, it is customary to state R in terms of a reinforcement matrix $[a_{ij}]_{1 \leq i, j \leq k}$, where $a_{ij} \geq 0$ denotes the number of balls of color x_j that will be added to the urn after color x_i has been observed.

If the observation process $(X_n)_{n \geq 1}$ is exchangeable, then we have from (2.4) that

$$p_i = \mathbb{P}(X_2 = x_i) = \sum_{j=1}^k \mathbb{P}(X_1 = x_j) \mathbb{P}(X_2 = x_i | X_1 = x_j) = \sum_{j=1}^k p_j \frac{\theta p_i + a_{j,i}}{\theta + \sum_{l=1}^k a_{j,l}};$$

thus, if $p_i = 0$ (i.e., there are initially no x_i balls in the urn), then $a_{ji} = 0$ for all $j : p_j > 0$, so color x_i never appears in the urn. The same argument can be repeated for any color

that is not initially present in the urn, so we may assume $p_i > 0$ for every i , without loss of generality. In that case, $R_x \ll \nu$ and $(X_n)_{n \geq 1}$ satisfies the conditions of Theorem 3.10.

Assuming $R_x(\mathbb{X}) = 1$, it follows from the particular form (3.3) of R that the reinforcement matrix of an exchangeable urn scheme with a finite number of colors has a block-diagonal design, i.e. \mathbb{X} can be partitioned into subsets of colors D_1, \dots, D_l , where a color reinforces and is reinforced only by the colors belonging to the same subset. Furthermore, the rows in each block are identical and equal to the conditional probability of ν given that a color from the same block has been observed. In other words, each D_j can be viewed as representing different *nuances* of the same color, so that the reinforcement (i) will be identical for, say, light or dark red; and (ii) will be equal to the initial number of balls for each nuance of red, normalized by the total number of red balls.

To illustrate the above points, let $\mathbb{X} = \{x_1, x_2, x_3\}$ and suppose that the urn has initial composition (w_1, w_2, w_3) , with $w_1, w_2, w_3 > 0$, so that $\nu(\cdot) = \sum_{i=1}^3 \frac{w_i}{\bar{w}} \delta_{x_i}(\cdot)$ where $\bar{w} = \sum_{i=1}^3 w_i$. Put $\bar{w}_i = \sum_{j \neq i} w_j$, $i = 1, 2, 3$. Then any MVPS $(X_n)_{n \geq 1}$ on \mathbb{X} with initial composition (w_1, w_2, w_3) is exchangeable if and only if its associated reinforcement matrix R has one of the following forms, up to some multiplicative constants $m, m_1, m_2, m_3 > 0$,

$$\begin{array}{c}
 \begin{array}{ccc}
 & x_1 & x_2 & x_3 \\
 x_1 & \begin{bmatrix} m_1 \frac{w_1}{\bar{w}} & m_1 \frac{w_2}{\bar{w}} & m_1 \frac{w_3}{\bar{w}} \end{bmatrix} & \begin{bmatrix} m \frac{w_1}{\bar{w}_3} & m \frac{w_2}{\bar{w}_3} & 0 \end{bmatrix} & \begin{bmatrix} m \frac{w_1}{\bar{w}_2} & 0 & m \frac{w_3}{\bar{w}_2} \end{bmatrix} \\
 x_2 & \begin{bmatrix} m_2 \frac{w_1}{\bar{w}} & m_2 \frac{w_2}{\bar{w}} & m_2 \frac{w_3}{\bar{w}} \end{bmatrix} & \begin{bmatrix} m \frac{w_1}{\bar{w}_3} & m \frac{w_2}{\bar{w}_3} & 0 \end{bmatrix} & \begin{bmatrix} 0 & m & 0 \end{bmatrix} \\
 x_3 & \begin{bmatrix} m_3 \frac{w_1}{\bar{w}} & m_3 \frac{w_2}{\bar{w}} & m_3 \frac{w_3}{\bar{w}} \end{bmatrix} & \begin{bmatrix} 0 & 0 & m \end{bmatrix} & \begin{bmatrix} m \frac{w_1}{\bar{w}_2} & 0 & m \frac{w_3}{\bar{w}_2} \end{bmatrix}
 \end{array} \\
 \\
 \begin{array}{ccc}
 \begin{bmatrix} m & 0 & 0 \\ 0 & m \frac{w_2}{\bar{w}_1} & m \frac{w_3}{\bar{w}_1} \\ 0 & m \frac{w_2}{\bar{w}_1} & m \frac{w_3}{\bar{w}_1} \end{bmatrix} & \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix}
 \end{array}
 \end{array}$$

Note that the first matrix corresponds to the case of an i.i.d. sequence of random variables with marginal distribution ν , and the last one to the three-color Pólya urn model with a Dirichlet prior distribution with parameters $(\frac{w_1}{m}, \frac{w_2}{m}, \frac{w_3}{m})$.

Example 3.12 (Discrete base measure). The previous example can be extended to MVPSs on more general spaces \mathbb{X} when the base measure ν is discrete, which is true especially if \mathbb{X} is countable. In this case, $\nu(\cdot) = \sum_{x \in \mathbb{X}_0} p(x) \delta_x(\cdot)$ for some countable subset $\mathbb{X}_0 \subseteq \mathbb{X}$ and positive weights $p(x) > 0$ such that $\sum_{x \in \mathbb{X}_0} p(x) = 1$. If $(X_n)_{n \geq 1}$ is one such process, then it satisfies, by virtue of (2.4),

$$1 = \sum_{x \in \mathbb{X}_0} p(x) = \mathbb{P}(X_1 \in \mathbb{X}_0) = \mathbb{P}(X_1 \in \mathbb{X}, X_2 \in \mathbb{X}_0) = \sum_{x \in \mathbb{X}_0} p(x) \frac{\theta + R_x(\mathbb{X}_0)}{\theta + R_x(\mathbb{X})};$$

thus, $R_x(\mathbb{X}_0^c) = 0$, and so $R_x \ll \nu$ for every $x \in \mathbb{X}$. Theorem 3.10 then implies the existence of a countable partition on \mathbb{X} such that $\frac{R_x}{R_x(\mathbb{X})}$ is a r.c.d for ν given the σ -algebra generated by said partition.

The form of the replacement kernel in (3.3) has further implications on the directing random measure \tilde{P} of the exchangeable process $(X_n)_{n \geq 1}$. In particular, it follows from (2.1) that, for every $A \in \mathcal{X}$, on a set of probability one,

$$\begin{aligned}
 \tilde{P}(A) &= \lim_{n \rightarrow \infty} \mathbb{P}(X_{n+1} \in A | X_1, \dots, X_n) \\
 &= \lim_{n \rightarrow \infty} \sum_k \frac{\theta \nu(D_k) + \sum_{i=1}^n R_{X_i}(\mathbb{X}) \cdot \mathbb{1}_{D_k}(X_i)}{\theta + \sum_{i=1}^n R_{X_i}(\mathbb{X})} \nu(A | D_k) \\
 &= \lim_{n \rightarrow \infty} \sum_k \mathbb{P}(X_{n+1} \in D_k | X_1, \dots, X_n) \nu(A | D_k) = \sum_k \tilde{P}(D_k) \nu(A | D_k);
 \end{aligned}$$

thus, as measures,

$$\tilde{P}(\cdot) \stackrel{a.s.}{=} \sum_k \tilde{P}(D_k) \nu(\cdot | D_k), \tag{3.4}$$

as \mathcal{X} is countably generated. The next theorem is a direct consequence of this fact and the distribution results in Theorem 3.9 (see Example 15 in [4] and Section 4.1.2 in [2]).

Theorem 3.13. *Let $(X_n)_{n \geq 1}$ be an exchangeable MVPS (θ, ν, R) with directing random measure \tilde{P} . If $R_x \ll \nu$ for ν -a.e. x , then \tilde{P} has the form (3.4) and, as $n \rightarrow \infty$,*

$$\sup_{A \in \mathcal{X}} |\mathbb{P}(X_{n+1} \in A | X_1, \dots, X_n) - \tilde{P}(A)| \xrightarrow{a.s.} 0.$$

Moreover, for every choice of indices k_1, \dots, k_j ,

$$(\tilde{P}(D_{k_1}), \dots, \tilde{P}(D_{k_j}), \tilde{P}(\cap_{i=1}^j D_{k_i}^c)) \sim \text{Dir}\left(\frac{\theta}{m} \nu(D_{k_1}), \dots, \frac{\theta}{m} \nu(D_{k_j}), \frac{\theta}{m} \nu(\cap_{i=1}^j D_{k_i}^c)\right),$$

using the notation in Example 2.2, where $m > 0$ is such that $R_x(\mathbb{X}) = m$ for ν -a.e. x .

The directing random measure (3.4) implies that the observations X_1, X_2, \dots will form clusters on the level of the sets D_1, D_2, \dots . Let $\tilde{P}^*(\cdot) = \sum_k \tilde{P}(D_k) \delta_k(\cdot)$ and $\nu^*(\cdot) = \sum_k \nu(D_k) \delta_k(\cdot)$. By Theorem 3.13, any exchangeable MVPS with absolutely continuous reinforcement is a Dirichlet process mixture model

$$\begin{aligned} X_n | \xi_n, \tilde{P}^* &\stackrel{i.i.d.}{\sim} \nu(\cdot | D_{\xi_n}) \\ \xi_n | \tilde{P}^* &\stackrel{i.i.d.}{\sim} \tilde{P}^* \\ \tilde{P}^* &\sim \text{DP}\left(\frac{\theta}{m}, \nu^*\right) \end{aligned}$$

using the notation in Example 2.2, where $(\xi_n)_{n \geq 1}$ is a sequence of exchangeable random labels such that $\xi_n = k$ if and only if $X_n \in D_k$. We refer to [12, Section 3.4] for more details on Dirichlet process mixtures.

3.3 Singular case

In general, we have the decomposition of the reinforcement

$$R_x = R_x^\perp + R_x^a,$$

for some finite measures R_x^\perp and R_x^a on \mathbb{X} such that R_x^\perp and ν are mutually singular, denoted by $R_x^\perp \perp \nu$, and $R_x^a \ll \nu$. The next theorem shows that, for fixed x , the replacement R_x of any exchangeable MVPS is either mutually singular or absolutely continuous with respect to ν . In addition, the support of any mutually singular component, R_x^\perp , does not intersect the support of any absolutely continuous one, R_y^a , implying a complete separation of the two regimes.

Theorem 3.14. *Let $(X_n)_{n \geq 1}$ be an exchangeable MVPS (θ, ν, R) . Then there exists a set $\mathcal{S} \in \mathcal{X}$ such that*

$$R_x = R_x^\perp \quad \text{and} \quad R_x^\perp(\mathcal{S}^c) = 0, \quad \text{for } \nu\text{-a.e. } x \text{ in } \mathcal{S},$$

and

$$R_x = R_x^a \quad \text{and} \quad R_x^a(\mathcal{S}) = 0, \quad \text{for } \nu\text{-a.e. } x \text{ in } \mathcal{S}^c.$$

Remark 3.15. It follows from Theorem 3.14 that whenever $(X_n)_{n \geq 1}$ is not i.i.d., then for ν -a.e. x , the support of R_x , say $S_x \in \mathcal{X}$ such that $R_x(S_x) = 1$, has a ν -measure strictly smaller than one, $\nu(S_x) < 1$. Indeed, this is obvious if we have an $\mathcal{S} \in \mathcal{X}$ as in Theorem 3.14 such that $0 < \nu(\mathcal{S}) < 1$. If instead $\nu(\mathcal{S}) = 0$, then $R_x \ll \nu$ for ν -a.e. x and R_x has a “block-diagonal” design by Theorem 3.10 with at least two disjoint blocks. On the other hand, if $\nu(\mathcal{S}) = 1$, then $R_x \perp \nu$ and there exists $S_x \in \mathcal{X}$ such that $R_x(S_x) = 1$ and $\nu(S_x) = 0$, for ν -a.e. x .

The following simple example demonstrates one such case.

Example 3.16. Let ν be a diffuse probability measure, i.e. $\nu(\{x\}) = 0$, for $x \in \mathbb{X}$. Fix $\theta = 1$. Take $S \in \mathcal{X}$ such that $0 < \nu(S) < 1$. Let us define, for $x \in \mathbb{X}$,

$$R_x(\cdot) := \begin{cases} \delta_x(\cdot), x \in S; \\ \nu(\cdot | S^c), x \in S^c. \end{cases}$$

Then $R_x \perp \nu$ and $R_x(S^c) = 0$, for $x \in S$, and $R_x \ll \nu$ and $R_x(S) = 0$, for $x \in S^c$.

Let $A, B \in \mathcal{X}$ and $x \in \mathbb{X}$. It is an easy exercise to check that

$$\int_A R_x(B)\nu(dx) = \int_B R_x(A)\nu(dx) \quad \text{and} \quad \int_A R_y(B)R_x(dy) = \int_B R_y(A)R_x(dy).$$

Therefore, by [4, Theorem 7], the balanced MVPS $(X_n)_{n \geq 1}$ with parameters (θ, ν, R) is exchangeable.

Define

$$\mathcal{G} := \{A \in \mathcal{X} : \nu(A|S^c) = 0 \text{ or } \nu(A|S^c) = 1\}.$$

Then \mathcal{G} is a σ -algebra.

Let $A \in \mathcal{G}$ and $B \in \mathcal{X}$. If $\nu(A|S^c) = 0$, then $\nu(A \cap S^c) = 0$, so $\nu(A \cap S^c)\nu(B|S^c) = \nu(A \cap B \cap S^c)$; else, if $\nu(A|S^c) = 1$, then $\nu(A \cap S^c) = \nu(S^c)$ and $\nu(A^c \cap S^c) = 0$, from where $\nu(A^c \cap B \cap S^c) = 0$, and so $\nu(A \cap S^c)\nu(B|S^c) = \nu(B \cap S^c) = \nu(A \cap B \cap S^c)$. As a result,

$$\begin{aligned} \int_A R_x(B)\nu(dx) &= \int_{A \cap S} \delta_x(B)\nu(dx) + \int_{A \cap S^c} \nu(B|S^c)\nu(dx) \\ &= \nu(A \cap B \cap S) + \nu(A \cap S^c)\nu(B|S^c) = \nu(A \cap B \cap S) + \nu(A \cap B \cap S^c) = \nu(A \cap B). \end{aligned}$$

On the other hand, for all $t \in [0, 1]$ and $B \in \mathcal{X}$, the set $\{x \in \mathbb{X} : R_x(B) \leq t\}$ is equal to one of $B^c \cap S$, $B^c \cup S^c$, S or \emptyset , all in \mathcal{G} . Therefore, $x \mapsto R_x(B)$ is \mathcal{G} -measurable and

$$R_x(\cdot) = \nu(\cdot | \mathcal{G})(x) \quad \text{for } \nu\text{-a.e. } x.$$

The decomposition implied in Theorem 3.14 can be used to analyze separately the singular and the absolutely continuous parts of R . Concentrating on R_x^\perp , let us assume that $R_x \perp \nu$ for all $x \in \mathbb{X}$. By Theorem 3.2, such an MVPS is necessarily balanced, say $R_x(\mathbb{X}) = 1$ for ν -a.e. x . Moreover, ν has to be diffuse; else, if $\nu(\{z\}) > 0$ for some $z \in \mathbb{X}$, then $R_x(\{z\}) = 0$, and using (2.4),

$$\nu(\{z\}) = \int_{\mathbb{X}} \frac{\theta \nu(\{z\}) + R_x(\{z\})}{\theta + R_x(\mathbb{X})} \nu(dx) = \frac{\theta}{\theta + 1} \nu(\{z\}),$$

absurd. We show in Proposition 3.17 below that if R_x is further discrete, then x is an atom of R_x for ν -a.e. x . Examples include the Pólya sequence from Example 2.2 with diffuse ν , and Example 2 in [4] where the authors consider an MVPS with reinforcement $R_x = \frac{1}{2}(\delta_x + \delta_{-x})$. In addition, Proposition 3.17 states for discrete R that either $R_x = R_y$ or $R_x \perp R_y$, which is reminiscent of the “block-diagonal” design in Section 3.2.

Proposition 3.17. Let $(X_n)_{n \geq 1}$ be an exchangeable MVPS (θ, ν, R) . If $R_x \perp \nu$ and R_x is discrete, then $R_x(\{x\}) > 0$ for ν -a.e. x . Moreover,

$$\text{for } \nu\text{-a.e. } x \text{ and } y, \text{ either } R_x = R_y \text{ or } R_x \perp R_y.$$

4 Proofs

Proof of Proposition 3.1. Let $(X_n)_{n \geq 1}$ be an i.i.d. MVPS (θ, ν, R) . Take $A, B \in \mathcal{X}$. It follows from (2.4) that

$$\mathbb{P}(X_1 \in A, X_2 \in B) = \int_A \mathbb{P}(X_2 \in B | X_1 = x) \mathbb{P}(X_1 \in dx) = \int_A \frac{\theta \nu(B) + R_x(B)}{\theta + R_x(\mathbb{X})} \nu(dx).$$

On the other hand,

$$\mathbb{P}(X_1 \in A, X_2 \in B) = \mathbb{P}(X_1 \in A)\mathbb{P}(X_2 \in B) = \nu(A)\nu(B) = \int_A \nu(B)\nu(dx).$$

Since A is arbitrary, then $\frac{\theta\nu(B)+R_x(B)}{\theta+R_x(\mathbb{X})} = \nu(B)$ for ν -a.e. x , and so $\frac{R_x(B)}{R_x(\mathbb{X})} = \nu(B)$ for ν -a.e. x . As \mathcal{X} is countably generated, we get, as measures, $\frac{R_x(dy)}{R_x(\mathbb{X})} = \nu(dy)$ for ν -a.e. x .

Conversely, if R is of the form (3.1), then, for any $A \in \mathcal{X}$,

$$\mathbb{P}(X_{n+1} \in A | X_1, \dots, X_n) = \frac{\theta\nu(A) + \sum_{i=1}^n R_{X_i}(\mathbb{X})\nu(A)}{\theta + \sum_{i=1}^n R_{X_i}(\mathbb{X})} = \nu(A) \quad \text{a.s.,}$$

implying that $(X_n)_{n \geq 1}$ is i.i.d. with marginal distribution ν . □

Proof of Theorem 3.2. Let $(X_n)_{n \geq 1}$ be an exchangeable MVPS (θ, ν, R) . Define

$$f(x) := R_x(\mathbb{X}), \quad \text{for } x \in \mathbb{X}, \quad \text{and} \quad H := \{(x, y) \in \mathbb{X}^2 : f(x) = f(y)\},$$

and let H_x be the x -section of H . It follows from exchangeability and the form of the predictive distributions (2.4) that, for every $A, B \in \mathcal{X}$,

$$\begin{aligned} \int_{(x,y) \in A \times B} \frac{\theta\nu(dy) + R_x(dy)}{\theta + f(x)} \nu(dx) &= \mathbb{P}(X_1 \in A, X_2 \in B) \\ &= \mathbb{P}(X_1 \in B, X_2 \in A) = \int_{(x,y) \in A \times B} \frac{\theta\nu(dx) + R_y(dx)}{\theta + f(y)} \nu(dy). \end{aligned} \tag{4.1}$$

We proceed by making some important observations from (4.1). Let $C \in \mathcal{X}$ be such that $\nu(C) = 0$. Applying (4.1) to $C \times \mathbb{X}$, we get $\nu(C) = \int_{\mathbb{X}} \frac{R_x(C)}{\theta+f(x)} \nu(dx)$, so $\frac{R_x(C)}{\theta+f(x)} = 0$ for ν -a.e. x , and hence

$$\text{for any } C \in \mathcal{X}, \text{ if } \nu(C) = 0, \text{ then } R_x(C) = 0 \text{ for } \nu\text{-a.e. } x. \tag{4.2}$$

On the other hand, (4.1) implies the equivalence of the measures

$$\frac{\theta\nu(dy) + R_x(dy)}{\theta + f(x)} \nu(dx) = \frac{\theta\nu(dx) + R_y(dx)}{\theta + f(y)} \nu(dy);$$

thus, $\nu(dx) = \int_{y \in \mathbb{X}} \frac{\theta\nu(dx)+R_y(dx)}{\theta+f(y)} \nu(dy)$, and by rearranging the terms,

$$\int_{y \in \mathbb{X}} \frac{R_y(dx)}{\theta + f(y)} \nu(dy) = \int_{y \in \mathbb{X}} \frac{f(y)}{\theta + f(y)} \nu(dy) \nu(dx). \tag{4.3}$$

Moreover, the following two measures, defined as the indefinite integrals of $(\theta + f(x))(\theta + f(y))$ with respect to $\frac{\theta\nu(dy)+R_x(dy)}{\theta+f(x)} \nu(dx)$ and $\frac{\theta\nu(dx)+R_y(dx)}{\theta+f(y)} \nu(dy)$, are equal

$$(\theta + f(y))(\theta\nu(dy) + R_x(dy))\nu(dx) = (\theta + f(x))(\theta\nu(dx) + R_y(dx))\nu(dy),$$

which after simplification becomes

$$\theta f(y)\nu(dx)\nu(dy) + (\theta + f(y))R_x(dy)\nu(dx) = \theta f(x)\nu(dx)\nu(dy) + (\theta + f(x))R_y(dx)\nu(dy). \tag{4.4}$$

In particular, restricting the above measures to H where $f(x) = f(y)$, we obtain

$$\mathbb{1}_H(x, y)R_x(dy)\nu(dx) = \mathbb{1}_H(x, y)R_y(dx)\nu(dy). \tag{4.5}$$

We now consider the implications that the invariance of the two-step-ahead predictive distributions has on R . Let $A, B \in \mathcal{X}$. Since $X_1 \sim \nu$, it follows from (2.2) in Theorem 2.4 that, for ν -a.e. x ,

$$\begin{aligned} & \int_{(y,z) \in A \times B} \frac{\theta \nu(dz) + R_x(dz) + R_y(dz)}{\theta + f(x) + f(y)} \frac{\theta \nu(dy) + R_x(dy)}{\theta + f(x)} = \mathbb{P}(X_2 \in A, X_3 \in B | X_1 = x) \\ & = \mathbb{P}(X_2 \in B, X_3 \in A | X_1 = x) = \int_{(y,z) \in A \times B} \frac{\theta \nu(dy) + R_x(dy) + R_z(dy)}{\theta + f(x) + f(z)} \frac{\theta \nu(dz) + R_x(dz)}{\theta + f(x)}. \end{aligned}$$

Since \mathcal{X} is countably generated, we may combine the non-null sets and get, for ν -a.e. x , the equivalence of the measures

$$\begin{aligned} & \frac{\theta \nu(dz) + R_x(dz) + R_y(dz)}{(\theta + f(x) + f(y))(\theta + f(x))} (\theta \nu(dy) + R_x(dy)) \\ & = \frac{\theta \nu(dy) + R_x(dy) + R_z(dy)}{(\theta + f(x) + f(z))(\theta + f(x))} (\theta \nu(dz) + R_x(dz)). \end{aligned}$$

Arguing as in (4.4), we get for ν -a.e. x

$$\begin{aligned} & (\theta + f(x) + f(z)) (\theta \nu(dz) + R_x(dz) + R_y(dz)) (\theta \nu(dy) + R_x(dy)) \\ & = (\theta + f(x) + f(y)) (\theta \nu(dy) + R_x(dy) + R_z(dy)) (\theta \nu(dz) + R_x(dz)), \end{aligned} \tag{4.6}$$

which after simplification² becomes

$$\begin{aligned} & \theta^2 \left\{ f(z) \nu(dy) \nu(dz) + R_y(dz) \nu(dy) \right\} \\ & \quad + \theta \left\{ f(z) R_x(dy) \nu(dz) + f(z) R_x(dz) \nu(dy) \right. \\ & \quad \quad \quad \left. + (f(x) + f(z)) R_y(dz) \nu(dy) + R_y(dz) R_x(dy) \right\} \\ & \quad + \left\{ f(z) R_x(dy) R_x(dz) + (f(x) + f(z)) R_y(dz) R_x(dy) \right\} \\ & = \theta^2 \left\{ f(y) \nu(dy) \nu(dz) + R_z(dy) \nu(dz) \right\} \\ & \quad + \theta \left\{ f(y) R_x(dz) \nu(dy) + f(y) R_x(dy) \nu(dz) \right. \\ & \quad \quad \quad \left. + (f(x) + f(y)) R_z(dy) \nu(dz) + R_z(dy) R_x(dz) \right\} \\ & \quad + \left\{ f(y) R_x(dy) R_x(dz) + (f(x) + f(y)) R_z(dy) R_x(dz) \right\}. \end{aligned}$$

Using (4.4) w.r.t. (y, z) , we may collect $\theta f(z) R_y(dz) \nu(dy)$ and the two terms with θ^2 on the left-hand side and cancel them with the sum of $\theta f(y) R_z(dy) \nu(dy)$ and the two terms with θ^2 on the right-hand side to obtain

$$\begin{aligned} & \theta f(z) R_x(dy) \nu(dz) + \theta f(z) R_x(dz) \nu(dy) + \theta f(x) R_y(dz) \nu(dy) \\ & \quad + f(z) R_x(dy) R_x(dz) + (\theta + f(x) + f(z)) R_y(dz) R_x(dy) \\ & = \theta f(y) R_x(dz) \nu(dy) + \theta f(y) R_x(dy) \nu(dz) + \theta f(x) R_z(dy) \nu(dz) \\ & \quad + f(y) R_x(dy) R_x(dz) + (\theta + f(x) + f(y)) R_z(dy) R_x(dz); \end{aligned} \tag{4.7}$$

²Note that the measures in (4.6), and in the sequel, need not be finite, so extra care should be taken when cancelling common terms. The measures are, however, σ -finite with respect to the product of the sets $F_n := \{x \in \mathbb{X} : n^{-1} \leq f(x) \leq n\}$, $n \geq 1$; hence, if one works with the measures in (4.6) restricted to $(x, y, z) \in F_n \times F_n \times F_n$, all quantities from there on will be finite. The general result then follows from monotone convergence, letting $\mathbb{1}_{F_n} \uparrow 1$ as $n \rightarrow \infty$. For exposition reasons, we will omit this procedure and work as if all quantities were finite to begin with.

thus, restricting the above measures to H where $f(y) = f(z)$,

$$\begin{aligned} & \mathbb{1}_H(y, z)\theta f(x)R_y(dz)\nu(dy) + \mathbb{1}_H(y, z)(\theta + f(x) + f(y))R_y(dz)R_x(dy) \\ & = \mathbb{1}_H(y, z)\theta f(x)R_z(dy)\nu(dz) + \mathbb{1}_H(y, z)(\theta + f(x) + f(y))R_z(dy)R_x(dz). \end{aligned}$$

By (4.5), we have $\mathbb{1}_H(y, z)R_y(dz)\nu(dy) = \mathbb{1}_H(y, z)R_z(dy)\nu(dz)$, and so, for ν -a.e. x ,

$$\mathbb{1}_H(y, z)(\theta + f(x) + f(y))R_y(dz)R_x(dy) = \mathbb{1}_H(y, z)(\theta + f(x) + f(y))R_z(dy)R_x(dz),$$

which after “dividing” by $(\theta + f(x) + f(y))$, i.e. using again the notion of indefinite integrals, becomes

$$\mathbb{1}_H(y, z)R_y(dz)R_x(dy) = \mathbb{1}_H(y, z)R_z(dy)R_x(dz). \tag{4.8}$$

Next, considering equation (2.2) for $n = 2$, we get that, as measures,

$$\mathbb{P}(X_3 \in dz, X_4 \in du | X_1 = x, X_2 = y) = \mathbb{P}(X_3 \in du, X_4 \in dz | X_1 = x, X_2 = y), \tag{4.9}$$

for $\mathbb{P}_{(X_1, X_2)}$ -a.e. (x, y) in \mathbb{X}^2 . It is an easy observation from (4.1) that $(\nu \times \nu) \ll \mathbb{P}_{(X_1, X_2)}$; thus, (4.9) holds true for $(\nu \times \nu)$ -a.e. (x, y) in \mathbb{X}^2 . Equation (4.9) implies the equivalence of the measures

$$\begin{aligned} & \frac{\theta\nu(du) + R_x(du) + R_y(du) + R_z(du)}{\theta + f(x) + f(y) + f(z)} \frac{\theta\nu(dz) + R_x(dz) + R_y(dz)}{\theta + f(x) + f(y)} \\ & = \frac{\theta\nu(dz) + R_x(dz) + R_y(dz) + R_u(dz)}{\theta + f(x) + f(y) + f(u)} \frac{\theta\nu(du) + R_x(du) + R_y(du)}{\theta + f(x) + f(y)}, \end{aligned}$$

so arguing as in (4.6),

$$\begin{aligned} & (\theta + f(x) + f(y) + f(u))(\theta\nu(du) + R_x(du) + R_y(du) + R_z(du)) \\ & \quad \times (\theta\nu(dz) + R_x(dz) + R_y(dz)) \\ & = (\theta + f(x) + f(y) + f(z))(\theta\nu(dz) + R_x(dz) + R_y(dz) + R_u(dz)) \\ & \quad \times (\theta\nu(du) + R_x(du) + R_y(du)). \end{aligned}$$

After some simple algebra, we obtain

$$\begin{aligned} & \theta^2 \{ f(u)\nu(dz)\nu(du) + R_z(du)\nu(dz) \} \\ & + \theta \{ f(u)R_x(du)\nu(dz) + f(u)R_y(du)\nu(dz) + (f(x) + f(y) + f(u))R_z(du)\nu(dz) \\ & \quad + f(u)R_x(dz)\nu(du) + f(u)R_y(dz)\nu(du) + R_z(du)R_x(dz) + R_z(du)R_y(dz) \} \\ & + \{ f(u)R_x(dz)R_x(du) + f(u)R_y(du)R_x(dz) + (f(x) + f(y) + f(u))R_z(du)R_x(dz) \\ & \quad + f(u)R_x(du)R_y(dz) + f(u)R_y(dz)R_y(du) + (f(x) + f(y) + f(u))R_z(du)R_y(dz) \} \\ & = \theta^2 \{ f(z)\nu(dz)\nu(du) + R_u(dz)\nu(du) \} \\ & + \theta \{ f(z)R_x(dz)\nu(du) + f(z)R_y(dz)\nu(du) + (f(x) + f(y) + f(z))R_u(dz)\nu(du) \\ & \quad + f(z)R_x(du)\nu(dz) + f(z)R_y(du)\nu(dz) + R_u(dz)R_x(du) + R_u(dz)R_y(du) \} \\ & + \{ f(z)R_x(dz)R_x(du) + f(z)R_y(dz)R_x(du) + (f(x) + f(y) + f(z))R_u(dz)R_x(du) \\ & \quad + f(z)R_x(dz)R_y(du) + f(z)R_y(dz)R_y(du) + (f(x) + f(y) + f(z))R_u(dz)R_y(du) \}, \end{aligned}$$

which after further rearrangement becomes

$$\begin{aligned} & \theta \{ \theta f(u)\nu(dz)\nu(du) + (\theta + f(u))R_z(du)\nu(dz) \} \\ & + \{ \theta f(u)R_x(dz)\nu(du) + \theta f(u)R_x(du)\nu(dz) + \theta f(x)R_z(du)\nu(dz) \\ & \quad + (\theta + f(x) + f(u))R_z(du)R_x(dz) + f(u)R_x(dz)R_x(du) \} \\ & + \{ \theta f(u)R_y(dz)\nu(du) + \theta f(u)R_y(du)\nu(dz) + \theta f(y)R_z(du)\nu(dz) \\ & \quad + (\theta + f(y) + f(u))R_z(du)R_y(dz) + f(u)R_y(dz)R_y(du) \} \\ & + \{ f(x)R_z(du)R_y(dz) + f(u)R_x(dz)R_y(du) + f(y)R_z(du)R_x(dz) \\ & \quad + f(u)R_x(du)R_y(dz) \} \end{aligned}$$

$$\begin{aligned}
 &= \theta \{ \theta f(z) \nu(dz) \nu(du) + (\theta + f(z)) R_u(dz) \nu(du) \} \\
 &\quad + \{ \theta f(z) R_x(du) \nu(dz) + \theta f(z) R_x(dz) \nu(du) + \theta f(x) R_u(dz) \nu(du) \\
 &\quad\quad + (\theta + f(x) + f(z)) R_u(dz) R_x(du) + f(z) R_x(dz) R_x(du) \} \\
 &\quad + \{ \theta f(z) R_y(du) \nu(dz) + \theta f(z) R_y(dz) \nu(du) + \theta f(y) R_u(dz) \nu(du) \\
 &\quad\quad + (\theta + f(y) + f(z)) R_u(dz) R_y(du) + f(z) R_y(dz) R_y(du) \} \\
 &\quad + \{ f(x) R_u(dz) R_y(du) + f(z) R_x(du) R_y(dz) + f(y) R_u(dz) R_x(du) \\
 &\quad\quad f(z) R_x(dz) R_y(du) \}.
 \end{aligned}$$

It follows from (4.4) w.r.t. (z, u) , (4.7) w.r.t. (x, z, u) , and (4.7) w.r.t. (y, z, u) that we may cancel the terms in the first, second and third brackets, respectively, to get, for $(\nu \times \nu)$ -a.e. (x, y) in \mathbb{X}^2 ,

$$\begin{aligned}
 &f(x) R_z(du) R_y(dz) + f(u) R_x(dz) R_y(du) + f(y) R_z(du) R_x(dz) + f(u) R_x(du) R_y(dz) \\
 &= f(x) R_u(dz) R_y(du) + f(z) R_x(du) R_y(dz) + f(y) R_u(dz) R_x(du) + f(z) R_x(dz) R_y(du),
 \end{aligned}$$

from where³

$$\begin{aligned}
 &f(x) R_z(du) R_y(dz) + f(y) R_z(du) R_x(dz) + (f(u) - f(z)) R_x(dz) R_y(du) \\
 &\quad + (f(u) - f(z)) R_x(du) R_y(dz) \tag{4.10} \\
 &= f(x) R_y(du) R_u(dz) + f(y) R_x(du) R_u(dz).
 \end{aligned}$$

Dividing both sides in (4.10) by $(\theta + f(x))(\theta + f(y))$ and integrating with respect to $\nu(dx)$ and $\nu(dy)$ yields

$$\begin{aligned}
 &\int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(x)}{(\theta + f(x))(\theta + f(y))} R_z(du) R_y(dz) \nu(dx) \nu(dy) \\
 &\quad + \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(y)}{(\theta + f(x))(\theta + f(y))} R_z(du) R_x(dz) \nu(dx) \nu(dy) \\
 &\quad + \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(u) - f(z)}{(\theta + f(x))(\theta + f(y))} R_x(dz) R_y(du) \nu(dx) \nu(dy) \\
 &\quad + \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(u) - f(z)}{(\theta + f(x))(\theta + f(y))} R_x(du) R_y(dz) \nu(dx) \nu(dy) \\
 &= \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(x)}{(\theta + f(x))(\theta + f(y))} R_y(du) R_u(dz) \nu(dx) \nu(dy) \\
 &\quad + \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(y)}{(\theta + f(x))(\theta + f(y))} R_x(du) R_u(dz) \nu(dx) \nu(dy).
 \end{aligned}$$

Now, using the identity in (4.3) w.r.t. $\int_{\mathbb{X}} \frac{R_x(dz)}{\theta + f(x)} \nu(dx)$, $\int_{\mathbb{X}} \frac{R_x(du)}{\theta + f(x)} \nu(dx)$, $\int_{\mathbb{X}} \frac{R_y(dz)}{\theta + f(y)} \nu(dy)$ and $\int_{\mathbb{X}} \frac{R_y(du)}{\theta + f(y)} \nu(dy)$ on, respectively, the second and third terms on the left, the fourth term on the left and the second on the right, the first and fourth terms on the left, the third term on the left and the first on the right of the equation, leads to

$$\begin{aligned}
 &2 \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(x) f(y)}{(\theta + f(x))(\theta + f(y))} R_z(du) \nu(dx) \nu(dy) \nu(dz) \\
 &\quad + 2 \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(x) f(y) (f(u) - f(z))}{(\theta + f(x))(\theta + f(y))} \nu(dx) \nu(dy) \nu(dz) \nu(du) \\
 &= 2 \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(x) f(y)}{(\theta + f(x))(\theta + f(y))} R_u(dz) \nu(dx) \nu(dy) \nu(du).
 \end{aligned}$$

³Again we refer to the footnote on p. 12 on how to deal with signed measures like those in (4.10) in order to avoid indeterminate forms.

Observe that the terms involving z and u can be pulled out of the integrals; thus, dividing both sides by $2 \int_{\mathbb{X}} \int_{\mathbb{X}} \frac{f(x)f(y)}{(\theta+f(x))(\theta+f(y))} \nu(dx)\nu(dy)$, we get

$$R_z(du)\nu(dz) + (f(u) - f(z))\nu(dz)\nu(du) = R_u(dz)\nu(du).$$

On the other hand, we have $R_u(dz)\nu(du) = \frac{\theta+f(u)}{\theta+f(z)}R_z(du)\nu(dz) + \theta \frac{f(u)-f(z)}{\theta+f(z)}\nu(dz)\nu(du)$ from (4.4), so substituting above,

$$R_z(du)\nu(dz) + (f(u) - f(z))\nu(dz)\nu(du) = \frac{\theta + f(u)}{\theta + f(z)}R_z(du)\nu(dz) + \theta \frac{f(u) - f(z)}{\theta + f(z)}\nu(dz)\nu(du),$$

from where

$$(f(u) - f(z))R_z(du)\nu(dz) = (f(u) - f(z))f(z)\nu(dz)\nu(du).$$

Therefore, working on H^c where $f(z) \neq f(u)$, we obtain

$$\mathbb{1}_{H^c}(z, u)R_z(du)\nu(dz) = \mathbb{1}_{H^c}(z, u)f(z)\nu(dz)\nu(du). \tag{4.11}$$

Going back to (4.7), we now focus on the triplets (x, y, z) in \mathbb{X}^3 for which $f(x) = f(y)$ and $f(x) \neq f(z)$, so taking the indefinite integral of the terms in (4.7) with respect to $\nu(dx)$, we get that, after some simple algebra,

$$\begin{aligned} & \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)\left\{\theta(f(z) - f(x))R_x(dy)\nu(dz)\nu(dx) + \theta(f(z) - f(x))R_x(dz)\nu(dy)\nu(dx) \right. \\ & \quad + (f(z) - f(x))R_x(dy)R_x(dz)\nu(dx) + \theta f(x)R_y(dz)\nu(dy)\nu(dx) \\ & \quad \left. + (\theta + f(x) + f(z))R_y(dz)R_x(dy)\nu(dx)\right\} \\ & = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)\left\{\theta f(x)R_z(dy)\nu(dz)\nu(dx) + (\theta + 2f(x))R_z(dy)R_x(dz)\nu(dx)\right\}. \end{aligned} \tag{4.12}$$

On the set where $f(y) = f(x) \neq f(z)$, equation (4.11) implies (i) $R_x(dz) = f(x)\nu(dz)$ for ν -a.e. x ; (ii) $R_y(dz) = f(y)\nu(dz) = f(x)\nu(dz)$ for ν -a.e. y ; and (iii) $R_z(dy) = f(z)\nu(dy)$ for ν -a.e. z . Moreover, by (4.2), (ii) implies for ν -a.e. x that (iv) $R_y(dz) = f(x)\nu(dz)$ for R_x -a.e. y . Thus,

$$\begin{aligned} & \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)R_x(dz)\nu(dy)\nu(dx) = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)f(x)\nu(dx)\nu(dy)\nu(dz) \quad \text{by (i),} \\ & \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)R_x(dy)R_x(dz)\nu(dx) = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)f(x)R_x(dy)\nu(dz)\nu(dx) \quad \text{by (i),} \\ & \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)R_y(dz)\nu(dy)\nu(dx) = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)f(x)\nu(dx)\nu(dy)\nu(dz) \quad \text{by (ii),} \\ & \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)R_y(dz)R_x(dy)\nu(dx) = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)f(x)R_x(dy)\nu(dz)\nu(dx) \quad \text{by (iv),} \\ & \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)R_z(dy)\nu(dz)\nu(dx) = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)f(z)\nu(dx)\nu(dy)\nu(dz) \quad \text{by (iii),} \end{aligned}$$

and, by (i) and (iii),

$$\mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)R_z(dy)R_x(dz)\nu(dx) = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)f(x)f(z)\nu(dx)\nu(dy)\nu(dz).$$

Substituting for the above terms in (4.12), we obtain

$$\begin{aligned} & \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)\left\{\theta(f(z) - f(x))R_x(dy)\nu(dx)\nu(dz) + \theta f(x)(f(z) - f(x))\nu(dx)\nu(dy)\nu(dz) \right. \\ & \quad + f(x)(f(z) - f(x))R_x(dy)\nu(dx)\nu(dz) + \theta f(x)^2\nu(dx)\nu(dy)\nu(dz) \\ & \quad \left. + f(x)(\theta + f(x) + f(z))R_x(dy)\nu(dx)\nu(dz)\right\} \\ & = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)\left\{\theta f(x)f(z)\nu(dx)\nu(dy)\nu(dz) + f(x)f(z)(\theta + 2f(x))\nu(dx)\nu(dy)\nu(dz)\right\}, \end{aligned}$$

which after simplification turns into

$$\begin{aligned} \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)\left\{f(z)(\theta + 2f(x))R_x(dy)\nu(dx)\nu(dz)\right\} \\ = \mathbb{1}_H(x, y)\mathbb{1}_{H^c}(x, z)\left\{f(x)f(z)(\theta + 2f(x))\nu(dx)\nu(dy)\nu(dz)\right\}. \end{aligned} \tag{4.13}$$

Dividing by $f(z)(\theta + 2f(x))$ and integrating z out leaves

$$\mathbb{1}_H(x, y)\nu(H_x^c)R_x(dy)\nu(dx) = \mathbb{1}_H(x, y)\nu(H_x^c)f(x)\nu(dx)\nu(dy). \tag{4.14}$$

Let us define

$$D^* := \{x \in \mathbb{X} : \nu(H_x^c) > 0\}.$$

Suppose that $\nu(D^*) > 0$. Take $y \in \mathbb{X}$. If $f(y) = f(x)$ for some $x \in D^*$, then $H_y^c = H_x^c$, and $\nu(H_y^c) = \nu(H_x^c) > 0$; otherwise, if $f(y) \neq f(x)$ for all $x \in D^*$, then $D^* \subseteq H_y^c$ and $\nu(H_y^c) \geq \nu(D^*) > 0$. In either case, $y \in D^*$, which implies that $\nu(D^*) = 1$. As a result, we can divide both sides of (4.14) by $\nu(H_x^c)$ to obtain

$$\mathbb{1}_{H_x}(y)R_x(dy)\nu(dx) = \mathbb{1}_{H_x}(y)f(x)\nu(dx)\nu(dy).$$

Summing this with (4.11), we get that, for ν -a.e. x ,

$$R_x(dy) = f(x)\nu(dy),$$

which implies from Proposition 3.1 that $(X_n)_{n \geq 1}$ is i.i.d.

If instead $\nu(D^*) = 0$, then $\nu(H_x^c) = 0$ for ν -a.e. x . It follows for some fixed x' with $\nu(H_{x'}) = 1$ that

$$f(x) = f(x') \quad \text{for } \nu\text{-a.e. } x,$$

implying that the model is balanced. This concludes the proof of the theorem. \square

Remark 4.1. The proof of Theorem 3.2 largely remains the same even if we assume $R_x(\mathbb{X}) = 0$ for all x in some $Z \in \mathcal{X}$ such that $0 < \nu(Z) < 1$. The only place of concern is in “dividing” (4.13) by $R_x(\mathbb{X})$, so by restricting (4.13) on Z^c we get

$$\mathbb{1}_H(x, y)\nu(H_x^c \cap Z^c)R_x(dy)\nu(dx) = \mathbb{1}_H(x, y)\nu(H_x^c \cap Z^c)f(x)\nu(dx)\nu(dy).$$

Now, letting

$$D^* := \{x \in \mathbb{X} : \nu(H_x^c \cap Z^c) > 0\},$$

we obtain under $\nu(D^* \cap Z^c) > 0$ that $(X_n)_{n \geq 0}$ is i.i.d., using the same logic as above. If on the other hand, $\nu(D^* \cap Z^c) = 0$, then $\nu((D^*)^c) = \nu(Z^c)$, since $(D^*)^c \subseteq Z^c$, and so

$$(\nu \times \nu)(H^c \cap (Z^c \times Z^c)) = \int_{Z^c} \nu(H_x^c \cap Z^c)\nu(dx) = \int_{(D^*)^c} \nu(H_x^c \cap Z^c)\nu(dx) = 0,$$

thus showing that $R_x(\mathbb{X}) = m > 0$ for ν -a.e. x in Z^c and some $m > 0$, when $(X_n)_{n \geq 1}$ is not i.i.d.

Proof of Theorem 3.7. If $(X_n)_{n \geq 1}$ is i.i.d., then Proposition 3.1 implies (3.2) with $\mathcal{G} = \{\emptyset, \mathbb{X}\}$. Suppose that $(X_n)_{n \geq 1}$ is exchangeable, but not i.i.d. It follows from Theorem 3.2 that the model is balanced, so we may assume, without loss of generality, that $R_x(\mathbb{X}) = 1$, for $x \in \mathbb{X}$; otherwise, we can normalize θ and R (see Remark 3.3) and proceed from there. In this case, (4.5) reduces to

$$R_x(dy)\nu(dx) = R_y(dx)\nu(dy), \tag{4.15}$$

and (4.8) implies, taking the indefinite integral of the terms with respect to $\nu(dx)$, that

$$R_y(dz)R_x(dy)\nu(dx) = R_z(dy)R_x(dz)\nu(dx). \tag{4.16}$$

Applying (4.15) and (4.16) repeatedly, we obtain that

$$\begin{aligned} R_y(dz)R_x(dy)\nu(dx) &= R_z(dy)R_x(dz)\nu(dx) = R_z(dy)R_z(dx)\nu(dz) \\ &= R_z(dx)R_y(dz)\nu(dy) = R_x(dz)R_y(dx)\nu(dy) = R_x(dz)R_x(dy)\nu(dx). \end{aligned}$$

As \mathcal{X} is countably generated, there exists $C_0 \in \mathcal{X}$ such that $\nu(C_0) = 1$ and, for all $x \in C_0$,

$$R_y(dz)R_x(dy) = R_x(dz)R_x(dy). \tag{4.17}$$

Define $C_1 := \{x \in C_0 : R_x(C_0) = 1\}$ and $C_n := \{x \in C_{n-1} : R_x(C_{n-1}) = 1\}$ for every $n \geq 2$. As $R_x(\mathbb{X}) = 1$, we get from (4.15) that

$$\begin{aligned} 1 &= \int_{C_0} R_x(\mathbb{X})\nu(dx) = \int_{\mathbb{X}} R_x(C_0)\nu(dx) \\ &= 1 - \int_{\mathbb{X}} (1 - R_x(C_0))\nu(dx) = 1 - \int_{C_1^c} (1 - R_x(C_0))\nu(dx). \end{aligned}$$

But $R_x(C_0) < 1$ for $x \in C_1^c$, so $\nu(C_1^c) = 0$, otherwise the term on the right-hand side of the equation becomes strictly less than 1. Proceeding by induction, we get $\nu(C_n) = 1$ for all $n \geq 1$; thus, letting $C^* = \bigcap_{n=1}^{\infty} C_n$, it holds $\nu(C^*) = 1$ and $R_x(C_n) = 1$, for all $x \in C^*$ and $n \geq 1$. As a result, for every $x \in C^*$, we have that $R_x(C^*) = 1$ and, since $C^* \subseteq C_0$, from (4.17),

$$R_x(dz) = R_y(dz) \quad \text{for } R_x\text{-a.e. } y. \tag{4.18}$$

Let us define

$$\mathcal{G} := \{A \in \mathcal{X} : R_x(A) = \delta_x(A) \text{ for all } x \in C^*\}.$$

It is not hard to see that \mathcal{G} is a σ -algebra. Next, fix $t \in [0, 1]$ and $B \in \mathcal{X}$. Denote

$$E_{t,B} = \{x \in \mathbb{X} : R_x(B) < t\}.$$

Let $x \in C^*$. If $x \in E_{t,B}$, then $R_y(B) = R_x(B) < t$ for R_x -a.e. y by (4.18), so $R_x(E_{t,B}) = 1$. If instead $x \in E_{t,B}^c$, then again $R_y(B) = R_x(B) \geq t$ for R_x -a.e. y by (4.18), and so $R_x(E_{t,B}) = 0$. As a result, $R_x(E_{t,B}) = \delta_x(E_{t,B})$, which implies that $E_{t,B} \in \mathcal{G}$; thus, $x \mapsto R_x(B)$ is \mathcal{G} -measurable.

Let $A \in \mathcal{G}$ and $B \in \mathcal{X}$. It follows from (4.15) and $\nu(C^*) = 1$ that

$$\int_A R_x(B)\nu(dx) = \int_B R_x(A)\nu(dx) = \int_{B \cap C^*} R_x(A)\nu(dx) = \int_B \delta_x(A)\nu(dx) = \nu(A \cap B).$$

Therefore, R is a regular version of $\nu(\cdot | \mathcal{G})$. Moreover, by construction,

$$R_x(A) = \delta_x(A) \quad \text{for all } A \in \mathcal{G} \text{ and } x \in C^*;$$

thus, R is a.e. proper and \mathcal{G} is c.g. under ν (see [3, p.649] and [5, Theorem 1]). □

Proof of Theorem 3.10. If $(X_n)_{n \geq 1}$ is i.i.d., then Proposition 3.1 implies (3.3) with respect to the partition $\{\mathbb{X}, \emptyset\}$, and vice versa. Suppose that $(X_n)_{n \geq 0}$ is exchangeable, but not i.i.d. Assume, without loss of generality, that $R_x(\mathbb{X}) = 1$ for all $x \in \mathbb{X}$. If R_x is of the form (3.3), then it is obvious that $R_x \ll \nu$ for ν -a.e. x .

Conversely, suppose that $R_x \ll \nu$ for ν -a.e. x . Then $R_x(dy)\nu(dx) \ll \nu(dx)\nu(dy)$, so there exists a jointly measurable function $r : \mathbb{X}^2 \rightarrow \mathbb{R}_+$ such that

$$R_x(dy)\nu(dx) = r(x, y)\nu(dx)\nu(dy),$$

and, since \mathcal{X} is countably generated, as measures,

$$R_x(dy) = r(x, y)\nu(dy) \quad \text{for } \nu\text{-a.e. } x. \tag{4.19}$$

Denote by

$$D := \{(x, y) \in \mathbb{X}^2 : r(x, y) > 0\},$$

and, letting D_x be the x -section of D for $x \in \mathbb{X}$, define

$$G := \{x \in \mathbb{X} : \nu(D_x) > 0\}.$$

Then $R_x(D_x) = \int_{D_x} r(x, y)\nu(dy) = 1$ for ν -a.e. x . Moreover,

$$1 = \int_{\mathbb{X}} R_x(D_x)\nu(dx) = \int_G \int_{D_x} r(x, y)\nu(dy)\nu(dx) + \int_{G^c} \int_{D_x} r(x, y)\nu(dy)\nu(dx) = \nu(G).$$

On the other hand, by (4.15), (4.16) and (4.17), we have the following identities

$$r(x, y)\nu(dx)\nu(dy) = r(y, x)\nu(dx)\nu(dy), \tag{4.20}$$

$$r(x, y)r(y, z)\nu(dy)\nu(dz) = r(x, z)r(z, y)\nu(dy)\nu(dz) \quad \text{for } \nu\text{-a.e. } x, \tag{4.21}$$

$$r(x, y)r(y, z)\nu(dx)\nu(dy)\nu(dz) = r(x, y)r(x, z)\nu(dx)\nu(dy)\nu(dz). \tag{4.22}$$

In particular, equations (4.20) and (4.21) imply that

$$\begin{aligned} \mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z)r(x, y)\nu(dx)\nu(dy)\nu(dz) &\stackrel{(a)}{=} \mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z)\mathbb{1}_{D_y}(z)r(x, y)\nu(dx)\nu(dy)\nu(dz) \\ &= \mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z)\mathbb{1}_{D_y}(z)r(x, y)\frac{r(y, z)}{r(y, z)}\nu(dx)\nu(dy)\nu(dz) \\ &= \mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z)\mathbb{1}_{D_y}(z)r(x, y)\frac{r(z, y)}{r(y, z)}\nu(dx)\nu(dy)\nu(dz) \\ &= \mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z)\mathbb{1}_{D_y}(z)r(x, z)\nu(dx)\nu(dy)\nu(dz) \\ &\stackrel{(b)}{=} \mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z)r(x, z)\nu(dx)\nu(dy)\nu(dz), \end{aligned} \tag{4.23}$$

where (a) and (b) follow from the fact that, for $(\nu \times \nu \times \nu)$ -a.e. (x, y, z) in \mathbb{X}^3 , if $y \in D_x$ and $z \in D_x$, then $r(x, y)r(x, z) > 0$, and so $r(x, y)r(y, z) > 0$ by (4.22), which implies that $r(y, z) > 0$ and, as a result, $\mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z) = \mathbb{1}_{D_x}(y)\mathbb{1}_{D_x}(z)\mathbb{1}_{D_y}(z)$.

Let $A, B \in \mathcal{X}$. Since $\nu(G) = 1$,

$$\begin{aligned} \int_A R_x(B)\nu(dx) &= \int_A \int_B \mathbb{1}_{D_x}(y) \cdot r(x, y)\nu(dy)\nu(dx) \\ &= \int_A \mathbb{1}_G(x) \left(\int_B \mathbb{1}_{D_x}(y) \cdot r(x, y) \frac{1}{\nu(D_x)} \left(\int_{\mathbb{X}} \mathbb{1}_{D_x}(z)\nu(dz) \right) \nu(dy) \right) \nu(dx) \\ &\stackrel{(a)}{=} \int_A \mathbb{1}_G(x) \left(\int_B \frac{1}{\nu(D_x)} \mathbb{1}_{D_x}(y) \left(\int_{\mathbb{X}} \mathbb{1}_{D_x}(z) \cdot r(x, z)\nu(dz) \right) \nu(dy) \right) \nu(dx) \\ &= \int_A \mathbb{1}_G(x) \left(\int_B \frac{1}{\nu(D_x)} \mathbb{1}_{D_x}(y)\nu(dy) \right) \nu(dx) \\ &= \int_A \nu(B|D_x)\nu(dx), \end{aligned} \tag{4.24}$$

where (a) follows from (4.23).

Given (4.24), the proof of Theorem 3.10 would be complete if we can find a countable partition among the sets D_x . As we demonstrate next, this is possible, provided we intersect each D_x with some a.s. sets.

First, observe that

$$\begin{aligned} 1 &= \int_{\mathbb{X}} \int_{\mathbb{X}} R_y(\mathbb{X}) R_x(dy) \nu(dx) = \int_{\mathbb{X}} \int_{D_x} \left(\int_{D_y} r(y, z) \nu(dz) \right) r(x, y) \nu(dy) \nu(dx) \\ &= \int_{\mathbb{X}} \int_{D_x} \left(\int_{D_x \cap D_y} r(y, z) \nu(dz) \right) r(x, y) \nu(dy) \nu(dx), \end{aligned} \quad (4.25)$$

where the last equality follows, similarly to (4.23), from the fact that $\mathbb{1}_{D_x}(y) \mathbb{1}_{D_y}(z) = \mathbb{1}_{D_x}(y) \mathbb{1}_{D_x}(z) \mathbb{1}_{D_y}(z)$ for $(\nu \times \nu \times \nu)$ -a.e. (x, y, z) in \mathbb{X}^3 . Since $\int_{D_x} r(x, y) \nu(dy) = 1$ for ν -a.e. x , the integrand in (4.25) w.r.t. $\nu(dx)$ is smaller or equal to 1, so that

$$\int_{D_x} r(x, y) \left(\int_{D_x \cap D_y} r(y, z) \nu(dz) \right) \nu(dx) = 1 \quad \text{for } \nu\text{-a.e. } x,$$

which itself implies that, for ν -a.e. y in D_x ,

$$\int_{D_x \cap D_y} r(y, z) \nu(dz) = 1.$$

But $\int_{D_y} r(y, z) \nu(dz) = 1$, thus $\nu(D_x^c \cap D_y) = 0$ for ν -a.e. y in D_x .

On the other hand,

$$\begin{aligned} 0 &= \int_{\mathbb{X}} \int_{D_x} \left(\int_{D_y^c} r(y, z) \nu(dz) \right) r(x, y) \nu(dy) \nu(dx) \\ &\stackrel{(a)}{=} \int_{\mathbb{X}} \int_{D_x} \left(\int_{D_y^c} r(x, z) \nu(dz) \right) r(x, y) \nu(dy) \nu(dx) \\ &\stackrel{(b)}{=} \int_{\mathbb{X}} \int_{D_x} \left(\int_{D_x \cap D_y^c} r(x, z) \nu(dz) \right) r(x, y) \nu(dy) \nu(dx), \end{aligned} \quad (4.26)$$

where (a) comes from (4.22), and (b) since $r(x, z) = 0$ for $z \in D_x^c$. Applying the same reasoning as in (4.25) to (4.26), we can show that, for ν -a.e. x , it holds $\nu(D_x \cap D_y^c) = 0$ for ν -a.e. y in D_x .

It follows from both results that, for some $E \in \mathcal{X} : \nu(E) = 1$ and every $x \in E$, there exists a further set $E_x \in \mathcal{X}$ such that $\nu(E_x) = 1$ and, for all $y \in E_x \cap D_x$, it holds

$$\nu(D_x^c \cap D_y) = 0 \quad \text{and} \quad \nu(D_x \cap D_y^c) = 0,$$

implying that

$$\nu(D_x) = \nu(D_x \cap D_y) = \nu(D_y). \quad (4.27)$$

Recall that $G = \{x \in \mathbb{X} : \nu(D_x) > 0\}$ and $\nu(G) = 1$. Let us define, for $x \in E$,

$$\bar{D}_x := E_x \cap D_x \quad \text{and} \quad \mathbb{X}_0 := G \cap E \cap \{x \in \mathbb{X} : R_x(dy) = r(x, y) \nu(dy)\}.$$

Then $\nu(\bar{D}_x) = \nu(D_x) > 0$ for all $x \in \mathbb{X}_0$, and $\nu(\mathbb{X}_0) = 1$. We will now show that the sets \bar{D}_x and \bar{D}_y , for different $x, y \in \mathbb{X}_0$, are either disjoint or their difference has ν -measure zero, which will allow us to create the aforementioned partition on \mathbb{X} .

Let $x, y \in \mathbb{X}_0$ be such that $\bar{D}_x \cap \bar{D}_y \neq \emptyset$. Pick $z \in \bar{D}_x \cap \bar{D}_y$. Then $z \in E_x$ and $z \in E_y$, so $\nu(D_x \cap D_z^c) = 0$ and $\nu(D_y^c \cap D_z) = 0$ from (4.27), which implies that

$$\begin{aligned} \nu(\bar{D}_x) &= \nu(D_x) = \nu(D_x \cap D_y) + \nu(D_x \cap D_y^c \cap D_z) + \nu(D_x \cap D_y^c \cap D_z^c) \\ &= \nu(D_x \cap D_y) = \nu(\bar{D}_x \cap \bar{D}_y), \end{aligned}$$

and, similarly,

$$\nu(\bar{D}_y) = \nu(\bar{D}_x \cap \bar{D}_y) = \nu(\bar{D}_x). \quad (4.28)$$

If there exists, in addition, $y' \in \mathbb{X}_0$ such that $\bar{D}_y \cap \bar{D}_{y'} \neq \emptyset$, then $\nu(\bar{D}_{y'}) = \nu(\bar{D}_y \cap \bar{D}_{y'})$ from (4.28) and

$$\begin{aligned} \nu(\bar{D}_x \cap \bar{D}_{y'}) &= \nu(\bar{D}_x \cap \bar{D}_y \cap \bar{D}_{y'}) \\ &= \nu(\bar{D}_x \cap \bar{D}_y \cap \bar{D}_{y'}) + \nu(\bar{D}_x^c \cap \bar{D}_y \cap \bar{D}_{y'}) = \nu(\bar{D}_y \cap \bar{D}_{y'}) = \nu(\bar{D}_{y'}) > 0; \end{aligned}$$

thus, $\bar{D}_x \cap \bar{D}_{y'} \neq \emptyset$. Therefore, we can partition \mathbb{X}_0 into equivalent classes according to whether $\bar{D}_x \cap \bar{D}_y \neq \emptyset$ or not, and we can pick one element from each class to create a family of subsets $\{\bar{D}_x\}_{x \in \mathbb{Y}}$, for some $\mathbb{Y} \subseteq \mathbb{X}_0$, such that $\bar{D}_x \cap \bar{D}_y = \emptyset$, for $x \neq y$, and $\nu(\bar{D}_x) > 0$, for all $x \in \mathbb{Y}$. In fact, \mathbb{Y} is at most countable as ν is finite.

If $x \in \mathbb{X}_0$, then $R_x \ll \nu$, and thus $R_x(\bar{D}_x) = R_x(D_x) = 1$; hence we get from (4.28) that, for every $x, y \in \mathbb{X}_0$,

$$R_x(\bar{D}_y) = R_x(\bar{D}_x \cap \bar{D}_y) = R_x(\bar{D}_x) = 1, \quad \text{if } \bar{D}_x \cap \bar{D}_y \neq \emptyset,$$

and

$$R_x(\bar{D}_y) = R_x(\bar{D}_x^c \cap \bar{D}_y) = 0, \quad \text{if } \bar{D}_x \cap \bar{D}_y = \emptyset.$$

The two equations together imply that $\sum_{x' \in \mathbb{Y}} R_x(\bar{D}_{x'}) = 1$ for each $x \in \mathbb{X}_0$, so using (4.26) and the fact that $\{\bar{D}_{x'}\}_{x' \in \mathbb{Y}}$ are disjoint,

$$\nu\left(\bigcup_{x' \in \mathbb{Y}} \bar{D}_{x'}\right) = \sum_{x' \in \mathbb{Y}} \nu(\bar{D}_{x'}) = \sum_{x' \in \mathbb{Y}} \int_{\mathbb{X}} R_x(\bar{D}_{x'}) \nu(dx) = \int_{\mathbb{X}_0} \sum_{x' \in \mathbb{Y}} R_x(\bar{D}_{x'}) \nu(dx) = \nu(\mathbb{X}_0) = 1.$$

Let $A, B \in \mathcal{X}$. Then

$$\begin{aligned} \int_A R_x(B) \nu(dx) &= \int_{A \cap (\bigcup_{x' \in \mathbb{Y}} \bar{D}_{x'})} R_x(B) \nu(dx) \\ &= \sum_{x' \in \mathbb{Y}} \int_{A \cap \mathbb{X}_0 \cap \bar{D}_{x'}} R_x(B) \nu(dx) \\ &\stackrel{(a)}{=} \sum_{x' \in \mathbb{Y}} \int_{A \cap \mathbb{X}_0 \cap \bar{D}_{x'}} \nu(B|D_x) \nu(dx) \\ &\stackrel{(b)}{=} \sum_{x' \in \mathbb{Y}} \int_{A \cap \mathbb{X}_0 \cap \bar{D}_{x'}} \nu(B|\bar{D}_{x'}) \nu(dx) = \int_A \sum_{x' \in \mathbb{Y}} \nu(B|\bar{D}_{x'}) \cdot \mathbb{1}_{\bar{D}_{x'}}(x) \nu(dx), \end{aligned}$$

where (a) follows from (4.24) with A replaced by $A \cap \mathbb{X}_0 \cap \bar{D}_{x'}$, and (b) from (4.27) since $\nu(\bar{D}_x^c \cap \bar{D}_{x'}) = 0 = \nu(\bar{D}_x \cap \bar{D}_{x'}^c)$ whenever $x \in \bar{D}_{x'} \cap \mathbb{X}_0$. Therefore, $R_x(B) = \sum_{x' \in \mathbb{Y}} \nu(B|\bar{D}_{x'}) \cdot \mathbb{1}_{\bar{D}_{x'}}(x)$ for ν -a.e. x and, as \mathcal{X} is countably generated, as measures,

$$R_x(\cdot) = \sum_{x' \in \mathbb{Y}} \nu(\cdot | \bar{D}_{x'}) \cdot \mathbb{1}_{\bar{D}_{x'}}(x) \quad \text{for } \nu\text{-a.e. } x.$$

□

Proof of Theorem 3.14. If $(X_n)_{n \geq 1}$ is i.i.d., then Proposition 3.1 implies that $R_x = R_x^a$, for ν -a.e. x . Suppose that $(X_n)_{n \geq 0}$ is exchangeable, but not i.i.d. Assume, without loss of generality, that $R_x(\mathbb{X}) = 1$ for $x \in \mathbb{X}$. By Lebesgue decomposition (Theorem 1 in [11]),

$$R_x = R_x^\perp + R_x^a,$$

for some finite transition kernels R_x^\perp and R_x^a on \mathbb{X} such that $R_x^\perp \perp \nu$ for some $S_x \in \mathcal{X}$ with $\nu(S_x) = 0 = R_x^\perp(S_x^c)$, and $R_x^a \ll \nu$, for $x \in \mathbb{X}$. Moreover,

$$R_x^\perp(\mathbb{X}) = R_x^\perp(S_x) = R_x(S_x) \quad \text{and} \quad R_x^a(\mathbb{X}) = R_x^a(S_x^c) = R_x(S_x^c). \quad (4.29)$$

Arguing as in (4.19), there exists some measurable function $r : \mathbb{X}^2 \rightarrow \mathbb{R}_+$ such that

$$R_x^a(dy) = r(x, y)\nu(dy) \quad \text{for } \nu\text{-a.e. } x. \tag{4.30}$$

In addition, we have from (4.15) that

$$R_x^\perp(dy)\nu(dx) + R_x^a(dy)\nu(dx) = R_y^\perp(dx)\nu(dy) + R_y^a(dx)\nu(dy), \tag{4.31}$$

and, from (4.16), for ν -a.e. x ,

$$\begin{aligned} R_y^\perp(dz)R_x^\perp(dy) + R_y^a(dz)R_x^\perp(dy) + R_y^\perp(dz)R_x^a(dy) + R_y^a(dz)R_x^a(dy) \\ = R_z^\perp(dy)R_x^\perp(dz) + R_z^a(dy)R_x^\perp(dz) + R_z^\perp(dy)R_x^a(dz) + R_z^a(dy)R_x^a(dz). \end{aligned} \tag{4.32}$$

Fix one such x . Since $\nu(S_x) = 0$, then $R_y^a(S_x) = 0$ for all $y \in \mathbb{X}$, and integrating (4.32) on $z \in S_x$ and $y \in \mathbb{X}$, we get

$$\begin{aligned} \int_{\mathbb{X}} R_y^\perp(S_x)R_x^\perp(dy) + \int_{\mathbb{X}} R_y^\perp(S_x)R_x^a(dy) \\ = \int_{S_x} R_z^\perp(\mathbb{X})R_x^\perp(dz) + \int_{S_x} R_z^a(\mathbb{X})R_x^\perp(dz) = R_x^\perp(\mathbb{X}), \end{aligned} \tag{4.33}$$

where we have used in the last equality that $R_x^\perp(S_x^c) = 0$. On the other hand, from (4.30),

$$\begin{aligned} \int_{\mathbb{X}} R_y^\perp(S_x)R_x^a(dy) &= \int_{S_x^c} R_y^\perp(S_x)R_x^a(dy) \\ &= \int_{y \in S_x^c} \int_{z \in S_x} r(x, y)R_y^\perp(dz)\nu(dy) \\ &\stackrel{(a)}{=} \int_{z \in S_x} \int_{y \in S_x^c} r(x, y)R_z^\perp(dy)\nu(dz) + \int_{z \in S_x} \int_{y \in S_x^c} r(x, y)R_z^a(dy)\nu(dz) \\ &\quad - \int_{y \in S_x^c} \int_{z \in S_x} r(x, y)R_y^a(dz)\nu(dy) \\ &\stackrel{(b)}{=} 0, \end{aligned}$$

where (a) follows from (4.31), and for (b) we have used again that $\nu(S_x) = 0$ and $R_y^a(S_x) = 0$ for all $y \in \mathbb{X}$.

Plugging this into (4.33), we get $\int_{\mathbb{X}} R_y^\perp(S_x)R_x^\perp(dy) = R_x^\perp(\mathbb{X})$ for ν -a.e. x , and thus

$$R_y^\perp(S_x) = 1 \quad \text{for } R_x^\perp\text{-a.e. } y. \tag{4.34}$$

Moreover, $1 = R_y^\perp(S_x) \leq R_y^\perp(\mathbb{X}) = R_y(S_y)$, so $R_y^a(\mathbb{X}) = 0$, by (4.29), and $R_y^a(dz)R_x^\perp(dy) = 0$ for ν -a.e. x , which implies the splitting of the measures

$$R_y(dz)R_x(dy)\nu(dx) = R_y^\perp(dz)R_x^\perp(dy)\nu(dx) + R_y(dz)R_x^a(dy)\nu(dx).$$

On the other hand, using the identity in (4.18), we obtain

$$R_y(dz)R_x(dy)\nu(dx) = R_x(dz)R_x(dy)\nu(dx) = R_x(dz)R_x^\perp(dy)\nu(dx) + R_x(dz)R_x^a(dy)\nu(dx).$$

Combining the two result, we get, for ν -a.e. x ,

$$R_y^\perp(dz)R_x^\perp(dy) + R_y(dz)R_x^a(dy) = R_x(dz)R_x^\perp(dy) + R_x(dz)R_x^a(dy).$$

But $R_x^\perp(dy) \perp R_x^a(dy)$ w.r.t. S_x and $R_x^\perp(S_x) = 1$, so it must be

$$R_y^\perp(dz)R_x^\perp(dy) = \mathbb{1}_{S_x}(y)R_y^\perp(dz)R_x^\perp(dy) = \mathbb{1}_{S_x}(y)R_x(dz)R_x^\perp(dy) = R_x(dz)R_x^\perp(dy).$$

It follows from this result, (4.29) and (4.34) that, for ν -a.e. x ,

$$\begin{aligned} R_x(S_x) &= R_x^\perp(S_x) = \int_{S_x} R_y^\perp(S_x) R_x^\perp(dy) \\ &= \int_{S_x} \int_{S_x} R_y^\perp(dz) R_x^\perp(dy) = \int_{S_x} R_x(S_x) R_x^\perp(dy) = R_x(S_x) R_x(S_x); \end{aligned}$$

thus, $R_x(S_x) \in \{0, 1\}$ for ν -a.e. x .

Let us define

$$\mathcal{S} := \{x \in \mathbb{X} : R_x^\perp(\mathbb{X}) = 1\},$$

which is \mathcal{X} -measurable since R_x^\perp is a transition kernel. By (4.29), we have $R_x(S_x) = R_x^\perp(\mathbb{X}) = 1$, for $x \in \mathcal{S}$. Therefore, using that $R_x(S_x) \in \{0, 1\}$ for ν -a.e. x , we get for ν -a.e. x that: if $x \in \mathcal{S}$, then $R_x = R_x^\perp$; if $x \in \mathcal{S}^c$, then $R_x = R_x^a$. Moreover, (4.29) and (4.34) imply for ν -a.e. x that $y \in \mathcal{S}$ for R_x^\perp -a.e. y ; in other words, $R_x^\perp(\mathcal{S}^c) = 0$ for ν -a.e. x . Finally, using (4.15) and the fact that $R_x^a = 0$ for ν -a.e. $x \in \mathcal{S}$ and $R_x^\perp = 0$ for ν -a.e. $x \in \mathcal{S}^c$, we get

$$\begin{aligned} \int_{\mathbb{X}} R_x^a(\mathcal{S}) \nu(dx) &= \int_{\mathcal{S}} R_x^a(\mathcal{S}) \nu(dx) + \int_{\mathcal{S}^c} R_x^a(\mathcal{S}) \nu(dx) = \int_{\mathcal{S}^c} R_x(\mathcal{S}) \nu(dx) \\ &= \int_{\mathcal{S}} R_x(\mathcal{S}^c) \nu(dx) = \int_{\mathcal{S}} R_x^\perp(\mathcal{S}^c) \nu(dx) + \int_{\mathcal{S}^c} R_x^\perp(\mathcal{S}^c) \nu(dx) = \int_{\mathbb{X}} R_x^\perp(\mathcal{S}^c) \nu(dx) = 0, \end{aligned}$$

and so $R_x^a(\mathcal{S}) = 0$ for ν -a.e. x . □

Proof of Proposition 3.17. Since $R_x \perp \nu$, then $(X_n)_{n \geq 1}$ is not i.i.d., and hence not unbalanced by Theorem 3.2, so from Corollary 3.4 we may assume that $R_x(\mathbb{X}) = 1$, without loss of generality.

Let us define, for $x \in \mathbb{X}$,

$$D_x := \{y \in \mathbb{X} : R_x(\{y\}) > 0\} \quad \text{and} \quad D := \{x \in \mathbb{X} : x \in D_x\}.$$

Then $R_x(\cdot) = \sum_{y \in D_x} p_x(y) \delta_y(\cdot)$ for some $p_x(y) > 0$ such that $\sum_{y \in D_x} p_x(y) = 1$.

It follows from (4.18) and the form of R_x that there exists $C \in \mathcal{X}$ such that $\nu(C) = 1$ and, for all $x \in C$,

$$R_x(dz) = R_y(dz) \quad \text{for } y \in D_x;$$

thus, $D_x = D_y$ for all $y \in D_x$, which implies that $y \in D_y$, i.e. $y \in D$. Therefore, by (4.15),

$$\nu(D) = \int_{\mathbb{X}} R_x(D) \nu(dx) = \int_C \left(\sum_{y \in D_x} p_x(y) \delta_y(D) \right) \nu(dx) = \int_C \left(\sum_{y \in D_x} p_x(y) \right) \nu(dx) = 1.$$

For the second part, let $x, y \in C \cap D$ be such that $y \notin D_x$. Assume $D_x \cap D_y \neq \emptyset$. Take $z \in D_x \cap D_y$. It follows from above that $D_x = D_z = D_y$. But $y \in D$, so $y \in D_y = D_x$, absurd. As a result, for ν -a.e. x and y , either $D_x = D_y$ or $D_x \cap D_y = \emptyset$. □

5 Discussion

The results in this paper allow us to state some universal facts about exchangeable MVPSs that were previously unknown: 1) all models are necessarily balanced; 2) the normalized reinforcement kernels are a.e. proper regular conditional distributions; 3) the prior distributions have the stick-breaking construction of a Dirichlet process. When \mathbb{X} is countable or the reinforcement is dominated by ν , then every MVPS is a Dirichlet process mixture model over a family of probability distributions with disjoint supports derived from ν . Relaxing parts of this structure while retaining exchangeability would

lead to a very different sampling scheme from the one that (1.2) entails. On the other hand, the fact that, for fixed x , R_x is either absolutely continuous or mutually singular with respect to ν , though surprising at first, seems very natural under exchangeability. Therefore, we expect that R can be further decomposed along other lines, e.g., R_x^\perp might be discrete only for some x and diffuse for the rest.

References

- [1] Aldous, D. J.: Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII 1983* **1117**, (1985), 1–198. MR0883646
- [2] Bandyopadhyay, A. and Thacker, D.: A new approach to Pólya urn schemes and its infinite color generalization. *Ann. Appl. Probab.* **32**(1), (2022), 46–79. MR4386521
- [3] Berti, P. and Rigo, P.: 0–1 laws for regular conditional distributions. *Ann. Probab.* **35**(2), (2007), 649–662. MR2308591
- [4] Berti, P., Dreassi, E., Leisen, F., Pratelli, L. and Rigo, P.: Kernel based Dirichlet sequences. *Bernoulli* **29**(2), (2023), 1321–1342. MR4550225
- [5] Blackwell, D. and Dubins, L. E.: On existence and non-existence of proper, regular, conditional distributions. *Ann. Probab.* **3**(5), (1975), 741–752. MR0400320
- [6] Blackwell, D. and MacQueen, J. B.: Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**(2), (1973), 353–355. MR0362614
- [7] Fortini, S. and Petrone, S.: Predictive construction of priors in Bayesian nonparametrics. *Braz. J. Probab. Stat.* **26**(4), (2012), 423–449. MR2949087
- [8] Fortini, S., Ladelli, L. and Regazzini, E.: Exchangeability, predictive distributions and parametric models. *Sankhyā Ser. A* **62**(1), (2000), 86–109. MR1769738
- [9] Fortini, S., Petrone, S. and Sariev, H.: Predictive constructions based on measure-valued Pólya urn processes. *Mathematics* **9**, (2021), 2845.
- [10] Janson, S.: Random replacements in Pólya urns with infinitely many colours. *Electron. Commun. Probab.* **24**, (2019), paper no. 23, 11 pp. MR3953824
- [11] Lange, K.: Decompositions of substochastic transition functions. *Proc. Amer. Math. Soc.* **37**(2), (1973), 575–580. MR0314124
- [12] Lijoi, A. and Prünster, I.: Models beyond the Dirichlet process. In Bayesian Nonparametrics, N. L. Hjort, C. Holmes, P. Müller and S. G. Walker, eds. *Cambridge University Press*, Cambridge, UK, 2010. 80–136. MR2730661
- [13] Mahmoud, H. M.: Pólya Urn Models. *CRC Press*, Boca Raton, FL, 2009. MR2435823
- [14] Mailler, C. and Marckert, J.-F.: Measure-valued Pólya urn processes. *Electron. J. Probab.* **22**, (2017), paper no. 26, 33 pp. MR3629870
- [15] Mailler, C. and Villemonais, D.: Stochastic approximation on non-compact measure spaces and application to measure-valued Pólya processes. *Ann. Appl. Probab.* **30**(5), (2020), 2393–2438. MR4149532
- [16] Pemantle, R.: A survey of random processes with reinforcement. *Probab. Surv.* **4**, (2007), 1–79. MR2282181
- [17] Sariev, H., Fortini, S. and Petrone, S.: Infinite-color randomly reinforced urns with dominant colors. *Bernoulli* **29**(1), (2023), 132–152. MR4497242

Acknowledgments. We are grateful to an Associate Editor and two anonymous referees for their valuable comments and helpful suggestions, and for pointing out inaccuracies in an earlier version of this work.

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System: <https://vtex.lt/services/ejms-peer-review/>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <https://imstat.org/shop/donation/>