RANDOM REPLACEMENTS IN PÓLYA URNS WITH INFINITELY MANY COLOURS

SVANTE JANSON

ABSTRACT. We consider the general version of Pólya urns recently studied by Bandyopadhyay and Thacker (2016+) and Mailler and Marckert (2017), with the space of colours being any Borel space S and the state of the urn being a finite measure on S. We consider urns with random replacements, and show that these can be regarded as urns with deterministic replacements using the colour space $S \times [0, 1]$.

1. INTRODUCTION

The original Pólya urn, studied already in 1917 by Markov [14] but later named after Pólya who studied it in Eggenberger and Pólya [6] (1923) and Pólya [17] (1930), contains balls of two colours. At discrete time steps, a ball is drawn at random from the urn (uniformly), and it is replaced together with a balls of the same colour, where $a \geqslant 1$ is some given constant. The (contents of the) urn is thus a Markov process $(X_n)_0^\infty$, with state space $\mathbb{Z}_{\geqslant 0}^2$. (The initial state X_0 is some arbitrary given non-zero state.)

This urn model has been generalized by various authors in a number of ways, all keeping the basic idea of a Markov process of sets of balls of different colours (types), where balls are drawn at random and the drawn balls determine the next step in the process. (The extensions are all usually called Pólya urns, or perhaps generalized Pólya urns.) These generalizations have been studied by a large number of authors, and have found a large number of applications, see for example [12], [8], [3], [13] and the references given there. The extensions include (but are not limited to) the following, in arbitrary combinations.

(i) The number of different colours can be any finite integer $d \geqslant 2$. The state space is thus $\mathbb{Z}_{\geqslant 0}^d$.

(ii) The new balls added to the urn can be of any colours. We have a (fixed) replacement matrix $(R_{i,j})_{i,j=1}^d$ of non-negative integers; when a ball of colour i is drawn, it is replaced together with $R_{i,j}$ new balls of colour j, for every $j = 1, \ldots, d$.

(iii) The replacements can be random. Instead of a fixed replacement matrix as in (ii), we have for each colour i a random vector $(R_{i,j})_{j=1}^d$. Each time a ball of colour i is drawn, replacements are made according to a new copy of this vector, independent of everything that has happened so far.

(iv) The "numbers of balls" of different colours can be arbitrary non-negative real numbers (which can be interpreted as the amount or mass of each colour). The state space is thus $\mathbb{R}_{\geqslant 0}^d$, and the replacement matrix $(R_{i,j})$ in (ii), or its random version in (iii), has arbitrary entries in $\mathbb{R}_{\geqslant 0}$.

(v) Balls may also be removed from the urn. This means that $R_{i,j}$ in (ii) or (iii) may be negative. (Some conditions are required in order to guarantee that we never remove balls that do not exist; the state space is still $\mathbb{Z}_{\geqslant 0}^d$ or $\mathbb{R}_{\geqslant 0}^d$.) The simplest case, which frequently appears in applications, is drawing without replacement; then $R_{i,i} = -1$ is allowed but $R_{i,j} \geqslant 0$ when $i \neq j$, this means that the drawn ball is not replaced (but balls of other colours are added).

In contrast to the many papers on Pólya urns with a finite number of colours, there has so far been very few studies of extensions to infinitely many colours. One example is Bandyopadhyay and Thacker [2, 1] who studied the case when the space of colours is $\mathbb{Z}^d$, and the replacements are translation invariant. A very general version of Pólya urns was introduced by Blackwell and MacQueen [4] in a special case (with every replacement having the colour of the drawn ball, as in the original Pólya urn, see Example 3.2), and much more generally (with rather arbitrary deterministic replacements) by Bandyopadhyay and Thacker [3] and Mailler and Marckert [13]; this version can be described by:

(vi) The space S of colours is a measurable space. The state space is now the space $\mathcal{M}(S)$ of finite measures on S; if the current state is μ, then the next ball is drawn with the distribution $\mu/\mu(S)$.

This version seems very powerful, and can be expected to find many applications in the future.

Remark 1.1. Note that the case when S is finite in (vi) is equivalent to the version (iv). Also with an infinite S in (vi), a state $\mu \in \mathcal{M}(S)$ of the process can be interpreted as the amount of different colours in the urn. (The amount is thus now described by a measure; note that the measure may be diffuse, meaning that each single colour has mass 0).

Remark 1.2. The colour space S is assumed to be a Polish topological space in [4] and [13], and for the convergence results in [3], while the representation results in [3] are stated for a general S. We too make our definitions for an arbitrary measurable space S, but we restrict to Borel spaces in our main result. (This includes the case of a Polish space, see Lemma 2.1 below. Our results do not use any topology on S.)

The purpose of the present note is to show that this model with a measure-valued Pólya urn and the results for it by [3] and [13] extend almost automatically to the case of random replacements, at least in the case with no removals. In fact, we show that the model is so flexible that a random replacement can be seen as a deterministic replacement using the larger colour space $S \times [0, 1]$, where the extra coordinate is used to simulate the randomization. Random replacement in this general setting was raised as an open problem in [13], and our results together with the results of [13] thus answer this question.

We give a precise definition of the measure-valued version of Pólya urns with random replacement in Section 3. We include there a detailed treatment of measurability questions, showing that there are no such problems. (This was omitted in [3] and [13], where the situation is simpler and straightforward. In our, technically more complex, situation, there is a need to verify measurability explicitly.)

The main theorem is the following. The proof is given in Section 4.

Theorem 1.3. *Consider a measure-valued Pólya urn process $(X_n)_0^\infty$ in a Borel space S, with random replacements. Then there exists a Pólya urn process $(\widetilde{X}_n)_0^\infty$ in $S \times [0,1]$ with deterministic replacements such that $\widetilde{X}_n = X_n \times \lambda$ and thus $X_n = \pi^\sharp(\widetilde{X}_n)$ for every $n \geqslant 0$, where λ is the Lebesgue measure, $\pi : X \times [0,1] \to X$ is the projection, and $\pi^\sharp$ the corresponding mapping of measures.*

Note that X_n is a measure on S; similarly $\widetilde{X}_n$ is a measure on $S \times [0,1]$, and the theorem says that $\widetilde{X}_n$ is a product measure $X_n \times \lambda$. (We find it convenient to use both the notation $\widetilde{X}_n$ and the explicit $X_n \times \lambda$ for this urn process in $S \times [0,1]$, depending on the context.)

Urns without replacement or with other removals, see (v), are treated in Section 5. We show that Theorem 1.3 holds in this case too, but the result in this case is less satisfactory than in the case without removals, and it cannot be directly applied to extend the results for this case in [13], see Section 5.

In Section 6 we, as an example, apply Theorem 1.3 to show that one of the main theorems in [13] extends from deterministic to random replacements.

Remark 1.4. Many papers, including [3] and [13], consider only *balanced* Pólya urns, i.e., urns where the total number of balls added to the urn each time is deterministic, and thus the total number of balls in the urn after n steps is a deterministic linear function of n; in the measure-valued context, this means that the total mass $X_n(S) = an + b$, where $b = X_0(S)$. (We may without loss of generality assume $a = 1$ by rescaling.) We have no need for this assumption in the present paper.

2. Preliminaries

We state some more or less well-known definitions and facts, adding a few technical details.

2.1. **Measurable spaces.** A measurable space $(S, \mathcal{S})$ is a set S equipped with a σ-field $\mathcal{S}$ of subsets of S. We often abbreviate $(S, \mathcal{S})$ to S when the σ-field is evident. When $S = [0,1]$ or another Polish topological space (i.e., a complete metric space), we tacitly assume $\mathcal{S} = \mathcal{B}(S)$, the Borel σ-field generated by the open subsets.

For a measurable space $S = (S, \mathcal{S})$, let $\mathcal{M}(S)$ be the set of finite measures on S, let $\mathcal{M}_*(S) := \{\mu \in \mathcal{M}(S) : \mu \neq 0\}$ and $\mathcal{P}(S) := \{\mu \in \mathcal{M}(S) : \mu(S) = 1\}$, the set of probability measures on S; furthermore, let $\mathcal{M}_\pm(S)$ be the space of finite signed measures on S. We regard $\mathcal{M}_\pm(S)$, $\mathcal{M}(S)$, $\mathcal{M}_*(S)$ and $\mathcal{P}(S)$ as measurable spaces, equipped with the σ-field generated by the mappings $\mu \mapsto \mu(B)$, $B \in \mathcal{S}$; note that $\mathcal{M}_*(S)$ and $\mathcal{P}(S)$ are measurable

subsets of $\mathcal{M}(S)$. (See e.g. [10, Chapter 1, p. 19], but note that $\mathcal{M}(S)$ there is larger than ours.)

If $f \geqslant 0$ is a measurable function on a measurable space S and μ is a measure on S, let $\mu(f) := \int_S f \, \mathrm{d}\mu \in [0, \infty]$. Note that the mapping $\mu \mapsto \mu(f)$ is measurable $\mathcal{M}(S) \to [0, \infty]$ for every fixed $f \geqslant 0$.

If S and T are measurable spaces, and $\varphi : S \to T$ is a measurable mapping, then, as in Theorem 1.3 above, $\varphi^{\sharp} : \mathcal{M}(S) \to \mathcal{M}(T)$ denotes the induced mapping of measures, defined by $\varphi^{\sharp}(\mu)(B) = \mu(\varphi^{-1}(B))$ for $\mu \in \mathcal{M}(S)$ and $B \in \mathcal{T}$.

If X is a random element of S, its distribution is an element of $\mathcal{P}(S)$, denoted by $\mathcal{L}(X)$.

A signed measure $\mu \in \mathcal{M}_{\pm}(S)$ has a Jordan decomposition $\mu = \mu^+ - \mu^-$ with $\mu^+, \mu^- \in \mathcal{M}(S)$, and the variation of μ is $|\mu| = \mu^+ + \mu^- \in \mathcal{M}(S)$; see [5, Chapter 4].

2.2. **Borel spaces.** A *Borel space* is a measurable space that is isomorphic to a Borel subset of $[0, 1]$. This can be reformulated by the following standard result.

Lemma 2.1. *The following are equivalent for a measurable space $(S, \mathcal{S})$, and thus each property characterizes Borel spaces.*
  (i) *$(S, \mathcal{S})$ is isomorphic to a Borel subset of a Polish space.*
 (ii) *$(S, \mathcal{S})$ is isomorphic to a Borel subset of $[0, 1]$.*
(iii) *$(S, \mathcal{S})$ is isomorphic to a Polish space.*
 (iv) *$(S, \mathcal{S})$ is isomorphic to a compact metric space.*
  (v) *$(S, \mathcal{S})$ is either countable (with all subsets measurable), or isomorphic to $[0, 1]$.*

For a proof, see e.g. [5, Theorem 8.3.6] or [15, Theorem I.2.12]. An essentially equivalent statement is that any two Borel spaces with the same cardinality are isomorphic.

In Theorem 1.3, we consider only Borel spaces; Lemma 2.1 shows that this is no great loss of generality for applications.

Lemma 2.2. *If S is a Borel space, then the mappings $\mu \mapsto |\mu|$, $\mu \mapsto \mu^+$ and $\mu \mapsto \mu^-$ are measurable $\mathcal{M}_{\pm}(S) \to \mathcal{M}(S)$. In particular, $\mathcal{M}(S)$ is a measurable subset of $\mathcal{M}_{\pm}(S)$.*

Proof. By Lemma 2.1, we may assume that S is a Borel subset of $[0, 1]$. Then, for every $B \in \mathcal{S}$ and $\mu \in \mathcal{M}_{\pm}(S)$,

$$|\mu|(B) = \lim_{n \to \infty} \sum_{i=0}^{n} \left| \mu\big(B \cap [\tfrac{i}{n}, \tfrac{i+1}{n})\big) \right|, \qquad (2.1)$$

which shows that $\mu \mapsto |\mu|(B)$ is measurable. Hence $\mu \mapsto |\mu|$ is measurable. Furthermore, $\mu^{\pm} = \frac{1}{2}(|\mu| \pm \mu)$, and $\mathcal{M}(S) = \{\mu \in \mathcal{M}_{\pm}(S) : \mu^- = 0\}$. $\square$

Lemma 2.3. *If S is a Borel space, then $\mathcal{M}_{\pm}(S)$, $\mathcal{M}(S)$, $\mathcal{M}_*(S)$ and $\mathcal{P}(S)$ are Borel spaces.*

Proof. $\mathcal{M}(S)$ is Borel as a special case of [11, Theorem 1.5]. Alternatively, by Lemma 2.1, we may assume that S is a compact metric space with its Borel σ-field. Then, see e.g. [10, Theorem A2.3], $\mathcal{M}(S)$ is a Polish space,

and its Borel σ-field equals the σ-field defined above for $\mathcal{M}(S)$; hence, $\mathcal{M}(S)$ is a Borel space.

Next, $\mathcal{M}_*(S)$ and $\mathcal{P}(S)$ are measurable subsets of $\mathcal{M}(S)$ and thus also Borel spaces.

Finally, the Jordan decomposition $\mu \mapsto (\mu^+, \mu^-)$ gives a bijection

$$\psi : \mathcal{M}_\pm(S) \leftrightarrow \mathcal{M}' := \big\{ (\mu_1, \mu_2) \in \mathcal{M}(S)^2 : |\mu_1 - \mu_2|(S) = |\mu_1|(S) + |\mu_2|(S) \big\}. \tag{2.2}$$

Lemma 2.2 shows that ψ is measurable, and so is trivially its inverse $\psi^{-1} :$ $(\mu_1, \mu_2) \mapsto \mu_1 - \mu_2$. Moreover, it follows from Lemma 2.2 that the set $\mathcal{M}'$ is a measurable subset of $\mathcal{M}(S)^2$; hence $\mathcal{M}'$ is a Borel space, and thus so is $\mathcal{M}_\pm(S)$. $\qquad \square$

Remark 2.4. Lemma 2.2 may fail if S is not a Borel space. For a counterexample, let $S = \{0,1\}^{\mathbb{R}}$, define for $A \subseteq \mathbb{R}$ the σ-field $\mathcal{S}_A$ on S consisting of all sets $\pi^{-1}(B)$ where $\pi : S \to S_A := \{0,1\}^A$ is the projection and B is a Borel set in $\{0,1\}^A$, and let $\mathcal{S} := \bigcup \{ \mathcal{S}_A : A \text{ countable} \}$. (This is the Baire σ-field on S.) Then any measurable function $F(\mu)$ on $\mathcal{M}_\pm(S)$ is a function of $(\mu(B_i))_0^\infty$ for some sequence $B_i \in \mathcal{S}$, which means that $B_i \in \mathcal{S}_{A_i}$ for some countable A_i. Choose some $x \in \mathbb{R} \setminus \bigcup_i A_i$, and define $s_0, s_1 \in S = \{0,1\}^{\mathbb{R}}$ by $s_j(x) = j$ and $s_j(y) = 0$ when $y \neq x$. Let $\nu := \delta_{s_0} - \delta_{s_1}$. Then $\nu(B_i) = 0$ for every i, and thus $F(\nu) = F(0)$. Hence, $F(\mu)$ cannot equal $|\mu|(S)$ for every $\mu \in \mathcal{M}_\pm(S)$. Consequently, $\mu \mapsto |\mu|(S)$ is not measurable on $\mathcal{M}_\pm(S)$. Similarly, $\mathcal{M}(S)$ is not a measurable subset of $\mathcal{M}_\pm(S)$.

2.3. **Kernels.** (See e.g. [10, pp. 20–21, 106–107, 116 and 141–142].) Given two measurable spaces $S = (S, \mathcal{S})$ and $T = (T, \mathcal{T})$, a *kernel* from S to T is a measurable mapping $\mu : S \to \mathcal{M}(T)$. We write the mapping as $s \mapsto \mu_s$; thus a kernel is, equivalently, a family $\{\mu_s\}_{s \in S}$ of finite measures on T such that $s \mapsto \mu_s(B)$ is measurable for every $B \in \mathcal{T}$. It follows that if $\mu : S \to \mathcal{M}(T)$ is a kernel and $f : T \to [0, \infty)$ is measurable, then $s \mapsto \mu_s(f)$ is measurable $S \to [0, \infty]$. A *probability kernel* is a kernel that maps S into $\mathcal{P}(T)$, i.e., a kernel μ such that μ_s is a probability measure for every $s \in S$.

If μ is a probability kernel from S to T and ν is a probability measure on S, then a probability measure $\nu \otimes \mu$ is defined on $S \times T$ by

$$\nu \otimes \mu(B) = \int_S \mathrm{d}\nu(s) \int_T \mathbf{1}_B(s,t) \, \mathrm{d}\mu_s(t), \qquad B \in \mathcal{S} \times \mathcal{T}. \tag{2.3}$$

Note that if the random element $(X, Y) \in S \times T$ has the distribution $\nu \otimes \mu$, then the marginal distribution of X is $\nu \in \mathcal{P}(S)$; we denote the marginal distribution of Y by $\nu \cdot \mu \in \mathcal{P}(T)$.

If X and Y are random elements of S and T, respectively, then a *regular conditional distribution* of Y given X is a probability kernel μ from S to T such that for each $B \in \mathcal{T}$, $\mathbb{P}\big(Y \in B \mid X\big) = \mu_X(B)$ a.s. (I.e., $\mu_X(B)$ is a version of the conditional expectation $\mathbb{P}\big(Y \in B \mid X\big)$.) This is easily seen to be equivalent to: μ is a probability kernel such that (X, Y) has the distribution $\mathcal{L}(X) \otimes \mu$ given by (2.3).

If μ is a probability kernel from a measurable space S to itself, and $\mu_0 \in \mathcal{P}(S)$ is any distribution, we can iterate (2.3) and define, for any $N \geqslant 1$, a probability measure $\mu_0 \otimes \mu \otimes \cdots \otimes \mu$ on S^{N+1} such that if

$(X_0, \ldots, X_N)$ has this distribution, then $X_0, \ldots, X_N$ is a Markov chain with initial distribution $X_0 \sim \mu_0$ and transitions given by the kernel μ, i.e., $\mathbb{P}\big(X_n \in B \mid X_0, \ldots, X_{n-1}\big) = \mathbb{P}\big(X_n \in B \mid X_{n-1}\big) = \mu_{X_{n-1}}(B)$ for any $B \in \mathcal{S}$ and $1 \leqslant n \leqslant N$. Moreover, these finite Markov chains extend to an infinite Markov chain $X_0, X_1, \ldots$ with the transition kernel μ.

Remark 2.5. The existence of an infinite Markov chain follows without any condition on S by a theorem by Ionescu Tulcea [10, Theorem 6.17]. (If S is a Borel space, we may also, as an alternative, use Kolmogorov's theorem [10, Theorem 6.16].)

The construction of an infinite Markov chain extends to any sequence of different measurable spaces $S_0, S_1, \ldots$ and probability kernels μ_i from S_{i-1} to S_i, $i \geqslant 1$, but we need here only the homogeneous case.

2.4. **Two lemmas.**

Lemma 2.6. *Let $S = (S, \mathcal{S})$ and $T = (T, \mathcal{T})$ be measurable spaces. A map $\mu : S \to \mathcal{M}(\mathcal{M}(T))$ is a kernel from S to $\mathcal{M}(T)$ if and only if, for every bounded measurable function $h : T \to [0, \infty)$,*

$$s \mapsto \int_{\mathcal{M}(T)} e^{-\nu(h)} \, \mathrm{d}\mu_s(\nu) \tag{2.4}$$

is measurable on S.

Proof. If $h : T \to [0, \infty)$ is measurable, then $\nu \to e^{-\nu(h)}$ is measurable $\mathcal{M}(T) \to [0, 1]$. Hence, if μ is a kernel, then (2.4) is measurable.

Conversely, let $B_+(T)$ be the set of all bounded measurable $h : T \to [0, \infty)$ and assume that (2.4) is measurable for every $h \in B_+(T)$. Let $\mathcal{A}$ be the set of bounded measurable functions $F : \mathcal{M}(T) \to \mathbb{R}$ such that $s \mapsto \int_{\mathcal{M}(T)} F(\nu) \, \mathrm{d}\mu_s(\nu)$ is measurable $S \to \mathbb{R}$. Furthermore, if $h \in B_+(T)$, let $\Psi_h : \mathcal{M}(T) \to [0, 1]$ be the function $\Psi_h(\nu) = e^{-\nu(h)}$, and let $\mathcal{C} := \big\{\Psi_h : h \in B_+(T)\big\}$. The assumption says that $\mathcal{C} \subset \mathcal{A}$. Furthermore, $\mathcal{C}$ is closed under multiplication, since $\Psi_{h_1} \Psi_{h_2} = \Psi_{h_1+h_2}$. It follows by the monotone class theorem, in e.g. the version given in [7, Theorem A.1], that $\mathcal{A}$ contains every bounded function that is measurable with respect to the σ-field $\mathcal{F}(\mathcal{C})$ generated by $\mathcal{C}$.

Let again $h \in B_+(T)$. Then, for every $\nu \in \mathcal{M}(T)$,

$$n\big(\Psi_0(\nu) - \Psi_{h/n}(\nu)\big) = n\big(1 - e^{-\nu(h)/n}\big) \to \nu(h) \qquad \text{as } n \to \infty. \tag{2.5}$$

Hence the mapping $\nu \mapsto h(\nu)$ is $\mathcal{F}(\mathcal{C})$-measurable. In particular, taking $h = \mathbf{1}_B$, it follows that $\nu \mapsto \nu(B)$ is $\mathcal{F}(\mathcal{C})$-measurable for every $B \in \mathcal{T}$. Since these maps generate the σ-field of $\mathcal{M}(T)$, it follows that if $D \subseteq \mathcal{M}(T)$ is measurable, then $\mathbf{1}_D$ is $\mathcal{F}(\mathcal{C})$-measurable, and thus $\mathbf{1}_D \in \mathcal{A}$. This means that

$$s \mapsto \int_{\mathcal{M}(T)} \mathbf{1}_D(\nu) \, \mathrm{d}\mu_s(\nu) = \mu_s(D) \tag{2.6}$$

is measurable for all such D, which means that $s \mapsto \mu_s$ is measurable. $\qquad \square$

We shall also use the following lemma from [10].

Lemma 2.7 ([10, Lemma 3.22]). *Let $(\mu_s)_{s \in S}$ be a probability kernel from a measurable space S to a Borel space T. Then there exists a measurable function $f : S \times [0,1] \to T$ such that if $U \sim \mathrm{U}(0,1)$, then $f(s,U)$ has the distribution μ_s for every $s \in S$.*

3. Pólya urns

In this section, we give formal definitions of the Pólya urn model with an arbitrary colour space S. The state space of the urn process is $\mathcal{M}(S)$, or more precisely $\mathcal{M}_*(S)$, since the process gets stuck and stops when there is no ball left in the urn.

In this section we consider for simplicity only urns with replacement and no removals, i.e., all replacements are positive. See Section 5 for the more general case.

We treat first the deterministic case defined and studied by [3] and [13]; our model is the same as theirs and we add only some technical details as a preparation for the random replacement case.

3.1. **Deterministic replacements.**
The replacements are described by a *replacement kernel*, which is a kernel $R = (R_s)_{s \in S}$ from S to itself, i.e., a measurable map $S \to \mathcal{M}(S)$; the interpretation is that if we draw a ball of colour s, then it is returned together with an additional measure R_s. More formally, we define, for $\mu \in \mathcal{M}_*(S)$, a function $\phi_\mu : S \to \mathcal{M}_*(S)$ by

$$\phi_\mu(s) := \mu + R_s; \tag{3.1}$$

thus if the composition of the urn is described by the measure μ, and we draw a ball of colour s, then the new composition of the urn is $\phi_\mu(s)$. Moreover, the ball is drawn with distribution $\mu' := \mu/\mu(S)$. Hence, letting $\phi_\mu^\sharp : \mathcal{P}(S) \to \mathcal{P}(\mathcal{M}_*(S))$ denote the mapping of probability measures induced by ϕ_μ, the composition after the draw has the distribution

$$\widehat{R}_\mu := \phi_\mu^\sharp(\mu') = \phi_\mu^\sharp(\mu/\mu(S)) \in \mathcal{P}(\mathcal{M}_*(S)). \tag{3.2}$$

Lemma 3.1. *The mapping $\mu \mapsto \widehat{R}_\mu$ defined by (3.1)–(3.2) is a measurable map $\mathcal{M}_*(S) \to \mathcal{P}(\mathcal{M}_*(S))$, i.e., a probability kernel from $\mathcal{M}_*(S)$ to itself.*

Proof. We use Lemma 2.6. Let $h : S \to [0,\infty)$ be measurable. Then, by (3.2),

$$\int_{\mathcal{M}(S)} e^{-\nu(h)} \, \mathrm{d}\widehat{R}_\mu(\nu) = \int_S e^{-\phi_\mu(s)(h)} \, \mathrm{d}\mu'(s) = \int_S e^{-(\mu(h)+R_s(h))} \, \mathrm{d}\mu'(s)$$

$$= \frac{e^{-\mu(h)}}{\mu(S)} \int_S e^{-R_s(h)} \, \mathrm{d}\mu(s). \tag{3.3}$$

Since R is a kernel, $s \mapsto R_s(h)$ is measurable, and thus $\mu \mapsto \int_S e^{-R_s(h)} \, \mathrm{d}\mu(s)$ is a measurable function on $\mathcal{M}(S)$. Hence, (3.3) shows that the left-hand side $\int_{\mathcal{M}(S)} e^{-\nu(h)} \, \mathrm{d}\widehat{R}_\mu(\nu)$ is a measurable function of $\mu \in \mathcal{M}_*(S)$, and thus Lemma 2.6 shows that $\widehat{R} : \mathcal{M}_*(S) \to \mathcal{M}(\mathcal{M}(S))$ is a kernel from $\mathcal{M}_*(S)$ to $\mathcal{M}(S)$.

Since $\mu \in \mathcal{M}_*(S)$ implies $\phi_\mu(s) \in \mathcal{M}_*(S)$ by (3.1), $\widehat{R}$ is also a kernel from $\mathcal{M}_*(S)$ to itself. Finally, $\widehat{R}$ is a probability kernel, since $\widehat{R}_\mu$ is a probability measure by (3.2). $\hfill\square$

The Pólya urn process $(X_n)_0^\infty$ is the Markov process with values in $\mathcal{M}_*(S)$ defined as in Section 2.3 by the probability kernel $\widehat{R}$ and an arbitrary initial state $X_0 \in \mathcal{M}_*(S)$. (In general X_0 may be random, but we assume for simplicity that X_0 is deterministic; this is also the case in most applications.)

Example 3.2. We illustrate the definition with a classical example.

Let S be any measurable space and let the replacement kernel be $R_s = \delta_s$, i.e., $R_s(B) = \mathbf{1}_B(s)$ for $s \in S$ and $B \in \mathcal{S}$. This means that the drawn ball is returned together with another ball of the same colour. (Note that δ_s is well defined even if $\{s\} \notin \mathcal{S}$.)

With $S = \{0,1\}$ and X_0 an integer-valued measure, this is the urn studied by Markov [14], Eggenberger and Pólya [6] and Pólya [17].

The case when S is an arbitrary Polish space and $X_0 \in \mathcal{M}(S)$ is arbitrary was studied by Blackwell and MacQueen [4]; they showed that $X_n/X_n(S)$ a.s. converges (in total variation) to a random discrete probability measure, with a so called Ferguson distribution. See also Pitman [16, Exercises 2.2.6 and 0.3.2, and Section 3.2] (the case $S = [0,1]$, which is no loss of generality by Lemma 2.1), which imply that the limit can be represented as $\sum_i P_i \delta_{\xi_i}$ with ξ_i i.i.d. with distribution $X_0/X_0(S)$ and $(P_i)_1^\infty$ with the Poisson–Dirichlet distribution $\mathrm{PD}(0, X_0(S))$. By Lemma 2.1, the result of [4] extends to any Borel space S. In fact, the result holds for an arbitrary measurable space S; this can for example be seen by considering the same process on $S \times [0,1]$, starting with $X_0 \times \lambda$, regarding the second coordinate as labels and using the result for $[0,1]$; we omit the details.

3.2. **Random replacement.** For the more general version with random replacement, the replacement measures R_s, $s \in S$ are random. We let $\mathcal{R}_s := \mathcal{L}(R_s) \in \mathcal{P}(\mathcal{M}(S))$ for every $s \in S$; $\mathcal{R}_s$ is thus the distribution of the replacement, and we assume that $s \mapsto \mathcal{R}_s$ is a given probability kernel $S \to \mathcal{P}(\mathcal{M}(S))$. This means that $\phi_\mu(s)$ in (3.1) is a random measure in $\mathcal{M}_*(S)$, with a distribution that we denote by $\Phi_\mu(s) \in \mathcal{P}(\mathcal{M}_*(S))$. Note that for a fixed $\mu \in \mathcal{M}_*(S)$, the map $\psi_\mu : \nu \mapsto \mu + \nu$ is measurable $\mathcal{M}(S) \to \mathcal{M}_*(S)$, and thus induces a measurable map $\psi_\mu^\sharp : \mathcal{P}(\mathcal{M}(S)) \to \mathcal{P}(\mathcal{M}_*(S))$; furthermore,

$$\Phi_\mu(s) = \psi_\mu^\sharp(\mathcal{R}_s). \tag{3.4}$$

Hence, $s \mapsto \Phi_\mu(s)$ is a probability kernel from S to $\mathcal{M}_*(S)$.

If we draw from an urn with composition $\mu \in \mathcal{M}_*(S)$, then the drawn colour s has as above distribution $\mu' := \mu/\mu(S)$, and the resulting urn has thus a distribution $\widehat{R}_\mu$ that is the corresponding mixture of the distributions $\Phi_\mu(s)$, i.e., in the notation of Section 2.3, see (2.3) and the comments after it,

$$\widehat{R}_\mu = \mu' \cdot \Phi_\mu. \tag{3.5}$$

Lemma 3.3. *The mapping $\mu \mapsto \widehat{R}_\mu$ defined by (3.4)–(3.5) is a measurable map $\mathcal{M}_*(S) \to \mathcal{P}(\mathcal{M}_*(S))$, i.e., a probability kernel from $\mathcal{M}_*(S)$ to itself.*

Proof. Let $h : S \to [0, \infty)$ be measurable. Then, extending (3.3) in the deterministic case, by (3.5) and (3.4),

$$\int_{\mathcal{M}(S)} e^{-\nu(h)} \, \mathrm{d}\widehat{R}_\mu(\nu) = \int_S \int_{\mathcal{M}(S)} e^{-\nu(h)} \, \mathrm{d}\Phi_\mu(s)(\nu) \, \mathrm{d}\mu'(s)$$

$$= \int_S \int_{\mathcal{M}(S)} e^{-\psi_\mu(\nu)(h)} \, \mathrm{d}\mathcal{R}_s(\nu) \, \mathrm{d}\mu'(s)$$

$$= \frac{e^{-\mu(h)}}{\mu(S)} \int_S \int_{\mathcal{M}(S)} e^{-\nu(h)} \, \mathrm{d}\mathcal{R}_s(\nu) \, \mathrm{d}\mu(s). \qquad (3.6)$$

Here $\nu \mapsto e^{-\nu(h)}$ is a measurable function $\mathcal{M}(S) \to [0, \infty)$, and thus $s \mapsto \int_{\mathcal{M}(S)} e^{-\nu(h)} \, \mathrm{d}\mathcal{R}_s(\nu)$ is a measurable function $S \to [0, \infty]$. Consequently, the right-hand side of (3.6) is a measurable function of μ, and Lemma 2.6 shows that $\widehat{R}$ is a kernel. The proof is completed as the proof of Lemma 2.7. $\qquad \square$

The Pólya urn process $(X_n)_0^\infty$ is, as in the deterministic case above, the Markov process with values in $\mathcal{M}_*(S)$ defined by the probability kernel $\widehat{R}$.

4. Proof of Theorem 1.3

Let $U \sim \mathrm{U}(0, 1)$. By Lemma 2.7, there exists a measurable function $f : S \times [0, 1] \to \mathcal{M}(S)$ such that $f(s, U) \sim \mathcal{R}_s$ for every s; i.e., $f(s, U) \overset{\mathrm{d}}{=} R_s$. In other words, we can use $f(s, U)$ as the replacement measure R_s for the urn $(X_n)_0^\infty$.

Let $\tilde{S} := S \times [0, 1]$ and define

$$\tilde{R}_{s,u} := f(s, u) \times \lambda \in \mathcal{M}(\tilde{S}). \qquad (4.1)$$

The mapping $\mu \mapsto \mu \times \lambda$ is measurable $\mathcal{M}(S) \to \mathcal{M}(\tilde{S})$; hence $\tilde{R}_{s,u}$ is measurable, and thus a kernel from $\tilde{S}$ to itself.

We now let $(\widetilde{X}_n)_0^\infty$ be the Pólya urn process in $\mathcal{M}(\tilde{S})$ defined by the replacement kernel $\tilde{R}$, with initial value $\widetilde{X}_0 = X_0 \times \lambda$. We claim that we can couple the processes such that $\widetilde{X}_n = X_n \times \lambda$ for every $n \geqslant 0$. We prove this by induction. Given $\widetilde{X}_n = \mu \times \lambda$, we draw a ball (s, u) with the distribution $(\mu \times \lambda)' = \mu' \times \lambda$, which means that s has distribution μ' and u is uniform and independent of s; hence, given s, $\tilde{R}_{s,u} = f(s, u) \times \lambda$ has the same distribution as $f(s, U) \times \lambda \overset{\mathrm{d}}{=} R_s \times \lambda$. We may thus assume (formally by the transfer theorem [10, Theorem 6.10]) that $\tilde{R}_{s,u} = R_s \times \lambda$, and thus $\widetilde{X}_{n+1} = \widetilde{X}_n + \tilde{R}_{s,u} = (X_n + R_s) \times \lambda = X_{n+1} \times \lambda$. $\qquad \square$

5. Urns without replacement or with other subtractions

The models in Section 3 can easily be extended to urns without replacement or with removals (subtractions) of other balls.

5.1. **Deterministic replacements.** In the deterministic case, we let the replacements R_s be given by a measurable map $S \to \mathcal{M}_\pm(S)$. We assume that we are given some measurable subset $\mathcal{M}_0$ of $\mathcal{M}_*(S)$ such that for every $\mu \in \mathcal{M}_0$,

$$R_s + \mu \in \mathcal{M}_0 \quad \text{for } \mu\text{-a.e. } s. \qquad (5.1)$$

I.e., by (3.1), $\phi_\mu(s) \in \mathcal{M}_0$ μ-a.e., which means that $\widehat{R}_\mu$ in (3.2) is a probability measure on $\mathcal{M}_0 \subseteq \mathcal{M}_*(S)$. Lemma 3.1 is modified to say that $\widehat{R}$ is a probability kernel from $\mathcal{M}_0$ to itself; the proof is the same. Then, assuming also $X_0 \in \mathcal{M}_0$, the Pólya urn process is defined by the kernel $\widehat{R}$ as before; we have $X_n \in \mathcal{M}_0$ for every n.

Example 5.1 (Drawing without replacement). Let S be a Borel space and let $\mathcal{M}_0$ be the set $\mathcal{N}_*(S)$ of non-zero finite integer-valued measures on S; these are the measures of the type $\sum_1^m \delta_{s_i}$ for some finite sequence $s_1, \ldots, s_m$ in S. (The set $\mathcal{N}_*(S)$ is a measurable set in $\mathcal{M}(S)$, e.g. as a consequence of [11, Theorem 1.6].) Assume that

$$R_s + \delta_s \in \mathcal{N}_*(S) \quad \text{for every } s \in S. \tag{5.2}$$

The interpretation is that the drawn ball is discarded, and instead we add a set of balls described by the (positive, integer-valued) measure $R_s + \delta_s$; this is thus the classical case of drawing without replacement, see (v) in Section 1. The (5.1) holds, and thus a Pólya urn process is defined for any initial $X_0 \in \mathcal{N}_*(S)$. If the urn is balanced, this is essentially the same as "κ-discrete MVPPs" in [13].

Example 5.2. Let S be a Borel space and consider an urn process with colour space $S \times [0,1]$. Let $\mathcal{M}_0 := \{\mu \times \lambda : \mu \in \mathcal{N}_*(S)\} \subset \mathcal{M}_*(S \times [0,1])$, amd assume that the replacements $R_{s,u}$ are such that

$$R_{s,u} + \delta_s \times \lambda \in \mathcal{M}_0 \quad \text{for every } s \in S \text{ and } u \in [0,1]. \tag{5.3}$$

Then (5.1) holds (in $S \times [0,1]$), and thus any $X_0 \in \mathcal{M}_0$ defines a Pólya urn process.

Remark 5.3. We may relax the condition $\mathcal{M}_0 \subseteq \mathcal{M}_*(S)$ to $\mathcal{M}_0 \subseteq \mathcal{M}(S)$; thus allowing $0 \in \mathcal{M}_0$ and consequently $R_s = -\mu$ in (5.1), which means that we remove all balls from the urn, leaving the urn empty, i.e., $X_{n+1} = 0$. In this case, we stop the process, and define $X_m = 0$ for all $m > n$. Formally, $\widehat{R}$ as defined in (3.2) then is a probability kernel from $\mathcal{M}_0 \setminus \{0\}$ to $\mathcal{M}_0$; we extend it to a kernel from $\mathcal{M}_0$ to $\mathcal{M}_0$ by defining $\widehat{R}_0 = \delta_0$. We leave further details for this case to the reader.

5.2. **Random replacements.** In the random case, we similarly assume that (5.1) holds a.s., for some measurable $\mathcal{M}_0 \subseteq \mathcal{M}_*(S)$, every $\mu \in \mathcal{M}_0$ and μ-a.e. s. We assume that S is a Borel space; then Lemma 2.2 implies that $\mathcal{M}_0$ is a measurable subset of $\mathcal{M}_\pm(S)$, and thus so is, for every μ,

$$\mathcal{M}_\mu := \{\nu \in \mathcal{M}_\pm(S) : \nu + \mu \in \mathcal{M}_0\}. \tag{5.4}$$

Hence the condition is that $\mathcal{R}$ is a probability kernel from S to $\mathcal{M}_\pm(S)$ such that for every $\mu \in \mathcal{M}_0$, $\mathcal{R}_s(\mathcal{M}_\mu) = 1$ for μ-a.e. s. Then the argument in Section 3.2 shows that $\widehat{R}$ is a probability kernel from $\mathcal{M}_0$ to itself, and thus defines a Pólya urn process for any initial $X_0 \in \mathcal{M}_0$.

Theorem 1.3 holds in this setting too, with the same proof given in Section 4; the deterministic urn $\widetilde{X}_n$ in $S \times [0,1]$ is defined as in Section 5.1 using $\widetilde{\mathcal{M}}_0 := \{\mu \times \lambda : \mu \in \mathcal{M}_0\} \subseteq \mathcal{M}_*(S \times [0,1])$.

Example 5.4 (Random drawing without replacement). Let $\mathcal{N}_*(S)$ be as in Example 5.1 and assume that R_s is a random replacement such that (5.2) holds a.s. for every $s \in S$; as always we assume also that $s \mapsto \mathcal{R}_s := \mathcal{L}(R_s) \in \mathcal{M}_\pm(S)$ is measurable. Then (5.1) holds a.s. for every $\mu \in \mathcal{N}_*(S)$ and μ-a.e. s, and thus R_s defines a Pólya urn process for any initial $X_0 \in \mathcal{N}_*(S)$.

Theorem 1.3 gives an equivalent urn in $S \times [0,1]$ with deterministic replacements. Note, however, that this deterministic urn is of the type in Example 5.2, and not of the simpler type in Example 5.1, as the random urn. Hence, Theorem 1.3 may be less useful in this setting.

6. An application

As an application, we show how Theorem 1.3 can be used to extend one of the main results in Mailler and Marckert [13] to random replacements.

Mailler and Marckert [13, Theorem 1.4] consider the case when S is a Polish space, and the replacement measures R_s are deterministic with total mass $R_s(S) = 1$, i.e., $R_s \in \mathcal{P}(S)$. Let $(W_n)_0^\infty$ be the Markov chain with values in S, some initial distribution $\mu_0 \in \mathcal{P}(S)$ and transitions given by the kernel R_s. Assume that there exist sequences $a(n) > 0$ and $b(n) \in S$ such that for any initial distribution μ_0,

$$a(n)^{-1}(W_n - b(n)) \xrightarrow{\text{d}} \gamma, \tag{6.1}$$

for some distribution $\gamma \in \mathcal{P}(S)$ not depending on μ_0.

In order for (6.1) to make sense, we assume, as in [13], that either S is a vector space, or that we are given some other definitions (satisfying standard axioms) of subtraction $s - b(n)$ and scalar multiplication $a(n)^{-1}s$ for $s \in S$ (for example the trivial definition if $b(n) = 0$ and $a(n) = 1$, for any space S).

Assume also that $f : \mathbb{R} \to S$ and $g : \mathbb{R} \to \mathbb{R}$ are two functions such that, for every $x \in \mathbb{R}$ and every sequence $x_n \to x$,

$$\frac{b(\lfloor n + x_n \sqrt{n} \rfloor) - b(n)}{a(n)} \to f(x), \tag{6.2}$$

$$\frac{a(\lfloor n + x_n \sqrt{n} \rfloor)}{a(n)} \to g(x). \tag{6.3}$$

(This implies that f and g are continuous, see [9, Lemma 3.3 and Remark 3.4].) Then [13, Theorem 1.4] says that

$$\Theta_{a(\log n), b(\log n)}\big(n^{-1} X_n\big) \xrightarrow{\text{p}} \nu := \mathcal{L}\big(g(\Lambda)\Gamma + f(\Lambda)\big), \tag{6.4}$$

where $\Theta_{a,b} : \mathcal{M}(S) \to \mathcal{M}(S)$ is the mapping induced by the map $s \mapsto a^{-1}(s - b)$, $\Lambda \sim N(0,1)$, and Γ is a random variable in S with distribution γ in (6.1). (See [13] for examples.)

Suppose now that the replacements R_s are random (with all other assumptions the same). We use Theorem 1.3 and construct an urn process $\widetilde{X}_n$ on $\widetilde{S} := S \times [0,1]$ with deterministic replacements. We may, as above, define a corresponding Markov chain $\widetilde{W}_n$ in $S \times [0,1]$. Write $\widetilde{W}_n = (W_n, U_n)$. Since (by Theorem 1.3) we have $\widetilde{X}_n = X_n \times \lambda$, each replacement measure $R_{(s,u)}$ is also a product measure of the type $\nu_{s,u} \times \lambda$. (This is also seen in (4.1) in the proof.) Since $R_{(s,u)}$ is the conditional distribution of (W_{n+1}, U_{n+1}) given

$(W_n, U_n) = (s, u)$, it follows that U_{n+1} is uniform on $[0, 1]$ and independent of W_{n+1} and of (W_m, U_m), $m \leqslant n$. It follows that $(W_n)_0^\infty$ is a Markov chain in S with transition kernel $\bar{R}_s := \mathbb{E}\, R_{s,U}$. However, by construction, $R_{s,U}$ equals the random replacement measure R_s, and thus $\bar{R}_s = \mathbb{E}\, R_s$. Hence, (W_n) is the Markov chain with kernel $\mathbb{E}\, R_s$. In other words, given $W_n = s$, take first a realization of R_s and then draw W_{n+1} from this distribution. We assume that (6.1) holds for this Markov chain, for any initial distribution.

In order to apply [13, Theorem 1.4] to the urn $\widetilde{X}_n$, we define subtraction and scalar multiplication on $S \times [0, 1]$ by

$$(s, x) - b := (s - b, x), \qquad b \in S, \tag{6.5}$$

$$a(s, x) := (as, x); \tag{6.6}$$

the operations thus acting trivially on the second coordinate. (We may identify S with $S \times \{0\} \subset S \times [0, 1]$. We have no need for subtracing $b \notin S$.) Then, since we have assumed (6.1), it follows that (6.1) holds also for $\widetilde{W}_n = (W_n, U_n)$, with the same $a(n)$ and $b(n)$ and with γ replaced by $\gamma \times \lambda$.

Hence [13, Theorem 1.4] applies to the urn $\widetilde{X}_n$, with $\widetilde{X}_n \in \mathcal{M}(S \times [0, 1])$. Projecting the measures by the projection $\pi : S \times [0, 1] \to [0, 1]$, we see that the conclusion of the theorem holds also for X_n.

In other words, [13, Theorem 1.4] holds also for random replacements.

Acknowledgement

References

[1] Antar Bandyopadhyay and Debleena Thacker, Rate of convergence and large deviation for the infinite color Pólya urn schemes. *Statist. Probab. Lett.* **92** (2014), 232–240.

[2] Antar Bandyopadhyay and Debleena Thacker, Pólya urn schemes with infinitely many colors. *Bernoulli* **23** (2017), no. 4B, 3243–3267.

[3] Antar Bandyopadhyay and Debleena Thacker, A new approach to Pólya urn schemes and its infinite color generalization. Preprint, 2016. `arXiv:1606.05317`

[4] David Blackwell and James B. MacQueen, Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** (1973), 353–355.

[5] Donald L. Cohn, *Measure Theory*, Birkhäuser, Boston, 1980.

[6] F. Eggenberger and G. Pólya, Über die Statistik verketteter Vorgänge. *Zeitschrift Angew. Math. Mech.* **3** (1923), 279–289.

[7] Svante Janson, *Gaussian Hilbert Spaces*. Cambridge Univ. Press, Cambridge, UK, 1997.

[8] Svante Janson, Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110** (2004), 177–245.

[9] Svante Janson and Lutz Warnke. On the critical probability in percolation. *Electronic J. Probability*, **23** (2018), paper no. 1, 25 pp.

[10] Olav Kallenberg. *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.

[11] Olav Kallenberg. *Random Measures, Theory and Applications.* Springer, Cham, Switzerland, 2017.

[12] Hosam M Mahmoud, *Pólya urn models.* CRC Press, Boca Raton, FL, 2009.

[13] Cécile Mailler and Jean-François Marckert. Measure-valued Pólya processes. *Electron. J. Probab.* **22** (2017), no. 26, 1–33.

[14] A. A. Markov, Sur quelques formules limites du calcul des probabilités. (Russian.) *Bulletin de l'Académie Impériale des Sciences* **11** (1917), no. 3, 177–186.

[15] K. R. Parthasarathy, *Probability Measures on Metric Spaces.* Academic Press, New York, 1967.

[16] Jim Pitman. *Combinatorial Stochastic Processes.* École d'Été de Probabilités de Saint-Flour XXXII – 2002. Lecture Notes in Math. 1875, Springer, Berlin, 2006.

[17] G. Pólya, Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* **1** (1930), no. 2, 117–161.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: svante.janson@math.uu.se

URL: http://www.math.uu.se/svante-janson