Check for updates

J. Appl. Probab. **62,** 925–949 (2025)
*doi:*10.1017/jpr.2024.109

SUFFICIENTNESS POSTULATES FOR MEASURE-VALUED PÓLYA URN SEQUENCES

HRISTO SARIEV,* *Bulgarian Academy of Sciences*
MLADEN SAVOV,** *Sofia University 'St. Kliment Ohridski'*

Abstract

In a recent paper, the authors studied the distribution properties of a class of exchangeable processes, called measure-valued Pólya sequences (MVPSs), which arise as the observation process in a generalized urn sampling scheme. Here we present several results in the form of 'sufficientness' postulates that characterize their predictive distributions. In particular, we show that exchangeable MVPSs are the unique exchangeable models whose predictive distributions are a mixture of the marginal distribution and the average of a probability kernel evaluated at past observations. When the latter coincides with the empirical measure, we recover a well-known result for the exchangeable model with a Dirichlet process prior. In addition, we provide a 'pure' sufficientness postulate for exchangeable MVPSs that does not assume a particular analytic form for the predictive distributions. Two other sufficientness postulates consider the case when the state space is finite.

Keywords: Pólya urns; predictive distributions; Bayesian nonparametrics; exchangeability; sufficientness postulate

2010 Mathematics Subject Classification: Primary 60G25; 60G09
Secondary 62G99; 62B99

1. Introduction

The exchangeable model with a Dirichlet process prior is a cornerstone of Bayesian nonparametric analysis and has inspired much of the subsequent work in the area (for an overview, see [22]). One remarkable feature of the Dirichlet process, due to [26] and later [23], is that it is the unique prior for which the predictive distributions of the underlying exchangeable process are a mixture of the marginal distribution and the empirical measure. More specifically, let $(X_n)_{n\geq 1}$ be a sequence of random variables taking values in some standard space, say $\mathbb{X} = [0, 1]$, and suppose that $(X_n)_{n\geq 1}$ is exchangeable, i.e. the order of $X_1, X_2, \ldots$ is probabilistically irrelevant (refer to Section 2 for a rigorous treatment). Denote by $\nu(\cdot) = \mathbb{P}(X_1 \in \cdot\,)$ the marginal distribution of the process. If, for each $n = 1, 2, \ldots$, the predictive distribution of X_{n+1} given $(X_1, \ldots, X_n)$ is the convex combination

Received 19 November 2023; accepted 17 November 2024.
* Postal address: Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 8 Acad. Georgi Bonchev Str., Sofia 1113, Bulgaria. Email: h.sariev@math.bas.bg
** Postal address: Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", 5 James Bourchier Blvd, Sofia 1164, Bulgaria. Email: msavov@fmi.uni-sofia.bg

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = (1 - a_n)\nu(\cdot) + a_n \cdot \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\cdot) \tag{1.1}$$

for some constant $a_n \in (0, 1)$, where δ_x is the unit mass at x, then [23, Theorem 1] implies that $a_n = n/(n + \theta)$, with $\theta = a_1^{-1} - 1$, and $(X_n)_{n\geq 1}$ has a Dirichlet process prior with parameters (θ, ν). The well-known Blackwell–MacQueen urn model [7] assumes (1.1) directly with $a_n = n/(n + \theta)$.

Predictive constructions of the type (1.1) have several attractive features. For one, they provide a simple and efficient way to simulate the sequence $X_1, X_2, \ldots$ Indeed, (1.1) leads to a recursive procedure

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = \frac{\theta + n - 1}{\theta + n}\, \mathbb{P}(X_n \in \cdot \mid X_1, \ldots, X_{n-1}) + \frac{1}{\theta + n}\, \delta_{X_i}(\cdot),$$

which allows for fast computations and limited storage of information for streaming data (see [12, Section 5] for more details on recursive algorithms and predictions). On the other hand, the system of predictive distributions $(\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n))_{n\geq 0}$ determines the law of the process $\mathbb{P}((X_1, X_2, \ldots) \in \cdot\,)$ without the need for prior elicitation. Such a predictive approach to Bayesian modeling, focused on observable quantities instead of unknown parameters, is deeply rooted in the philosophical foundations of Bayesian analysis [11, 12], and has recently enjoyed renewed interest; see, e.g., [5, 9, 12].

A particular class of predictive characterizations, to which (1.1) belongs (see below), are the so-called postulates of *sufficientness*, dating back to the writings of Cambridge philosopher W. E. Johnson [19] (for more historical details, see [29] and [14, Chapter 5]). Unlike the usual notions of sufficiency (see, e.g., [10, Section 5]), sufficientness postulates provide conditions on the functional form of the predictive distributions arising from natural judgments on the inductive process of learning. In other words, model specification is based on general assessments about the form and type of sample information used to make predictions. As an example, the original Johnson's sufficientness postulate assumes an exchangeable multinomial model on $\mathbb{X} = \{1, \ldots, k\}$ with k types such that, for each $n = 1, 2, \ldots$, the probability of observing type $j \in \mathbb{X}$ on the $(n + 1)$th trial,

$$\mathbb{P}(X_{n+1} = j \mid X_1, \ldots, X_n) = f_{n,j}(N_{n,j}), \tag{1.2}$$

depends on the sample $(X_1, \ldots, X_n)$ only by the number of times $N_{n,j}$ that j has been observed in the past. By [29, Theorem 2.1], any exchangeable process with predictive distributions (1.2) has a Dirichlet distribution prior (see also Proposition 3.1 in Section 3). Similar considerations about the relevance of different pieces of sample information have led to sufficientness postulates for the two-parameter Poisson–Dirichlet process [31], Gibbs-type priors [2], and feature-sampling models [8]. Sufficientness postulates have also been proposed for neutral to the right priors [28] and recurrent Markov exchangeable sequences [30].

Within this paradigm, the predictive rule (1.1) can be shown to be motivated by the same principle behind (1.2), extended to more general state spaces $\mathbb{X}$, without necessarily assuming that the number of observed types is limited or known in advance. To that end, let K_n and $X_1^*, \ldots, X_{K_n}^*$ denote, respectively, the number and the distinct values of $(X_1, \ldots, X_n)$ in *order of appearance*. Following [31], we assume that the probability of observing the jth value to appear,

$$\mathbb{P}(X_{n+1} = X_j^* \mid X_1, \ldots, X_n) = p_n(N_{n,j}), \tag{1.3}$$

is a function of the number of times it has been previously observed, while the probability of observing (discovering) a new type depends only on the sample size,

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

$$\mathbb{P}(X_{n+1} \notin \{X_1^*, \ldots, X_{K_n}^*\} \mid X_1, \ldots, X_n) = r_n. \tag{1.4}$$

Note that unlike (1.2), the function p_n does not change depending on j, which is a consequence of the framework, since type j is not known to exist in advance [31, p. 381, Footnote 5]. Moreover, the original proposal in [31] allows $r_n = r_n(K_n)$ to depend further on the number of distinct values, leading to a general characterization result for the two-parameter Poisson–Dirichlet process. On the other hand, the model in (1.3) and (1.4) describes an (exchangeable) random partition process (in fact, the Chinese restaurant process and the related Ewens' partition structure [31, p. 380, Remarks]), but does not discuss the mechanism of 'labeling' $X_1^*, X_2^*, \ldots$ Assuming that their values are assigned independently, from a *diffuse* distribution ν, i.e. $\nu(\{x\}) = 0$ for all $x \in \mathbb{X}$, we can express (1.3) and (1.4) jointly as

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = \sum_{j=1}^{K_n} p_n(N_{n,j}) \delta_{X_j^*}(\cdot) + r_n \nu(\cdot). \tag{1.5}$$

Exchangeable processes with predictive distributions of the kind (1.5) belong to the class of *species sampling sequences* [16, 25] and are closely related to the theory of exchangeable random partitions (for an excellent overview, see [15, Chapter 14]). It follows from (1.5) that, for any $k = 1, \ldots, n$ and all counts $n_1, \ldots, n_k \in \{1, \ldots, n\}$ such that $\sum_{j=1}^{k} n_j = n$,

$$\sum_{j=1}^{k} p_n(n_j) + r_n = 1;$$

thus, assuming that every partition of $(X_1, \ldots, X_n)$ by type has a positive probability of occurring, we can easily deduce that $p_n(n_j) = (1 - r_n) n_j / n$. As a consequence,

$$\begin{aligned} \mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) &= (1 - r_n) \cdot \frac{1}{n} \sum_{j=1}^{K_n} N_{n,j} \delta_{X_j^*}(\cdot) + r_n \nu(\cdot) \\ &= (1 - r_n) \cdot \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\cdot) + r_n \nu(\cdot), \end{aligned}$$

suggesting that the analytic form in (1.1) arises from a 'pure' sufficientness principle.

An important limitation of (1.1) and species sampling models, in general, is that they are designed to work with categorical data, taking into account only the frequencies of the observed types. In this work, we study an extension of (1.1) by replacing δ_x with a general probability measure R_x, so that

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = (1 - a_n) \nu(\cdot) + a_n \cdot \frac{1}{n} \sum_{i=1}^{n} R_{X_i}(\cdot) \tag{1.6}$$

for some constant $a_n \in (0, 1)$. Depending on the particular R, the predictive rule (1.6) can more efficiently exploit sample information coming from continuous data by considering, for example, where each observation falls inside $\mathbb{X}$. In fact, for diffuse R and ν, all $X_1, \ldots, X_n$ will be distinct, so notions like random partition and species discovery become practically meaningless. Our goal is to characterize the class of exchangeable processes that are governed by (1.6), and thus motivate their use in Bayesian nonparametric analysis from an inductive learning perspective.

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Notice that when $a_n = n/(n+\theta)$ for some $\theta > 0$, (1.6) becomes

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = \frac{\theta \nu(\cdot) + \sum_{i=1}^{n} R_{X_i}(\cdot)}{\theta + n},$$

and, following [27], the process $(X_n)_{n\geq 1}$ is a member of the class of *measure-valued Pólya sequences* (MVPSs). We might expect from (1.1), where $R_x = \delta_x$, that the only exchangeable processes having predictive rules (1.6) will be exchangeable MVPSs, in which case [27] provides a complete description of their prior distributions. In particular, it would follow from [27, Theorem 3.13] that when $\mathbb{X}$ is countable or R is dominated by ν, the exchangeable sequence $(X_n)_{n\geq 1}$ has a Dirichlet process mixture prior (see Section 2.2). Our main result, Theorem 3.1, states that this is indeed the case, but the proof is fundamentally different from that for (1.1), making extensive use of the measure-valued nature of (1.6) as the theory of random partitions is generally not available.

The rest of the paper is structured as follows. In the next section, we introduce MVPSs and provide notation that will be used later on. Theorem 3.1 is stated in Section 3 along with a 'pure' sufficientness postulate for exchangeable MVPSs that does not assume a particular analytic form for the predictive distributions, as does (1.6). In addition, we provide two other sufficientness postulates that generalize (1.2) for the case when $\mathbb{X}$ is finite. Proofs are postponed to Section 4.

2. The model

2.1. Preliminaries

From now on, $\mathbb{X}$ is a complete separable metric space endowed with the Borel σ-algebra $\mathcal{X}$, which from standard results is countably generated. All random quantities are defined on a common probability space $(\Omega, \mathcal{H}, \mathbb{P})$ unless otherwise specified. For all unexplained measure-theoretic details, we refer to [20].

Given two arbitrary measures μ, π on $\mathbb{X}$, we say that μ is *absolutely continuous* with respect to π, denoted $\mu \ll \pi$, if $\mu(A) = 0$ whenever $\pi(A) = 0$ for any $A \in \mathcal{X}$, and we say that μ and π are *mutually singular*, denoted $\mu \perp \pi$, if there exists $S \in \mathcal{X}$ such that $\mu(S) = 0 = \pi(S^c)$.

A *transition kernel on* $\mathbb{X}$ is a function $R\colon \mathbb{X} \times \mathcal{X} \to \mathbb{R}_+$ such that (i) the map $x \mapsto R(x, B) \equiv R_x(B)$ is $\mathcal{X}$-measurable, for all $B \in \mathcal{X}$; and (ii) the map $B \mapsto R_x(B)$ is a measure on $\mathbb{X}$, for all $x \in \mathbb{X}$. In addition, R is said to be *finite* if $R_x(\mathbb{X}) < \infty$, and is called a *probability kernel* if $R_x(\mathbb{X}) = 1$ for all $x \in \mathbb{X}$.

Let ν be a probability measure on $\mathbb{X}$, and $\mathcal{G} \subseteq \mathcal{X}$ a sub-σ-algebra on $\mathbb{X}$. A probability kernel R on $\mathbb{X}$ 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

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

if (i) the map $x \mapsto R_x(B)$ is $\mathcal{G}$-measurable, for all $B \in \mathcal{X}$; and (ii) $\int_A R_x(B)\, \nu(\mathrm{d}x) = \nu(A \cap B)$ for all $A \in \mathcal{G}$ and $B \in \mathcal{X}$. The assumptions on $(\mathbb{X}, \mathcal{X})$ guarantee that an r.c.d. for ν given $\mathcal{G}$ exists and is unique up to a ν-null set, and all conditional distributions must be understood as regular versions. Moreover, a regular version R of $\nu(\cdot \mid \mathcal{G})$ is called *almost everywhere proper* if there exists $F \in \mathcal{G}$ such that $\nu(F) = 1$ and

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

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Following [6, p. 649], an almost everywhere proper r.c.d. for ν given $\mathcal{G}$ exists if and only if $\mathcal{G}$ is *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.

An $\mathbb{X}$-valued sequence $(X_n)_{n\geq 1}$ of random variables is said to be *exchangeable* if, for each $n = 1, 2, \ldots$ and all permutations σ of $\{1, \ldots, n\}$,

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

In that case, by the de Finetti representation theorem for exchangeable sequences (see, e.g., [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 $X_1, X_2, \ldots$ are conditionally independently and identically distributed (i.i.d.) according to $\tilde{P}$,

$$\begin{aligned} X_n \mid \tilde{P} &\overset{\text{i.i.d.}}{\sim} \tilde{P}, \\ \tilde{P} &\sim Q, \end{aligned}$$

where Q is the (nonparameteric) *prior* distribution of $\tilde{P}$. Moreover, $\tilde{P}$ is the almost sure weak limit of the empirical measure,

$$1/n \sum_{i=1}^{n} \delta_{X_i} \xrightarrow{\text{w}} \tilde{P} \quad \text{almost surely (a.s.)},$$

and the predictive distributions of $(X_n)_{n\geq 1}$,

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) \xrightarrow{\text{w}} \tilde{P} \quad \text{(a.s.)}.$$

The most popular prior distribution in Bayesian nonparametric analysis is arguably the *Dirichlet process*, $\mathrm{DP}(\theta, \nu)$, which is parameterized by a probability measure ν on $\mathbb{X}$ and a constant $\theta > 0$. We recall that $\tilde{P} \sim \mathrm{DP}(\theta, \nu)$ if, for every finite partition $\{B_1, \ldots, B_n\} \subseteq \mathcal{X}$ of $\mathbb{X}$, the random vector $(\tilde{P}(B_1), \ldots, \tilde{P}(B_n))$ has a Dirichlet distribution with parameters $(\theta\nu(B_1), \ldots, \theta\nu(B_n))$. Equivalently (see, e.g., [15, Theorem 4.12]), the Dirichlet process is the distribution of an a.s. discrete random probability measure with so-called stick-breaking weights,

$$\sum_{j=1}^{\infty} V_j \delta_{Z_j} \sim \mathrm{DP}(\theta\nu), \tag{2.1}$$

where $V_j = W_j \prod_{i=1}^{j-1} W_i$, and $W_1, W_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathrm{Beta}(1, \theta)$ and $Z_1, Z_2, \ldots \overset{\text{i.i.d.}}{\sim} \nu$ are independent random variables.

Exchangeable processes having directing random measures $\tilde{P}$ as in (2.1), for arbitrary sequences of nonnegative random weights $(V_j)_{j\geq 1}$ and independent atoms $(Z_j)_{j\geq 1}$ coming from a diffuse distribution ν, are called (proper) *species sampling sequences* and have predictive distributions of the form (1.5) for more general functions p_n and r_n; see [16, 25] for a complete characterization. Because of the discrete nature of their $\tilde{P}$ and the fact that the atoms $Z_1, Z_2, \ldots$ are a.s. all distinct, species sampling models are, by Kingman's theory, in one-to-one correspondence with exchangeable random partitions, up to a specification of ν (for more information, we refer to [25] and [15, Chapter 14]). In fact, the exchangeable partition generated by a random sample from a Dirichlet process (2.1) with diffuse ν is the well-known Chinese restaurant process.

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

2.2. Exchangeable MVPS

Following [27], we call any sequence $(X_n)_{n\geq 1}$ of $\mathbb{X}$-valued random variables a *measure-valued Pólya (urn) sequence* with parameters θ, ν, and R, denoted MVPS(θ, ν, R), if $\nu(\cdot) = \mathbb{P}(X_1 \in \cdot\,)$ and, for each $n = 1, 2, \ldots$,

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = \frac{\theta\nu(\cdot) + \sum_{i=1}^{n} R_{X_i}(\cdot)}{\theta + \sum_{i=1}^{n} R_{X_i}(\mathbb{X})}, \tag{2.2}$$

where R is a finite transition kernel on $\mathbb{X}$, called the *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. When $R_x(\mathbb{X}) = m$ for ν-a.e. x and some $m > 0$, both R and the corresponding MVPS are said to be *balanced*.

The predictive structure (2.2) admits a physical interpretation in terms of a generalized urn sampling scheme as follows. Suppose initially that we have an urn whose contents are described by the measure $\theta\nu$ in the sense that, for each $B \in \mathcal{X}$, the quantity $\theta\nu(B)$ records the total *mass* of balls whose colors lie in B. At time $n = 1$, a ball is drawn at random from the urn with probability distribution ν. Depending on its color, X_1, we reinforce the urn with additional balls according to the measure R_{X_1}, so that the updated composition becomes $\theta\nu + R_{X_1}$. At each subsequent time $n > 1$, a ball with color X_n is drawn with probability proportional to $\theta\nu + \sum_{i=1}^{n-1} R_{X_i}$, and the contents of the urn get 'reinforced' by R_{X_n}. The model with $|\mathbb{X}| = k$ and $R_x = \delta_x$, i.e. we only add one additional ball of the observed color, corresponds to the classical k-color Pólya urn scheme, and measure-valued Pólya urn processes have been proposed by [3, 13, 18, 24] as extensions to infinite colors and more general reinforcement mechanisms. The case where the observation process $(X_n)_{n\geq 1}$ is exchangeable has been investigated by [27], and the main results from that paper are summarized in the following theorem.

Theorem 2.1. ([27, Theorems 3.2, 3.7, and 3.9].) *Let $(X_n)_{n\geq 1}$ be an exchangeable, but not i.i.d., MVPS(θ, ν, R) with directing random measure $\tilde{P}$. Then:*

(i) *R is balanced and there exists a (c.g. under ν) sub-σ-algebra $\mathcal{G}$ of $\mathcal{X}$ such that, normalized, R is an r.c.d. for ν given $\mathcal{G}$,*

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

(ii) *$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n)$ converges a.s. in total variation, as $n \to \infty$, to $\tilde{P}$;*

(iii) *$\tilde{P}$ is equal in law to*

$$\sum_{j=1}^{\infty} V_j \frac{R_{Z_j}(\cdot)}{R_{Z_j}(\mathbb{X})}, \tag{2.3}$$

where $(V_j)_{j\geq 1}$ and $(Z_j)_{j\geq 1}$ are given in (2.1).

Independently from the theory on measure-valued urn processes, [4] introduced the *kernel-based Dirichlet sequence*, which is a balanced MVPS with reinforcement kernel $R(\cdot) = \nu(\,\cdot \mid \mathcal{G})$, for some sub-$\sigma$-algebra $\mathcal{G}$ of $\mathcal{X}$. In [4], the authors prove that kernel-based Dirichlet sequences are exchangeable, having directing random measures (2.3). Theorem 2.1 provides a converse statement and shows that, under exchangeability, R need not be assumed to be balanced a priori. In fact, it follows from Theorem 2.1 that any exchangeable MVPS(θ, ν, R) can

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

be reparameterized as a balanced exchangeable MVPS($\tilde{\theta}$, ν, $\tilde{R}$) for some probability kernel $\tilde{R}$ and constant $\tilde{\theta}$, so that

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = \frac{\tilde{\theta}}{\tilde{\theta}+n}\nu(\cdot) + \frac{n}{\tilde{\theta}+n} \cdot \frac{1}{n}\sum_{i=1}^{n} \tilde{R}_{X_i}(\cdot). \tag{2.4}$$

Therefore, exchangeable MVPSs satisfy the sufficientness postulate (1.6), and in Section 3 we show that the opposite statement is also true. On the other hand, by Theorem 2.1(iii), the directing random measure $\tilde{P}$ of any MVPS inherits its properties directly from the reinforcement kernel $\tilde{R}$, so that $\tilde{P}$ is a.s. diffuse and/or absolutely continuous with respect to some reference probability measure whenever $\tilde{R}$ is. This is in stark contrast to the discrete nature of the Dirichlet process (2.1) and species sampling models, in general, all while keeping the same simple form of the predictive distributions (2.4).

Theorem 2.1 can be refined when $\mathbb{X}$ is countable or R is dominated by ν, in which case exchangeable MVPSs turn out to have Dirichlet process mixture priors.

Theorem 2.2. *([27, Theorems 3.10 and 3.13].) Let $(X_n)_{n\geq 1}$ be as in Theorem 2.1. If $R_x \ll \nu$ for ν-a.e. x, then:*

(i) *there exists a countable partition $D_1, D_2, \ldots$ of $\mathbb{X}$ in $\mathcal{X}$ such that, for ν-a.e. x,*

$$\frac{R_x(\cdot)}{R_x(\mathbb{X})} = \sum_{k\geq 1} \nu(\,\cdot \mid D_k) \cdot \mathbf{1}_{D_k}(x);$$

(ii) *letting $m = R_x(\mathbb{X})$,*

$$\begin{aligned} X_n \mid \tilde{P} &\overset{\text{i.i.d.}}{\sim} \sum_{k\geq 1} \tilde{P}(D_k)\, \nu(\,\cdot \mid D_k), \\ (\tilde{P}(D_k))_{k\geq 1} &\sim \mathrm{DP}(\theta/m, \nu). \end{aligned}$$

3. Results

Theorem 3.1, together with the discussion around (2.4), states that the family of exchangeable MVPSs coincides with the class of exchangeable processes satisfying the sufficientness postulate (1.6).

Theorem 3.1. *Let $(X_n)_{n\geq 1}$ be an exchangeable process with marginal distribution ν and predictive distributions (1.6), that is, for each $n = 1, 2, \ldots$,*

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = (1-a_n)\nu(\cdot) + a_n \cdot \frac{1}{n}\sum_{i=1}^{n} R_{X_i}(\cdot), \tag{3.1}$$

where R is a probability kernel on $\mathbb{X}$, and $a_n \in (0, 1)$. Then $a_n = n/(a_1^{-1} - 1 + n)$ for all $n \geq 1$. Moreover, $(X_n)_{n\geq 1}$ is an MVPS with parameters $(a_1^{-1} - 1, \nu, R)$.

Remark 3.1. Since $(X_n)_{n\geq 1}$ in Theorem 3.1 is an exchangeable MVPS, Theorems 2.1 and 2.2 provide additional guidance for the choice of the parameter R. In particular, there exists a (c.g. under ν) sub-σ-algebra $\mathcal{G}$ of $\mathcal{X}$ such that R is an r.c.d. for ν given $\mathcal{G}$. Therefore, in a Bayesian nonparametric setting, when we make the initial assessment that our exchangeable model is

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

characterized by (3.1), instead of choosing a specific probability kernel R, we actually have to decide on a predefined body of knowledge in the form of the σ-algebra $\mathcal{G}$, making the model 'essentially' parameterized by $\mathcal{G}$. If, in addition, R is dominated by ν, the above amounts to a decision on a particular partition of $\mathbb{X}$ (see Examples 3.1 and 3.2).

Example 3.1. (*Mixture model.*) Let $\theta > 0$ and $\nu, P_1, P_2, \ldots$ be probability measures on $\mathbb{X}$ such that $P_k(D_k) = 1$, $k \geq 1$, where $D_1, D_2, \ldots \in \mathcal{X}$ forms a countable partition of $\mathbb{X}$. For each $x \in \mathbb{X}$, we define $P_{(x)}(\cdot) := P_k(\cdot)$ if and only if $x \in D_k$. Let $(X_n)_{n\geq 1}$ be an exchangeable MVPS with predictive distributions

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) = \frac{\theta}{\theta + n}\nu(\cdot) + \frac{n}{\theta + n} \cdot \frac{1}{n}\sum_{i=1}^{n} P_{(X_i)}(\cdot).$$

By [27, Theorem 3.14], the reinforcement kernel of any MVPS admits a Lebesgue decomposition with respect to a universal set, so in the present example, there exist sets $C, S \in \mathcal{X}$ such that $\nu(C) = 1$, $P_{(x)} \perp \nu$ and $P_{(x)}(S) = 1$ for every $x \in C \cap S$, and $P_{(x)} \ll \nu$ and $P_{(x)}(S^{\mathrm{c}}) = 1$ for every $x \in C \cap S^{\mathrm{c}}$. Let $\{D_{j_1}, D_{j_2}, \ldots\}$ be the smallest cover of $C \cap S^{\mathrm{c}}$ extracted from $\{D_1, D_2, \ldots\}$. Since $P_{(x)} \equiv P_k$ when $x \in D_k$, we have $P_{(x)} \ll \nu$ and $P_{(x)}(\cup_{i\geq 1} D_{j_i}) = 1$ for all $x \in \bigcup_{i\geq 1} D_{j_i}$. Furthermore, $\nu(D_{j_i}) > 0$ for each $i = 1, 2, \ldots$, and

$$\mathbb{P}(X_{n+1} \in \cup_{i\geq 1} D_{j_i} \mid X_1, \ldots, X_n) \geq \frac{\theta}{\theta + n}\nu(\cup_{i\geq 1} D_{j_i}),$$

which implies that $\sum_{n=1}^{\infty} \mathbb{P}(X_{n+1} \in \cup_{i\geq 1} D_{j_i} \mid X_1, \ldots, X_n) = \infty$ and, from the conditional Borel–Cantelli lemma, $\mathbb{P}(X_n \in \cup_{i\geq 1} D_{j_i} \text{ infinitely often}) = 1$. It follows that the subsequence of $(X_n)_{n\geq 1}$ restricted on $\bigcup_{i\geq 1} D_{j_i}$ is itself an exchangeable MVPS with parameters $\theta^* = \theta\nu(\cup_{i\geq 1} D_{j_i})$, $\nu^*(\cdot) = \nu(\cdot \mid \cup_{i\geq 1} D_{j_i})$ and $R_x = P_{(x)}$ for $x \in \bigcup_{i\geq 1} D_{j_i}$, so by Theorem 2.2, as applied to the subsequence, we get

$$P_{j_i}(\cdot) = P_{(x)}(\cdot) = \nu^*(\cdot \mid D_{j_i}) = \nu(\cdot \mid D_{j_i}),$$

using some $x \in D_{j_i}$ as reference. The singular part of the reinforcement kernel has no such simple representation, yet by Theorem 2.1 it is an r.c.d. for ν on S.

If $\nu(S) = 0$ or, equivalently, $P_k \ll \nu$ for all $k \geq 1$, then

$$P_{(x)}(\cdot) = \sum_{k\geq 1} \nu(\cdot \mid D_k) \cdot \mathbf{1}_{D_k}(x) \quad \text{for every } x \in \mathbb{X},$$

and, from Theorem 2.2,

$$\begin{aligned} X_n \mid \tilde{P} &\overset{\text{i.i.d.}}{\sim} \sum_{k\geq 1} \tilde{P}(D_k)P_k(\cdot), \\ (\tilde{P}(D_k))_{k\geq 1} &\sim \mathrm{DP}(\theta, \nu). \end{aligned}$$

Such models are commonly used for density estimation (see, e.g., [15, Chapter 5]), and exchangeable MVPSs with dominated reinforcement kernels offer a particular set of model choices. For example, when $|\mathbb{X}| < \infty$ and $\nu(A) = |A|/|\mathbb{X}|$, $A \in \mathcal{X}$, is uniform on $\mathbb{X}$, the directing random measure of $(X_n)_{n\geq 1}$ becomes

$$\sum_{k\geq 1} \tilde{P}(D_k)\frac{|\cdot \cap D_k|}{|D_k|},$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

which is a random histogram with nonuniform widths given by the $|D_k|$. At the other extreme, when $\nu(\mathrm{d}x) = f(x)\mathrm{d}x$ is absolutely continuous (with respect to the Lebesgue measure) on $\mathbb{X} = \mathbb{R}$, the directing random measure of $(X_n)_{n\geq 1}$ is itself absolutely continuous with random density

$$\sum_{k\geq 1} \tilde{P}(D_k) f_k(x),$$

where $f_k(x) = \mathbf{1}_{D_k}(x) f(x) / \int_{D_k} f(y)\,\mathrm{d}y$, $x \in \mathbb{X}$, is the section of f over D_k, normalized.

Example 3.2. (k-color urn model) Let $(X_n)_{n\geq 1}$ be an exchangeable MVPS(θ, ν, R) on $\mathbb{X} = \{1, \ldots, k\}$, where, without loss of generality, we assume that R is a probability kernel (see Section 2.2). By Theorem 2.2, there exists a partition $\{D_1, \ldots, D_m\} \subseteq \mathcal{X}$ of $\mathbb{X}$, for some $1 \leq m \leq k$, such that

$$R_x(\cdot) = \sum_{l=1}^{m} \nu(\,\cdot \mid D_l) \cdot \mathbf{1}_{D_l}(x) \qquad \text{for every } x \in \mathbb{X}. \tag{3.2}$$

It follows that, with ν and θ fixed, the total number of exchangeable k-color MVPSs is equal to the number of possible partitions of $\mathbb{X}$, also known as Bell's number, which is given by the formula

$$B_k = \sum_{j=0}^{k} \frac{1}{j!} \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} i^k,$$

with $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, etc. The classical k-color Pólya urn model corresponds to the case with the finest partition $D_j = \{j\}$, $j = 1, \ldots, k$, while we obtain an i.i.d. sequence when $D_1 = \mathbb{X}$. On the other hand, intermediate cases can be appealing from an application point of view and can be used, for example, in a clinical trial setting to target both fixed asymptotic allocations (within the D_j) and random asymptotic allocations (across the D_j), simultaneously.

3.1. 'Pure' sufficientness postulate for MVPSs

Recall from the introduction that (1.1) is equivalent to (1.3) and (1.4), up to a relabeling of the observations. In this section, we investigate whether exchangeable MVPSs can be characterized by statements like those in (1.3) and (1.4), which one might argue reflect a 'pure' sufficientness postulate since they do not assume a particular analytic form for the predictive distributions. Based on the discussion in Remark 3.1, such a result should be stated in terms of the sub-σ-algebra $\mathcal{G}$ for which $R(\cdot) = \nu(\,\cdot \mid \mathcal{G})$. Here, we follow the approach from [12, Proposition 3.7] (see also Example 3.3), which gives an alternative formulation of (1.3) and (1.4) via predictive sufficiency. To that end, we introduce, for each $n = 1, 2, \ldots$ and every $A \in \mathcal{X}$, the notation

$$N_{n,A} := \sum_{i=1}^{n} \mathbf{1}_A(X_i).$$

We also define, for each $n = 1, 2, \ldots$,

$$\mathcal{G}_n := X_n^{-1}(\mathcal{G}).$$

The next theorem states that an exchangeable process $(X_n)_{n\geq 1}$ is an MVPS with reference to some c.g. under the ν sub-σ-algebra $\mathcal{G}$ of $\mathcal{X}$ if (i) the probability that X_{n+1} belongs to a set $A \in \mathcal{G}$ depends only on the number of observations $N_{n,A}$ in A; and (ii) given that it belongs to $\mathcal{G}$, the random variable X_{n+1} is conditionally independent from $(X_1, \ldots, X_n)$. Note that condition

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

(i) becomes ambiguous when $\mathcal{G}$ is a *simple* σ-algebra generated by a single subset $A \in \mathcal{X}$, since $N_{n,A^c} = n - N_{n,A}$ (see also [29, Remark 1] and the discussion after Proposition 3.1). For that purpose, we introduce the notion of a *ν-essentially not simple* σ-algebra, which is a σ-algebra $\mathcal{G}$ containing a finite partition $\{B_1, \dots, B_k\} \subseteq \mathcal{G}$ of $\mathbb{X}$ with $|k| > 2$ elements such that at least three of them have positive ν-probability, $\#\{j\colon \nu(B_j) > 0\} > 2$.

Theorem 3.2. *Let $(X_n)_{n\geq 1}$ be an exchangeable process with marginal distribution ν, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{X}$ on $\mathbb{X}$, which is c.g. under ν and ν-essentially not simple. If, for each $n = 1, 2, \dots$,*

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in A \mid N_{n,A}) \qquad \text{for all } A \in \mathcal{G}, \tag{3.3}$$

$$\mathbb{P}(X_{n+1} \in B \mid X_1, \dots, X_n, \mathcal{G}_{n+1}) = \mathbb{P}(X_{n+1} \in B \mid \mathcal{G}_{n+1}) \qquad \text{for all } B \in \mathcal{X}, \tag{3.4}$$

then $(X_n)_{n\geq 1}$ is an MVPS with reinforcement kernel $R(\cdot) = \nu(\cdot \mid \mathcal{G})$. The converse statement is also true.

Example 3.3. (*Dirichlet process*.) The current example essentially repeats [12, Proposition 3.7]. Suppose in Theorem 3.2 that we have $\mathcal{G} = \mathcal{X}$, which is c.g. by assumption. Then (3.4) is trivially true, so any exchangeable process $(X_n)_{n\geq 1}$ satisfying

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in A \mid N_{n,A})$$

for all $A \in \mathcal{X}$ is an MVPS with reinforcement kernel

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

and, following [7], has a Dirichlet process prior.

Example 3.4 (*Mixture model*.) In Theorem 3.2, suppose that $\mathcal{G} = \sigma(D_1, D_2, \dots)$, where $D_1, D_2, \dots \in \mathcal{X}$ forms a countable partition of $\mathbb{X}$. Then (3.3) and (3.4) can be equivalently formulated as

$$\mathbb{P}(X_{n+1} \in D_k \mid X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in D_k \mid N_{n,D_k}) \quad \text{for each } k = 1, 2, \dots$$

and, with a slight abuse of notation,

$$\mathbb{P}(X_{n+1} \in B \mid X_1, \dots, X_n; X_{n+1} \in D_k) = \mathbb{P}(X_{n+1} \in B \mid X_{n+1} \in D_k)$$

for all $B \in \mathcal{X}$. In that case, $(X_n)_{n\geq 1}$ is an MVPS with reinforcement kernel

$$R_x(\cdot) = \nu(\cdot \mid \sigma(D_1, D_2, \dots))(x) = \sum_{k\geq 1} \nu(\cdot \mid D_k) \cdot \mathbf{1}_{D_k}(x)$$

for ν-a.e. x, so that $R_x \ll \nu$ and, from Theorem 2.2, $(X_n)_{n\geq 1}$ has a Dirichlet process mixture prior.

Example 3.5. (*Dirichlet invariant process*.) Suppose that $\mathbb{X} = \mathbb{R}$. Define

$$\mathcal{G} := \{A \in \mathcal{X}\colon -A = A\} \equiv \{A \cup (-A), A \in \mathcal{X}\},$$

where $-A = \{-x, x \in A\}$. Let $(X_n)_{n\geq 1}$ be an exchangeable process such that, for each $n = 1, 2, \dots$ and every $A \in \mathcal{X}$,

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in A \mid N_{n,A} + N_{n,-A}), \tag{3.5}$$

$$\mathbb{P}(X_{n+1} \in -A \mid X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in A \mid X_1, \dots, X_n). \tag{3.6}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

If we denote by ν the marginal distribution of $(X_n)_{n\geq 1}$, we get $\nu(-A)=\nu(A)$ from (3.6), for every $A\in\mathcal{X}$. Moreover, (3.5) implies (3.3) when $A\in\mathcal{G}$, since $-A=A$.

Let $A_1,\dots,A_n,A\in\mathcal{X}$ and $B\in\mathcal{G}$. Since ν is invariant with respect to $\mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}[\mathbf{1}_B(X_{n+1})\cdot\mathbb{P}(X_{n+1}\in A\mid\mathcal{G}_{n+1})]=\nu(A\cap B)&=\tfrac{1}{2}(\nu(A\cap B)+\nu((-A)\cap B))\\&=\mathbb{E}\big[\mathbf{1}_B(X_{n+1})\cdot\tfrac{1}{2}(\mathbf{1}_A(X_{n+1})+\mathbf{1}_{-A}(X_{n+1}))\big];\end{aligned}$$

thus, $\mathbb{P}(X_{n+1}\in A\mid\mathcal{G}_{n+1})=\frac{1}{2}(\delta_{X_{n+1}}(A)+\delta_{-X_{n+1}}(A))$ a.s. If, in addition, $A\cap(-A)=\emptyset$, then $\mathbb{P}(X_{n+1}\in A\mid\mathcal{G}_{n+1})=\frac{1}{2}\cdot\mathbf{1}_{A\cup(-A)}(X_{n+1})$ a.s., and as $\mathbb{P}(X_{n+1}\in A\cap B\mid X_1,\dots,X_n)=\mathbb{P}(X_{n+1}\in(-A)\cap B\mid X_1,\dots,X_n)$ from (3.6), we get

$$\begin{aligned}&\mathbb{E}[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\mathbf{1}_A(X_{n+1})\mathbf{1}_B(X_{n+1})]\\&\quad=\mathbb{E}\big[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\cdot\mathbb{P}(X_{n+1}\in A\cap B\mid X_1,\dots,X_n)\big]\\&\quad=\tfrac{1}{2}\cdot\mathbb{E}\big[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\cdot\mathbb{P}(X_{n+1}\in(A\cup(-A))\cap B\mid X_1,\dots,X_n)\big]\\&\quad=\tfrac{1}{2}\cdot\mathbb{E}[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\mathbf{1}_{A\cup(-A)}(X_{n+1})\mathbf{1}_B(X_{n+1})]\\&\quad=\mathbb{E}\big[\mathbb{P}(X_1\in A_1,\dots,X_n\in A_n\mid\mathcal{G}_{n+1})\cdot\mathbb{P}(X_{n+1}\in A\mid\mathcal{G}_{n+1})\cdot\mathbf{1}_B(X_{n+1})\big].\end{aligned}$$

By varying $A\in\mathcal{C}:=\{(-b,-a],(a,b]\colon a,b\in\mathbb{R}_+,a<b\}\cup\{\emptyset\}$ and using the fact that $\mathcal{C}$ forms a π-class generating $\mathcal{X}=\sigma(\mathcal{C})$, we can conclude that X_{n+1} and $(X_1,\dots,X_n)$ are conditionally independent given $\mathcal{G}_{n+1}$. Hence, $(X_n)_{n\geq 1}$ satisfies the conditions of Theorem 3.2 and is therefore an exchangeable MVPS with reinforcement kernel

$$R_x(\cdot)=\nu(\cdot\mid\mathcal{G})(x)=\tfrac{1}{2}(\delta_x(\cdot)+\delta_{-x}(\cdot))$$

for ν-a.e. x. This model was studied in [4], which showed that $(X_n)_{n\geq 1}$ has a so-called Dirichlet invariant process prior (for more information, see [4, Examples 3 and 17]).

3.2. *k*-color MVPS

In this section, we prove a generalization of Johnson's sufficientness postulate (1.2) in the context of the k-color urn model from Example 3.2. Although some of the results can be deduced from previous sections, for completeness we present a separate proof in the spirit of [29], exploiting the discrete nature of the process.

Let $\mathbb{X}=\{1,\dots,k\}$, and $\{D_1,\dots,D_m\}\subseteq\mathcal{X}$ be a partition of $\mathbb{X}$, for some $1\leq m\leq k$. For each $j\in\mathbb{X}$, let us denote by $m(j)$ the index $l\in\{1,\dots,m\}$ such that $j\in D_l$, and let us define $N_{0,l}:=0$ and $N_{n,l}:=\sum_{i=1}^n\mathbf{1}_{D_l}(X_i)$ for $n\geq 1$, $l=1,\dots,m$. If $(X_n)_{n\geq 1}$ is an exchangeable MVPS with reinforcement kernel given by (3.2), then, for each $j=1,\dots,k$,

$$\begin{aligned}\mathbb{P}(X_{n+1}=j\mid X_1,\dots,X_n)&=\frac{\theta\nu(\{j\})+\sum_{i=1}^n\nu(\{j\}\mid D_{m(j)})\cdot\mathbf{1}_{D_{m(j)}}(X_i)}{\theta+n}\\&=\frac{\theta\nu(\{j\})+\nu(\{j\}\mid D_{m(j)})\cdot N_{n,m(j)}}{\theta+n},\end{aligned}$$

that is, the probability of observing type j on the $(n+1)$th trial depends on the sample $(X_1,\dots,X_n)$ only by the number of times that each type in $D_{m(j)}$ has been observed in the past. This fact motivates the next proposition, which extends Johnson's sufficientness postulate (1.2) by considering a grouping of types according to $D_1,\dots,D_m$.

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

For ease of notation, we will write $\mathbb{P}_0(\cdot) = \mathbb{P}(X_1 \in \cdot\,)$ and $\mathbb{P}_n(\cdot) = \mathbb{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n)$ for $n = 1, 2, \ldots$ In addition, we will work under the assumption that any sequence of types has a positive probability of being observed,

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) > 0 \quad \text{for all } (x_1, \ldots, x_n) \in \mathbb{X}^n \text{ and } n = 1, 2, \ldots \tag{3.7}$$

Proposition 3.1. *Let $D_1, \ldots, D_m$ form a partition of $\mathbb{X} = \{1, \ldots, k\}$, with $3 \le m \le k$, and let $(X_n)_{n\ge1}$ be an exchangeable sequence of $\mathbb{X}$-valued random variables satisfying (3.7). Using the established notation, if, for each $j = 1, \ldots, k$ and $n = 0, 1, \ldots$,*

$$\mathbb{P}_n(\{j\}) = f_{n,j}(N_{n,m(j)})$$

for some functions $f_{n,j}\colon \{0, \ldots, n\} \to [0, 1]$, then there exist $w_1, \ldots, w_k > 0$ such that $(X_n)_{n\ge1}$ is an MVPS with constant $\bar{w} = \sum_{j=1}^k w_j$, base measure $\nu(\cdot) = \sum_{j=1}^k (w_j/\bar{w})\delta_j(\cdot)$, and reinforcement kernel $R(\cdot) = \nu(\cdot \mid \sigma(D_1, \ldots, D_m))$.

If $m = 1$ in Proposition 3.1, it is easy to see that $(X_n)_{n\ge1}$ is i.i.d. with marginal distribution $\mathbb{P}_0$. When $m = 2$, however, additional assumptions on P_n are needed to obtain the stated characterization result (see Remark 4.1). A different approach, following [17], is presented in the next proposition, which is more in line with Proposition 3.1 and covers the case $m = 2$. The original result [17] is itself recovered by taking $D_j = \{j\}$ for $j = 1, \ldots, k$.

Proposition 3.2. *Let $D_1, \ldots, D_m$ form a partition of $\mathbb{X} = \{1, \ldots, k\}$, with $1 \le m \le k$, and let $(X_n)_{n\ge1}$ be an exchangeable, but not i.i.d., sequence of $\mathbb{X}$-valued random variables satisfying (3.7). Fix $w_1, \ldots, w_k > 0$, and define $\bar{w} := \sum_{j=1}^k w_j$ and $\bar{w}_l := \sum_{j\in D_l} w_j$ for $l = 1, \ldots, m$. Using the established notation, if, for each $j = 1, \ldots, k$ and $n = 0, 1, \ldots$,*

$$\mathbb{P}_n(\{j\}) = f_j\left(\frac{\bar{w}_{m(j)} + N_{n,m(j)}}{\bar{w} + n}\right)$$

for some functions $f_j\colon [0, 1] \to [0, 1]$, then $(X_n)_{n\ge1}$ is an MVPS with constant $\bar{w}$, base measure $\nu(\cdot) = \sum_{j=1}^k (w_j/\bar{w})\delta_j(\cdot)$, and reinforcement kernel $R(\cdot) = \nu(\cdot \mid \sigma(D_1, \ldots, D_m))$.

4. Proofs

Proof of Theorem 3.1. Theorem 3.1 will be proved in several steps. We start by deriving some relations between ν, R, and $(a_n)_{n\ge1}$ as a consequence of exchangeability.

Step 1. Let $A, B \in \mathcal{X}$. It follows from the form of the predictive distributions (3.1) and $\mathbb{P}(X_1 \in \mathrm{d}x) = \nu(\mathrm{d}x)$ that

$$\begin{aligned}\mathbb{P}(X_1 \in A, X_2 \in B) &= \int_A \mathbb{P}(X_2 \in B \mid X_1 = x)\mathbb{P}(X_1 \in \mathrm{d}x) \\ &= \int_A \{(1 - a_1)\nu(B) + a_1 R_x(B)\}\, \nu(\mathrm{d}x) \\ &= (1 - a_1)\nu(A)\nu(B) + a_1 \int_A R_x(B)\, \nu(\mathrm{d}x).\end{aligned}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Since $(X_1, X_2) \overset{\mathrm{d}}{=} (X_2, X_1)$ by exchangeability, we get

$$\begin{aligned}(1-a_1)\nu(A)\nu(B) + a_1 \int_A R_x(B)\,\nu(\mathrm{d}x) &= \mathbb{P}(X_1 \in A, X_2 \in B)\\ &= \mathbb{P}(X_1 \in B, X_2 \in A)\\ &= (1-a_1)\nu(A)\nu(B) + a_1 \int_B R_x(A)\,\nu(\mathrm{d}x);\end{aligned}$$

therefore, $\int_{x\in A}\int_{y\in B} R_x(\mathrm{d}y)\,\nu(\mathrm{d}x) = \int_{x\in B}\int_{y\in A} R_x(\mathrm{d}y)\,\nu(\mathrm{d}x)$ or, as measures on $\mathbb{X}^2$,

$$R_x(\mathrm{d}y)\nu(\mathrm{d}x) = R_y(\mathrm{d}x)\nu(\mathrm{d}y). \tag{4.1}$$

Next, for ν-a.e. x,

$$\begin{aligned}&\mathbb{P}(X_3 \in B \mid X_1 = x)\\ &\quad= \int_{\mathbb{X}} \mathbb{P}(X_3 \in B \mid X_1 = x, X_2 = y)\mathbb{P}(X_2 \in \mathrm{d}y \mid X_1 = x)\\ &\quad= \int_{\mathbb{X}} \left\{(1-a_2)\nu(B) + \frac{a_2}{2}R_x(B) + \frac{a_2}{2}R_y(B)\right\}((1-a_1)\,\nu(\mathrm{d}y) + a_1\,R_x(\mathrm{d}y))\\ &\quad= (1-a_2)\nu(B) + \frac{a_2}{2}R_x(B) + \frac{(1-a_1)a_2}{2}\int_{\mathbb{X}} R_y(B)\,\nu(\mathrm{d}y) + \frac{a_1a_2}{2}\int_{\mathbb{X}} R_y(B)\,R_x(\mathrm{d}y)\\ &\quad\overset{\text{(a)}}{=} (1-a_2)\nu(B) + \frac{a_2}{2}R_x(B) + \frac{(1-a_1)a_2}{2}\nu(B) + \frac{a_1a_2}{2}\int_{\mathbb{X}} R_y(B)\,R_x(\mathrm{d}y)\end{aligned}$$

for every $B \in \mathcal{X}$, where (a) follows from (4.1) as $\int_{\mathbb{X}} R_y(B)\,\nu(\mathrm{d}y) = \nu(B)$. Since $(X_1, X_2) \overset{\mathrm{d}}{=} (X_1, X_3)$ by exchangeability we get, for ν-a.e. x,

$$\begin{aligned}(1-a_1)\nu(B) + a_1R_x(B) &= \mathbb{P}(X_2 \in B \mid X_1 = x)\\ &= \mathbb{P}(X_3 \in B \mid X_1 = x)\\ &= (1-a_2)\nu(B) + \frac{a_2}{2}R_x(B) + \frac{(1-a_1)a_2}{2}\nu(B) + \frac{a_1a_2}{2}\int_{\mathbb{X}} R_y(B)\,R_x(\mathrm{d}y);\end{aligned}$$

thus, rearranging the terms,

$$\int_{\mathbb{X}} R_y(B)\,R_x(\mathrm{d}y) = \left(1 - \frac{a_2 - 2a_1 + a_1a_2}{a_1a_2}\right)R_x(B) + \frac{a_2 - 2a_1 + a_1a_2}{a_1a_2}\nu(B).$$

Using that $\mathcal{X}$ is countably generated, we obtain, for ν-a.e. x, that, as measures on $\mathbb{X}$,

$$\int_{y\in\mathbb{X}} R_y(\mathrm{d}z)\,R_x(\mathrm{d}y) = (1-c^*)\,R_x(\mathrm{d}z) + c^*\,\nu(\mathrm{d}z), \tag{4.2}$$

where $c^* := (a_2 - 2a_1 + a_1a_2)/(a_1a_2)$.

Let $n \geq 2$. It follows from (3.1) that, for a.e. $(x_1, \ldots, x_{n-1}, x)$ in $\mathbb{X}^n$ with respect to the marginal distribution $\mathbb{P}_{(X_1,\ldots,X_n)}$ of $(X_1, \ldots, X_n)$ on $\mathcal{X}^n$,

$$\begin{aligned}&\mathbb{P}(X_{n+1} \in B \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x)\\ &\qquad= (1-a_n)\,\nu(\mathrm{d}y) + \frac{a_n}{n}\sum_{i=1}^{n-1} R_{x_i}(B) + \frac{a_n}{n}R_x(B)\end{aligned} \tag{4.3}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

for every $B \in \mathcal{X}$. On the other hand, disintegrating $\mathbb{P}_{(X_1,\dots,X_n)}$ repeatedly via (3.1), we obtain (see also (4.12)) $\mathbb{P}(X_1 \in \mathrm{d}x_1, \dots, X_n \in \mathrm{d}x_n) \geq \prod_{i=1}^{n-1}(1-a_i)\,\nu(\mathrm{d}x_1)\cdots\nu(\mathrm{d}x_n)$. Thus, (4.3) remains true for $(\nu \times \cdots \times \nu)$-a.e. $(x_1, \dots, x_{n-1}, x)$ in $\mathbb{X}^n$, so integrating (4.3) with respect to $\nu(\mathrm{d}x_1), \dots, \nu(\mathrm{d}x_{n-1})$ we get, for ν-a.e. x,

$$\begin{aligned}&\int_{\mathbb{X}} \cdots \int_{\mathbb{X}} \mathbb{P}(X_{n+1} \in B \mid X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x)\,\nu(\mathrm{d}x_{n-1}) \cdots \nu(\mathrm{d}x_1)\\ &= (1-a_n)\nu(B) + \frac{a_n}{n}\sum_{i=1}^{n-1}\nu(B) + \frac{a_n}{n}R_x(B) = (1-a_n)\nu(B) + \frac{(n-1)a_n}{n}\nu(B) + \frac{a_n}{n}R_x(B),\end{aligned} \tag{4.4}$$

where we have used that $\int_{\mathbb{X}} R_{x_i}(B)\,\nu(\mathrm{d}x_i) = \nu(B)$ from (4.1), for $i = 1, \dots, n-1$.

On the other hand, for $(\nu \times \cdots \times \nu)$-a.e. $(x_1, \dots, x_{n-1}, x)$ in $\mathbb{X}^n$,

$$\begin{aligned}&\mathbb{P}(X_{n+2} \in B \mid X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x)\\ &= \int_{\mathbb{X}} \left\{(1-a_{n+1})\nu(B) + \frac{a_{n+1}}{n+1}\sum_{i=1}^{n-1} R_{x_i}(B) + \frac{a_{n+1}}{n+1}R_x(B) + \frac{a_{n+1}}{n+1}R_y(B)\right\}\\ &\qquad \left((1-a_n)\,\nu(\mathrm{d}y) + \frac{a_n}{n}\sum_{i=1}^{n-1} R_{x_i}(\mathrm{d}y) + \frac{a_n}{n}\,R_x(\mathrm{d}y)\right)\\ &= (1-a_{n+1})\nu(B) + \frac{a_{n+1}}{n+1}\sum_{i=1}^{n-1} R_{x_i}(B) + \frac{a_{n+1}}{n+1}R_x(B) + \frac{(1-a_n)a_{n+1}}{n+1}\int_{\mathbb{X}} R_y(B)\,\nu(\mathrm{d}y)\\ &\quad + \frac{a_n a_{n+1}}{n(n+1)}\sum_{i=1}^{n-1}\int_{\mathbb{X}} R_y(B)\,R_{x_i}(\mathrm{d}y) + \frac{a_n a_{n+1}}{n(n+1)}\int_{\mathbb{X}} R_y(B)\,R_x(\mathrm{d}y)\\ &\overset{\text{(a)}}{=} (1-a_{n+1})\nu(B) + \frac{a_{n+1}}{n+1}\sum_{i=1}^{n-1} R_{x_i}(B) + \frac{a_{n+1}}{n+1}R_x(B) + \frac{(1-a_n)a_{n+1}}{n+1}\nu(B)\\ &\quad + \frac{a_n a_{n+1}}{n(n+1)}\sum_{i=1}^{n-1}\int_{\mathbb{X}} R_y(B)\,R_{x_i}(\mathrm{d}y) + (1-c^*)\frac{a_n a_{n+1}}{n(n+1)}R_x(B) + c^*\frac{a_n a_{n+1}}{n(n+1)}\nu(B)\end{aligned} \tag{4.5}$$

for every $B \in \mathcal{X}$, where (a) follows from (4.1) and (4.2). Note that, for each $i = 1, \dots, n-1$, the following identity is true from (4.1):

$$\begin{aligned}\int_{\mathbb{X}}\int_{\mathbb{X}} R_y(B)\,R_{x_i}(\mathrm{d}y)\,\nu(\mathrm{d}x_i) &= \int_{\mathbb{X}}\int_{\mathbb{X}} R_y(B)\,R_y(\mathrm{d}x_i)\,\nu(\mathrm{d}y)\\ &= \int_{\mathbb{X}} R_y(B)R_y(\mathbb{X})\,\nu(\mathrm{d}y) = \int_{\mathbb{X}} R_y(B)\,\nu(\mathrm{d}y) = \nu(B).\end{aligned}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Thus, integrating (4.5) with respect to $\nu(\mathrm{d}x_1), \ldots, \nu(\mathrm{d}x_{n-1})$, we get, for ν-a.e. x,

$$\begin{aligned}
&\int_{\mathbb{X}} \cdots \int_{\mathbb{X}} \mathbb{P}(X_{n+2} \in B \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x)\, \nu(\mathrm{d}x_{n-1}) \cdots \nu(\mathrm{d}x_1) \\
&\quad = (1 - a_{n+1})\nu(B) + \frac{a_{n+1}}{n+1} \sum_{i=1}^{n-1} \nu(B) + \frac{a_{n+1}}{n+1} R_x(B) + \frac{(1-a_n)a_{n+1}}{n+1} \nu(B) \\
&\qquad + \frac{a_n a_{n+1}}{n(n+1)} \sum_{i=1}^{n-1} \nu(B) + (1 - c^*) \frac{a_n a_{n+1}}{n(n+1)} R_x(B) + c^* \frac{a_n a_{n+1}}{n(n+1)} \nu(B) \\
&\quad = (1 - a_{n+1})\nu(B) + \frac{(n-1)a_{n+1}}{n+1} \nu(B) + \frac{a_{n+1}}{n+1} R_x(B) + \frac{(1-a_n)a_{n+1}}{n+1} \nu(B) \\
&\qquad + \frac{(n-1)a_n a_{n+1}}{n(n+1)} \nu(B) + (1 - c^*) \frac{a_n a_{n+1}}{n(n+1)} R_x(B) + c^* \frac{a_n a_{n+1}}{n(n+1)} \nu(B). \qquad (4.6)
\end{aligned}$$

Since $(X_1, \ldots, X_n, X_{n+1}) \overset{\mathrm{d}}{=} (X_1, \ldots, X_n, X_{n+2})$ by exchangeability, we can equate (4.4) and (4.6) to obtain, for ν-a.e. x,

$$\begin{aligned}
&(1 - a_n)\nu(B) + \frac{(n-1)a_n}{n} \nu(B) + \frac{a_n}{n} R_x(B) \\
&\quad = \int_{\mathbb{X}} \cdots \int_{\mathbb{X}} \mathbb{P}(X_{n+1} \in B \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x)\, \nu(\mathrm{d}x_{n-1}) \cdots \nu(\mathrm{d}x_1) \\
&\quad = \int_{\mathbb{X}} \cdots \int_{\mathbb{X}} \mathbb{P}(X_{n+2} \in B \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x)\, \nu(\mathrm{d}x_{n-1}) \cdots \nu(\mathrm{d}x_1) \\
&\quad = (1 - a_{n+1})\nu(B) + \frac{(n-1)a_{n+1}}{n+1} \nu(B) + \frac{a_{n+1}}{n+1} R_x(B) + \frac{(1-a_n)a_{n+1}}{n+1} \nu(B) \\
&\qquad + \frac{(n-1)a_n a_{n+1}}{n(n+1)} \nu(B) + (1 - c^*) \frac{a_n a_{n+1}}{n(n+1)} R_x(B) + c^* \frac{a_n a_{n+1}}{n(n+1)} \nu(B)
\end{aligned}$$

for every $B \in \mathcal{X}$; therefore, after some simple algebra,

$$(na_{n+1} - (n+1)a_n + (1 - c^*)a_n a_{n+1})\nu(B) = (na_{n+1} - (n+1)a_n + (1 - c^*)a_n a_{n+1}) R_x(B).$$

Using that $\mathcal{X}$ is countably generated, we get, for ν-a.e. x, that, as measures on $\mathbb{X}$,

$$\begin{aligned}
&(na_{n+1} - (n+1)a_n + (1 - c^*)a_n a_{n+1})\, \nu(\mathrm{d}y) \\
&\qquad\qquad = (na_{n+1} - (n+1)a_n + (1 - c^*)a_n a_{n+1})\, R_x(\mathrm{d}y). \qquad (4.7)
\end{aligned}$$

Step 2. We proceed by showing that, for different specifications of R, (4.2) and (4.7) together imply the prescribed form of the coefficients a_n and, in particular, that $c^* = 0$. To that end, let us consider, for each $x \in \mathbb{X}$, the Lebesgue decomposition of the reinforcement kernel,

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

where $R_x^{\perp}$ and R_x^{a} are substochastic measures on $\mathbb{X}$ such that $R_x^{\perp} \perp \nu$ for some $S_x \in \mathcal{X}$ with $R_x^{\perp}(S_x^{\mathrm{c}}) = 0 = \nu(S_x)$, and $R_x^{a} \ll \nu$. It follows from [21, Theorem 1] that the mappings $x \mapsto R_x^{\perp}$ and $x \mapsto R_x^{a}$ are measurable, hence the following sets are well-defined in $\mathcal{X}$, since $\mathcal{X}$ is countably generated:

$$D := \{x \in \mathbb{X} \colon R_x \neq \nu\}, \qquad F := \{x \in \mathbb{X} \colon R_x^{\perp}(S_x) > 0\}.$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Case 1: $\nu(D) > 0$. By definition, for every $x \in D$, there exists $B_x \in \mathcal{X}$ such that $\nu(B_x) \neq R_x(B_x)$. It follows from (4.7) that, for ν-a.e. x in D,

$$(na_{n+1} - (n+1)a_n + (1-c^*)a_n a_{n+1})\nu(B_x) = (na_{n+1} - (n+1)a_n + (1-c^*)a_n a_{n+1})R_x(B_x),$$

which is true if and only if $na_{n+1} - (n+1)a_n + (1-c^*)a_n a_{n+1} = 0$, or, equivalently,

$$\frac{a_{n+1}}{n+1} = \frac{a_n}{n + (1-c^*)a_n}.$$

Using the substitution $b_n = n/a_n$, we obtain $b_{n+1} = b_n + (1-c^*)$; thus, $b_n = b_1 + (n-1)(1-c^*)$ for each $n = 1, 2, \ldots$, and so

$$a_n = \frac{n}{(n-1)(1-c^*) + (1/a_1)} \quad \text{for } n = 1, 2, \ldots \tag{4.8}$$

If $c^* = 1$ then $a_n = na_1$, which leads to a contradiction since, by assumption, $0 < a_n < 1$ for all $n = 1, 2, \ldots$; otherwise,

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{(1-c^*) - ((1-c^*)/n) + (1/a_1 n)} = \frac{1}{1-c^*},$$

implying that $a_n \notin (0, 1)$ for n large enough when $c^* \in (0, 1) \cup (1, \infty)$ – absurd. Therefore, $c^* \leq 0$.

Sub-case 1a: $\nu(D) > 0$ *and* $\nu(F) > 0$. Let $G \in \mathcal{X}$ be the ν-almost sure set in (4.2). Define $G_0 := \{x \in G\colon R_x(G) = 1\}$ and $G_n := \{x \in G_{n-1}\colon R_x(G_{n-1}) = 1\}$ for $n \geq 1$. It follows from (4.1) that

$$1 = \int_{\mathbb{X}} R_x(\mathbb{X})\, \nu(\mathrm{d}x) = \int_G R_x(\mathbb{X})\, \nu(\mathrm{d}x) = \int_{\mathbb{X}} R_x(G)\, \nu(\mathrm{d}x),$$

which implies that $R_x(G) = 1$ for ν-a.e. x, and so $\nu(G_0) = 1$. Proceeding by induction, we obtain $\nu(G_n) = 1$ for all $n \geq 1$; thus, letting $G^* := \bigcap_{n=0}^{\infty} G_n$, we have $\nu(G^*) = 1$ and $R_x(G_n) = 1$ for all $x \in G^*$ and $n \geq 1$. As a result, for every $x \in G^*$ we get $R_x(G^*) = 1$ and, since $G^* \subseteq G$,

$$\int_{y\in\mathbb{X}} R_y(\mathrm{d}z)\, R_x(\mathrm{d}y) = (1-c^*)\, R_x(\mathrm{d}z) + c^*\, \nu(\mathrm{d}z). \tag{4.9}$$

Recall that $F = \{x \in \mathbb{X}\colon R_x^{\perp}(S_x) > 0\}$. Let us define $\beta := \sup_{x\in F\cap G^*} R_x^{\perp}(S_x)$. Assuming that $\nu(F) > 0$, we have $\beta > 0$. Fix $\varepsilon \in (0, \beta)$. From the properties of suprema, there exists $x_0 \in F \cap G^*$ such that $R_{x_0}^{\perp}(S_{x_0}) \geq \beta - \varepsilon$. Moreover, $\nu(S_{x_0}) = 0$, which implies that $R_y^a(S_{x_0}) = 0$ for all $y \in \mathbb{X}$. It follows that, first,

$$(1-c^*)R_{x_0}(S_{x_0}) + c^*\nu(S_{x_0}) = (1-c^*)R_{x_0}^{\perp}(S_{x_0}) \geq (1-c^*)(\beta - \varepsilon),$$

since $c^* \leq 0$ from the previous part, and second,

$$\begin{aligned}\int_{\mathbb{X}} R_y(S_{x_0})\, R_{x_0}(\mathrm{d}y) &= \int_{\mathbb{X}} R_y^{\perp}(S_{x_0})\, R_{x_0}(\mathrm{d}y) \\ &\overset{(a)}{\leq} \int_{\mathbb{X}} R_y^{\perp}(S_y)\, R_{x_0}(\mathrm{d}y) \overset{(b)}{=} \int_{F\cap G^*} R_y^{\perp}(S_y)\, R_{x_0}(\mathrm{d}y) \leq \beta,\end{aligned}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

where we have used in (a) that $R_y^{\perp}(S_y^{\mathrm{c}}) = 0$, so $R_y^{\perp}(S_{x_0}) = R_y^{\perp}(S_{x_0} \cap S_y) \le R_y^{\perp}(S_y)$ for $y \in \mathbb{X}$; and in (b) that (i) $R_y^{\perp}(S_y) = 0$ for $y \in F^{\mathrm{c}}$ by construction, and (ii) $R_{x_0}(G^*) = 1$, since $x_0 \in G^*$. Therefore, by (4.9),

$$\beta \ge \int_{\mathbb{X}} R_y(S_{x_0})\, R_{x_0}(\mathrm{d}y) = (1 - c^*) R_{x_0}(S_{x_0}) + c^* \nu(S_{x_0}) \ge (1 - c^*)(\beta - \varepsilon).$$

Letting $\varepsilon \downarrow 0$, we get $1 - c^* \le \beta/(\beta - \varepsilon) \to 1$, and so $c^* \ge 0$. But $c^* \le 0$ from the previous part, hence $c^* = 0$, which implies from (4.8) that

$$a_n = \frac{n}{n + (1/a_1) - 1}.$$

Writing $\theta := a_1^{-1} - 1$, we obtain $\theta > 0$ and $a_n = n/(n + \theta)$ for each $n = 1, 2, \ldots$, so

$$\mathbb{P}(X_{n+1} \in \mathrm{d}x \mid X_1, \ldots, X_n) = \frac{\theta}{\theta + n}\, \nu(\mathrm{d}x) + \frac{n}{\theta + n} \cdot \frac{1}{n} \sum_{i=1}^{n} R_{X_i}(\mathrm{d}x) = \frac{\theta\, \nu(\mathrm{d}x) + \sum_{i=1}^{n} R_{X_i}(\mathrm{d}x)}{\theta + n},$$

that is, $(X_n)_{n \ge 1}$ is an MVPS with parameters (θ, ν, R). *Sub-case 1b:* $\nu(D) > 0$ *and* $\nu(F) = 0$. If $\nu(F) = 0$ instead, then $R_x \ll \nu$ for ν-a.e. x, so $R_x(\mathrm{d}y)\, \nu(\mathrm{d}x) \ll \nu(\mathrm{d}x)\, \nu(\mathrm{d}y)$ and there exists a jointly measurable function $r\colon \mathbb{X}^2 \to \mathbb{R}_+$ such that

$$R_x(\mathrm{d}y)\, \nu(\mathrm{d}x) = r(x, y)\, \nu(\mathrm{d}x)\, \nu(\mathrm{d}y). \tag{4.10}$$

In that case, (4.1) implies that, as measures on $\mathbb{X}^2$,

$$r(x, y)\, \nu(\mathrm{d}x)\, \nu(\mathrm{d}y) = r(y, x)\, \nu(\mathrm{d}x)\, \nu(\mathrm{d}y). \tag{4.11}$$

Let $A, B, C \in \mathcal{X}$. It follows from the form of the predictive distributions (3.1) that

$$\begin{aligned}
&\mathbb{P}(X_1 \in A, X_2 \in B, X_3 \in C) \\
&= \int_A \int_B \mathbb{P}(X_3 \in C \mid X_1 = x, X_2 = y) \mathbb{P}(X_2 \in \mathrm{d}y \mid X_1 = x)\, \nu(\mathrm{d}x) \\
&= \int_A \int_B \left\{ (1 - a_2)\nu(C) + \frac{a_2}{2} R_x(C) + \frac{a_2}{2} R_y(C) \right\} ((1 - a_1)\, \nu(\mathrm{d}y) + a_1\, R_x(\mathrm{d}y))\, \nu(\mathrm{d}x) \\
&= \int_A \Bigg\{ (1 - a_1)(1 - a_2)\nu(B)\nu(C) + a_1(1 - a_2)\nu(C) R_x(B) + \frac{(1 - a_1)a_2}{2} \nu(B) R_x(C) \\
&\qquad + \frac{a_1 a_2}{2} R_x(B) R_x(C) + \frac{(1 - a_1)a_2}{2} \int_B R_y(C)\, \nu(\mathrm{d}y) + \frac{a_1 a_2}{2} \int_B R_y(C)\, R_x(\mathrm{d}y) \Bigg\}\, \nu(\mathrm{d}x) \\
&= (1 - a_1)(1 - a_2)\nu(A)\nu(B)\nu(C) + a_1(1 - a_2)\nu(C) \int_A R_x(B)\, \nu(\mathrm{d}x) \\
&\quad + \frac{(1 - a_1)a_2}{2} \nu(B) \int_A R_x(C)\, \nu(\mathrm{d}x) + \frac{a_1 a_2}{2} \int_A R_x(B) R_x(C)\, \nu(\mathrm{d}x) \\
&\quad + \frac{(1 - a_1)a_2}{2} \nu(A) \int_B R_y(C)\, \nu(\mathrm{d}y) + \frac{a_1 a_2}{2} \int_A \int_B R_y(C)\, R_x(\mathrm{d}y)\, \nu(\mathrm{d}x).
\end{aligned} \tag{4.12}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

By exchangeability, $\mathbb{P}(X_1 \in A, X_2 \in B, X_3 \in C) = \mathbb{P}(X_1 \in A, X_2 \in C, X_3 \in B)$, so from (4.12) we obtain, after some simplification,

$$\begin{aligned}
&a_1(1-a_2)\nu(C)\int_A R_x(B)\,\nu(\mathrm{d}x) + \frac{(1-a_1)a_2}{2}\nu(B)\int_A R_x(C)\,\nu(\mathrm{d}x)\\
&\quad + \frac{(1-a_1)a_2}{2}\nu(A)\int_B R_y(C)\,\nu(\mathrm{d}y) + \frac{a_1a_2}{2}\int_A\int_B R_y(C)\,R_x(\mathrm{d}y)\,\nu(\mathrm{d}x)\\
&= a_1(1-a_2)\nu(B)\int_A R_x(C)\,\nu(\mathrm{d}x) + \frac{(1-a_1)a_2}{2}\nu(C)\int_A R_x(B)\,\nu(\mathrm{d}x)\\
&\quad + \frac{(1-a_1)a_2}{2}\nu(A)\int_C R_y(B)\,\nu(\mathrm{d}y) + \frac{a_1a_2}{2}\int_A\int_C R_y(B)\,R_x(\mathrm{d}y)\,\nu(\mathrm{d}x).
\end{aligned}$$

Using that $\int_B R_y(C)\,\nu(\mathrm{d}y) = \int_C R_y(B)\,\nu(\mathrm{d}y)$ from (4.1), we further get

$$\begin{aligned}
&a_1(1-a_2)\nu(C)\int_A R_x(B)\,\nu(\mathrm{d}x) + \frac{(1-a_1)a_2}{2}\nu(B)\int_A R_x(C)\,\nu(\mathrm{d}x)\\
&\quad + \frac{a_1a_2}{2}\int_A\int_B R_y(C)\,R_x(\mathrm{d}y)\,\nu(\mathrm{d}x)\\
&= a_1(1-a_2)\nu(B)\int_A R_x(C)\,\nu(\mathrm{d}x) + \frac{(1-a_1)a_2}{2}\nu(C)\int_A R_x(B)\,\nu(\mathrm{d}x)\\
&\quad + \frac{a_1a_2}{2}\int_A\int_C R_y(B)\,R_x(\mathrm{d}y)\,\nu(\mathrm{d}x),
\end{aligned}$$

which after some simple algebra becomes

$$\begin{aligned}
&\int_A\int_B R_y(C)\,R_x(\mathrm{d}y)\,\nu(\mathrm{d}x) - c^*\cdot\nu(C)\int_A R_x(B)\,\nu(\mathrm{d}x)\\
&\qquad = \int_A\int_C R_y(B)\,R_x(\mathrm{d}y)\,\nu(\mathrm{d}x) - c^*\cdot\nu(B)\int_A R_x(C)\,\nu(\mathrm{d}x),
\end{aligned}$$

where $c^* = (a_2 - 2a_1 + a_1a_2)/a_1a_2 \le 0$ from before. In terms of the representation in (4.10), the above becomes

$$\begin{aligned}
&\int_A\int_B\int_C (r(x,y)r(y,z) - c^*r(x,y))\,\nu(\mathrm{d}z)\,\nu(\mathrm{d}y)\,\nu(\mathrm{d}x)\\
&\quad = \int_A\int_B\int_C (r(x,z)r(z,y) - c^*r(x,z))\,\nu(\mathrm{d}z)\,\nu(\mathrm{d}y)\,\nu(\mathrm{d}x)\\
&\quad = \int_A\int_B\int_C (r(x,z)r(y,z) - c^*r(x,z))\,\nu(\mathrm{d}z)\,\nu(\mathrm{d}y)\,\nu(\mathrm{d}x),
\end{aligned}$$

where the last equation follows from (4.11). As a result, for $(\nu\times\nu\times\nu)$-a.e. (x,y,z),

$$r(x,y)(r(y,z) - c^*) = r(x,z)(r(y,z) - c^*).$$

Suppose by contradiction that $c^* < 0$. In such a case, after canceling the common term above, we get, for $(\nu\times\nu\times\nu)$-a.e. (x,y,z) in $\mathbb{X}^3$, $r(x,y) = r(x,z)$. Together with (4.11), this implies that, for $(\nu\times\nu\times\nu)$-a.e. (x,y,z) in $\mathbb{X}^3$,

$$r(x,y)r(y,z) = r(x,z)r(z,y). \tag{4.13}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Now, recall that for every $x \in D$ there exists $B_x \in \mathcal{X}$ such that $\nu(B_x) \neq R_x(B)$. It follows from (4.13) that, for ν-a.e. x in D,

$$\begin{aligned}\int_{\mathbb{X}} R_y(B_x)\, R_x(\mathrm{d}y) &= \int_{y\in\mathbb{X}} \int_{z\in B_x} r(x, y) r(y, z)\, \nu(\mathrm{d}z)\, \nu(\mathrm{d}y) \\ &= \int_{y\in B_x} \int_{z\in\mathbb{X}} r(x, y) r(y, z)\, \nu(\mathrm{d}z)\, \nu(\mathrm{d}y) = R_x(B_x).\end{aligned}$$

On the other hand, by (4.2), for ν-a.e. x in D,

$$\int_{\mathbb{X}} R_y(B_x)\, R_x(\mathrm{d}y) = (1 - c^*) R_x(B_x) + c^* \nu(B_x);$$

therefore, $R_x(B_x) = \nu(B_x)$ – absurd. As a result, $c^* = 0$, which implies as in the previous subcase that $(X_n)_{n\geq 1}$ is an MVPS with parameters $(a^{-1} - 1, \nu, R)$.

Case 2: $\nu(D) = 0$. If $\nu(D) = 0$, then $R_x \equiv \nu$ for ν-a.e. x. Since $\mathbb{P}(X_1 \in \cdot\,) = \nu(\cdot)$, it follows that, for every $B_1, B_2 \in \mathcal{X}$,

$$\begin{aligned}\mathbb{P}(X_1 \in B_1, X_2 \in B_2) &= \int_{B_1} \mathbb{P}(X_2 \in B_2 \mid X_1 = x)\, \nu(\mathrm{d}x) \\ &= \int_{B_1} \{(1 - a_1)\nu(B_2) + a_1 R_x(B_2)\}\, \nu(\mathrm{d}x) = \nu(B_1)\nu(B_2);\end{aligned}$$

thus, $\mathbb{P}_{(X_1,X_2)}(\cdot) = (\nu \times \nu)(\cdot)$. By induction, for every $B_1, \ldots, B_n \in \mathcal{X}$,

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \nu(B_i);$$

therefore, $(X_n)_{n\geq 1}$ is i.i.d. with marginal distribution ν, and thus trivially an MVPS with parameters $(a^{-1} - 1, \nu, R)$. This concludes the proof of Theorem 3.1. □

Proof of Theorem 3.2. Suppose that $(X_n)_{n\geq 1}$ satisfies (3.3) and (3.4). By assumption, $\mathcal{G}$ is c.g. under ν, so there exists (see [6, p. 649]) an almost everywhere proper regular version R of $\nu(\cdot \mid \mathcal{G})$, in the sense that, for some $F \in \mathcal{G}$ such that $\nu(F) = 1$,

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

Moreover, since $(X_n)_{n\geq 1}$ is identically distributed with marginal distribution ν, we have $\mathbb{P}(X_1 \in F, \ldots, X_n \in F) = 1$ for each $n = 1, 2, \ldots$, which implies that, on a set of probability 1,

$$\sum_{i=1}^{n} \delta_{X_i}(A) = \sum_{i=1}^{n} R_{X_i}(A) \quad \text{for all } A \in \mathcal{G}. \tag{4.14}$$

Let $B_1, \ldots, B_k \in \mathcal{G}$ form a partition of $\mathbb{X}$ with $k > 2$ and $\#\{j\colon \nu(B_j) > 0\} > 2$. It follows from (3.3) that the sequence $(X_n)_{n\geq 1}$ on the level of $\{B_1, \ldots, B_k\}$ satisfies Johnson's sufficientness postulate (1.2), so from [29, Theorem 2.1 and Corollary 2.1], there exist $\alpha_{B_1}, \ldots, \alpha_{B_k} \geq 0$ such that $\sum_{i=1}^{k} \alpha_{B_i} > 0$ and, for each $j = 1, \ldots, k$,

$$\mathbb{P}(X_{n+1} \in B_j \mid X_1, \ldots, X_n) = \frac{\alpha_{B_j} + N_{n,B_j}}{\sum_{i=1}^{k} \alpha_{B_i} + n} \quad \text{a.s.}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

In particular, $\mathbb{P}(X_1 \in B_j) = \alpha_{B_j} / \sum_{i=1}^k \alpha_{B_i}$, which implies that $a_{B_j} = \nu(B_j) \sum_{i=1}^k \alpha_{B_i}$. Moreover, $\theta := \sum_{i=1}^k \alpha_{B_i}$ does not depend on the particular partition chosen. Indeed, let $\{C_1, \ldots, C_m\} \subseteq \mathcal{G}$ be another partition of $\mathbb{X}$ with $m > 2$ and $\#\{l\colon \nu(C_l) > 0\} > 2$. If $\{C_1, \ldots, C_m\}$ is finer than $\{B_1, \ldots, B_k\}$ and, say, $B_j = \bigcup_{l=1}^L C_{j_l}$ for some j with $\nu(B_j) > 0$, we obtain, using the aforementioned results in [29],

$$\begin{aligned}\frac{\theta \nu(B_j) + N_{n,B_j}}{\theta + n} &= \sum_{l=1}^L \mathbb{P}(X_{n+1} \in C_{j_l} \mid X_1, \ldots, X_n) \\ &= \sum_{l=1}^L \frac{\nu(C_{j_l}) \sum_{i=1}^m \alpha_{C_i} + N_{n,C_{j_l}}}{\sum_{i=1}^m \alpha_{C_i} + n} = \frac{\nu(B_j) \sum_{i=1}^m \alpha_{C_i} + N_{n,B_j}}{\sum_{i=1}^m \alpha_{C_i} + n} \quad \text{a.s.}\end{aligned}$$

After simplification, we get $n\big(\theta - \sum_{i=1}^m \alpha_{C_i}\big)\nu(B_j) = \big(\theta - \sum_{i=1}^m \alpha_{C_i}\big)N_{n,B_j}$ a.s. for all $n = 1, 2, \ldots$, which implies that $\theta = \sum_{i=1}^m \alpha_{C_i}$. In any other case, we can consider the partition $\{B_i \cap C_l, i = 1, \ldots, k, l = 1, \ldots, m\}$ and conclude that the three partitions produce the same constant.

Let $A \in \mathcal{G}$. Arguing through the partition $\{A \cap B_j, A^c \cap B_j, j = 1, \ldots, k\}$, we get

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \ldots, X_n) = \sum_{j=1}^k \frac{\theta \nu(A \cap B_j) + N_{n,A\cap B_j}}{\theta + n} = \frac{\theta \nu(A) + \sum_{i=1}^n \delta_{X_i}(A)}{\theta + n} \quad \text{a.s.}$$

Therefore, using that $\mathcal{G}$ is c.g. under ν, we obtain, as measures on $(\mathbb{X}, \mathcal{G})$,

$$\mathbb{P}(X_{n+1} \in \mathrm{d}x \mid X_1, \ldots, X_n) = \frac{\theta\, \nu(\mathrm{d}x) + \sum_{i=1}^n \delta_{X_i}(\mathrm{d}x)}{\theta + n} \quad \text{a.s.} \tag{4.15}$$

Let $A \in \mathcal{G}$ and $B \in \mathcal{X}$. Recall that $\mathcal{G}_n = X_n^{-1}(\mathcal{G})$ for $n \geq 1$. Then

$$\mathbb{E}[\mathbf{1}_A(X_{n+1}) \cdot \mathbb{P}(X_{n+1} \in B \mid \mathcal{G}_{n+1})] = \nu(A \cap B) = \int_A R_x(B)\, \nu(\mathrm{d}x) = \mathbb{E}[\mathbf{1}_A(X_{n+1}) \cdot R_{X_{n+1}}(B)],$$

and, since $\omega \mapsto R_{X_{n+1}(\omega)}(B)$ is $\mathcal{G}_{n+1}$-measurable,

$$\mathbb{P}(X_{n+1} \in B \mid \mathcal{G}_{n+1}) = R_{X_{n+1}}(B) \quad \text{a.s.} \tag{4.16}$$

As a result, using (3.4), (4.15), and (4.16), we obtain, on a set of probability 1,

$$\begin{aligned}\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n) &= \mathbb{E}[\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n, \mathcal{G}_{n+1}) \mid X_1, \ldots, X_n] \\ &= \mathbb{E}[\mathbb{P}(X_{n+1} \in B \mid \mathcal{G}_{n+1}) \mid X_1, \ldots, X_n] \\ &= \mathbb{E}[R_{X_{n+1}}(B) \mid X_1, \ldots, X_n] \\ &= \int_{\mathbb{X}} R_x(B) \left\{ \frac{\theta\, \nu(\mathrm{d}x) + \sum_{i=1}^n \delta_{X_i}(\mathrm{d}x)}{\theta + n} \right\} \\ &= \frac{\theta \int_{\mathbb{X}} R_x(B)\, \nu(\mathrm{d}x) + \sum_{i=1}^n \int_{\mathbb{X}} R_x(B)\, \delta_{X_i}(\mathrm{d}x)}{\theta + n} \\ &= \frac{\theta \nu(B) + \sum_{i=1}^n R_{X_i}(B)}{\theta + n}.\end{aligned}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Note that the above result does not contradict (4.15) when $B \in \mathcal{G}$ because of (4.14). Therefore, we can conclude that $(X_n)_{n\geq 1}$ is an MVPS with parameters (θ, ν, R).

Conversely, suppose that $(X_n)_{n\geq 1}$ is an exchangeable MVPS with reinforcement kernel $R(\cdot) = \nu(\cdot \mid \mathcal{G})$. Recall from Theorem 2.1 and the discussion around (2.4) that any exchangeable MVPS admits such a representation. Moreover, we may assume that R is almost everywhere proper, without loss of generality. Arguing as in (4.14), we obtain, for all $A \in \mathcal{G}$, on a set of probability 1,

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \ldots, X_n) = \frac{\theta\nu(A) + \sum_{i=1}^n R_{X_i}(A)}{\theta + n} = \frac{\theta\nu(A) + \sum_{i=1}^n \delta_{X_i}(A)}{\theta + n} = \frac{\theta\nu(A) + N_{n,A}}{\theta + n};$$

thus, $(X_n)_{n\geq 1}$ satisfies (3.3). Let $A_1, \ldots, A_n, B \in \mathcal{X}$ and $A \in \mathcal{G}$. Then

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\mathbf{1}_A(X_{n+1})\mathbf{1}_B(X_{n+1})] \\
&\quad= \mathbb{E}[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\cdot\mathbb{P}(X_{n+1}\in A\cap B \mid X_1,\ldots,X_n)] \\
&\quad= \mathbb{E}\left[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\frac{\theta\nu(A\cap B)+\sum_{i=1}^n R_{X_i}(A\cap B)}{\theta+n}\right] \\
&\quad\overset{(a)}{=} \mathbb{E}\left[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\frac{\theta\nu(A\cap B)+\sum_{i=1}^n R_{X_i}(B)\delta_{X_i}(A)}{\theta+n}\right] \\
&\quad= \mathbb{E}\left[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\int_A R_x(B)\left\{\frac{\theta\,\nu(\mathrm{d}x)+\sum_{i=1}^n\delta_{X_i}(\mathrm{d}x)}{\theta+n}\right\}\right] \\
&\quad= \mathbb{E}[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\cdot\mathbb{E}[\mathbf{1}_A(X_{n+1})R_{X_{n+1}}(B)\mid X_1,\ldots,X_n]] \\
&\quad= \mathbb{E}[\mathbf{1}_{A_1}(X_1)\cdots\mathbf{1}_{A_n}(X_n)\mathbf{1}_A(X_{n+1})\cdot R_{X_{n+1}}(B)] \\
&\quad= \mathbb{E}[\mathbb{P}(X_1\in A_1,\ldots,X_n\in A_n\mid\mathcal{G}_{n+1})\cdot\mathbf{1}_A(X_{n+1})\cdot R_{X_{n+1}}(B)] \\
&\quad\overset{(b)}{=} \mathbb{E}[\mathbb{P}(X_1\in A_1,\ldots,X_n\in A_n\mid\mathcal{G}_{n+1})\cdot\mathbf{1}_A(X_{n+1})\cdot\mathbb{P}(X_{n+1}\in B\mid\mathcal{G}_{n+1})],
\end{aligned}$$

where (a) follows from $R_x(A\cap B) = \nu(A\cap B\mid\mathcal{G})(x) = \nu(B\mid\mathcal{G})(x)\delta_x(A) = R_x(B)\delta_x(A)$ for ν-a.e. x, since $A\in\mathcal{G}$; and for (b) we have used that $\mathbb{P}(X_{n+1}\in B\mid\mathcal{G}_{n+1}) = R_{X_{n+1}}(B)$ a.s., arguing as in (4.16). Therefore, X_{n+1} and $(X_1,\ldots,X_n)$ are conditionally independent given $\mathcal{G}_{n+1}$. This concludes the proof of Theorem 3.2. □

Proof of Proposition 3.1. Let us define, for every $n\geq 0$, $l = 1,\ldots,m$, and $n_l = 0, 1, \ldots, n$, the quantity

$$g_{n,l}(n_l) := \sum_{j\in D_l} f_{n,j}(n_l).$$

It follows from (3.7) that $f_{n,j}(n_l) > 0$, and so $g_{n,l}(n_l) > 0$. Recall that $m(j)$ denotes, for each $j = 1,\ldots,k$, the index $l\in\{1,\ldots,m\}$ such that $j\in D_l$. Define $Y_n := m(X_n)$ for $n\geq 1$. Then $(Y_n)_{n\geq 1}$ is an exchangeable process with predictive distributions

$$\mathbb{P}(Y_{n+1} = l \mid Y_1,\ldots,Y_n) = P_n(D_l) = g_{n,l}(N_{n,l}) \quad \text{a.s., for } l = 1,\ldots,m. \tag{4.17}$$

Therefore, $(Y_n)_{n\geq 1}$ satisfies Johnson's sufficientness postulate (1.2), so from [29, Theorem 2.1 and Corollary 2.1] there exist $\bar{w}_1,\ldots,\bar{w}_m > 0$ such that

$$g_{n,l}(n_l) = \frac{\bar{w}_l + n_l}{\bar{w} + n}, \tag{4.18}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

where $\bar{w} := \sum_{l=1}^{m} \bar{w}_l$; thus, in particular,

$$\mathbb{P}_0(D_l) = \mathbb{P}(Y_1 = l) = g_{0,l}(0) = \frac{\bar{w}_l}{\bar{w}} \quad \text{for } l = 1, \dots, m. \tag{4.19}$$

Let $j \in \{1, \dots, k\}$ be such that $D_{m(j)} = \{j\}$. Then $f_{n,j}(n_{m(j)}) = g_{n,m(j)}(n_{m(j)})$, so from (4.18) and (4.19),

$$\mathbb{P}_n(\{j\}) = g_{n,m(j)}(N_{n,m(j)}) = \frac{\bar{w} \cdot \mathbb{P}_0(\{j\}) + \sum_{i=1}^{n} \mathbf{1}_{D_{m(j)}}(X_i)}{\bar{w} + n} \quad \text{a.s.} \tag{4.20}$$

Let $i, j \in \{1, \dots, k\}$ be such that $m(i) = m(j)$. Fix $n \geq 0$, where for $n = 0$ the expressions involving $\mathbb{P}(\cdot \mid X_1, \dots, X_n)$ are meant as unconditional statements. It follows that, for every $h = 1, 2, \dots$,

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = j, X_{n+2} \notin D_{m(j)}, \dots, X_{n+h+1} \notin D_{m(j)} \mid X_1, \dots, X_n) \\
&\qquad = f_{n,j}(N_{n,m(j)}) \prod_{s=1}^{h} (1 - g_{n+s,m(j)}(N_{n,m(j)} + 1)) \quad \text{a.s.}, \\
&\mathbb{P}(X_{n+1} \notin D_{m(j)}, \dots, X_{n+h} \notin D_{m(j)}, X_{n+h+1} = j \mid X_1, \dots, X_n) \\
&\qquad = f_{n+h,j}(N_{n,m(j)}) \prod_{s=0}^{h-1} (1 - g_{n+s,m(j)}(N_{n,m(j)})) \quad \text{a.s.}
\end{aligned}$$

By exchangeability, the two expressions are equal, so we obtain

$$\frac{f_{n,j}(N_{n,m(j)})}{f_{n+h,j}(N_{n,m(j)})} = \frac{\prod_{s=0}^{h-1} (1 - g_{n+s,m(j)}(N_{n,m(j)}))}{\prod_{s=1}^{h} (1 - g_{n+s,m(j)}(N_{n,m(j)} + 1))} \quad \text{a.s.}$$

Since the right-hand side depends on j only through $m(j)$, the same ratio $f_{n,i}/f_{n+h,i}$ for i is equal to the same quantity; therefore, with $n_{m(j)} \equiv n_{m(i)} \in \{0, 1, \dots, n\}$ fixed,

$$\frac{f_{n,i}(n_{m(i)})}{f_{n,j}(n_{m(j)})} = \frac{f_{n+h,i}(n_{m(i)})}{f_{n+h,j}(n_{m(j)})} \quad \text{for every } h = 1, 2, \dots \tag{4.21}$$

On the other hand, for every $h = 1, \dots, n$,

$$\begin{aligned}
&\mathbb{P}(X_1 \notin D_{m(j)}, \dots, X_n \notin D_{m(j)}, X_{n+1} = j, X_{n+2} \in D_{m(j)}, \dots, X_{2n+1} \in D_{m(j)}) \\
&\qquad = f_{n,j}(0) \prod_{s=0}^{n-1} (1 - g_{s,m(j)}(0)) \prod_{s=1}^{n} g_{n+s,m(j)}(s) \quad \text{a.s.}
\end{aligned}$$

and, applying the permutation σ of $\{1, \dots, 2n+1\}$ that swaps the indices $(1, \dots, h)$ with $(n+2, \dots, n+h+1)$,

$$\begin{aligned}
&\mathbb{P}(X_{\sigma(1)} \notin D_{m(j)}, \dots, X_{\sigma(n)} \notin D_{m(j)}, X_{\sigma(n+1)} = j, \dots, X_{\sigma(2n+1)} \in D_{m(j)}) \\
&\qquad = f_{n,j}(h) \prod_{s=0}^{h-1} g_{s,m(j)}(s) \prod_{s=0}^{n-h-1} (1 - g_{h+s,m(j)}(h)) \\
&\qquad\quad \times \prod_{s=1}^{h} (1 - g_{n+s,m(j)}(h)) \prod_{s=1}^{n-h} g_{n+h+s,m(j)}(h + s - 1) \quad \text{a.s.}
\end{aligned}$$

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Since $(X_1, \ldots, X_{2n+1}) \overset{\mathrm{d}}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(2n+1)})$ by exchangeability, we obtain, arguing as in (4.21), with n fixed,

$$\frac{f_{n,i}(0)}{f_{n,j}(0)} = \frac{f_{n,i}(h)}{f_{n,j}(h)} \quad \text{for every } h = 1, \ldots, n. \tag{4.22}$$

It follows from (4.21) and (4.22) that, for any such pair i, j, the ratio

$$\frac{f_{n,i}(n_{m(i)})}{f_{n,j}(n_{m(j)})} =: d_{i,j}$$

is independent of n and $n_{m(j)} \equiv n_{m(i)}$. Summing over all i in $D_{m(j)}$ and letting $w_j := \bar{w}_{m(j)} / \sum_{i \in D_{m(j)}} d_{i,j}$ gives

$$f_{n,j}(n_{m(j)}) = \frac{1}{\sum_{i \in D_{m(j)}} d_{i,j}} g_{n,m(j)}(n_{m(j)}) = \frac{w_j}{\bar{w}_{m(j)}} g_{n,m(j)}(n_{m(j)});$$

thus, in particular, from (4.19),

$$\mathbb{P}_0(\{j\}) = f_{0,j}(0) = \frac{w_j}{\bar{w}_{m(j)}} g_{0,m(j)}(0) = \frac{w_j}{\bar{w}}.$$

In that case, $\mathbb{P}_0(\{j\} \mid D_{m(j)}) = w_j / \bar{w}_{m(j)}$; therefore, using (4.18),

$$\begin{aligned} \mathbb{P}_n(\{j\}) = f_{n,j}(N_{n,m(j)}) &= \frac{w_j}{\bar{w}_{m(j)}} \frac{\bar{w}_{m(j)} + N_{n,m(j)}}{\bar{w} + n} \\ &= \frac{\bar{w} \cdot \mathbb{P}_0(\{j\}) + \mathbb{P}_0(\{j\} \mid D_{m(j)}) \sum_{i=1}^{n} \mathbf{1}_{D_{m(j)}}(X_i)}{\bar{w} + n} \quad \text{a.s.} \end{aligned} \tag{4.23}$$

We can now conclude from (4.20) and (4.23) that, as measures on $\mathbb{X}$,

$$\begin{aligned} \mathbb{P}(X_{n+1} \in \mathrm{d}y \mid X_1, \ldots, X_n) &= \frac{\bar{w} \cdot \mathbb{P}_0(\mathrm{d}y) + \sum_{i=1}^{n} \sum_{l=1}^{m} \mathbb{P}_0(\mathrm{d}y \mid D_l) \cdot \mathbf{1}_{D_l}(X_i)}{\bar{w} + n} \\ &= \frac{\bar{w} \cdot \mathbb{P}_0(\mathrm{d}y) + \sum_{i=1}^{n} \mathbb{P}_0(\mathrm{d}y \mid \sigma(D_1, \ldots, D_m))(X_i)}{\bar{w} + n} \quad \text{a.s.}, \end{aligned}$$

which implies that $(X_n)_{n \geq 1}$ is an MVPS with parameters $(\bar{w}, \mathbb{P}_0, \mathbb{P}_0(\cdot \mid \sigma(D_1, \ldots, D_m)))$. □

Remark 4.1. Notice that when $m = 2$, the sufficientness postulate (4.17) at the level of the partition does not necessarily imply (4.18), and so $(X_n)_{n \geq 1}$ need not be an MVPS (see [29, Remark 1] and references therein). To obtain (4.18) when $m = 2$, the authors in [29] suggest assuming that the function $g_{n,l}$ in (4.17) is linear in n_l.

Proof of Proposition 3.2. For every $l = 1, \ldots, m$ and $p_l \in \{(\bar{w}_l + n_l)/(\bar{w} + n) \colon n_l = 0, \ldots, n; n = 0, 1, \ldots\}$, let us define the quantity $g_l(p_l) := \sum_{j \in D_l} f_j(p_l)$. Then a straightforward extension of the arguments in [17] to m colors implies that $g_l(p_l) = p_l$. The rest of the proof follows the same steps as in the proof of Proposition 3.1, starting from (4.18). □

Acknowledgements

We are grateful to an Associate Editor and two anonymous referees for their valuable comments and helpful suggestions, which improved the quality of this work.

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

Funding Information

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.

Competing Interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII – 1983* (Lect. Notes *Math.* 1117), eds D. J. Aldous, I. A. Ibragimov and J. Jacod. Springer, Berlin, pp. 1–198.
[2] BACALLADO, S., BATTISTON, M., FAVARO, S. AND TRIPPA, L. (2017). Sufficientness postulates for Gibbs-type priors and hierarchical generalizations. *Statist. Sci.* **32**, 487–500.
[3] BANDYOPADHYAY, A. AND THACKER, D. (2022). A new approach to Pólya urn schemes and its infinite color generalization. *Ann. Appl. Prob.* **32**, 46–79.
[4] BERTI, P., DREASSI, E., LEISEN, F., PRATELLI, L. AND RIGO, P. (2023). Kernel-based Dirichlet sequences. *Bernoulli* **29**, 1321–1342.
[5] BERTI, P., DREASSI, E., LEISEN, F., PRATELLI, L. AND RIGO, P. (2023). A probabilistic view on predictive constructions for Bayesian learning. To appear in *Statist. Sci.*
[6] BERTI, P. AND RIGO, P. (2007). 0–1 laws for regular conditional distributions. *Ann. Prob.* **35**, 649–662.
[7] BLACKWELL, D. AND MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353–355.
[8] CAMERLENGHI, F. AND FAVARO, S. (2021). On Johnson's 'sufficientness' postulates for feature-sampling models. *Mathematics* **9**, 2891.
[9] FONG, E., HOLMES, C. AND WALKER, S. G. (2023). Martingale posterior distributions. *J. R. Statist. Soc. B* **85**, 1357–1391.
[10] FORTINI, S., LADELLI, L. AND REGAZZINI, E. (2000). Exchangeability, predictive distributions and parametric models. *Sankhyā A* **62**, 86–109.
[11] FORTINI, S. AND PETRONE, S. (2023). Prediction-based uncertainty quantification for exchangeable sequences. *Phil. Trans. R. Soc. A* **381**, 20220142.
[12] FORTINI, S. AND PETRONE, S. (2024). Exchangeability, prediction and predictive modeling in Bayesian statistics. *Statist. Sci.* **40**, 40–67.
[13] FORTINI, S., PETRONE, S. AND SARIEV, H. (2021). Predictive constructions based on measure-valued Pólya urn processes. *Mathematics* **9**, 2845.
[14] GARIBALDI, U. AND SCALAS, E. (2010). *Finitary Probabilistic Methods in Econophysics*. Cambridge University Press.
[15] GHOSAL, S. AND VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.
[16] HANSEN, B. AND PITMAN, J. (2000). Prediction rules for exchangeable sequences related to species sampling. *Statist. Prob. Lett.* **46**, 251–256.
[17] HILL, B. M., LANE, D. AND SUDDERTH, W. (1987). Exchangeable urn processes. *Ann. Prob.* **15**, 1586–1592.
[18] JANSON, S. (2019). Random replacements in Pólya urns with infinitely many colours. *Electron. Commun. Prob.* **24**, 23.
[19] JOHNSON, W. E. (1932). Probability: The deductive and inductive problems. *Minds* **41**, 409–423.
[20] KALLENBERG, O. (2021). *Foundations of Modern Probability*, 3rd edn. Springer, New York.
[21] LANGE, K. (1973). Decompositions of substochastic transition functions. *Proc. Amer. Math. Soc*. 37, 575–580.
[22] LIJOI, A. AND PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics*, eds N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker. Cambridge University Press, pp. 80–136.
[23] LO, A. Y. (1991). A characterization of the Dirichlet process. *Statist. Prob. Lett.* **13**, 185–187.
[24] MAILLER, C. AND MARCKERT, J.-F. (2017). Measure-valued Pólya urn processes. *Electron. J. Prob.* **22**, 26.
[25] PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell* (IMS Lect. Notes Monogr. *Ser*. 30), eds T. S. Ferguson, L. S. Shapley and J. B. MacQueen. Institute of Mathematical Statistics, 245–267.

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press

[26] REGAZZINI, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità. *G. Ist. Ital. Attuari* **41**, 77–89.
[27] SARIEV, H. AND SAVOV, M. (2024). Characterization of exchangeable measure-valued Pólya urn sequences. *Electron. J. Prob.* **29**, 73.
[28] WALKER, S. G. AND MULIERE, P. (1999). A characterization of a neutral to the right prior via an extension of Johnson's sufficientness postulate. *Ann. Statist.* **27**, 589–599.
[29] ZABELL, S. L. (1982). W. E. Johnson's 'sufficientness' postulate. *Ann. Statist.* 10, 1090–1099.
[30] ZABELL, S. L. (1995). Characterizing Markov exchangeable sequences. *J. Theoret. Prob.* **8**, 175–178.
[31] ZABELL, S. L. (1997). The continuum of inductive methods revisited. In *The Cosmos of Science: Essays in Exploration*, eds. J. Earman and J. D. Norton. University of Pittsburgh Press, pp. 351–385.

https://doi.org/10.1017/jpr.2024.109 Published online by Cambridge University Press