Infinite-color randomly reinforced urns with dominant colors

HRISTO SARIEV [1,*], SANDRA FORTINI [2,†] and SONIA PETRONE[2,‡]

[1]*Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, "Akad. Georgi Bonchev" bl. 8, Sofia 1113, Bulgaria. E-mail:* [*]h.sariev@math.bas.bg
[2]*Department of Decision Sciences, Bocconi University, via Roentgen 1, 20136 Milano, Italy.*
E-mail: [†]sandra.fortini@unibocconi.it; [‡]sonia.petrone@unibocconi.it

We define and prove limit results for a class of *dominant Pólya sequences*, which are randomly reinforced urn processes with *color-specific* random weights and *unbounded* number of possible colors. Under fairly mild assumptions on the expected reinforcement, we show that the predictive and the empirical distributions converge almost surely (a.s.) in total variation to the same random probability measure $\tilde{P}$; moreover, $\tilde{P}(\mathcal{D}) = 1$ a.s., where $\mathcal{D}$ denotes the set of *dominant* colors for which the expected reinforcement is maximum. In the general case, the predictive probabilities and the empirical frequencies of any δ-neighborhood of $\mathcal{D}$ converge a.s. to one. That is, although *non*-dominant colors continue to be regularly observed, their distance to $\mathcal{D}$ converges in probability to zero. We refine the above results with rates of convergence. We further hint potential applications of dominant Pólya sequences in randomized clinical trials and species sampling, and use our central limit results for Bayesian inference.

Keywords: Reinforced processes; Pólya sequences; Random probability measures; Species sampling; Bayesian nonparametrics

1. Introduction

Random processes with reinforcement constitute an important class of mathematical models that are of interest to probabilistis and practitioners alike; see [29] for a review on the subject. A classic example is the k-color *Pólya urn* model, which describes the composition of an urn with balls of k different colors that is sequentially being sampled and reinforced with an additional ball of the observed color. Denote by $X_n \in \{1, \ldots, k\}$ the color of the additional ball at step n. Then, X_1 is chosen from a discrete distribution with $\mathbb{P}(X_1 = j) = m_j / \sum_{l=1}^{k} m_l, j = 1, \ldots, k$, where $m_j > 0$ is the initial number of balls of color j in the urn, and, for every $n \geq 1$,

$$\mathbb{P}(X_{n+1} = j | X_1, \ldots, X_n) = \frac{m_j + \sum_{i=1}^{n} \delta_{X_i}(\{j\})}{\sum_{l=1}^{k} m_l + n},$$

where δ_x is the Dirac measure at the point x. It is well-known (see, e.g., [29, p.30]) that, for each $j = 1, \ldots, k$, as n approaches infinity, the proportion of j-colored balls in the urn converges almost surely (a.s.) to a random variable $\tilde{p}_j$, that is,

$$\mathbb{P}(X_{n+1} = j | X_1, \ldots, X_n) \xrightarrow{a.s.} \tilde{p}_j.$$

Moreover, the random vector $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_k)$ has a Dirichlet distribution with parameters $(m_j)_{j=1}^{k}$ and the process $(X_n)_{n \geq 1}$ is exchangeable; thus, by de Finetti representation theorem for exchangeable

1

sequences, see, e.g., [1, Theorem 3.1], given $\tilde{p}$, the X_n are conditionally independent and identically distributed (i.i.d.) according to $\tilde{p}$.

In this paper, we study the generalization of the above urn sampling scheme towards infinite colors and random reinforcement. Let $(\mathbb{X}, \mathcal{X})$ be a measurable space, ν a probability measure on $\mathbb{X}$, and $\theta > 0$ a positive constant. Formally, we consider a sequence of $\mathbb{X}$-valued random variables $(X_n)_{n \geq 1}$ with predictive structure given by $\mathbb{P}(X_1 \in \cdot) = \nu(\cdot)$ and, for every $n \geq 1$,

$$\mathbb{P}(X_{n+1} \in \cdot \mid X_1, W_1, \ldots, X_n, W_n) = \sum_{i=1}^{n} \frac{W_i}{\theta + \sum_{j=1}^{n} W_j} \delta_{X_i}(\cdot) + \frac{\theta}{\theta + \sum_{j=1}^{n} W_j} \nu(\cdot), \qquad (1.1)$$

where $(W_n)_{n \geq 1}$ is a sequence of non-negative random weights. The important extension of the k-color Pólya urn model to infinite colors, also known as *Pólya sequence*, was proposed by Blackwell and MacQueen [11] and corresponds to the case of constant reinforcement $W_n = 1$ and ν possibly diffuse. By Theorem 1 in [11], the predictive distributions (1.1) of a Pólya sequence converge weakly with probability one to a random probability measure $\tilde{P}$ on $\mathbb{X}$, which has a Dirichlet process distribution with parameters (θ, ν). Moreover, given $\tilde{P}$, the random variables $X_1, X_2, \ldots$ are i.i.d. with marginal distribution $\tilde{P}$; thus, they are exchangeable. We recall that, for any exchangeable sequence, the predictive and the empirical distributions converge to the same random probability measure, which is called the *directing random measure* of the process; see [1, Section 2].

Because of their simple predictive structure, Pólya sequences play a fundamental role in the construction of more complex procedures for nonparametric Bayesian inference; see [16] for a review. On the other hand, there is a growing interest from different areas of research, such as signalling theory [3], economics [6], clinical trial design [32], in extending the basic Pólya urn model along the lines of time-dependent or random reinforcement. The predictive construction (1.1) thus represents a natural and important contribution in this direction. However, the weighing scheme in (1.1) may lead to an unbalanced design, which can violate the condition of exchangeability. In general, an exchangeable process $(X_n)_{n \geq 1}$ with prediction rule (1.1) is a member of the family of *species sampling sequences* [31], which implies by [20, Theorem 1] a rather restrictive form of dependence between X_n and W_n.

Bassetti and coauthors [5] define a class of *generalized* species sampling sequences that have predictive distributions of the kind (1.1), and the process they generate is *conditionally identically distributed* (c.i.d.). By [9], a stochastic process $(Y_n)_{n \geq 1}$ is c.i.d. if, for every $n \geq 1$, given $Y_1, \ldots, Y_n$, the random variables $Y_{n+1}, Y_{n+2}, \ldots$ are identically distributed. Therefore, any exchangeable process is c.i.d., while the converse is generally not true; indeed, by [24, Proposition 2.1], a sequence of random variables is exchangeable if and only if it is both stationary and c.i.d. Nevertheless, c.i.d. processes are asymptotically exchangeable and thus preserve basic limit properties of exchangeable sequences; see [9] for more details. In (1.1), the presence of the random weights W_n can lead to a temporary disequilibrium in the observation process. Then it may be reasonable to assume that $(X_n)_{n \geq 1}$ is c.i.d., and a sufficient condition due to [5] is that W_n and X_n are conditionally independent given $(X_1, W_1, \ldots, X_{n-1}, W_{n-1})$.

In this work, we focus on the case where the random reinforcement W_n in (1.1) *depends* explicitly on the color X_n, and the process $(X_n)_{n \geq 1}$ is in general neither exchangeable nor c.i.d. This situation occurs in many contexts; for example, in problems of species sampling, color-specific random weights describe different levels of species adaptability. Then we expect that colors which have a higher expected reinforcement will tend to *dominate* the observation process, just as better-adapted species have a greater chance at survival. To the best of our knowledge, existing studies consider only the model with k colors, which is known as a randomly reinforced urn (RRU), and show that the probability of drawing a color whose expected weight is maximum goes indeed to one, see, e.g., [7, 27, 28, 33].

We thus explore the extension of RRUs to an unbounded set of possible colors, which parallels the extension provided by exchangeable Pólya sequences for k-color Pólya urns.

Recently, [4, 23, 26] introduced a general class of random processes with reinforcement, called measure-valued Pólya urn processes, which formalize the idea of urn composition as a random finite measure. In this setting, the conditional distribution (1.1) can be regarded as the (normalized) composition of an urn, which is "reinforced" at each time n with the random measure $W_n \delta_{X_n}$; see [18] for more details. A proper urn model to illustrate the sampling procedure implied by (1.1) is a weighted version of Hoppe's urn [22]. Imagine an urn initially containing θ black balls. Then, at each step n, a ball is picked at random from the urn. If the ball is black, we generate a new color from ν and update the urn with a random number of balls of the new color; if a non-black ball is picked instead, it is replaced together with a random number of balls of the same color. It follows that the parameter θ controls the probability of observing a new color, and ν is the color distribution. The above sampling scheme can be visualized as a randomly reinforced version of the Chinese restaurant process, which models the random partition generated by a Pólya sequence.

We study the limit behavior of the proposed class of randomly reinforced Pólya sequences under minimal assumptions on the reinforcement. Let $\bar{w}$ be the sup of the *expected* reinforcement as it varies across colors, and $\mathcal{D} \subseteq \mathbb{X}$ the set of *dominant* colors whose expected weight is $\bar{w}$. Let us denote the predictive and the empirical distributions of $(X_n)_{n \geq 1}$ by

$$P_n(\cdot) \equiv \mathbb{P}(X_{n+1} \in \cdot \mid X_1, W_1, \ldots, X_n, W_n) \qquad \text{and} \qquad \hat{P}_n(\cdot) \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\cdot).$$

We show in Theorem 3.3 that, under certain conditions of compactness, the probability of observing *non*-dominant colors may not vanish, though their distance to $\mathcal{D}$ converges in probability to zero. If $\bar{w}$ is significantly larger than the expected reinforcement of any $x \in \mathcal{D}^c$, then there exists a random probability measure $\tilde{P}$ on $\mathbb{X}$ with $\tilde{P}(\mathcal{D}) = 1$ a.s. such that

$$\|P_n - \tilde{P}\| \xrightarrow{a.s.} 0 \qquad \text{and} \qquad \|\hat{P}_n - \tilde{P}\| \xrightarrow{a.s.} 0, \tag{1.2}$$

where $\|\alpha - \beta\| \equiv \sup_{A \in \mathcal{X}} |\alpha(A) - \beta(A)|$ is the total variation distance between two probability measures α, β on $\mathbb{X}$. The latter implies, in particular, that $\tilde{P}$ is a.s. discrete on the set $\mathcal{D}$. On the other hand, one can show from (1.2) that $(X_n)_{n \geq 1}$ is asymptotically exchangeable with limit directing random measure $\tilde{P}$; that is, as $n \to \infty$, $(X_{n+1}, X_{n+2}, \ldots)$ converges in distribution to an exchangeable sequence with directing random measure $\tilde{P}$, see [1, Lemma 8.2].

In addition, we provide set-wise rates of convergence for P_n and $\hat{P}_n$ in the form of a central limit theorem; namely, Theorem 4.2 gives conditions under which, for every $A \in \mathcal{X}$ such that $A \cap \mathcal{D} \neq \emptyset$,

$$\sqrt{n}(\hat{P}_n(A) - P_n(A)) \qquad \text{and} \qquad \sqrt{n}(P_n(A) - \tilde{P}(A)),$$

converge to non-degenerate Gaussian limits in the sense of stable convergence and a.s. conditional convergence, respectively. We illustrate the statistical use of these results by obtaining marginal asymptotic credible intervals for $\tilde{P}$ under a Bayesian approach.

The paper is organized as follows. The formal definition of the proposed Pólya sequences with dominant colors is given in Section 2. Their first-order convergence properties are derived in Section 3, which also includes examples of their potential use in species sampling and clinical trials. Section 4 contains results on rates of convergence and central limit theorems. The proofs are postponed to Section 5.

2. Pólya sequences with dominant colors

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space, $(\mathbb{X}, d)$ a complete separable metric space, endowed with its Borel σ-algebra $\mathcal{X}$, and $(X_n)_{n \geq 1}$ a sequence of $\mathbb{X}$-valued random variables on $(\Omega, \mathcal{H}, \mathbb{P})$ with predictive distributions (1.1). We assume that the random reinforcement in (1.1) is color-specific in the sense that

$$W_n = h(X_n, U_n), \tag{2.1}$$

where $(U_n)_{n \geq 1}$ is a sequence of i.i.d. random variables such that U_n is independent of $(X_1, \ldots, X_n)$, and h is a measurable function from $\mathbb{X} \times \mathbb{R}$ into $\mathbb{R}_+$.

Let us define the expected weight of color $x \in \mathbb{X}$ by

$$w(x) = \mathbb{E}[W_n | X_n = x],$$

and denote

$$\bar{w} = \sup_{x \in \mathbb{X}} w(x).$$

We further let $\mathcal{F}_0 = \{\Omega, \emptyset\}$, $\mathcal{F}_n = \sigma(X_1, U_1, \ldots, X_n, U_n)$, $n \geq 1$, and

$$\mathcal{D} = \{x \in \mathbb{X} : w(x) = \bar{w}\}, \qquad \mathcal{D}_\delta = \{x \in \mathbb{X} : d(x, \mathcal{D}) < \delta\},$$

for every $\delta > 0$, where $d(x, \mathcal{D}) \equiv \inf\{d(x, y) : y \in \mathcal{D}\}$ denotes the distance of x from $\mathcal{D}$. By construction, w is $\mathcal{X}$-measurable, so $\mathcal{D} \in \mathcal{X}$, and we call $\mathcal{D}$ the set of *dominant* colors. We make the following assumptions:

$$
\begin{aligned}
&0 \leq W_n \leq \beta \text{ for some constant } \beta; \\
&\bar{w} \in \text{supp}(\nu_w); \\
&\bar{w} > \sup_{x \in \mathcal{D}_\delta^c} w(x) \text{ for every } \delta > 0,
\end{aligned}
\tag{2.2}
$$

where ν_w is the image measure of ν under w, and $\text{supp}(\nu_w)$ is the support of ν_w on $\mathbb{R}_+$, defined by

$$\text{supp}(\nu_w) = \{u \geq 0 : \forall \epsilon > 0, \nu_w((u - \epsilon, u + \epsilon)) > 0\}.$$

The assumptions (2.2) essentially require that the colors x, for which the expected weight $w(x)$ is arbitrarily close to $\bar{w}$, lie in a compact set that we are able to sample from. If $\mathcal{D}$ is non-empty and $\nu(\mathcal{D}) > 0$, then $\bar{w} \in \text{supp}(\nu_w)$ is automatically satisfied, in which case we shall be interested in the relationship between $\bar{w}$ and the next highest value in the support of ν_w,

$$\bar{w}^c = \sup\{u \geq 0 : u \in \text{supp}((\nu_{|\mathcal{D}^c})_w)\},$$

where $\nu_{|\mathcal{D}^c}(\cdot) = \nu(\cdot \,|\, \mathcal{D}^c)$. We call any process characterized by (1.1) with (2.1) a *dominant Pólya sequence* (DPS) if it satisfies the assumptions (2.2).

3. First-order convergence

We first consider the DPS with a constant expected weight function, $w(x) \equiv \bar{w}$. Then all colors are dominant and one can show that W_n and X_n are conditionally uncorrelated given $\mathcal{F}_{n-1}$. Recall from

Section 1 that $(X_n)_{n \geq 1}$ is c.i.d. whenever W_n and X_n are conditionally independent. In that case, general c.i.d. theory implies that the predictive and empirical distributions converge set-wise to the same random limit; see [9, Section 2]. Here, $(X_n)_{n \geq 1}$ is generally not c.i.d., yet P_n and $\hat{P}_n$ converge to a common random probability measure, and convergence is in total variation.

Proposition 3.1. *Assume $\mathbb{X} = \mathcal{D}$. Then there exists a random probability measure $\tilde{P}$ on $\mathbb{X}$ such that*

$$\|P_n - \tilde{P}\| \xrightarrow{a.s.} 0 \qquad and \qquad \|\hat{P}_n - \tilde{P}\| \xrightarrow{a.s.} 0.$$

The next result concerns the opposite extreme case of a strictly increasing w function on $\mathbb{X} = [0,1]$. Then "1" is the only dominant color and we show that the predictive and the empirical distributions converge weakly with probability one to a probability measure degenerate at "1". Even so, the probability of drawing a *non*-dominant color may not vanish.

Proposition 3.2. *Let $\mathbb{X} = [0,1]$ and $(X_n)_{n \geq 1}$ be a DPS with a continuous and strictly increasing expected weight function w. Then, as $n \to \infty$, $\frac{1}{n} \sum_{i=1}^{n} W_i \xrightarrow{a.s.} w(1)$ and*

$$P_n \xrightarrow{w} \delta_1 \quad a.s. [\mathbb{P}] \qquad and \qquad \hat{P}_n \xrightarrow{w} \delta_1 \quad a.s. [\mathbb{P}].$$

Moreover, there exists a $[0,1]$-valued random variable η such that $P_n(\{1\}^c) \xrightarrow{a.s.} \eta$.

The sampling scheme implied by a strictly increasing w can be described as follows: colors are picked and reinforced until a new color closer to "1" is drawn. This color is then increasingly preferred by the model until a better alternative appears, which happens almost surely. In fact, as we will show in Proposition 4.1, *non*-dominant colors are forgotten quicker the more distant they are from "1".

Given Propositions 3.1 and 3.2, we are ready to state our main result of this section, which involves a DPS with an arbitrary expected weight function $w(\cdot)$ and a general space of colors $\mathbb{X}$. We show that, as n increases, the observations tend to concentrate around the set of dominant colors. If further $\bar{w} > \bar{w}^c$, then the probability of observing *non*-dominant colors converges a.s. to zero.

Theorem 3.3. *For any DPS $(X_n)_{n \geq 1}$, as $n \to \infty$, $\frac{1}{n} \sum_{i=1}^{n} W_i \xrightarrow{a.s.} \bar{w}$ and*

$$P_n(\mathcal{D}_\delta^c) \xrightarrow{a.s.} 0 \qquad for \ every \ \delta > 0.$$

Moreover, there exists a $[0,1]$-valued random variable η such that

$$P_n(\mathcal{D}^c) \xrightarrow{a.s.} \eta.$$

If, in addition, it holds that $\bar{w} > \bar{w}^c$, then $\eta = 0$ and there exists a random probability measure $\tilde{P}$ on $\mathbb{X}$ with $\tilde{P}(\mathcal{D}) = 1$ a.s. [$\mathbb{P}$] such that

$$\|P_n - \tilde{P}\| \xrightarrow{a.s.} 0 \qquad and \qquad \|\hat{P}_n - \tilde{P}\| \xrightarrow{a.s.} 0.$$

If $\bar{w} = \bar{w}^c$, then $P_n(\mathcal{D}_\delta) \xrightarrow{a.s.} 1$ and, by Theorem 3.3, $\liminf_n P_n(\mathcal{D}_\delta \backslash \mathcal{D})$ can be strictly positive. In that case, *non*-dominant colors continue to be regularly observed yet get closer and closer to $\mathcal{D}$. In fact, using dominated convergence,

$$d(X_n, \mathcal{D}) \xrightarrow{p} 0. \tag{3.1}$$

If $\bar{w} > \bar{w}^c$, then, in particular, $P_n \xrightarrow{w} \tilde{P}$ a.s.$[\mathbb{P}]$ and $\hat{P}_n \xrightarrow{w} \tilde{P}$ a.s.$[\mathbb{P}]$. It follows from a generalized martingale convergence theorem (see [10, Theorem 2]) that the conditional distribution of X_{n+1} given $(X_1, \ldots, X_n)$ converges weakly with probability one to $\tilde{P}$. Thus, by Lemma 8.2 in [1], $(X_n)_{n \geq 1}$ is asymptotically exchangeable with limit directing random measure $\tilde{P}$.

The reinforced urn scheme (1.1) implies that there will be ties in a sample $(X_1, \ldots, X_n)$ of size n with positive probability. Let us denote the number of distinct colors in $(X_1, \ldots, X_n)$ by

$$L_n = \max\{k \in \{1, \ldots, n\} : X_k \notin \{X_1, \ldots, X_{k-1}\}\}.$$

The next result shows that L_n is approximately $\bar{w}^{-1}\theta \log n$ for large n.

Proposition 3.4. *For any DPS $(X_n)_{n \geq 1}$, if ν is diffuse, then, as $n \to \infty$,*

$$\frac{L_n}{\log n} \xrightarrow{a.s.} \frac{\theta}{\bar{w}}.$$

3.1. Examples

We close this section with a few examples and hints of potential areas of application.

Example **3.5** (k-color RRU). Let $F_1, \ldots F_k$ be probability distribution functions on $[0, \beta]$. Denote $w_j = \int t dF_j(t)$, $j = 1, \ldots, k$, and suppose that, for some $k_0 \in \{1, \ldots, k\}$,

$$w_1 = \cdots = w_{k_0} = \bar{w} > \bar{w}^c \equiv \max_{k_0 < j \leq k} w_j. \tag{3.2}$$

The k-color RRU is a DPS that associates $W_n|\{X_n = j\} \sim F_j$ to color $j = 1, \ldots, k$. Then $w(j) = w_j$, $j = 1, \ldots, k$ and $\mathcal{D} = \{1, \ldots, k_0\}$. RRUs with $k_0 = 1$ and $k_0 = k$ have been studied by [27, 28], while [7, 33] consider the case $1 < k_0 < k$, although [7] assumes (3.2) only asymptotically.

Denote by $P_n(\cdot|\mathcal{D}) = P_n(\cdot \cap \mathcal{D})/P_n(\mathcal{D})$ the predictive distribution of X_{n+1} restricted to $\mathcal{D}$. It follows from Proposition 3.1 that $P_n(\{j\}|\mathcal{D}) \xrightarrow{a.s.} \tilde{p}_j$, $j = 1, \ldots, k_0$, for some non-negative random variables $\tilde{p}_1, \ldots, \tilde{p}_{k_0}$ such that $\sum_{j=1}^{k_0} \tilde{p}_j = 1$ a.s.$[\mathbb{P}]$. As $P_n(\mathcal{D}) \xrightarrow{a.s.} 1$ by Theorem 3.3, then

$$P_n(\{j\}) = P_n(\mathcal{D})P_n(\{j\}|\mathcal{D}) + P_n(\mathcal{D}^c)P_n(\{j\}|\mathcal{D}^c) \xrightarrow{a.s.} \tilde{p}_j.$$

Therefore, $\|P_n - \tilde{P}\| \xrightarrow{a.s.} 0$ and, similarly, $\|\hat{P}_n - \tilde{P}\| \xrightarrow{a.s.} 0$, where $\tilde{P} = \sum_{j=1}^{k_0} \tilde{p}_j \delta_j$.

Example **3.6** (Unimodal DPS). Suppose that we have a DPS with a unique dominant color, *i.e.*, $\mathcal{D} = \{x_0\}$ for some $x_0 \in \mathbb{X}$. It follows from (3.1) that

$$X_n \xrightarrow{p} x_0.$$

Moreover,

$$P_n \xrightarrow{w} \delta_{x_0} \qquad \text{a.s.}[\mathbb{P}].$$

Indeed, for every open set $G \in \mathcal{X}$, we have that, if $x_0 \in G$, then there exists $\delta > 0$ such that $x_0 \in \mathcal{D}_\delta \subseteq G$; therefore, by Theorem 3.3,

$$\liminf_{n \to \infty} P_n(G) \geq \liminf_{n \to \infty} P_n(\mathcal{D}_\delta) = 1 = \delta_{x_0}(G) \qquad \text{a.s.}[\mathbb{P}].$$

As $\delta_{x_0}(G) = 0$ otherwise, a.s. weak convergence of P_n to δ_{x_0} follows.

Unimodal DPSs can be used in the context of response-adaptive randomization procedures for clinical trials (see [32] for a general reference) to provide an ethical design that can deal with an *unbounded* number or even a *continuum* of potential treatments. Consider a stylized trial for determining the optimal dose of a drug. Let X_n represent the dose given to the nth patient who enters the trial, and denote by W_n the patient's response to it. Then the reinforced predictive scheme (1.1) with a unimodal $w(\cdot)$ function provides a response-adaptive rule for selecting the dose for the next patient. It follows from Theorem 3.3 that such a design is ethical in the sense that, for any $\delta > 0$, $P_n(\mathcal{D}_\delta) \xrightarrow{a.s.} 1$; that is, even in situations where it is practically impossible to select x_0 exactly (e.g., $\nu(\{x_0\}) = 0$ with ν is diffuse), the probability of assigning the next patient a dose arbitrarily close to x_0 converges a.s. to one.

***Example* 3.7** (Species sampling model with dominant species)**.**   Exchangeable species sampling models are of interest in many fields and, in particular, in the broad area of Bayesian nonparametrics (see [31]), yet forms of competition between species can easily break the symmetry imposed by exchangeability. Dominant Pólya sequences can thus provide an extension to the framework of species sampling models by addressing situations where exchangeability holds only asymptotically *and* a restricted set of dominant species tends to prevail. Another related direction of application is in random graph theory, where conciliating between exchangeability and graph density is a challenging issue (see [12]), and dominant Pólya sequences can suggest 'sparse' and asymptotically exchangeable graph structures.

Consider again the sampling scheme (1.1). In this case, interest lies on the *categorical* species and X_n plays the role of a label chosen, say, in $\mathbb{X} = [0, 1]$ from a uniform ν. Take $0 < w_1 < w_2 < \infty$ and $p \in (0, 1)$. Let $W_n = w(X_n)$ where

$$w(x) = w_1 \cdot \mathbb{1}_{[0,p)}(x) + w_2 \cdot \mathbb{1}_{[p,1]}(x) \quad \text{for } x \in \mathbb{X}.$$

Then $(X_n)_{n \geq 1}$ is a DPS with $\mathcal{D} = [p, 1]$. According to this scheme, when a *new* species appears, it has probability p to be *non*-dominant and $(1 - p)$ to be dominant, independently over n. As $p > 0$, then $\sum_{i=1}^{n} \delta_{X_i}(\mathcal{D}^c) = \infty$ a.s.[$\mathbb{P}$] and we discover *non*-dominant species infinitely often. However, $\hat{P}_n([p, 1]) \xrightarrow{a.s.} 1$ by Theorem 3.3, so the discovery rate of *non*-dominant species is of order less than n (see also Proposition 4.1).

4. Rates and central limit results

Let $(X_n)_{n \geq 1}$ be a DPS such that $\bar{w} > \bar{w}^c$, and let $\tilde{P}$ be a random probability measure as in Theorem 3.3. It follows for every $A \in \mathcal{X}$ that

$$P_n(A) - \tilde{P}(A) \xrightarrow{a.s.} 0 \qquad \text{and} \qquad \hat{P}_n(A) - P_n(A) \xrightarrow{a.s.} 0.$$

We proceed to study the rate of convergence by first considering $A = \mathcal{D}^c$.

4.1. Rate of convergence for dominated colors

Assume $\bar{w} > \bar{w}^c$. By Theorem 3.3, both $\hat{P}_n(\mathcal{D}^c)$ and $P_n(\mathcal{D}^c)$ converge a.s. to zero. The next result shows that $\hat{P}_n(\mathcal{D}^c)$ and $P_n(\mathcal{D}^c)$ are $O(n^{-(\bar{w} - \bar{w}^c)/\bar{w}})$, as $n \to \infty$.

Proposition 4.1. *Suppose $\bar{w} > \bar{w}^c$. Then there exists a random variable $\xi \in [0, \infty)$ such that, letting $\gamma = 1 - \bar{w}^c/\bar{w}$, as $n \to \infty$,*

$$n^\gamma \cdot \hat{P}_n(\mathcal{D}^c) \xrightarrow{a.s.} \xi \qquad and \qquad n^\gamma \cdot P_n(\mathcal{D}^c) \xrightarrow{a.s.} \frac{\bar{w}^c}{\bar{w}}\xi.$$

Moreover, $n^\gamma \cdot \hat{P}_n(\mathcal{D}^c) \xrightarrow{a.s.} \infty$ and $n^\gamma \cdot P_n(\mathcal{D}^c) \xrightarrow{a.s.} \infty$ when $\gamma > 1 - \bar{w}^c/\bar{w}$.

4.2. Central limit theorem

Let $A \in \mathcal{X}$ be such that $A \cap \mathcal{D} \neq \emptyset$. We provide conditions under which

$$C_n(A) = \sqrt{n}\big(\hat{P}_n(A) - P_n(A)\big) \qquad and \qquad D_n(A) = \sqrt{n}\big(P_n(A) - \tilde{P}(A)\big),$$

converge to non-degenerate Gaussian limits. The results are given in terms of stable and almost sure (a.s.) conditional convergence, which we briefly recall.

Almost sure conditional convergence. Let $\mathcal{G} = (\mathcal{G}_n)_{n \geq 0}$ be a filtration on $\mathcal{H}$, and $\tilde{Q}$ a random probability measure on $\mathbb{X}$. A sequence $(Y_n)_{n \geq 1}$ is said to converge to $\tilde{Q}$ in the sense of a.s. conditional convergence with respect to (w.r.t.) $\mathcal{G}$ if, as $n \to \infty$,

$$\mathbb{P}(Y_n \in \cdot \mid \mathcal{G}_n) \xrightarrow{w} \tilde{Q}(\cdot) \quad \text{a.s.}[\mathbb{P}].$$

We refer to [14] for more details.

Stable convergence. Stable convergence is a strong form of convergence in distribution, albeit weaker than a.s. conditional convergence. A sequence $(Y_n)_{n \geq 1}$ converges stably to $\tilde{Q}$ if, as $n \to \infty$,

$$\mathbb{E}\big[V f(Y_n)\big] \longrightarrow \mathbb{E}\Big[V \int_{\mathbb{X}} f(x)\tilde{Q}(dx)\Big],$$

for all continuous bounded functions f and any integrable random variable V. The main application of stable convergence is in central limit theorems that allow for mixing variables in the limit. See [21] for a complete reference on stable convergence.

In the sequel, the stable and a.s. conditional limits will be some Gaussian law, which we denote by $\mathcal{N}(\mu, \sigma^2)$ for parameters (μ, σ^2), where $\mathcal{N}(\mu, 0) = \delta_\mu$. Recall from Section 2 that $\mathcal{F} = (\mathcal{F}_n)_{n \geq 0}$ is the filtration on $\mathcal{H}$ generated by $(X_n, U_n)_{n \geq 1}$.

Lastly, we can show using standard arguments that $\mathbb{E}[f(X_{n+1})|\mathcal{F}_n] \xrightarrow{a.s.} \int_{\mathbb{X}} f(x)\tilde{P}(dx)$ for every bounded measurable function f. Then the following limit exists with probability one

$$q_A = \lim_{n \to \infty} \mathbb{E}[W_{n+1}^2 \delta_{X_{n+1}}(A)|\mathcal{F}_n].$$

Theorem 4.2. *Suppose $\bar{w} > 2\bar{w}^c$. Let $A \in \mathcal{X}$ be such that $A \cap \mathcal{D} \neq \emptyset$. Define*

$$V(A) = \frac{1}{\bar{w}^2}\big\{(\tilde{P}(A^c))^2 q_A + (\tilde{P}(A))^2 q_{A^c}\big\} \quad and \quad U(A) = V(A) - \tilde{P}(A)\tilde{P}(A^c).$$

Then

$$C_n(A) \xrightarrow{stably} \mathcal{N}\big(0, U(A)\big) \qquad and \qquad D_n(A) \xrightarrow{a.s.cond.} \mathcal{N}\big(0, V(A)\big) \quad w.r.t. \ \mathcal{F}. \qquad (4.1)$$

It follows from (4.1) and Lemma 1 in [8] that

$$\sqrt{n}\big(\hat{P}_n(A) - \tilde{P}(A)\big) = C_n(A) + D_n(A) \xrightarrow{stably} \mathcal{N}\big(0, U(A) + V(A)\big).$$

Moreover, the a.s. conditional convergence of $D_n(A)$ implies that the distribution of $\tilde{P}(A)$ has no point masses when $0 < \nu(A) < 1$; see the proof of Theorem 3.2 in [2]. Then $0 < \tilde{P}(A) < 1$ a.s.$[\mathbb{P}]$, and $V(A)$ and $U(A)$ are strictly positive with probability one.

If $\mathcal{D} = \mathbb{X}$, the assumptions of Theorem 4.2 are automatically satisfied and (4.1) follows. In the general case, conditionally on $\mathcal{D}$, we have effectively a constant expected weight function. Let us define, for $n \geq 1$,

$$P_n(A|\mathcal{D}) = \frac{P_n(A \cap \mathcal{D})}{P_n(\mathcal{D})} \qquad \text{and} \qquad \hat{P}_n(A|\mathcal{D}) = \frac{\hat{P}_n(A \cap \mathcal{D})}{\hat{P}_n(\mathcal{D})}.$$

Under $\bar{w} > 2\bar{w}^c$, it holds $\nu(\mathcal{D}) > 0$, so $P_n(\mathcal{D}) > 0$. Moreover, $\liminf_{n \to \infty} \hat{P}_n(\mathcal{D}) > 0$ a.s.$[\mathbb{P}]$ by Theorem 3.3; thus, $\hat{P}_n(\mathcal{D}) > 0$ a.s.$[\mathbb{P}]$ for large n. Then, working with the sequence $(X_n)_{n \geq 1}$ restricted to $\mathcal{D}$, we get

$$\sqrt{n}\big(\hat{P}_n(A|\mathcal{D}) - P_n(A|\mathcal{D})\big) \xrightarrow{stably} \mathcal{N}\big(0, U(A)\big),$$

$$\sqrt{n}\big(P_n(A|\mathcal{D}) - \tilde{P}(A)\big) \xrightarrow{a.s.cond.} \mathcal{N}\big(0, V(A)\big) \quad \text{w.r.t. } \mathcal{F}. \tag{4.2}$$

In order to show (4.1), the assumption $\bar{w} > 2\bar{w}^c$ in Theorem 4.2 is critical. Indeed, under $\bar{w} \leq 2\bar{w}^c$, Proposition 4.1 implies that

$$D_n(A) - \sqrt{n}\big(P_n(A|\mathcal{D}) - \tilde{P}(A)\big) = \sqrt{n}\, P_n(\mathcal{D}^c) P_n(A|\mathcal{D}) \xrightarrow{a.s.} \infty;$$

thus, $D_n(A)$ (and, similarly, $C_n(A)$) fails to converge as $\sqrt{n}(P_n(A|\mathcal{D}) - \tilde{P}(A))$ converges by (4.2).

***Example* 4.3** (k-color RRU (Continued)). Consider again the k-color RRU from Example 3.5. It follows that $q_j = \lim_n \mathbb{E}[W_{n+1}^2 \delta_{X_{n+1}}(\{j\})|\mathcal{F}_n]$, $j = 1, \ldots, k_0$ exists a.s.$[\mathbb{P}]$. Then, arguing as in (4.2),

$$\sqrt{n}\big(\hat{P}_n(\{j\}|\mathcal{D}) - P_n(\{j\}|\mathcal{D})\big) \xrightarrow{stably} \mathcal{N}\Big(0, \frac{\tilde{p}_j}{\bar{w}^2}\Big\{(1-\tilde{p}_j)^2 q_j + \tilde{p}_j \sum_{i \leq k_0, i \neq j} q_i \tilde{p}_i\Big\} - \tilde{p}_j(1-\tilde{p}_j)\Big),$$

$$\sqrt{n}\big(P_n(\{j\}|\mathcal{D}) - \tilde{p}_j\big) \xrightarrow{a.s.cond.} \mathcal{N}\Big(0, \frac{\tilde{p}_j}{\bar{w}^2}\Big\{(1-\tilde{p}_j)^2 q_j + \tilde{p}_j \sum_{i \leq k_0, i \neq j} q_i \tilde{p}_i\Big\}\Big) \quad \text{w.r.t. } \mathcal{F}.$$

In addition, $\mathbb{P}(\tilde{p}_j = p) = 0$, $p \in [0, 1]$, so that $p_j > 0$ a.s.$[\mathbb{P}]$ and $\tilde{p}_i \neq \tilde{p}_j$ a.s.$[\mathbb{P}]$, $i \neq j$.

4.3. An application: asymptotic credible intervals for $\tilde{P}$

We close with an application of Theorem 4.2 to a problem of statistical inference on the limit structure of the process. We assume that the random weights W_n are observable, as it might be the case, for example, in the context of clinical trials (see Example 3.6). Suppose $\mathbb{X} = \mathbb{R}$, $\mathcal{D} = [a, b]$, $\bar{w} > 2\bar{w}^c$. Define

$F_n(x) = P_n((-\infty, x])$, $x \in \mathbb{R}$, $n \geq 1$. By Theorem 3.3, there exists a random distribution function $\tilde{F}$ on $[a, b]$ such that

$$\|F_n - \tilde{F}\| \xrightarrow{a.s.} 0.$$

The process $\tilde{F}$ is measurable w.r.t. $\sigma(X_1, W_1, X_2, W_2, \ldots)$ and so cannot be calculated on the basis of a finite sample $(X_i, W_i)_{i=1}^n$. Still, one can do inference on it; under a Bayesian approach, one does so by computing the conditional distribution of $\tilde{F}$ given $(X_i, W_i)_{i=1}^n$. The latter is usually summarized by a credible band or, in practice, by plotting *marginal* credible intervals around $\tilde{F}(x)$ for x varying on a fine grid. Take $x \in [a, b]$ such that $0 < \nu((-\infty, x]) < 1$. The following limits exist a.s.

$$q_1 = \lim_{n \to \infty} \mathbb{E}\big[W_{n+1}^2 \delta_{X_{n+1}}([a, x]) \big| \mathcal{F}_n\big], \qquad q_0 = \lim_{n \to \infty} \mathbb{E}\big[W_{n+1}^2 \delta_{X_{n+1}}((x, b]) \big| \mathcal{F}_n\big]. \qquad (4.3)$$

It follows from Theorem 4.2 that

$$\sqrt{n}\big(F_n(x) - \tilde{F}(x)\big) \xrightarrow{a.s.cond.} \mathcal{N}(0, V_x) \quad \text{w.r.t. } \mathcal{F},$$

where $V_x = \bar{w}^{-2}((1 - \tilde{F}(x))^2 q_1 + \tilde{F}(x)^2 q_0)$. In order to obtain asymptotic credible intervals for $\tilde{F}(x)$, we need a convergent estimator of V_x. Let us define, for $n \geq 1$,

$$V_{n,x} = \frac{1}{m_n^2}\big((1 - F_n(x))^2 s_{n,1} + (F_n(x))^2 s_{n,0}\big),$$

$$m_n = \frac{1}{n} \sum_{i=1}^n W_i, \quad s_{n,1} = \frac{1}{n} \sum_{i=1}^n W_i^2 \delta_{X_i}([a, x]), \quad s_{n,0} = \frac{1}{n} \sum_{i=1}^n W_i^2 \delta_{X_i}((x, b]).$$

By Theorem 3.3, $m_n \xrightarrow{a.s.} \bar{w}$. Moreover, $\sum_{n=1}^\infty \mathbb{E}[W_n^2]/n^2 \leq \sum_{n=1}^\infty \beta^2/n^2 < \infty$, so Lemma 2 in [8] implies from (4.3) that $s_{n,1} \xrightarrow{a.s.} q_1$ and $s_{n,0} \xrightarrow{a.s.} q_0$. Thus,

$$V_{n,x} \xrightarrow{a.s.} V_x.$$

A generalized Slutsky's theorem (see also [17, Theorem 6]) implies, for every $t \in \mathbb{R}$,

$$\mathbb{P}\Big(\frac{\tilde{F}(x) - F_n(x)}{\sqrt{V_{n,x}/n}} \leq t \Big| \mathcal{F}_n\Big) \xrightarrow{a.s.} \mathcal{N}(0, 1)((-\infty, t]).$$

As a consequence, denoting by z_α the $(1 - \alpha/2)$-quantile of the standard Normal distribution, we get

$$\liminf_{n \to \infty} \mathbb{P}\Big(F_n(x) - z_\alpha \sqrt{\frac{V_{n,x}}{n}} \leq \tilde{F}(x) \leq F_n(x) + z_\alpha \sqrt{\frac{V_{n,x}}{n}} \Big| \mathcal{F}_n\Big) \geq 1 - \alpha.$$

Thus, an asymptotic marginal credible interval for $\tilde{F}(x)$ at level $1 - \alpha$ is

$$\Big[F_n(x) - z_\alpha \sqrt{\frac{V_{n,x}}{n}}, \ F_n(x) + z_\alpha \sqrt{\frac{V_{n,x}}{n}}\Big].$$

5. Proofs

We shall use the following notation throughout this section,

$$N_n(A) = \theta\nu(A) + \sum_{i=1}^{n} W_i \delta_{X_i}(A), \quad M_n(A) = \sum_{i=1}^{n} \delta_{X_i}(A) + 1, \quad N_n = \theta + \sum_{i=1}^{n} W_i,$$

so that $P_n(A) = N_n(A)/N_n$ for every $n \geq 0$ and $A \in \mathcal{X}$, with the convention $\sum_{i=1}^{0} a_i = 0$, and we will omit the parenthesis when A is an interval for the sake of clarity. All random quantities are defined on the probability space $(\Omega, \mathcal{H}, \mathbb{P})$ unless otherwise specified.

The following identity, given for any $A \in \mathcal{X}$ and $n \geq 0$, is a consequence of (2.1) and will be applied repeatedly,

$$\mathbb{E}[W_{n+1}\delta_{X_{n+1}}(A)|\mathcal{F}_n] = \mathbb{E}[w(X_{n+1})\delta_{X_{n+1}}(A)|\mathcal{F}_n]. \tag{5.1}$$

The proof of Proposition 3.1 uses the following general fact for DPSs with $\mathbb{X} = \mathcal{D}$.

Lemma 5.1. *Under the conditions in Proposition 3.1, as $n \to \infty$,*

$$\frac{n}{N_n} \longrightarrow \frac{1}{\bar{w}} \qquad a.s.[\mathbb{P}] \text{ and in } L^p \text{ for all } p \geq 1.$$

Proof. By hypothesis, $\mathbb{E}[W_{n+1}|\mathcal{F}_n] = \bar{w}$ and $\sum_{n=1}^{\infty} \mathbb{E}[W_n^2]/n^2 \leq \sum_{n=1}^{\infty} \beta^2/n^2 < \infty$, so [8, Lemma 2] implies that $N_n/n \xrightarrow{a.s.} \bar{w}$. Define $N_n^* = \sum_{i=1}^{n}\{W_i - \mathbb{E}[W_i]\}$, $n \geq 1$. By a classical martingale inequality (see [25, Lemma 1.5] and [8, Lemma 3]), we have

$$\mathbb{P}(|N_n^*| > x) \leq 2\exp\{-x^2/2\beta n^2\} \qquad \text{for all } x > 0.$$

Fix $p > 0$. It follows that

$$\mathbb{E}[N_n^{-p}] = p \int_{\theta}^{\infty} \frac{1}{t^{p+1}}\mathbb{P}(N_n < t)dt \leq \frac{p}{\theta^{p+1}} \int_{\theta}^{\theta + \frac{n\bar{w}}{2}} \mathbb{P}(N_n < t)dt + p \int_{\theta + \frac{n\bar{w}}{2}}^{\infty} \frac{1}{t^{p+1}}dt.$$

Clearly, $p \int_{\theta + \frac{n\bar{w}}{2}}^{\infty} t^{-p-1}dt = (\theta + \frac{n\bar{w}}{2})^{-p} = O(n^{-p})$. On the other hand, for $t < \theta + \frac{n\bar{w}}{2}$,

$$\mathbb{P}(N_n < t) = \mathbb{P}(N_n^* < t - \theta - n\bar{w}) \leq \mathbb{P}(|N_n^*| > \theta + n\bar{w} - t) \leq 2\exp\left\{-\frac{(\theta + n\bar{w} - t)^2}{2\beta^2 n}\right\};$$

thus, $\int_{\theta}^{\theta + \frac{n\bar{w}}{2}} \mathbb{P}(N_n < t)dt \leq n\bar{w}\exp\{-n\bar{w}^2/4\beta^2\}$ and, ultimately, $\mathbb{E}[N_n^{-p}] = O(n^{-p})$. □

Proof of Proposition 3.1. Let $A \in \mathcal{X}$ and $n \geq 0$. Then

$$P_{n+1}(A) - P_n(A) = \frac{W_{n+1}/N_n}{1 + W_{n+1}/N_n}\left(\delta_{X_{n+1}}(A) - P_n(A)\right)$$

$$\leq \frac{W_{n+1}}{N_n}P_n(A^c)\delta_{X_{n+1}}(A) - \left(\frac{W_{n+1}}{N_n} - \frac{W_{n+1}^2}{N_n^2}\right)P_n(A)\delta_{X_{n+1}}(A^c),$$

where we have used that $x - x^2 \leq \frac{x}{1+x} \leq x$ for $x \geq 0$; therefore,

$$\mathbb{E}[P_{n+1}(A) - P_n(A)|\mathcal{F}_n] \leq \beta^2 P_n(A)P_n(A^c)N_n^{-2}. \tag{5.2}$$

and, by Lemma 5.1,

$$\sum_{n=0}^{\infty} \mathbb{E}[P_{n+1}(A) - P_n(A)|\mathcal{F}_n] \leq \sum_{n=0}^{\infty} \frac{\beta^2}{N_n^2} < \infty \qquad \text{a.s.}[\mathbb{P}], \tag{5.3}$$

$$\sum_{n=0}^{\infty} \mathbb{E}[(P_{n+1}(A) - P_n(A))^2|\mathcal{F}_n] \leq \sum_{n=0}^{\infty} \mathbb{E}\left[\frac{W_{n+1}^2}{N_n^2}\Big|\mathcal{F}_n\right] \leq \sum_{n=0}^{\infty} \frac{\beta^2}{N_n^2} < \infty \quad \text{a.s.}[\mathbb{P}]. \tag{5.4}$$

Given (5.3) and (5.4), there exists by [30, Lemma 3.2] a random variable $\tilde{p}_A$ such that

$$P_n(A) \xrightarrow{a.s.} \tilde{p}_A.$$

Define $\tilde{Q}(A) = \tilde{p}_A$, $A \in \mathcal{X}$. Then $\tilde{Q}(\mathbb{X}) = 1$, $\tilde{Q}(\emptyset) = 0$ and $\tilde{Q}(A) \geq 0$, $A \in \mathcal{X}$, a.s.[$\mathbb{P}$]. Moreover, $\omega \mapsto \tilde{Q}(A)(\omega)$ is $\mathcal{H}$-measurable for every $A \in \mathcal{X}$, and $\tilde{Q}(A_1 \cup A_2) = \tilde{Q}(A_1) + \tilde{Q}(A_2)$ a.s.[$\mathbb{P}$] for disjoint $A_1, A_2 \in \mathcal{X}$. In addition, the map $A \mapsto \mathbb{E}[\tilde{Q}(A)]$ is a probability measure on $\mathbb{X}$. Indeed, let $\{A_m\}_{m \geq 1} \subseteq \mathcal{X}$ be such that $A_k \cap A_l = \emptyset$, $k \neq l$. Put $A = \bigcup_{m=1}^{\infty} A_m$. It follows from (5.2) that

$$\mathbb{E}[P_n(A_m)] = \sum_{i=1}^{n} \mathbb{E}[P_i(A_m) - P_{i-1}(A_m)] + \nu(A_m) \leq \beta^2 \sum_{n=0}^{\infty} \mathbb{E}[P_n(A_m)N_n^{-2}] + \nu(A_m) \equiv M(m).$$

By dominated convergence theorem w.r.t. the counting measure, $\lim_{n \to \infty} \sum_{m=1}^{\infty} \mathbb{E}[P_n(A_m)] = \sum_{m=1}^{\infty} \lim_{n \to \infty} \mathbb{E}[P_n(A_m)]$, whenever $\sum_{m=1}^{\infty} M(m) < \infty$. The latter follows from Lemma 5.1 as

$$\sum_{m=1}^{\infty} M(m) = \beta^2 \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} \mathbb{E}[P_n(A_m)N_n^{-2}] + \nu(A) \leq \beta^2 \sum_{n=0}^{\infty} \mathbb{E}[N_n^{-2}] + \nu(A) < \infty.$$

Therefore,

$$\mathbb{E}[\tilde{Q}(A)] = \lim_{n \to \infty} \sum_{m=1}^{\infty} \mathbb{E}[P_n(A_m)] = \sum_{m=1}^{\infty} \lim_{n \to \infty} \mathbb{E}[P_n(A_m)] = \sum_{m=1}^{\infty} \mathbb{E}[\tilde{Q}(A_m)]. \tag{5.5}$$

By [19, Theorem 3.1], there exists a random probability measure $\tilde{P}$ on $\mathbb{X}$ such that $\tilde{Q}(A) = \tilde{P}(A)$ a.s.[$\mathbb{P}$], for every $A \in \mathcal{X}$; thus,

$$P_n(A) \xrightarrow{a.s.} \tilde{P}(A).$$

Using standard arguments, $\mathbb{E}[f(X_{n+1})|\mathcal{F}_n] \xrightarrow{a.s.} \int_{\mathbb{X}} f(x)\tilde{P}(dx)$ for every bounded measurable function f, which implies that $P_n \xrightarrow{w} \tilde{P}$ a.s.[$\mathbb{P}$] since $\mathbb{X}$ is separable.

Define $S_n = \{X_1, \ldots, X_n\}$, $n \geq 1$. Then $S_n \uparrow \{X_1, X_2, \ldots\} \equiv S$ and, by Portmanteau theorem,

$$\tilde{P}(S) = \lim_{n \to \infty} \tilde{P}(S_n) \geq \lim_{n \to \infty} \limsup_{k \to \infty} P_k(S_n) \geq \lim_{n \to \infty} \limsup_{k \to \infty} \left(1 - \frac{\theta}{N_k}\right) = 1 \qquad \text{a.s.}[\mathbb{P}],$$

that is, $\tilde{P}$ is $\mathbb{P}$-a.s. discrete with random support S. On the other hand, proceeding as in (5.3) and (5.4), we can show that $P_n(\{X_m\}) \xrightarrow{a.s.} \tilde{p}_m$ for some random variable $\tilde{p}_m$, $m \geq 1$. By Portmanteau theorem,

$$\tilde{P}(\{X_m\}) \geq \limsup_{n \to \infty} P_n(\{X_m\}) = \tilde{p}_m \qquad \text{a.s.}[\mathbb{P}], \text{ for } m \geq 1.$$

But $1 = \tilde{P}(S) = \sum_{m=1}^{\infty} \tilde{P}(\{X_m\}) \geq \sum_{m=1}^{\infty} \tilde{p}_m$ a.s.$[\mathbb{P}]$ and, using the same arguments as in (5.5),

$$\mathbb{E}\Big[\sum_{m=1}^{\infty} \tilde{p}_m\Big] = \sum_{m=1}^{\infty} \lim_{n \to \infty} \mathbb{E}[P_n(\{X_m\})] = \lim_{n \to \infty} \mathbb{E}[P_n(S)] = 1 \qquad \text{a.s.}[\mathbb{P}],$$

so $\sum_{m=1}^{\infty} \tilde{p}_m = 1$ a.s.$[\mathbb{P}]$, and thus $\tilde{P}(\{X_m\}) = \tilde{p}_m$ a.s.$[\mathbb{P}]$, $m \geq 1$. As a result, we have

$$\mathbb{P}\big(P_n(\{x\}) \to \tilde{P}(\{x\}) \text{ for all } x \in S\big) = \mathbb{P}\big(P_n(\{X_m\}) \to \tilde{P}(\{X_m\}) \text{ for all } m \geq 1\big) = 1,$$

and, by Scheffe's lemma,

$$\|P_n - \tilde{P}\| \leq \sup_{x \in S}\big|P_n(\{x\}) - \tilde{P}(\{x\})\big| + P_n(S^c) \leq \sum_{x \in S}\big|P_n(\{x\}) - \tilde{P}(\{x\})\big| + \frac{\theta}{N_n}\nu(S^c) \xrightarrow{a.s.} 0.$$

Finally, [8, Lemma 2] implies that $\hat{P}_n(\{X_m\}) \xrightarrow{a.s.} \tilde{P}(\{X_m\})$, $m \geq 1$, and so $\|\hat{P}_n - \tilde{P}\| \xrightarrow{a.s.} 0$. $\qquad \square$

The proof of Proposition 3.2 makes use of the following three preliminary results.

Lemma 5.2. *Under the conditions in Proposition 3.2, as $n \to \infty$, $N_n(t,1] \xrightarrow{a.s.} \infty$ for every $t \in (0,1)$.*

Proof. Fix $t \in (0,1)$. Denote $A = (t,1]$. It follows from (2.2) that $P_n(A) \geq \theta\nu(A)/(\theta + n\beta) > 0$, so $\sum_{n=1}^{\infty} P_n(A) = \infty$, and thus $\sum_{n=1}^{\infty} \delta_{X_n}(A) = \infty$ a.s.$[\mathbb{P}]$ by [15, Theorem 1]. Since $N_n(A) = \sum_{i=1}^{n}(N_i(A) - N_{i-1}(A)) + \theta\nu(A)$, then, by [13, Theorem 1], $N_n(A) \xrightarrow{a.s.} \infty$ if we can show that $\sum_{n=1}^{\infty} \mathbb{E}[N_n(A) - N_{n-1}(A)|\mathcal{F}_{n-1} \vee \sigma(X_n)] = \infty$ a.s.$[\mathbb{P}]$. But w is strictly increasing, so

$$\sum_{i=1}^{n} \mathbb{E}[N_i(A) - N_{i-1}(A)|\mathcal{F}_{i-1} \vee \sigma(X_i)] = \sum_{i=1}^{n} w(X_i)\delta_{X_i}(A) \geq w(t)\sum_{i=1}^{n} \delta_{X_i}(A) \xrightarrow{a.s.} \infty.$$

$\qquad \square$

Lemma 5.3. *Under the conditions in Proposition 3.2, for every $t \in (0,1)$, there exist $s \in (t,1)$ and $\lambda_t \in (0,1)$ such that, for all $\lambda \in (\lambda_t, 1]$, the ratio $N_n[0,t]/N_n(s,1]^\lambda$ converges $\mathbb{P}$-a.s. to a finite limit.*

Proof. Fix $t \in (0,1)$. Put $\lambda_t = (w(1) + w(t))/2w(1)$. Take $\lambda \in (\lambda_t, 1]$ and $s \in (t,1)$ such that $w(s) > w(t)/\lambda_t$; such an s exists as $w([0,1])$ is connected and $w(t) < w(t)/\lambda_t < w(1)$. Then, by (5.1),

$$\mathbb{E}\Big[\frac{N_{n+1}[0,t]}{N_{n+1}(s,1]^\lambda} - \frac{N_n[0,t]}{N_n(s,1]^\lambda}\Big|\mathcal{F}_n\Big] = \frac{N_n[0,t]}{N_n(s,1]^\lambda}\mathbb{E}\Big[\frac{N_{n+1}[0,t]}{N_n[0,t]}\Big(1 - \frac{W_{n+1}\mathbb{1}_{\{X_{n+1}>s\}}}{N_{n+1}(s,1]}\Big)^\lambda - 1\Big|\mathcal{F}_n\Big]$$

$$\leq \frac{N_n[0,t]}{N_n(s,1]^\lambda N_n}\Big\{w(t) - \lambda w(s)\frac{N_n(s,1]}{N_n(s,1] + \beta}\Big\},$$

where we have used that $(1+x)^\lambda \le 1 + \lambda x$ for $-1 \le x \le 0$. Let us define, for $n, m \ge 0$,

$$R_n = w(t) - \lambda w(s)\frac{N_n(s,1]}{N_n(s,1] + \beta} \qquad \text{and} \qquad T_m = \inf\{n \ge m : R_n \ge 0\}.$$

Each T_m is a $\mathcal{F}$-stopping time and $(N_{n \wedge T_m}[0,t]/N_{n \wedge T_m}(s,1]^\lambda)_{n \ge m}$ a non-negative supermatingale, so it converges to a random limit $L_m < \infty$, say, on $\Omega_m \in \mathcal{H} : \mathbb{P}(\Omega_m) = 1$. Put $\Omega^* = \{\limsup_n R_n < 0\} \cap \bigcup_m \Omega_m$. By [Lemma 5.2](#) and as $w(t) < \lambda w(s)$, we have $\limsup_n R_n < 0$ a.s.$[\mathbb{P}]$, and thus $\mathbb{P}(\Omega^*) = 1$. Take $\omega \in \Omega^*$. There exists $m_0 \in \mathbb{N}$ such that $R_n(\omega) < 0$, $n \ge m_0$. Then $T_{m_0}(\omega) = \infty$ and $N_n[0,t](\omega)/N_n(s,1](\omega)^\lambda \to L_{m_0}(\omega)$. Therefore, $\lim_{n \to \infty} N_n[0,t]/N_n(s,1]^\lambda < \infty$ a.s.$[\mathbb{P}]$. $\square$

Lemma 5.4. *Under the conditions in [Proposition 3.2](#), if $\nu(t,1) > 0$ for every $t \in (0,1)$, then* $N_n(\{1\}^c)/M_n(\{1\}^c) \xrightarrow{a.s.} w(1)$.

Proof. Fix $t \in (0,1)$. It follows from $\nu(t,1) > 0$ and a slightly modified version of [Lemma 5.3](#) that $N_n[0,t)/N_n(\{1\}^c) \xrightarrow{a.s.} 0$. Then

$$\liminf_{n \to \infty} \frac{1}{P_n(\{1\}^c)}\mathbb{E}[W_{n+1}\delta_{X_{n+1}}(\{1\}^c)|\mathcal{F}_n] \ge \liminf_{n \to \infty} w(t)\frac{P_n[t,1)}{P_n(\{1\}^c)} = w(t) \qquad \text{a.s.}[\mathbb{P}].$$

Since $\limsup_{n \to \infty}\mathbb{E}[W_{n+1}\delta_{X_{n+1}}(\{1\}^c)|\mathcal{F}_n]/P_n(\{1\}^c) \le w(1)$, letting $t \uparrow 1$, we get

$$\frac{1}{P_n(\{1\}^c)}\mathbb{E}[W_{n+1}\delta_{X_{n+1}}(\{1\}^c)|\mathcal{F}_n] \xrightarrow{a.s.} w(1). \tag{5.6}$$

Define $T_0 = 0$ and, for every $n \ge 1$,

$$T_n = \inf\{m > T_{n-1} : X_m \ne 1\}.$$

As $P_n(\{1\}^c) > 0$, then $M_n(\{1\}^c) \xrightarrow{a.s.} \infty$ by [15, Theorem 1], and so $\mathbb{P}(X_n \ne 1 \text{ i.o.}) = 1$. Therefore, each T_n is a $\mathcal{F}$-stopping time such that $T_n < \infty$ a.s.$[\mathbb{P}]$, and $T_n \xrightarrow{a.s.} \infty$.

We further let, for $n \ge 1$,

$$X_n^* = X_{T_n}, \quad W_n^* = W_{T_n}, \quad P_n^*(\cdot) = \frac{P_{T_{n+1}-1}(\cdot \cap \{1\}^c)}{P_{T_{n+1}-1}(\{1\}^c)}, \quad \mathcal{F}_n^* = \sigma(X_1^*, W_1^*, \ldots, X_n^*, W_n^*),$$

which are well-defined on the probability space $(\Omega^*, \mathcal{H}^*, \mathbb{P}^*) \equiv (\Omega^*, \mathcal{H} \cap \Omega^*, \mathbb{P}(\cdot|\Omega^*))$, where $\Omega^* \equiv \bigcap_{n=0}^\infty \{T_n < \infty\}$ and $\mathbb{P}(\Omega^*) = 1$. Fix $n \in \mathbb{N}$. Take $B \in \mathcal{B}[0,1]$. Then

$$P_n^*(B) = \frac{\theta\nu(B \cap \{1\}^c) + \sum_{i=1}^{T_{n+1}-1} W_i\delta_{X_i}(B \cap \{1\}^c)}{\theta\nu(B \cap \{1\}^c) + \sum_{i=1}^{T_{n+1}-1} W_i\delta_{X_i}(\{1\}^c)} = \frac{\theta\nu(B \cap \{1\}^c) + \sum_{k=1}^{n} W_k^*\delta_{X_k^*}(B)}{\theta\nu(\{1\}^c) + \sum_{k=1}^{n} W_k^*};$$

thus, $P_n^*(B)$ is $\mathcal{F}_n^*$-measurable. Let $A \in \mathcal{F}_n^*$. Since $\mathbb{1}_A = f(X_1^*, W_1^*, \ldots, X_n^*, W_n^*)$ for some measurable function f, we can show that

$$\mathbb{P}^*\big(A \cap \{X_{n+1}^* \in B\}\big) = \mathbb{E}\big[f(X_{T_1}, W_{T_1}, \ldots, X_{T_n}, W_{T_n})\mathbb{1}_{\{X_{T_{n+1}} \in B\} \cap \Omega^*}\big]$$

$$= \mathbb{E}\Big[f(X_{T_1}, W_{T_1}, \ldots, X_{T_n}, W_{T_n})\frac{P_{T_{n+1}-1}(B \cap \{1\}^c)}{P_{T_{n+1}-1}(\{1\}^c)}\mathbb{1}_{\Omega^*}\Big] = \mathbb{E}^*[\mathbb{1}_A \cdot P_n^*(B)].$$

Therefore, P_n^* is a version of the conditional distribution of X_{n+1}^* given $\mathcal{F}_n^*$. Then, by (5.1) and (5.6),

$$\mathbb{P}^*\Big(\lim_{n\to\infty}\mathbb{E}^*[w(X_{n+1}^*)|\mathcal{F}_n^*] = w(1)\Big) = 1.$$

Moreover, $\sum_{n=1}^{\infty}\mathbb{E}^*[(W_n^*)^2]/n^2 < \infty$, so $\mathbb{P}^*(\frac{1}{n}\sum_{k=1}^{n}W_k^* \to w(1)) = 1$ by [8, Lemma 2]. As a result,

$$\frac{N_n(\{1\}^c)}{M_n(\{1\}^c)} = \frac{\theta\nu(\{1\}^c)}{M_n(\{1\}^c)} + \frac{\sum_{k=1}^{M_n(\{1\}^c)-1}W_{T_k}}{M_n(\{1\}^c)} \longrightarrow w(1) \qquad \text{a.s.}[\mathbb{P}].$$

$\square$

Proof of Proposition 3.2.

Part I. $(P_n \xrightarrow{w} \delta_1$ a.s.$[\mathbb{P}]$, $\hat{P}_n \xrightarrow{w} \delta_1$ a.s.$[\mathbb{P}]$, $\frac{N_n}{n} \xrightarrow{a.s.} w(1))$. Let $t \in (0,1)$. Take $s \in (t,1)$ and $\lambda \in (0,1)$ as in the statement of Lemma 5.3. It follows from Lemmas 5.2 and 5.3 that

$$\limsup_{n\to\infty} P_n[0,t] \leq \limsup_{n\to\infty}\frac{N_n[0,t]}{N_n(t,1]} \leq \limsup_{n\to\infty}\frac{1}{N_n(s,1]^{1-\lambda}}\frac{N_n[0,t]}{N_n(s,1]^{\lambda}} = 0 \quad \text{a.s.}[\mathbb{P}];$$

thus, $P_n(t,1] \xrightarrow{a.s.} 1$ for every $t \in (0,1)$, and so $P_n \xrightarrow{w} \delta_1$ a.s.$[\mathbb{P}]$. By [8, Lemma 2], $\hat{P}_n(t,1] \xrightarrow{a.s.} 1$ for every $t \in (0,1)$ and, similarly, $\hat{P}_n \xrightarrow{w} \delta_1$ a.s.$[\mathbb{P}]$.

On the other hand, (5.1) and the a.s. weak convergence of P_n to δ_1 imply $\mathbb{E}[W_{n+1}|\mathcal{F}_n] \xrightarrow{a.s.} w(1)$. As $\sum_{n=1}^{\infty}\mathbb{E}[W_n^2]/n^2 \leq \sum_{n=1}^{\infty}\beta^2/n^2 < \infty$, then, by [8, Lemma 2],

$$\frac{N_n}{n} \xrightarrow{a.s.} w(1). \tag{5.7}$$

Part II. $(P_n(\{1\}^c) \xrightarrow{a.s.} \eta)$. Suppose $0 < \nu(\{1\}) < 1$; else, if $\nu(\{1\}) = 0$ (or $\nu(\{1\}) = 1$), then $P_n(\{1\}^c) = 1$ $(P_n(\{1\}^c) = 0)$, and so $P_n(\{1\}^c) \xrightarrow{a.s.} \eta$ with $\eta = 1$ $(\eta = 0)$.

Case 1: $\nu(t,1) = 0$ *for some* $t \in (0,1)$. It follows that $P_n(\{1\}^c) = P_n[0,t]$, so Part I implies $\limsup_{n\to\infty} P_n[0,t] \leq \delta_1([0,t]) = 0$ a.s.$[\mathbb{P}]$, and thus $P_n(\{1\}^c) \xrightarrow{a.s.} 0$.

Case 2: $\nu(t,1) > 0$ *for each* $t \in (0,1)$. By Lemma 5.4, $N_n(\{1\}^c)/M_n(\{1\}^c) \xrightarrow{a.s.} w(1)$ and, similarly, $N_n(\{1\})/M_n(\{1\}) \xrightarrow{a.s.} w(1)$. Let $\psi \in (0,1)$. Then

$$\mathbb{E}\Big[\frac{M_{n+1}(\{1\})^{\psi}}{M_{n+1}(\{1\}^c)} - \frac{M_n(\{1\})^{\psi}}{M_n(\{1\}^c)}\Big|\mathcal{F}_n\Big] = \frac{M_n(\{1\})^{\psi}}{M_n(\{1\}^c)}\mathbb{E}\Big[\Big(1+\frac{\delta_{X_{n+1}}(\{1\})}{M_n(\{1\})}\Big)^{\psi}\frac{M_n(\{1\}^c)}{M_{n+1}(\{1\}^c)} - 1\Big|\mathcal{F}_n\Big]$$

$$\leq \frac{M_n(\{1\})^{\psi}}{M_n(\{1\}^c)N_n}\Big\{\psi\frac{N_n(\{1\})}{M_n(\{1\})} - \frac{N_n(\{1\}^c)}{M_n(\{1\}^c)+1}\Big\},$$

using that $(1+x)^{\psi} \leq 1 + \psi x$ for $0 \leq x \leq 1$. Since

$$\limsup_{n\to\infty}\Big(\psi\frac{N_n(\{1\})}{M_n(\{1\})} - \frac{N_n(\{1\}^c)}{M_n(\{1\}^c)+1}\Big) < 0 \qquad \text{a.s.}[\mathbb{P}],$$

we can show, arguing as in the proof of Lemma 5.3, that $(M_n(\{1\})^\psi/M_n(\{1\}^c))_{n\geq 1}$ converges a.s.$[\mathbb{P}]$. But ψ is arbitrary and $0 < (1+\psi)/2 < 1$, so it holds that

$$\frac{M_n(\{1\})^\psi}{M_n(\{1\}^c)} = M_n(\{1\})^{\frac{\psi-1}{2}}\frac{M_n(\{1\})^{\frac{1+\psi}{2}}}{M_n(\{1\}^c)} \xrightarrow{a.s.} 0.$$

Analogously, $M_n(\{1\}^c)^\psi/M_n(\{1\}) \xrightarrow{a.s.} 0$. Therefore, for every $\psi < 1$,

$$\frac{n^\psi}{N_n(\{1\})} = \Big(\frac{M_n(\{1\}^c) + M_n(\{1\}) - 2}{M_n(\{1\})^{1/\psi}}\Big)^\psi \frac{M_n(\{1\})}{N_n(\{1\})} \xrightarrow{a.s.} 0 \quad \text{and} \quad \frac{n^\psi}{N_n(\{1\}^c)} \xrightarrow{a.s.} 0. \qquad (5.8)$$

Define $Z_n = \log\frac{N_n(\{1\}^c)}{N_n(\{1\})}$, $n \geq 0$. By [30, Lemma 3.2], $(Z_n)_{n\geq 0}$ converges $\mathbb{P}$-a.s. in $[-\infty, \infty)$ if

$$\sum_{n=0}^\infty \mathbb{E}[Z_{n+1} - Z_n|\mathcal{F}_n] < \infty \text{ a.s.}[\mathbb{P}] \quad \text{and} \quad \sum_{n=0}^\infty \mathbb{E}[(Z_{n+1} - Z_n)^2|\mathcal{F}_n] < \infty \text{ a.s.}[\mathbb{P}]. \qquad (5.9)$$

In that case, $(N_n(\{1\}^c)/N_n(\{1\}))_{n\geq 0}$ converges $\mathbb{P}$-a.s. in $[0, \infty)$ and there exists a random variable $\eta \in [0, 1)$ such that

$$P_n(\{1\}^c) = \frac{N_n(\{1\}^c)}{N_n(\{1\}^c) + N_n(\{1\})} \xrightarrow{a.s.} \eta.$$

To prove (5.9), observe that

$$\mathbb{E}[Z_{n+1} - Z_n|\mathcal{F}_n] = \mathbb{E}\Big[\log\frac{N_{n+1}(\{1\}^c)}{N_n(\{1\}^c)}\delta_{X_{n+1}}(\{1\}^c) - \log\frac{N_{n+1}(\{1\})}{N_n(\{1\})}\delta_{X_{n+1}}(\{1\})\Big|\mathcal{F}_n\Big]$$

$$= \mathbb{E}\Big[\int_0^{W_{n+1}}\frac{\delta_{X_{n+1}}(\{1\}^c)}{N_n(\{1\}^c) + t}dt - \int_0^{W_{n+1}}\frac{\delta_{X_{n+1}}(\{1\})}{N_n(\{1\}) + t}dt\Big|\mathcal{F}_n\Big]$$

$$\leq \mathbb{E}\Big[\delta_{X_{n+1}}(\{1\}^c)\Big(\frac{W_{n+1}}{N_n(\{1\}^c)} - \frac{W_{n+1}^2}{2N_n(\{1\}^c)^2} + k_1\frac{W_{n+1}^3}{3N_n(\{1\}^c)^3}\Big)\Big|\mathcal{F}_n\Big] -$$

$$- \mathbb{E}\Big[\delta_{X_{n+1}}(\{1\})\Big(\frac{W_{n+1}}{N_n(\{1\})} - k_2\frac{W_{n+1}^2}{2N_n(\{1\})^2}\Big)\Big|\mathcal{F}_n\Big]$$

$$\leq \frac{1}{N_n}\Big(k_1\frac{\beta^3}{3N_n(\{1\}^c)^2} + k_2\frac{\beta^2}{2N_n(\{1\})}\Big),$$

where we have used a Taylor expansion of the function $f(x) = 1/(x + t)$, $x > 0$ with $t \in \mathbb{R}_+$ fixed, for some constants $k_1, k_2 \geq 0$. Then (5.7)-(5.8) imply $\sum_{n=0}^\infty \mathbb{E}[Z_{n+1} - Z_n|\mathcal{F}_n] < \infty$ a.s.$[\mathbb{P}]$. On the other hand,

$$\mathbb{E}[(Z_{n+1} - Z_n)^2|\mathcal{F}_n] \leq \mathbb{E}\Big[\frac{\delta_{X_{n+1}}(\{1\}^c)}{N_n(\{1\}^c)^2}\Big(\int_0^{W_{n+1}} dt\Big)^2 + \frac{\delta_{X_{n+1}}(\{1\})}{N_n(\{1\})^2}\Big(\int_0^{W_{n+1}} dt\Big)^2\Big|\mathcal{F}_n\Big]$$

$$\leq P_n(\{1\}^c)\frac{\beta^2}{N_n(\{1\}^c)^2} + P_n(\{1\})\frac{\beta^2}{N_n(\{1\})^2},$$

and thus $\sum_{n=0}^\infty \mathbb{E}[(Z_{n+1} - Z_n)^2|\mathcal{F}_n] < \infty$ a.s.$[\mathbb{P}]$. $\qquad\qquad\square$

Proof of Theorem 3.3. Let us define $g(x) = w(x)/\bar{w}$, $x \in \mathbb{X}$ and, for $n \geq 1$ and $A \in \mathcal{B}[0,1]$,

$$\tilde{X}_n = g(X_n), \qquad \tilde{W}_n = W_n, \qquad \tilde{P}_n(A) = P_n(g^{-1}(A)), \qquad \tilde{\nu}(A) = \nu(g^{-1}(A)).$$

Then $\mathbb{E}[\tilde{W}_n | \tilde{X}_n] = \bar{w}\tilde{X}_n$ and, for every $t \in (0,1)$, by (2.2),

$$\tilde{\nu}(t,1] = \nu(\{x \in \mathbb{X} : w(x) > t\bar{w}\}) = \nu_w(t\bar{w}, \bar{w}] > 0;$$

thus, $1 \in \text{supp}(\tilde{\nu})$. On the other hand, for every $A \in \mathcal{B}[0,1]$,

$$\tilde{P}_n(A) = \sum_{i=1}^{n} \frac{\tilde{W}_i}{\theta + \sum_{j=1}^{n} \tilde{W}_j} \delta_{\tilde{X}_i}(A) + \frac{\theta}{\theta + \sum_{j=1}^{n} \tilde{W}_j} \tilde{\nu}(A).$$

Therefore, $(\tilde{X}_n)_{n \geq 1}$ is[1] a $[0,1]$-valued DPS with the continuous and strictly increasing expected weight function $\tilde{w}(t) = \bar{w}t$, $t \in [0,1]$. It follows immediately from Proposition 3.2 that

$$\frac{1}{n} \sum_{i=1}^{n} W_i = \frac{1}{n} \sum_{i=1}^{n} \tilde{W}_i \xrightarrow{a.s.} \tilde{w}(1) = \bar{w}.$$

Part I. $(P_n(\mathcal{D}_\delta^c) \xrightarrow{a.s.} 0)$. Fix $\delta > 0$. Let $\epsilon = (\bar{w} - \sup_{x \in \mathcal{D}_\delta^c} w(x))/2$. Then $\epsilon > 0$ by (2.2), so Proposition 3.2 implies

$$P_n(\mathcal{D}_\delta^c) \leq \mathbb{P}\big(w(X_{n+1}) < \bar{w} - \epsilon | \mathcal{F}_n\big) = \tilde{P}_n[0, 1 - \epsilon/\bar{w}] \xrightarrow{a.s.} 0.$$

Part II. $(P_n(\mathcal{D}^c) \xrightarrow{a.s.} \eta)$. By Proposition 3.2, there exists a random variable $\eta \in [0,1]$ such that

$$P_n(\mathcal{D}^c) = \tilde{P}_n(\{1\}^c) \xrightarrow{a.s.} \eta.$$

If it holds $\bar{w} > \bar{w}^c$, then there exists $t \in (0,1)$ such that $\nu_w(t\bar{w}, \bar{w}) = 0$. But $\tilde{\nu}(t,1) = \nu_w(t\bar{w}, \bar{w}) = 0$, so $P_n(\mathcal{D}^c) = \tilde{P}_n[0,t] \xrightarrow{a.s.} 0$.

Part III. $(\|P_n - \tilde{P}\| \xrightarrow{a.s.} 0, \|\hat{P}_n - \tilde{P}\| \xrightarrow{a.s.} 0)$. Suppose $\bar{w} > \bar{w}^c$. It follows from Part II that $P_n(\mathcal{D}) \xrightarrow{a.s.} 1$, so $\hat{P}_n(\mathcal{D}) \xrightarrow{a.s.} 1$ by [8, Lemma 2], and thus $\mathbb{P}(X_n \in \mathcal{D} \text{ i.o.}) = 1$ by [15, Theorem 1].

Let us define $T = 0$ and, for $n \geq 1$,

$$T_n = \inf\{m \in \mathbb{N} : m > T_{n-1}, X_m \in \mathcal{D}\} \qquad \text{and} \qquad P_n(\cdot | \mathcal{D}) = P_n(\cdot \cap \mathcal{D})/P_n(\mathcal{D}).$$

Put $\Omega^* = \bigcap_{n=0}^{\infty} \{T_n < \infty\}$. Each T_n is a $\mathcal{F}$-stopping time such that $T_n < \infty$ a.s.$[\mathbb{P}]$, and $T_n \xrightarrow{a.s.} \infty$; thus, $\mathbb{P}(\Omega^*) = 1$. Then, arguing as in Lemma 5.4, one can show that $P_n^*(\cdot) \equiv P_{T_{n+1}-1}(\cdot | \mathcal{D})$ is a version of the conditional distribution of $X_{T_{n+1}}$ given $\sigma(X_{T_1}, W_{T_1}, \ldots, X_{T_n}, W_{T_n})$ on the probability space $(\Omega^*, \mathcal{H}^*, \mathbb{P}^*) \equiv (\Omega^*, \mathcal{H} \cap \Omega^*, \mathbb{P}(\cdot | \Omega^*))$. Moreover, the sequence $(X_{T_n})_{n \geq 1}$ is a DPS with a constant expected weight function, so there exists by Proposition 3.1 a random probability measure $\tilde{P}^*$ on $\mathcal{D}$ such that

$$\mathbb{P}^*\big(\lim_{n \to \infty} \|P_n^* - \tilde{P}^*\| = 0\big) = 1.$$

---

[1] Strictly speaking, $(\tilde{X}_n)_{n \geq 1}$ differs from the definition of a DPS in that $\tilde{W}_n$ is a function of (X_n, U_n) but not of $(\tilde{X}_n, U_n)$, and $\tilde{P}_n$ is the conditional distribution of $\tilde{X}_{n+1}$ given $\mathcal{F}_n$ instead of $\sigma(\tilde{X}_1, U_1, \ldots, \tilde{X}_n, U_n)$. Nevertheless, the conclusions of Proposition 3.2 continue to hold for the process $(\tilde{X}_n)_{n \geq 1}$ since $\mathbb{E}[\tilde{W}_n | \mathcal{F}_{n-1} \vee \sigma(X_n)] = \tilde{w}(\tilde{X}_n)$.

Define $\tilde{P}(A)(\omega) = \tilde{P}^*(A \cap \mathcal{D})(\omega)$, $\omega \in \Omega^*$ and $\tilde{P}(A)(\omega) = 0$, $\omega \notin \Omega^*$, $A \in \mathcal{X}$. Then $\tilde{P}$ is a random probability measure on $\mathbb{X}$ such that $\tilde{P}(\mathcal{D}) = 1$ a.s.$[\mathbb{P}]$. On the other hand, $P_n(\cdot|\mathcal{D}) \equiv P^*_{m_n}(\cdot)$ on Ω^*, where $m_n = M_n(\mathcal{D}) - 1$, so

$$\mathbb{P}\Big(\lim_{n\to\infty}\|P_n(\cdot|\mathcal{D}) - \tilde{P}(\cdot)\| = 0\Big) = \mathbb{P}^*\Big(\lim_{n\to\infty}\|P^*_{m_n} - \tilde{P}^*\| = 0\Big) = 1.$$

Therefore,

$$\|P_n - \tilde{P}\| \le 2\,P_n(\mathcal{D}^c) + \|P_n(\cdot|\mathcal{D}) - \tilde{P}(\cdot)\| \xrightarrow{a.s.} 0.$$

Finally, denote by $\hat{P}_n(\cdot|\mathcal{D}) = \hat{P}_n(\cdot \cap \mathcal{D})/\hat{P}_n(\mathcal{D})$, $n \ge 0$ the relative frequency of $(X_n)_{n\ge 1}$ restricted to $\mathcal{D}$, which is well-defined for large n. Then, arguing as before, we have

$$\|\hat{P}_n - \tilde{P}\| \le 2\,\hat{P}_n(\mathcal{D}^c) + \|\hat{P}_n(\cdot|\mathcal{D}) - \tilde{P}(\cdot)\| \xrightarrow{a.s.} 0.$$

$\square$

Proof of Proposition 3.4. Define $\theta_n = \theta/N_n$ and $Z_n = L_n - L_{n-1}$, $n \ge 1$, with $L_0 = 0$ and $\theta_0 = 1$. As ν is diffuse and $Z_n \in \{0,1\}$, then $L_n = \sum_{i=1}^n \mathbb{1}_{\{Z_i=1\}}$ and $\theta_n = \mathbb{P}(Z_{n+1} = 1|\mathcal{F}_n)$. Moreover, $\theta_n \ge \theta/(\theta + \beta n)$, so $\sum_{n=1}^\infty \mathbb{P}(Z_n = 1|\mathcal{F}_{n-1}) = \infty$ a.s.$[\mathbb{P}]$ and, by [15, Theorem 1],

$$\frac{L_n}{\sum_{k=1}^n \theta_{k-1}} = \frac{\sum_{i=1}^n \mathbb{1}_{\{Z_i=1\}}}{\sum_{k=1}^n \mathbb{P}(Z_k = 1|\mathcal{F}_{k-1})} \xrightarrow{a.s.} 1.$$

As $N_n/n \xrightarrow{a.s.} \bar{w}$ by Theorem 3.3, we obtain

$$\frac{L_n}{\log n} = \frac{L_n}{\sum_{k=1}^n \theta_{k-1}}\left(\frac{1}{\log n} + \frac{\theta}{\log n}\sum_{k=1}^{n-1}\frac{1}{k}\Big(\frac{k}{N_k}\Big)\right) \xrightarrow{a.s.} \frac{\theta}{\bar{w}},$$

using the fact that, for any $(a_n)_{n\ge 1} \subseteq \mathbb{R}_+ : a_n \to a$, it holds $\frac{1}{\log n}\sum_{k=1}^n k^{-1}a_k \to a$. $\square$

Proof of Proposition 4.1. If $\nu(\mathcal{D}^c) = 0$, the result is immediate with $\xi = 0$. Suppose $\nu(\mathcal{D}^c) > 0$. Then $\bar{w}^c \in \mathrm{supp}((\nu_{|\mathcal{D}^c})_w)$ from the definition of $\bar{w}^c$. Moreover, $\nu(\{x \in \mathcal{D}^c : w(x) > \bar{w}^c\}) = 0$; otherwise, $(\bar{w}^c, \bar{w}) \cap \mathrm{supp}((\nu_{|\mathcal{D}^c})_w) \ne \emptyset$ and there exists $u > \bar{w}^c$ such that $u \in \mathrm{supp}((\nu_{|\mathcal{D}^c})_w)$, absurd. As a result, $\bar{w}^c$ plays the same role for the subsequence of non-dominant observations as $\bar{w}$ does for $(X_n)_{n\ge 1}$. Therefore, by Theorem 3.3, $N_n(\mathcal{D}^c)/M_n(\mathcal{D}^c) \xrightarrow{a.s.} \bar{w}^c$. Proceeding as in Part II of the proof of Proposition 3.2, we can show that, for every $\psi < \bar{w}^c/\bar{w}$,

$$\frac{M_n(\mathcal{D})^\psi}{M_n(\mathcal{D}^c)} \xrightarrow{a.s.} 0, \qquad \text{and then} \qquad \frac{M_n(\mathcal{D}^c)}{n^\psi} \xrightarrow{a.s.} \infty.$$

In addition, $\log \frac{N_n(\mathcal{D}^c)}{N_n(\mathcal{D})^{\bar{w}^c/\bar{w}}}$ converges $\mathbb{P}$-a.s. in $[-\infty, \infty)$, so there exists a random variable $\xi \in [0, \infty)$ such that, letting $\gamma = 1 - \bar{w}^c/\bar{w}$,

$$n^\gamma \cdot \hat{P}_n(\mathcal{D}^c) = \frac{M_n(\mathcal{D}^c) - 1}{N_n(\mathcal{D}^c)}\Big(\frac{N_n}{n}\frac{N_n(\mathcal{D}^c)^{\bar{w}/\bar{w}^c}}{N_n(\mathcal{D}) + N_n(\mathcal{D}^c)}\Big)^{\bar{w}^c/\bar{w}} \xrightarrow{a.s.} \xi,$$

$$n^\gamma \cdot P_n(\mathcal{D}^c) = n^\gamma \cdot \Big(\hat{P}_n(\mathcal{D}^c) - \frac{1}{n}\Big)\frac{n}{N_n}\frac{N_n(\mathcal{D}^c)}{M_n(\mathcal{D}^c)} \xrightarrow{a.s.} \frac{\bar{w}^c}{\bar{w}}\xi.$$

$\square$

Proof of Theorem 4.2. By Theorem 3.3, $P_n(\mathcal{D}) \xrightarrow{a.s.} 1$, so we can proceed as in Theorem 3.3 to derive results for the subsequence of dominant observations. Let us define $T_0 = 0$ and, for $n \geq 1$

$$T_n = \inf\{n > T_{n-1} : X_n \in \mathcal{D}\}, \qquad P_n(\cdot|\mathcal{D}) = \frac{P_n(\cdot \cap \mathcal{D})}{P_n(\mathcal{D})}, \qquad \hat{P}_n(\cdot|\mathcal{D}) = \frac{\hat{P}_n(\cdot \cap \mathcal{D})}{\hat{P}_n(\mathcal{D})},$$

which are well-defined for large n. We further let $m_n = M_n(\mathcal{D}) - 1$ and $\mathcal{G} = (\mathcal{G}_n)_{n \geq 0}$ be the filtration on $\mathcal{H}$, given by $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $\mathcal{G}_n = \sigma(X_{T_1}, U_{T_1}, \ldots, X_{T_{m_n}}, U_{T_{m_n}})$, $n \geq 1$. With a slight abuse of notation, $(X_{T_{m_n}})_{n \geq 1}$ forms a subsequence of dominant observations,

$$P_n(\cdot|\mathcal{D}) = \frac{\theta\nu(\cdot \cap \mathcal{D}) + \sum_{k=1}^{m_n} W_{T_k}\delta_{X_{T_k}}(\cdot)}{\theta\nu(\mathcal{D}) + \sum_{k=1}^{m_n} W_{T_k}},$$

is a version of the conditional distribution of $X_{T_{m_{n+1}}}$ given $\mathcal{G}_n$ (see the proof of Theorem 3.3), and $(\hat{P}_n(\cdot|\mathcal{D}))_{n \geq 1}$ is the relative frequency of $(X_{T_{m_n}})_{n \geq 1}$ restricted to $\mathcal{D}$. Observe that $P_n(\cdot|\mathcal{D})$ and $\hat{P}_n(\cdot|\mathcal{D})$ remain unchanged unless $m_n > m_{n-1}$. It follows from Proposition 3.1 that, for every $A \in \mathcal{X}$,

$$P_n(A|\mathcal{D}) \xrightarrow{a.s.} \tilde{P}(A) \qquad \text{and} \qquad \hat{P}_n(A|\mathcal{D}) \xrightarrow{a.s.} \tilde{P}(A).$$

We can prove the convergence to Gaussian limits of

$$\sqrt{m_n}\big(\hat{P}_n(A|\mathcal{D}) - P_n(A|\mathcal{D})\big) \qquad \text{and} \qquad \sqrt{m_n}\big(P_n(A|\mathcal{D}) - \tilde{P}(A)\big), \tag{5.10}$$

by applying [8, Theorem 1] and [8, Proposition 1], respectively. Indeed, one can easily show that (5.10) satisfy the assumptions of [8, Theorem 1, Proposition 1] using techniques from [7, Theorem 4] and [8, Corollary 3]. As a result,

$$\sqrt{m_n}\big(\hat{P}_n(A|\mathcal{D}) - P_n(A|\mathcal{D})\big) \xrightarrow{stably} \mathcal{N}\big(0, U(A)\big),$$

$$\sqrt{m_n}\big(P_n(A|\mathcal{D}) - \tilde{P}(A)\big) \xrightarrow{a.s.cond.} \mathcal{N}\big(0, V(A)\big) \quad \text{w.r.t. } \mathcal{G}.$$

Since $M_n(\mathcal{D})/n \xrightarrow{a.s.} 1$, a generalized Slutsky's theorem (see also [17, Theorem 6]) implies

$$\sqrt{n}\big(\hat{P}_n(A|\mathcal{D}) - P_n(A|\mathcal{D})\big) \xrightarrow{stably} \mathcal{N}\big(0, U(A)\big),$$

$$\sqrt{n}\big(P_n(A|\mathcal{D}) - \tilde{P}(A)\big) \xrightarrow{a.s.cond.} \mathcal{N}\big(0, V(A)\big) \quad \text{w.r.t. } \mathcal{G}.$$

But $\tilde{P}(A)$ is $\sigma(X_{T_{m_1}}, X_{T_{m_2}}, \ldots)$-measurable from $\hat{P}_n(A|\mathcal{D}) \xrightarrow{a.s.} \tilde{P}(A)$ and, using (2.1), we can show that $(X_{T_{m_{n+1}}}, X_{T_{m_{n+2}}}, \ldots)$ and $\mathcal{F}_n$ are conditionally independent given $\mathcal{G}_n$; therefore,

$$\sqrt{n}\big(P_n(A|\mathcal{D}) - \tilde{P}(A)\big) \xrightarrow{a.s.cond.} \mathcal{N}\big(0, V(A)\big) \quad \text{w.r.t. } \mathcal{F}.$$

On the other hand, Proposition 3.2 implies that

$$\big|\sqrt{n}(P_n(A|\mathcal{D}) - D_n(A)\big| = \sqrt{n}|P_n(A|\mathcal{D}) - P_n(A)| \leq 2\sqrt{n} \cdot P_n(\mathcal{D}^c) \xrightarrow{a.s.} 0.$$

Then, using again the generalized Slutsky's theorem, $D_n(A) \overset{a.s.cond.}{\longrightarrow} \mathcal{N}\big(0, V(A)\big)$ w.r.t. $\mathcal{F}$.

Regarding the second result, we have

$$\big|\sqrt{n}(\hat{P}_n(A|\mathcal{D}) - P_n(A|\mathcal{D})) - C_n(A)\big| \leq 2\sqrt{n} \cdot P_n(\mathcal{D}^c) + 2\sqrt{n} \cdot \hat{P}_n(\mathcal{D}^c) \overset{a.s.}{\longrightarrow} 0;$$

therefore, $C_n(A) \overset{stably}{\longrightarrow} \mathcal{N}\big(0, U(A)\big)$. □

Acknowledgments

Funding

References

[1] ALDOUS, D. J. (1985). Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII 1983* **1117** 1-198.

[2] ALETTI, G., MAY, C. and SECCHI, P. (2009). A central limit theorem, and related results, for a two-color randomly reinforced urn. *Adv. Appl. Probab.* **41** 829-844.

[3] ALEXANDER, J. M., SKYRMS, B. and ZABELL, S. L. (2012). Inventing new signals. *Dyn. Games Appl.* **2** 129-145.

[4] BANDYOPADHYAY, A. and THACKER, D. (2017). Pólya urn schemes with infinitely many colors. *Bernoulli* **23** 3243-3267.

[5] BASSETTI, F., CRIMALDI, I. and LEISEN, F. (2010). Conditionally identically distributed species sampling sequences. *Adv. Appl. Probab.* **42** 433-459.

[6] BEGGS, A. W. (2005). On the convergence of reinforcement learning. *J. Econ. Theory* **122** 1-36.

[7] BERTI, P., CRIMALDI, I., PRATELLI, L. and RIGO, P. (2010). Central limit theorems for multi-color urns with dominated colors. *Stoch. Process. Appl.* **120** 1473-1491.

[8] BERTI, P., CRIMALDI, I., PRATELLI, L. and RIGO, P. (2011). A central limit theorem and its applications to multicolor randomly reinforced urns. *J. Appl. Probab.* **48** 527-546.

[9] BERTI, P., PRATELLI, L. and RIGO, P. (2004). Limit theorems for a class of identically distributed random variables. *Ann. Probab.* **32** 2029-2052.

[10] BLACKWELL, D. and DUBINS, L. (1962). Merging of opinions with increasing information. *Ann. Math. Stat.* **33** 882-886.

[11] BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Stat.* **1** 353-355.

[12] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. Royal Stat. Soc., Series B* **79** 1295-1366.

[13] CHEN, L. H. Y. (1978). A short note on the conditional Borel-Cantelli lemma. *Ann. Probab.* **6** 699-700.

[14] CRIMALDI, I. (2009). An almost sure conditional convergence result and an application to a generalized Pólya urn. *Int. Math. Forum* **4** 1139-1156.

[15] DUBINS, L. and FREEDMAN, D. (1965). A sharper form of the Borel-Cantelli lemma and the strong law. *Ann. Math. Stat.* **36** 800-807.

[16] FORTINI, S. and PETRONE, S. (2012). Predictive construction of priors in Bayesian nonparametrics. *Braz. J. Probab. Stat.* **26** 423-449.

[17] FORTINI, S. and PETRONE, S. (2020). Quasi-Bayes properties of a procedure for sequential learning in mixture models. *J. Royal Stat. Soc., Series B* **82** 1087-1114.

[18] FORTINI, S., PETRONE, S. and SARIEV, H. (2021). Predictive constructions based on measure-valued Pólya urn processes. *Math.* **9** 2845.

[19] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge, UK.

[20] HANSEN, B. and PITMAN, J. (2000). Prediction rules for exchangeable sequences related to species sampling. *Stat. Probab. Letters* **46** 251-256.

[21] HÄUSLER, E. and LUSCHGY, H. (2015). *Stable Convergence and Stable Limit Theorems*. Springer, Cham, Switzerland.

[22] HOPPE, F. M. (1984). Pólya-like urns and the Ewens' sampling formula. *J. Math. Biol.* **20** 91-94.

[23] JANSON, S. (2019). Random replacements in Pólya urns with infinitely many colours. *Electron. Commun. Probab.* **24**. paper no. 23, 11 pp.

[24] KALLENBERG, O. (1988). Spreading and predictable sampling in exchangeable sequences and processes. *Ann. Probab.* **16** 508-534.

[25] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*. Springer, Berlin.

[26] MAILLER, C. and MARCKERT, J. F. (2017). Measure-valued Pólya urn processes. *Electron. Commun. Probab.* **22**. paper no. 26, 33 pp.

[27] MAY, C. and FLOURNOY, N. (2009). Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *Ann. Stat.* **37** 1058-1078.

[28] MULIERE, P., PAGANONI, A. M. and SECCHI, P. (2006). A randomly reinforced urn. *J. Stat. Plann. Inference* **136** 1853-1874.

[29] PEMANTLE, R. (2007). A survey on random processes with reinforcement. *Probab. Surv.* **4** 1-79.

[30] PEMANTLE, R. and VOLKOV, S. (1999). Vertex-reinforced random walk on $\mathbb{Z}$ has finite range. *Ann. Probab.* **27** 1368-1388.

[31] PITMAN, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *LNM*, **30** 245-267. IMS, Hayward, CA.

[32] ROSENBERGER, W. F. and LACHIN, J. (2016). *Randomization in Clinical Trials*, 2 ed. Wiley, Hoboken, NJ.

[33] ZHANG, L. X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2014). Asymptotic properties of multicolor randomly reinforced Pólya urns. *Adv. Appl. Probab.* **46** 585-602.