Urn schemes for constructing priors

Donato Michele Cifarelli e Pietro Muliere
Istituto di Metodi Quantitativi
Università "Luigi Bocconi" - Milano

Piercesare Secchi
Dipartimento di Matematica "Francesco Brioschi"
Politecnico di Milano

September, 2000

Abstract

We consider urn models for generating infinite exchangeable sequences of random variables whose de Finetti measure is either a neutral to the right process or a tailfree process.

This work is part of a book on 'Bayesian theory in a completely predictive perspective' that the authors are writing under contract with John Wiley & Sons, Ltd.

1 Introduction

Urn models have a long and relevant tradition in probability and statistics. They serves the purpose of making concrete the ideas generating a probabilistic model, yet keeping a level of abstraction which frees them from the particular application for which they were originally conceived and makes them suitable for purely theoretical consideration. The urn scheme introduced by Polya in 1923 represents the perfect historical example of this precious quality of urn models: it was introduced for describing contagious diseases and became soon the prototype for many probabilistic models for treating *aftereffects* (Feller, 1957). In the following pages it will be the cornerstone for the construction of widely used priors in Bayesian nonparametrics.

The next section reviews some classic two-color urn schemes, the Polya urn being probably the most famous. The latter generates an infinite exchangeable sequence of Bernoulli random variables which are conditionally independent and identically distributed given the random probability of success which is Beta distributed. This has been generalized by the elegant result of Blackwell and MacQueen (1973), described in Section 3, who consider a Polya urn with a continuum of colors generating an infinite exchangeable sequence of real random variables whose de Finetti measure is a Dirichlet process. Inspired by this result, Muliere, Secchi and Walker (2000) introduced a class of partially exchangeable processes on a state space of Polya urns called RUPs for reinforced urn processes; they will be instrumental for the construction of beta-Stacy and Polya trees priors in Sections 5 and 6 respectively. The last two sections treat extensions of this approach for the introduction of neutral to the right and tailfree priors.

2 Background on two-color urn processes

The leading character of this section is an urn initially containing $n_0(b) \geq 0$ black balls and $n_0(w) \geq 0$ white balls. At stage $n = 1, 2, ...$ we sample a ball from the urn and we then modify the composition of the urn following certain specific rules which characterize the urn process under consideration. This generates a sequence or random variables $\{X_n\}$, each equal to 0 or 1 according to the color white or black respectively of the ball sampled at stage n. Moreover, for all $n \geq 1$, we set B_n and W_n equal to the number of black balls and white balls respectively in the urn before the $n+1$-th stage. We are naturally interested in the law of the process $\{X_n\}$ as well as in the limit behavior of quantities which are functions of B_n and W_n, for instance the proportion of black balls contained in the urn before the $(n+1)$-th stage:

$$Z_n = \frac{B_n}{B_n + W_n}.$$

Our prototypical two-color urn process is also one of the simplest and of the historically oldest; it is known under the name of Polya urn after the 1923 paper by Eggenberger and Polya.

At each stage $n \geq 1$, a ball is chosen at random from the urn and put back into it along with other $m \geq 1$ balls of the same color.

For a Polya urn, the dynamics of the processes $\{X_n\}, \{B_n\}$ and $\{W_n\}$ are governed by the following: X_1 has distribution

$$\text{Bernoulli}(\frac{n_0(b)}{n_0(b)+n_0(w)}).$$

For all $n \geq 1$, conditionally on $X_1, ..., X_n$,

$$X_{n+1} = \begin{cases} 0 & \text{with probability } \frac{W_n}{W_n+B_n} \\ 1 & \text{with probability } \frac{B_n}{W_n+B_n} \end{cases}$$

whereas

$$(B_{n+1}, W_{n+1}) = \begin{cases} (B_n, W_n + m) & \text{with probability } \frac{W_n}{W_n+B_n} \\ (B_n + m, W_n) & \text{with probability } \frac{B_n}{W_n+B_n}. \end{cases}$$

There are two basic theorems about the processes $\{X_n\}$ and $\{(B_n, W_n)\}$ which are both consequences of de Finetti's Representation Theorem.

2.1 **Theorem.** *The sequence $\{X_n\}$ generated by a Polya urn is exchangeable and its de Finetti measure is a Beta with parameters $(\frac{n_0(b)}{m}, \frac{n_0(w)}{m})$.*

Proof. Let $1 \leq k \leq n$ and $(i_1, ..., i_n)$ such that $i_j \in \{0,1\}$ and $\sum_{j=1}^{n} i_j = k$. Then

$$\begin{aligned} P(X_1 = i_1, ..., X_n = i_n) &= \frac{\prod_{j=0}^{k-1}(n_0(b)+mj) \cdot \prod_{j=0}^{n-k-1}(n_0(w)+jm)}{\prod_{j=0}^{n-1}(n_0(b)+n_0(w)+jm)} \\ &= \frac{\Gamma(\frac{n_0(b)}{m}+\frac{n_0(w)}{m})}{\Gamma(\frac{n_0(b)}{m})\Gamma(\frac{n_0(w)}{m})} \frac{\Gamma(\frac{n_0(b)+k}{m})\Gamma(\frac{n_0(w)+n-k}{m})}{\Gamma(\frac{n_0(b)}{m}+\frac{n_0(w)}{m}+n)} \\ &= \int_0^1 \theta^k(1-\theta)^{n-k}[\frac{\Gamma(\frac{n_0(b)}{m}+\frac{n_0(w)}{m})}{\Gamma(\frac{n_0(b)}{m})\Gamma(\frac{n_0(w)}{m})}\theta^{\frac{n_0(b)}{m}-1}(1-\theta)^{\frac{n_0(w)}{m}-1}]d\theta. \end{aligned}$$

Because of de Finetti's Representation Theorem, this shows that the sequence is exchangeable. Moreover , the unicity of the representation implies that the de Finetti measure of the sequence $\{X_n\}$ is a Beta with parameters $(\frac{n_0(b)}{m}, \frac{n_0(w)}{m})$. ◇

The second theorem considers the limit behavior of the proportion of black balls contained in the urn.

2.2 **Theorem.** *In a Polya urn, as n grows to infinity, the proportion of black balls*

$$Z_n = \frac{B_n}{B_n + W_n}$$

converges almost surely to a random limit. Moreover, the distribution of the limit is a Beta with parameters $(n_0(b)/m, n_0(w)/m)$.

Proof. For proving the result we could use the fact that the de Finetti measure of the exchangeable sequence $\{X_n\}$ is the weak limit of the sequence of empirical distributions generated by $\{X_n\}$. We prefer to follow instead a different approach which is a direct proof of the above mentioned fact in this particular situation.

We claim that the sequence $\{Z_n\}$ is a bounded martingale. In fact, for all $n \geq 0$, $0 \leq Z_n \leq 1$. Moreover

$$\begin{aligned} E[Z_{n+1}|Z_1, ..., Z_n] &= \frac{B_n + m}{B_n + W_n + m}\frac{B_n}{B_n + W_n} + \frac{B_n}{B_n + W_n + m}\frac{W_n}{B_n + W_n} \\ &= \frac{B_n}{B_n + W_n} \\ &= Z_n. \end{aligned}$$

Therefore, from Doob's convergence theorem for martingales, the sequence $\{Z_n\}$ converges almost surely and L^1 to a random variable Z_∞.

However Theorem 2.1 and the law of large numbers imply that, for n growing to infinity, the distribution of the random variable $\frac{1}{n}\sum_{i=1}^n X_i$ converges to a Beta with parameters $(n_0(b)/m, n_0(w)/m)$.

Since, for all $n \geq 1$,

$$Z_n = \frac{n_0(b) + m\sum_{i=1}^n X_i}{n_0(b) + n_0(w) + nm}$$

for any given $z \in [0, 1]$,

$$P(Z_n \leq z) = P(\frac{1}{n}\sum_{i=1}^n X_i \leq z(\frac{n_0(b)}{nm} + \frac{n_0(w)}{nm} + 1) - \frac{n_0(b)}{nm}).$$

Therefore

$$\lim_{n\to\infty} P(Z_n \leq z) = \int_0^z \frac{\Gamma(\frac{n_0(b)}{m} + \frac{n_0(w)}{m})}{\Gamma(\frac{n_0(b)}{m})\Gamma(\frac{n_0(w)}{m})}\theta^{\frac{n_0(b)}{m}-1}(1-\theta)^{\frac{n_0(w)}{m}-1}d\theta.$$

◇

2.3 **Remark.** Note that the quantities $n_0(b), n_0(w)$ and m can be taken to be arbitrary strictly positive real numbers. Taking them to be integers has the only advantage of permitting the physical description of the processes involved in terms of balls of two colors sampled from an urn. ◇

A generalization of Polya urn is the following urn scheme proposed by Friedman (1949); an urn initially contains $n_0(b)$ black balls and $n_0(w)$ white balls. At stage n, with $n \geq 1$, a ball is sampled from the urn and $1 + \eta$ balls of the same color together with θ balls of the other color are put back in the urn. If $\eta = 0$ and $\theta = 0$ we have the usual extraction with replacement scheme which generates a sequence of i.i.d colors. If $\eta = m - 1 > 0$ and $\theta = 0$ we recover the Polya urn scheme considered above. For $\eta = -1$ and $\theta = 1$ we obtain the celebrated Erhenfest's urn scheme; note that in this case, at every stage n, the total number of balls in the urn is constant and equal to $n_0(b) + n_0(w)$. Moreover, it is an easy exercise with Markov chains to prove the following result.

2.4 **Theorem.** *If $\eta = -1$ and $\theta = 1$, the sequence $\{B_n\}$ is an homogeneous, irreducible Markov chain on $\{0, ..., n_0(b) + n_0(w)\}$ and its stationary distribution is a Binomial($n_0(w) + n_0(b), 1/2$).*

Therefore, in the case of Erhenfest's urn, the proportion of black balls,

$$Z_n = \frac{B_n}{B_n + W_n}$$

converges to a random limit which has a discrete distribution with support on $\{k/[n_0(b) + n_0(w)], k = 0, ..., n_0(b) + n_0(w)\}$ and masses equal to those of a Binomial($n_0(w) + n_0(b), 1/2$).

The situation is radically different when $\eta \geq 0$ and $\theta > 0$; this case has been considered by Freedman (1965) who proved the following result which we cite without proof.

2.5 **Theorem.** *If $\eta \geq 0$ and $\theta > 0$, the proportion of black balls*

$$Z_n = \frac{B_n}{B_n + W_n}$$

converges to 1/2 with probability one whatever the initial composition $(n_0(b), n_0(w))$ of the urn.

From our perspective a central question is the following: for which η and θ the sequence of colors $\{X_n\}$ produced by a Friedman urn is infinite and exchangeable? When this happens, de Finetti's Representation Theorem implies that the de Finetti measure of the sequence is that of the limit of the sequence $\{Z_n\}$.

2.6 **Theorem.** *Let $\{X_n\}$ be the sequence of colors produced by a Friedman urn. Then $\{X_n\}$ is infinite and exchangeable if and only if one the following is satisfied:*

(1) $\eta \geq 0$ and $\theta = 0$;

(2) $\eta = \theta$ and $n_0(b) = n_0(w)$.

Proof. Let us compute

$$P(X_1 = 1, X_2 = 0) = \frac{n_0(b)}{n_0(b) + n_0(w)} \frac{n_0(w) + \theta}{n_0(w) + n_0(b) + \eta + \theta}$$

and

$$P(X_1 = 0, X_2 = 1) = \frac{n_0(w)}{n_0(b) + n_0(w)} \frac{n_0(b) + \eta}{n_0(w) + n_0(b) + \eta + \theta}.$$

For exchangeability, we require $P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1)$. This happens either if $\theta = 0$ or if $n_0(b) = n_0(w)$.

If $\theta = 0$ and $\eta < 0$, after a finite number of stages there are no balls left in the urn; therefore the sequence $\{X_n\}$ is not infinite in this case. If $\theta = 0$ and $\eta = 0$ then $\{X_n\}$ is a sequence of i.i.d. Bernoulli$(n_0(b)(n_0(b) + n_0(w))^{-1})$ random variables which is clearly exchangeable. For $\theta = 0$ and $\eta > 0$, we fall back in the Polya urn scheme and we proved already that $\{X_n\}$ is in this case exchangeable.

If $\theta \neq 0$ and $n_0(b) = n_0(w)$, in order to have

$$P(X_1 = 0, X_2 = 1, X_3 = 0) = P(X_1 = 0, X_2 = 0, X_3 = 1)$$

we need $\eta = \theta$. Note that in this case the random variables of the sequence $\{X_n\}$ are i.i.d. Bernoulli$(1/2)$. ◇

Another generalization of the Polya urn scheme has been proposed by Hill, Lane and Sudderth (1987). Let $f : [0,1] \to [0,1]$ and for all $n \geq 0$ indicate, as before, with Z_n the proportion at stage n of black balls in an urn initially containing $n_0(b)$ black balls and $n_0(w)$ white balls. Assume that, at each stage $n \geq 1$, a black ball is added to the urn with probability $f(Z_{n-1})$ and a white ball is added with probability $1 - f(Z_{n-1})$. Again we consider the process $\{X_n\}$ of random variables generated by the urn scheme, with $X_n = 1$ if the ball added at the n-th stage is black and $X_n = 0$ if it is white. Note that if $f(x) = x$ for all $x \in [0,1]$ we recover the Polya urn scheme whereas a constant f generates a process $\{X_n\}$ of independent and identically distributed Bernoulli random variables; both of this processes are exchangeable. Another trivially exchangeable process can be generated with an f such as

$$f(x) = \begin{cases} 0 & \text{if } x < \frac{n_0(b)}{n_0(b)+n_0(w)}, \\ p & \text{if } x = \frac{n_0(b)}{n_0(b)+n_0(w)}, \\ 1 & \text{if } x > \frac{n_0(b)}{n_0(b)+n_0(w)}, \end{cases}$$

with $p \in [0,1]$. This f produces a process $\{X_n\}$ where $X_1 = 1$ with probability p and $X_n = X_1$ with probability one for all $n \geq 1$; such a process is termed deterministic.

Hill, Lane and Sudderth prove that the only exchangeable processes $\{X_n\}$ produced by their urn scheme are the Polya's process, the i.i.d. Bernoulli process and the deterministic one.

3 A Polya urn for the Dirichlet process

We will now extend the Polya urn scheme of the previous section by considering an urn which possibly contains initially a "continuum" of colors with a given distribution. This generalization is due to Blackwell and MacQueen (1973) who introduced it in order to generate an exchangeable sequence of random variables whose de Finetti measure is a Dirichlet process.

Let $\mathbf{X}$ be a metric space, separable and complete, endowed with its Borel σ-field $\mathcal{X}$ and let α be a finite measure on $(\mathbf{X}, \mathcal{X})$. The space $\mathbf{X}$ will be that for the colors of the balls in the urn; in fact we are imagining an urn containing balls labeled by the elements of $\mathbf{X}$ in such a way that the initial distribution of balls among the colors is that obtained by normalizing the finite measure α. At each stage $n = 1, 2, ...,$ a ball is sampled from the urn and replaced in it along with another ball of the same color. For $n \geq 1$, let $X_n = x$ if x is the color sampled at the n-th stage. Therefore, for all $B \in \mathcal{X}$,

$$P(X_1 \in B) = \frac{\alpha(B)}{\alpha(\mathbf{X})}$$

whereas for all $n \geq 1$, the conditional distribution of X_{n+1} given $X_1, ..., X_n$, is such that

$$P(X_{n+1} \in B | X_1, ..., X_n) = \frac{\alpha(B) + \sum_{i=1}^{n} \delta_{X_i}}{\alpha(\mathbf{X}) + n}$$

for all $B \in \mathcal{X}$.

3.1 **Definition.** *The sequence $\{X_n\}$ is called a Polya sequence on $\mathbf{X}$ with parameter α.*

The initial distribution of colors in the urn is described by the probability distribution

$$m_0 = \frac{1}{\alpha(\mathbf{X})} \cdot \alpha.$$

For all $n \geq 1$, set

$$\alpha_n = \alpha + \sum_{i=1}^{n} \delta_{X_i} \quad \text{and} \quad m_n = \frac{1}{\alpha_n(\mathbf{X})} \cdot \alpha_n.$$

Hence, m_n describes the random probability distribution of colors in the urn after the n-th ball has been sampled; we will say that m_n is the probability distribution on $\mathbf{X}$ generated by the Polya sequence $\{X_n\}$ at stage n.

When $\mathbf{X}$ is discrete and finite its easy to extend the result of Theorem 2.1; without loss of generality we may assume that $\mathbf{X} = \{0, ..., r\}$.

3.2 **Lemma.** *Let $\mathbf{X} = \{0, ..., r\}$ and $\{X_n\}$ be a Polya sequence on $\mathbf{X}$ with parameter α. Then:*

(a) As n grows to infinity, the random vector $(m_n(\{0\}), ..., m_n(\{r\}))$ converges with probability one to a random vector $\alpha^* = (\alpha^*(\{0\}), ..., \alpha^*(\{r\}))$;

(b) The random vector α^* has Dirichlet distribution with parameters $(\alpha(\{0\}), ..., \alpha(\{r\}))$;

(c) Given α^*, the random variables of the Polya sequence $\{X_n\}$ are independent and identically distributed according to α^*.

Proof. Note that, for all $n \geq 1$ and $x_1, ..., x_n \in \mathbf{X}$,

$$P(X_1 = x_1, ..., X_n = x_n) = \prod_{j=0}^{r} \frac{\alpha(\{j\})^{[n(j)]}}{\alpha(\mathbf{X})^{[\mathrm{n}]}}$$

where $n(j)$ is equal to the number of x_i equal to j, for $i = 1, ..., n$, and we set $a^{[k]} = a(a + 1) \cdots (a + k - 1)$ if $k \geq 1$ and $a^{[0]} = 1$. This probability does not change for any permutation of $x_1, ..., x_n$. Therefore the Polya sequence $\{X_n\}$ is exchangeable and de Finetti's Representation Theorem implies that there is a unique distribution G on the simplex

$$S_r = \{(y_1, ..., y_r) \in [0, 1]^r : y_i \geq 0 \text{ for } i = 1, ..., r \text{ and } \sum_{i=1}^{r} y_i < 1\}$$

such that, for all $n \geq 1$ and $x_1, ..., x_n \in \mathbf{X}$,

$$P(X_1 = x_1, ..., X_n = x_n) = \int_{S_r} (1 - \sum_{i=1}^{r} y_i)^{n(0)} \prod_{i=1}^{r} y_i^{n(i)} dG(y_1, ..., y_r).$$

However, if G is the distribution induced on S_r by a Dirichlet distribution on $[0, 1]^{r+1}$ with parameters $(\alpha(\{0\}), ..., \alpha(\{r\}))$, then

$$\int_{S_r} (1 - \sum_{i=1}^{r} y_i)^{n(0)} \prod_{i=1}^{r} y_i^{n(i)} dG(y_1, ..., y_r) = \prod_{j=0}^{r} \frac{\alpha(\{j\})^{[n(j)]}}{\alpha(\mathbf{X})^{[\mathrm{n}]}}.$$

Since the de Finetti measure of the Polya sequence $\{X_n\}$ is unique, this shows that it must be a Dirichlet with parameters $(\alpha(\{0\}), ..., \alpha(\{r\}))$. Moreover the same de Finetti's Representation Theorem implies that the random vector

$$(\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\{0\}), ..., \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\{r\}))$$

converges with probability one to a random vector $\alpha^* = (\alpha^*(\{0\}), ..., \alpha^*(\{r\}))$ whose distribution is Dirichlet with parameters $(\alpha(\{0\}), ..., \alpha(\{r\}))$. The proof of the lemma is concluded by noticing that the limiting behavior of

$$(m_n(\{0\}), ..., m_n(\{r\}))$$

is the same as that of

$$(\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\{0\}), ..., \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\{r\})).$$

◇

The previous lemma is extended by the following result of Blackwell and MacQueen (1973) which holds for a Polya sequence on a generic metric space $\mathbf{X}$ separable and complete. It immediately implies a concrete procedure for constructing an infinite exchangeable sequence

$\{X_n\}$ of random variables with values on $\mathbf{X}$ whose de Finetti measure is a Dirichlet process; in Bayesian inference this serves the purpose of a good analogy for deciding whether the statistical model for the exchangeable sequence $\{X_n\}$ implied by the Dirichlet process prior is appropriate.

3.3 **Theorem**. *Let $\{X_n\}$ be a Polya sequence on $\mathbf{X}$ with parameter α. Then:*

(a) As n grows to infinity, with probability one the random probability distribution m_n weakly converges to a discrete random probability distribution α^*;

(b) The distribution of α^* is $\mathcal{D}_\alpha$, a Dirichlet process with parameter α;

(c) The Polya sequence $\{X_n\}$ is exchangeable and its de Finetti measure is $\mathcal{D}_\alpha$.

Proof. For $n \geq 1$, let

$$f_{n,j} = \begin{cases} \mathcal{I}_j m_n(\{X_j\}) & \text{if } 1 \leq j \leq n, \\ 0 & \text{if } j > n, \end{cases}$$

where

$$\mathcal{I}_j = \begin{cases} 1 & \text{if } X_j \neq X_i \text{ for all } i < j, \\ 0 & \text{otherwise.} \end{cases}$$

Hence $f_{n,j}$ is equal to the random mass assigned to the point X_j by m_n when X_j is a color different from those previously extracted from the urn. We claim that:

(i) For all $j \geq 1$, the sequence $\{f_{n,j}\}$ converges with probability one to a random limit f_j^*;

(ii) with probability one, $\sum_j f_j^* = 1$.

Given $r \geq 1$, let $\{U_n\}$ be a sequence of random variables defined by setting, for all $n \geq 1$,

$$U_n = \begin{cases} j & \text{if } 1 \leq j \leq r, \mathcal{I}_j = 1, X_{r+n} = X_j, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, we consider the distinct colors produced by the urn during the first r stages. If, when at stage $r+n$, one of these colors, say that appeared at the j-th extraction among the first r, appears again, then U_n is set to be j. Otherwise U_n is set to be 0. It easy to check that, conditionally on $X_1, ..., X_r$, the sequence $\{U_n\}$ is Polya on $\{0, ..., r\}$ with parameter α' where

$$\alpha'(\{0\}) = (\alpha(\mathbf{X}) + r)(1 - \sum_{j=1}^{r} f_{r,j})$$

and, for $1 \leq j \leq r$,

$$\alpha'(\{j\}) = (\alpha(\mathbf{X}) + r) f_{r,j}.$$

For $n \geq 1$, let m'_n be the random probability distribution generated on $\{0, ..., r\}$ by the Polya sequence $\{U_n\}$ at stage n. Notice that, for $1 \leq j \leq r$, $m'_n(\{j\}) = f_{r+n,j}$ whereas $m'_n(\{0\}) = 1 - \sum_{j=1}^{r} f_{r+n,j}$. It follows from Lemma 3.2 that, conditionally on $X_1, ..., X_r$, $(m'_n(\{0\}), ..., m'_n(\{r\}))$ converges with probability one to a random vector with Dirichlet distribution with parameter $(\alpha'(\{0\}), ..., \alpha'(\{r\}))$; let this limit vector be $(f_0^*, ..., f_r^*)$.

The previous argument proves that, for all $j \geq 1$, the sequence $\{f_{n,j}\}$ converges with probability one to f_j^*. Moreover, for $r \geq 1$,

$$\begin{aligned} E[1 - \sum_{j=1}^{r} f_j^*] &= E[E[1 - \sum_{j=1}^{r} f_j^* | X_1, ..., X_r]] \\ &= E[\frac{\alpha'(\{0\})}{\alpha(\mathbf{X}) + r}] \\ &\leq \frac{\alpha(\mathbf{X})}{\alpha(\mathbf{X}) + r}. \end{aligned}$$

Therefore, by dominated convergence,

$$E[1 - \sum_{j=1}^{\infty} f_j^*] = \lim_{r \to \infty} E[1 - \sum_{j=1}^{r} f_j^*] = 0.$$

Since, for all $r \geq 1$, $0 \leq \sum_{j=1}^{r} f_j^* \leq 1$ with probability one, the previous equation implies that $\sum_{j=1}^{\infty} f_j^* = 1$ with probability one, thus proving (ii).

Let $B \in \mathcal{X}$ be given and set, for all $n, r \geq 1$,

$$s_{n,r} = \sum_{j=1}^{r} I[X_j \in B] f_{n,j} \text{ and } t_{n,r} = \sum_{j=1}^{r} f_{n,j}.$$

If A_r is the set of distinct colors among those extracted from the urn in the first r stages, then

$$s_{n,r} = m_n(A_r \bigcap B) \leq m_n(B) \leq m_n(A_r \bigcap B) + m_n(\bar{A}_r) = s_{n,r} + (1 - t_{n,r}).$$

By letting first $n \to \infty$ and then $r \to \infty$,with probability one has that

$$\lim_{n \to \infty} m_n(B) = \sum_{j=1}^{\infty} I[X_j \in B] f_j^*. \tag{3.4}$$

Let α^* be the random probability measure defined on $\mathcal{X}$ by setting

$$\alpha^*(B) = \sum_{j=1}^{\infty} I[X_j \in B] f_j^*$$

for all $B \in \mathcal{X}$. Since $\mathbf{X}$ is separable and complete, $\mathcal{X}$ is countably generated and equation (3.4) implies that m_n converges with probability one to α^*. This proves part (a) of the theorem.

Now let $r \geq 1$ and $\nabla_r = \{B_0, ..., B_r\}$ be a finite measurable partition of $\mathbf{X}$ and set $\phi(x) = \sum_{i=0}^{r} i I[x \in B_i]$. It's easy to check that $\{\phi(X_n)\}$ is a Polya sequence on $\{0, ..., r\}$ with parameter

$(\alpha(B_0), ..., \alpha(B_r))$. Since the random probability distribution generated on $\{0, ..., r\}$ by $\{\phi(X_n)\}$ at stage n gives mass equal to $m_n(B_i)$ to the point i, for $i = 0, ..., r$, part (b) of Lemma 3.2 implies that $(\alpha^*(B_1), ..., \alpha^*(B_r))$ has Dirichlet distribution with parameter $(\alpha(B_0), ..., \alpha(B_r))$. Being this true for any $r \geq 1$ and any finite measurable partition ∇_r of $\mathbf{X}$, the law of α^* must be that of a Dirichlet process with parameter α proving part (b) of the theorem.

The previous argument also helps to show the last part of the theorem: in fact, for any $r \geq 1$ and any $\nabla_r = \{B_0, ..., B_r\}$ finite measurable partition of $\mathbf{X}$, part (c) of Lemma 3.2 applied to the Polya sequence $\{\phi(X_n)\}$implies that, for $C_1, ..., C_n \in \nabla_r$,

$$P(X_1 \in C_1, ..., X_n \in C_n) = E[\prod_{j=1}^{n} \alpha^*(C_j)]. \tag{3.5}$$

Given any finite sequence $D_1, ..., D_n$ of sets in $\mathcal{X}$, let $\nabla_r = \{B_1, ..., B_r\}$ be the measurable partition generated by $D_1, ..., D_n$, that is the class of measurable sets obtained by taking all the possible nonempty intersections of the sets D_j and their complements. Then

$$\begin{aligned} P(X_1 \in D_1, ..., X_n \in D_n) &= \sum_{i_1 \in (1)} \cdots \sum_{i_n \in (n)} P(X_1 \in B_{i_1}, ..., X_n \in B_{i_n}) \\ &= \sum_{i_1 \in (1)} \cdots \sum_{i_n \in (n)} E[\prod_{j=1}^{n} \alpha^*(B_{i_j})] \\ &= E[\prod_{j=1}^{n} \alpha^*(D_j)] \end{aligned}$$

where, for $j = 1, ..., n$, $(j) = \{k \in \{1, ..., r\} : B_k \subseteq D_j\}$. This shows that the random variables of the sequence $\{X_n\}$ are independent and identically distributed according to α^* concluding the proof of the theorem. ◇

4 Reinforced urn processes

In the context of Bayesian nonparametric inference, the importance of Blackwell and MacQueen's result described in the previous section is that it gives a simple and concrete procedure for constructing an infinite exchangeable sequence of random variables with Dirichlet process as de Finetti measure. The procedure has the additional advantage of making intuitively clear some of the mathematical properties of the Dirichlet process, like its being discrete with probability one, as well as some of its inferential prerogatives, like its conjugacy property or the form of the predictive distribution of the $(n+1)$-th random variable generated by a Dirichlet process conditionally on the values of the first n variables. In the spirit of Blackwell and MacQueen we introduce in this section a class of stochastic processes defined on a countable space of Polya urns which will be convenient for constructing more general classes of priors commonly used in Bayesian nonparametric inference like Polya trees and beta-Stacy processes. Much of the material for this and the following sections is taken from Muliere, Secchi and Walker (2000).

Let S be a countable space and $E = \{c_1, ..., c_k\}$ be a finite set of colors with cardinality $k \geq 1$. We imagine that a function U associates with every $x \in S$ a Polya urn whose initial composition is $U(x) = (n_x(c_1), ..., n_x(c_k))$, where $n_x(c)$ represents the number of balls of color c contained in the urn, for $c \in E$. In fact, U is called *urn composition function* if it maps S into the set of k-tuples of nonnegative real numbers whose sum is strictly positive. Hence, S is the space of labels for a countable number of Polya urns, each with a possibly different initial composition of balls colored with the same finite palette.

Let $q : S \times E \to S$ be a *law of motion* with the property that, for all $x, y \in S$, there is at most one color $c(x, y) \in E$ such that $q(x, c(x, y)) = y$.

We have now all the ingredients necessary for the construction of a stochastic process $\{X_n\}$ on S whose recursive definition goes as follows: fix an initial state $x_0 \in S$ and set $X_0 = x_0$. For all $n \geq 1$, if $X_{n-1} = x \in S$, a ball is picked at random from the urn associated with x and returned to it along with another of the same color. If $c \in E$ is the color of the sampled ball, set $X_n = q(x, c)$.

4.1 **Definition.** *We will say that* $\{X_n\} \in RUP(S, E, U, q)$ *with initial state* x_0, *where* RUP *is the acronym for Reinforced Urn Process.*

It's easy to compute an expression for the finite dimensional laws of the process $\{X_n\}$. Before stating this result we need some further notation.

If $\sigma = (s_0, \ldots, s_n)$ is a finite sequence of elements of S, say that σ is *admissible* if $s_0 = x_0$ and there is a finite sequence $(c_0, \ldots, c_{n-1})$ of elements of E such that $q(s_i, c_i) = s_{i+1}$ for $i = 0, \ldots, n-1$. Given an admissible sequence σ, for each $x, y \in S$ we count with $t(x, y)$ the number of transitions in σ from state x to state y. We then set, for all $x \in S$ and $c \in E$

$$l_x(c) = t(x, q(x, c))$$

and

$$t(x) = \sum_{y \in S} t(x, y).$$

4.2 **Theorem.** *For all* $n \geq 0$ *and all finite sequences* $(s_0, \ldots, s_n)$ *of elements of* S, $P[X_0 = s_0, \ldots, X_n = s_n] = 0$ *if* $(s_0, \ldots, s_n)$ *is not admissible; otherwise*

$$P[X_0 = s_0, \ldots, X_n = s_n] = \prod_{x \in S} \left[\frac{\prod_{c \in E} \prod_{i=0}^{l_x(c)-1} (n_x(c) + i)}{\prod_{j=0}^{t(x)-1} (j + \sum_{c \in E} n_x(c))} \right] \tag{4.3}$$

where the convention is made that $\prod_0^{-1} = 1$.

Proof. Let $\sigma = (s_0, \ldots, s_n)$ be a finite sequence of elements of S. If σ is not admissible, it is evident that $P[X_0 = s_0, \ldots, X_n = s_n] = 0$. If σ is admissible, then

$$P[X_0 = s_0, \ldots, X_n = s_n] = \prod_{i=0}^{n-1} \left[\frac{n_{s_i}(c(s_i, s_{i+1})) + m_{s_i}(c(s_i, s_{i+1}))}{\sum_{c \in E}(n_{s_i}(c) + m_{s_i}(c))} \right] \tag{4.4}$$

where

$$m_{s_i}(c) = \sum_{r=0}^{i-1} I[s_r = s_i, c(s_r, s_{r+1}) = c]$$

for $c \in E$ and $i = 0, \ldots, n-1$.

Given an element x of the sequence $(s_0, \ldots, s_{n-1})$, let $(y_1, \ldots, y_{t(x)})$ be the ordered sequence of elements which follow x in σ. Then the group of factors in (4.4) relative to x can be rewritten as

$$\frac{n_x(c(x, y_1))}{\sum_{c \in E} n_x(c)} \cdot \frac{n_x(c(x, y_2)) + I[y_1 = y_2]}{1 + \sum_{c \in E} n_x(c)} \cdots \frac{n_x(c(x, y_{t(x)})) + \sum_{i=1}^{t(x)-1} I[y_i = y_{t(x)-1}]}{t(x) - 1 + \sum_{c \in E} n_x(c)}$$
$$= \frac{\prod_{c \in E} \prod_{i=0}^{l_x(c)-1} (n_x(c) + i)}{\prod_{j=0}^{t(x)-1} (j + \sum_{c \in E} n_x(c))}.$$

From this, (4.3) follows. ◇

A consequence of the previous result is that a reinforced urn process is partially exchangeable in the sense of Diaconis and Freedman (1980). These authors called two finite sequences σ and ϕ of elements of S *equivalent* if they begin with the same element and, for every $x, y \in S$, the number of transitions from x to y is the same in both sequences. In addition, they defined partially exchangeable any process $\{Y_n\}$ defined on S such that, for all $n \geq 0$ and all equivalent sequences $\sigma = (s_0, \ldots, s_n)$ and $\phi = (y_0, \ldots, y_n)$ of elements of S,

$$P[Y_0 = s_0, \ldots, Y_n = s_n] = P[Y_0 = y_0, \ldots, Y_n = y_n]. \tag{4.5}$$

4.6 **Corollary.** *A reinforced urn process $\{X_n\}$ is partially exchangeable.*

Proof. Fix $n \geq 1$. If $\sigma = (s_0, \ldots, s_n)$ and $\phi = (y_0, \ldots, y_n)$ are two equivalent and admissible finite sequences of elements of S, then $P[X_0 = s_0, \ldots, X_n = s_n]$ depends only on the transition counts $t(x, y)$, $x, y \in S$, and hence it is the same as $P[X_0 = y_0, \ldots, Y_n = y_n]$. If, on the contrary, σ and ϕ are both not admissible, then $P[X_0 = s_0, \ldots, X_n = s_n] = P[X_0 = y_0, \ldots, X_n = y_n] = 0$. ◇

For a process $\{Y_n\}$ defined on S with initial state $Y_0 = y_0$, set $\tau_0 = 0$ and, for all $n \geq 1$, define

$$\tau_n = \inf\{j > \tau_{n-1} : Y_j = y_0\}. \tag{4.7}$$

Then $\{Y_n\}$ is said to be *recurrent* if

$$P\left[Y_n = Y_0 \text{ for infinitely many } n\right] = 1$$

or, equivalently, if

$$P\left[\bigcap_{n=0}^{\infty}\{\tau_n < \infty\}\right] = 1. \tag{4.8}$$

The following representation result of Diaconis and Freedman (1980), which we state without proof, will be essential for exploring properties of recurrent RUPs.

4.9 **Theorem.** *Let $\{Y_n\}$ be a recurrent and partially exchangeable process defined on S. Then $\{Y_n\}$ is a mixture of Markov chains.*

Corollary 4.6 and the previous representation theorem imply that a recurrent process $\{X_n\} \in RUP(S, E, U, q)$ is a mixture of Markov chains. In order to make the statement more precise, suppose that $X_0 = x_0$, and for all $x \in S$, set

$$R_x = \{y \in S : n_x(c(x, y)) > 0\}.$$

Then define $R^0 = \{x_0\}$ and, for all $n \geq 1$, set

$$R^n = \bigcup_{x \in R^{n-1}} R_x.$$

The states in R^n are those that the process $\{X_n\}$ with initial state x_0 reaches with positive probability in n steps. Let

$$R = \bigcup_{n=0}^{\infty} R_n. \tag{4.10}$$

Then it follows from Theorem 4.2 that, for all $n \geq 0$,

$$P[X_0 \in R, \ldots, X_n \in R] = 1.$$

Theorem 4.9 and Corollary 4.6 then imply that there is a probability distribution μ on the set $\mathcal{M}$ of stochastic matrices on $R \times R$ such that, for all $n \geq 1$ and all finite sequences $(x_1, \ldots, x_n)$ of elements of R,

$$P[X_0 = x_0, \ldots, X_n = x_n] = \int_{\mathcal{M}} \prod_{j=0}^{n-1} m(x_j, x_{j+1})\, \mu(dm). \tag{4.11}$$

Let M be a random element of $\mathcal{M}$ with probability distribution μ. The next theorem describes the measure μ by showing that the rows of M are independent Dirichlet processes.

For all $x \in R$, let $M(x)$ be the x-th row of M and $\alpha(x)$ be the measure on R which assigns mass $n_x(c)$ to $q(x, c)$ for each $c \in E$, and mass 0 to all other elements of R; finally set

$$\phi(x) = P[X_n = x \text{ for infinitely many } n]$$

4.12 **Lemma**. *Assume that $\{X_n\}$ is recurrent and let $x_1, \ldots, x_m$, $m \geq 1$, be distinct elements of R such that $\phi(x_i) = 1$ for $i = 1, \ldots, m$. Then $M(x_1), \ldots, M(x_m)$ are mutually independent random distributions on R and, for $i = 1, \ldots, m$, the law of $M(x_i)$ is that of a Dirichlet process with parameter $\alpha(x_i)$.*

Proof. Let $x \in \mathcal{R}$ be such that $\phi(x) = 1$. We begin by showing that $M(x)$ is a Dirichlet process of parameter $\alpha(x)$. For all $n \geq 1$, let Y_n be the color of the ball picked from urn $U(x)$ the n-th time it is sampled. Then $\{Y_n\}$ is a Polya sequence of parameter $U(x)$. Let us order the elements of E by writing $E = \{c_1, \ldots, c_k\}$. Then, as a special case of Lemma 3.2, there is a random vector $\Theta(x)$ such that, with probability one,

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{i=1}^{\infty} I[Y_i = c_1], \ldots, \frac{1}{n} \sum_{i}^{\infty} I[Y_i = c_k] \right) = \Theta(x) \qquad (4.13)$$

and $\Theta(x)$ has Dirichlet distribution with parameter $(n_x(c_1), \ldots, n_x(c_k))$. Furthermore, conditionally on $\Theta(x)$, the random variables of the sequence $\{Y_n\}$ are indipendent and identically distributed with distribution $\Theta(x)$.

For all $j \geq 1$, let $M^{(j)}$ be the doubly infinite matrix indexed by elements of R whose (s, y) entry $M^{(j)}(s, y)$ is the number of transitions from state $s \in R$ to state $y \in R$ divided by the number of transitions from state s made by the process $\{X_n\}$ up to time j. Diaconis and Freedman (1980) showed that, when $\{X_n\}$ is recurrent, $M^{(j)}$ converges almost surely in the topology of coordinate-wise convergence to a stochastic matrix M whose probability distribution is the measure μ appearing in (4.11). However, with probability one,

$$\lim_{j\to\infty} M^{(j)}(x, q(x, c)) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I[Y_i = c]$$

for all $c \in E$, whereas $\lim_{j\to\infty} M^{(j)}(x, y) = 0$ for all $y \in S$ such that $q(x, c) \neq y$ for all $c \in E$. This fact and (4.13) show that $M(x)$ is a Dirichlet process on S with parameter $\alpha(x)$. Notice that if $x_1, \ldots, x_m$ are different elements of $\mathcal{R}$ such that $\phi(x_i) = 1$ for $i = 1, \ldots, m$, then $M(x_1), \ldots, M(x_m)$ are independent since the sequences of colors produced by urns $U(x_1), \ldots, U(x_m)$ are independent. $\diamond$

The next lemma shows that every state in R is visited infinitely often with probability one by the process $\{X_n\}$ when it is recurrent.

4.14 **Lemma**. *Assume that $\{X_n\}$ is recurrent. Then, for every $x \in R$, $\phi(x) = 1$.*

Proof. We prove by induction on n that for all $n \geq 0$ and $x \in R^n$, $\phi(x) = 1$. In fact $\phi(x_0) = 1$ by assumption. Assume that $\phi(x) = 1$ for every state in R^n and let $x \in R^{n+1}$. Then there is a $y \in R^n$ such that $n_y(c(y, x)) > 0$. Since $\phi(y) = 1$, it follows from Lemma 4.12 that

$M(y)$ is a Dirichlet process with parameter $\alpha(y)$. Hence there is a measurable subset $\mathcal{A}_y$ of $\mathcal{M}$ such that $\mu(\mathcal{A}_y) = 1$ and for all $m \in \mathcal{A}_y$,

$$P[X_n = y \text{ for infinitely many } n|M = m] = 1 \tag{4.15}$$

and

$$m(y, x) > 0. \tag{4.16}$$

Equation (4.15) says that y is a recurrent state for a Markov chain on R with transition matrix $m \in \mathcal{A}_x$. Moreover, since the class of recurrent states of a Markov chain is closed, (4.16) implies that x is also recurrent for the same Markov chain; that is

$$P[X_n = x \text{ for infinitely many } n|M = m] = 1$$

if $m \in \mathcal{A}_x$. Therefore

$$\phi(x) = \int_{\mathcal{A}_x} P[X_n = x \text{ for infinitely many } n|M = m]\mu(dm) = 1.$$

◇

Lemma 4.12 and Lemma 4.14 immediately imply the following theorem which characterizes the measure μ appearing in (4.11).

4.17 **Theorem.** *If $\{X_n\}$ is recurrent, the rows of M are mutually independent random probability distributions on R and, for all $x \in R$, the law of $M(x)$ is that of a Dirichlet process with parameter $\alpha(x)$.*

Let $(s_0, \ldots, s_n)$ be an admissible sequence. Given that we observed $X_0 = s_0, \ldots, X_n = s_n$, for all $x \in S$ the urn $U(x)$ contains

$$n_x(c) + l_x(c) \tag{4.18}$$

balls of color c, for each $c \in E$; set $\tilde{U}(x)$ to be an urn whose composition is that described in (4.18). The following result is almost obvious but has important implications when one uses the process $\{X_n\}$ for predictive purposes.

4.19 **Theorem.** *Assume that $\{X_n\}$ is recurrent. Let $(s_0, \ldots, s_n)$ be an admissible sequence such that $P[X_0 = s_0, \ldots, X_n = s_n] > 0$. Then, given $X_0 = s_0, \ldots, X_n = s_n$, the process $(X_n, X_{n+1}, \ldots) \in RUP(S, E, \tilde{U}, q)$ with initial state s_n and is recurrent.*

The probability distribution μ has a nice 'conjugacy' property, as follows from Theorem 4.17 and the previous result.

4.20 **Corollary**. *Assume that $\{X_n\}$ is recurrent. Let $(s_0, \ldots, s_n)$ be an admissible sequence such that $P[X_0 = s_0, \ldots, X_n = s_n] > 0$. Then, given $X_0 = s_0, \ldots, X_n = s_n$, the rows of M are mutually independent random probability distributions and, for all $x \in R$, the law of $M(x)$ is that of a Dirichlet process on R with parameter*

$$\alpha(x) + \sum_{c \in E} l_x(c)\delta(q(x,c))$$

where $\delta(y)$ is the measure which gives mass 1 to y and 0 to all other elements of R, for each $y \in R$.

Following Diaconis and Freedman (1980), we define an *y_0-block* for process $\{Y_n\}$ defined on S to be a finite sequence of states which begins by y_0 and contains no further y_0. Endow the countable space S^* of all finite sequences of elements of S with the discrete topology. When $\{Y_n\}$ is recurrent, let $B_1 \in S^*, B_2 \in S^*, \ldots$, be the sequence of the successive y_0-blocks in $\{Y_n\}$. The next result was also proved by Diaconis and Freedman (1980) and we state it without proof.

4.21 **Theorem**. *If $\{Y_n\}$ is recurrent and partially exchangeable with initial state y_0, the sequence $\{B_n\}$ of its y_0-blocks is exchangeable.*

Hence the sequence $\{B_n\}$ of x_0-blocks of a recurrent $\{X_n\} \in RUP(S, E, U, q)$ is exchangeable. This implies that if ψ is a function which maps measurably S^* into another space, the sequence $\{\psi(B_n)\}$ is also exchangeable. For example: for all $p \in S^*$, let $\lambda(p)$ be the length of p and $\zeta(p)$ be the last coordinate of p. Then $\{\lambda(B_n)\}$ and $\{\zeta(B_n)\}$ are exchangeable.

It's often the case that the de Finetti measure of the exchangeable sequence $\{\psi(B_n)\}$ is simply characterized by the properties of the recurrent urn process which generates it. Important examples are, for instance, beta-Stacy priors and Polya tree priors which we will introduce in the following sections as de Finetti measures of specific exchangeable sequences embedded in a reinforced urn process. As for the Dirichlet process generated by means of Blackwell and McQueen's Polya urn scheme, the advantage of these constructions based on RUPs consists in that many characteristics of these priors become then intuitively clear: for instance, Corollary 4.20 easily implies that they are conjugate priors, a nice and usefull property for Bayesian analysis.

5 Reinforced urn processes for the beta-Stacy prior

A beta-Stacy process on $S = \{0, 1, 2, \ldots\}$ is a random distribution function constructed on S such that, for all $k \geq 0$, the random mass assigned to $\{0, ..., k\}$ is given by

$$1 - \prod_{j=0}^{k}(1 - W_j)$$

where the random variables of the sequence $\{W_j\}$ are independent with Beta distribution. The beta-Stacy process has been introduced by Walker and Muliere (1997) and finds its applications in nonparametric survival studies. In order to construct it by means of a RUP consider the set of colors $E = \{w, b\}$ where w is for 'white' and b for 'black'. Let $x_0 = 0$ and assume that $n_0(w) = 0$ whereas, for all $j \geq 1$, $n_j(w) > 0$. Finally define the law of motion q for $\{X_n\}$ by setting

$$q(x, b) = x + 1 \text{ and } q(x, w) = 0$$

for all $x \in S$. Let $\{X_n\} \in RUP(S, E, U, q)$ with initial state $x_0 = 0$.

Therefore, the process $\{X_n\}$ begins at 0 and, given that at stage $n \geq 0$ it is at state $x \in S$, it moves to state $x + 1$ if the ball sampled from urn $U(x)$ is black and to state 0 if it is white. Urn compositions are updated in the usual Polya's way. If $n_j(b) = 0$ for some $j \geq 1$, let $N = \min\{j \geq 1 : n_j(b) = 0\}$ and note that, with probability one, the process $\{X_n\}$ visits only the states $\{0, \ldots, N\}$.

By applying (4.3) we get for all admissible finite sequences $(x_0, \ldots, x_n)$ of elements of S

$$P[X_0 = x_0, \ldots, X_n = x_n] = \prod_{j=1}^{\infty} \frac{B(n_j(w) + t(j, 0), n_j(b) + t(j, j+1))}{B(n_j(w), n_j(b))}$$

where, for all $a, b > 0$, $B(a, b)$ is the usual beta integral.

For all $n \geq 1$, let $T_n = X_{\tau_n - 1}$ where the sequence of stopping times $\{\tau_n\}$ is defined according to (4.7); that is $\tau_0 = 0$ and

$$\tau_n = \inf\{j > \tau_{n-1} : X_j = 0\}$$

for $n \geq 1$. When $\{X_n\}$ is recurrent, T_n is the last state of the nth 0-block; that is

$$T_n = \zeta(B_n).$$

Equivalently, T_n measures the length of the sequence of states strictly between the n-th zero and the $(n + 1)$th zero in the sequence $\{X_n\}$.

5.1 **Lemma**. *The process $\{X_n\}$ is recurrent if and only if*

$$\lim_{n\to\infty} \prod_{i=0}^{n} \frac{n_i(b)}{n_i(b) + n_i(w)} = 0. \tag{5.2}$$

Proof. For all $n \geq 1$,

$$P[\tau_1 > n] = \prod_{i=0}^{n-1} \frac{n_i(b)}{n_i(b) + n_i(w)}.$$

Hence, $P[\tau_1 < \infty] = 1$ if (5.2) holds.

By induction on n, assume that for an $n \geq 1$, $P[\bigcap_{i=1}^{n}\{\tau_i < \infty\}] = 1$. Then

$$\begin{aligned} P[\tau_{n+1} < \infty] &= \int_{\bigcap_{i=1}^{n}\{\tau_i < \infty\}} P[\tau_{n+1} < \infty | T_1, \ldots, T_n] dP \\ &= \int_{\bigcap_{i=1}^{n}\{\tau_i < \infty\}} (1 - \lim_{k \to \infty} P[\tau_{n+1} > k | T_1, \ldots, T_n]) dP. \end{aligned}$$

However, for $k > max(T_1, \ldots, T_n) + 1 = l$

$$P[\tau_{n+1} > k | T_1, \ldots, T_n] \leq \prod_{j=l}^{k-1} \frac{n_j(b)}{n_j(b) + n_j(w)}$$

and the last quantity goes to 0 for $k \to \infty$ because of (5.2). Hence $P[\tau_{n+1} < \infty] = 1$ and this proves that (5.2) is a sufficient condition for the recurrence of $\{X_n\}$.

In order to prove that the condition is also necessary, notice that $P[\tau_1 < \infty] = 1$ if $\{X_n\}$ is recurrent. Therefore

$$\lim_{n \to \infty} \prod_{i=0}^{n} \frac{n_i(b)}{n_i(b) + n_i(w)} = \lim_{n \to \infty} P[\tau_1 > n + 1] = 0.$$

◇

5.3 **Corollary.** *If*

$$\sum_{i=0}^{\infty} \frac{n_i(w)}{n_i(b) + n_i(w)} = \infty \tag{5.4}$$

the process $\{X_n\}$ *is recurrent.*

Proof. Use the inequality $1 - x \leq \exp^{-x}$ for $x \in [0, 1]$ to show that, for all $n \geq 0$,

$$\prod_{i=0}^{n} \frac{n_i(b)}{n_i(b) + n_i(w)} \leq \exp\left[-\sum_{i=0}^{n} \frac{n_i(w)}{n_i(b) + n_i(w)}\right].$$

Hence

$$\lim_{n \to \infty} \prod_{i=0}^{n} \frac{n_i(b)}{n_i(b) + n_i(w)} = 0$$

if

$$\sum_{i=0}^{\infty} \frac{n_i(w)}{n_i(b) + n_i(w)} = \infty.$$

◇

Therefore, whenever (5.2) holds, the sequence $\{T_n\}$ is exchangeable and, by de Finetti's Representation Theorem, there exists a random distribution function F such that, given F, the random variables of the sequence $\{T_n\}$ are i.i.d. with distribution F. With the next theorem we prove that F is a beta-Stacy process on S with parameters $\{(n_j(w), n_j(b))\}$; that is we show that F is a random distribution function on S such that $F(0) = 0$ with probability one and, for $j \geq 1$, $[F(j) - F(j-1)]$ has the same distribution as

$$P_j = W_j \prod_{i=1}^{j-1}(1 - W_i) \tag{5.5}$$

where $\{W_j\}$ is a sequence of independent random variables such that, for all $j \geq 1$, W_j has distribution $\mathrm{Beta}(n_j(w), n_j(b))$. Note that in (5.5) we assume $\prod_{i=1}^{0} = 1$ and that $\mathrm{Beta}(a, 0)$ is the point mass at 1 for all $a > 0$.

5.6 **Theorem**. *F is a beta-Stacy process on S with parameter*

$$\{n_j(w), n_j(b)\}.$$

Proof. Fix $n \geq 1$ and integers $0 = k_0 < k_1 \leq \ldots \leq k_n$. Then,

$$\begin{aligned}
&P[T_1 = k_1, \ldots, T_n = k_n] \\
&\quad = P[X_1 = 1, \ldots, X_{k_1} = k_1, X_{k_1+1} = 0, \ldots \\
&\quad \ldots X_{\sum_{j=1}^{n-1} k_j + n - 1} = 0, \ldots, X_{\sum_{j=1}^{n} k_j + n - 1} = k_n, X_{\sum_{j=1}^{n} k_j + n} = 0] \\
&\quad = \prod_{j=1}^{\infty} \frac{B(n_j(w) + t(j,0), n_j(b) + t(j, j+1))}{B(n_j(w), n_j(b))} \\
&\quad = E\left[\prod_{j=1}^{\infty} W_j^{t(j,0)}(1 - W_j)^{t(j,j+1)}\right]
\end{aligned}$$

where $\{W_j\}$ is a sequence of independent random variables such that, for all $j \geq 1$, W_j has distribution $\mathrm{Beta}(n_j(w), n_j(b))$. However

$$t(j, 0) = \begin{cases} \sum_{i=1}^{n} I[k_i = j] & \text{if } j \in \{k_1, \ldots, k_n\} \\ 0 & \text{otherwise.} \end{cases}$$

Also, because the k_i's are ordered, for all $j \geq 1$,

$$t(j, j+1) = \sum_{i=0}^{n-1} (n - i) I[k_i \leq j \leq k_{i+1} - 1]$$

assuming that the indicator of an empty set is 0. Therefore

$$P[T_1 = k_1, \ldots, T_n = k_n]$$

$$= E\left[\prod_{j=1}^{n} W_{k_j} \prod_{j=0}^{k_1-1}(1-W_j)^n \prod_{j=k_1}^{k_2-1}(1-W_j)^{n-1} \cdots \prod_{j=k_{n-1}}^{k_n-1}(1-W_j)\right]$$

$$= E\left[W_{k_1}\prod_{j=0}^{k_1-1}(1-W_j)\cdots W_{k_n}\prod_{j=0}^{k_n-1}(1-W_j)\right]$$

$$= E\left[\prod_{j=1}^{n} P_{k_j}\right]$$

where, for all $k \geq 1$, we defined

$$P_k = W_k \prod_{i=1}^{k-1}(1-W_i)$$

and we used the convention that $\prod_j^{j-1} = 1$. Since $\{T_n\}$ is exchangeable, this shows that, for all $n \geq 1$ and all sequences $(k_1, \ldots, k_n)$ of strictly positive elements of S,

$$P[T_1 = k_1, \ldots, T_n = k_n] = E\left[\prod_{j=1}^{n} P_{k_j}\right]. \tag{5.7}$$

If one or more elements of $(k_1, \ldots, k_n)$ are 0, then trivially

$$P[T_1 = k_1, \ldots, T_n = k_n] = 0. \tag{5.8}$$

Equations (5.7) and (5.8) are sufficient to prove that the de Finetti measure of the sequence $\{T_n\}$ is that for the beta-Stacy process on S with parameters $\{(n_j(w), n_j(b))\}$. Note that F is a random distribution function since we assumed the recurrence condition (5.2) holds. $\diamond$

As a final remark before concluding the section, note that, for all $k \geq 1$,

$$P[T_1 = k] = \frac{n_k(w)}{n_k(b) + n_k(w)} \prod_{j=0}^{k-1} \frac{n_j(b)}{n_j(b) + n_j(w)}.$$

Also, for all $n \geq 1$ and $k \geq 1$,

$$P[T_{n+1} = k | T_1, T_2, \cdots, T_n] = \frac{n_k(w) + n_k}{n_k(b) + n_k(w) + n_k + m_k} \prod_{j=1}^{k-1} \frac{n_j(b) + m_j}{n_j(b) + n_j(w) + n_j + m_j} \tag{5.9}$$

on the set $\bigcap_{i=1}^{\infty}\{\tau_i < \infty\}$ if, for all $0 \leq j \leq k-1$, we define $n_j = \sum_{i=1}^{n} I[T_i = j]$ and $m_j = \sum_{i=1}^{n} I[T_i > j]$. Therefore, if (5.2) holds, (5.9) is true with probability one.

6 Reinforced urn processes for Polya trees

By specializing the state space S and the law of motion of a reinforced urn process we may describe an urn scheme for constructing Polya trees: this will be the aim of this section.

Let $E = \{0, ..., k\}$ be a finite, nonempty set of colors and define S to be the set of all finite sequences of elements of E including the empty sequence $\emptyset$. Let U be an urn composition function defined on S.

Given $s \geq 1$, we define the law of motion q_s on $S \times E$ by setting, for all $n \geq 1, c_1, ..., c_n, c \in E$,

$$q_s(\emptyset, c) = (c)$$

and

$$q_s((c_1, ..., c_n), c) = \begin{cases} (c_1, ..., c_n, c) & \text{if } 1 \leq n < s \\ \emptyset & \text{if } n \geq s \end{cases}$$

We now consider $\{X_n^{(s)}\} \in RUP(S, E, U, q_s)$ with initial state $X_0^{(s)} = \emptyset$. Therefore, for instance for $s = 3$, $X_1^{(s)} = (c_1)$ where c_1 is the color first extracted from urn $\emptyset$, $X_2^{(s)} = (c_1, c_2)$ is the state of S obtained by concatenating to (c_1) the color c_2 first extracted from urn (c_1), $X_3^{(s)} = (c_1, c_2, c_3)$ is the state of S obtained by concatenating to (c_1, c_2) the color c_3 first extracted from urn (c_1, c_2). The next state is $X_4^{(s)} = X_{s+1}^{(s)} = \emptyset$ whatever the color extracted from urn (c_1, c_2, c_3). A this point we start again a sequence of $s = 3$ successive extractions, the first from urn $\emptyset$ and then from the urns dictated by the law of motion q_s and corresponding to states of S of length respectively 1 and $2 = s - 1$ whose composition has been previously updated according to the Polya's rule. And so on.

Trivially $\{X_n^{(s)}\}$ is recurrent. Therefore the sequence of $\emptyset$-blocks of $\{X_n^{(s)}\}$ is exchangeable as well as the sequence $\{X_{n(s+1)-1}^{(s)}\}$, for $n \geq 1$, of the last elements of the $\emptyset$-blocks. This last sequence has been defined *s-stage Polya with parameter U* by Mauldin, Sudderth and Williams (1992) who proved its exchangeability directly without appealing to the partial exchangeability and recurrence of the sequence $\{X_n^{(s)}\}$ in which it is embedded.

Let now $R^{(s)}$ be the set of states of S visited by the RUP process $\{X_n^{(s)}\}$ with probability one and defined in (4.10). Then, as we showed in section 4, there is a stochastic matrix $M^{(s)}$ on $R^{(s)} \times R^{(s)}$ such that, for all $n \geq 1$ and $x_1, ..., x_n \in R^{(s)}$,

$$P[X_1 = x_1, ..., X_n = x_n] = E[\prod_{j=1}^{n-1} M^{(s)}(x_j, x_{j+1})].$$

Theorem 4.17 states that, for $x = (c_1, ..., c_k) \in R^{(s)}$ and $U(x) = (n_x(0), ..., n_x(k))$, the x-th row $M^{(s)}(x)$ of $M^{(s)}$ is a Dirichlet process on $R^{(s)}$ with parameter $\alpha(x)$, a measure which assigns mass $n_x(c)$ to $q(x, c)$ for all $c \in E$ and mass 0 to all other elements of $R^{(s)}$. Furthermore, for $x, y \in R^{(s)}, x \neq y$, $M^{(s)}(x)$ is independent of $M^{(s)}(y)$. Moreover, it is crucial to observe that, for all $s \geq 1$ and $x \in R^{(s)}$, the random vectors $M^{(s)}(x), M^{(s+1)}(x), ...$ have all the same distribution.

For ease of notation, we now define a collection of random variables indexed by elements of

S as follows: set $\Theta_\emptyset = 1$ and, for all $n \geq 1$, $x = (c_1, ..., c_n) \in S$ and $c \in E$,

$$\Theta_x(c) = \begin{cases} M^{(n+1)}(x, (c_1, ..., c_n, c)) & \text{if } x \in R^{(n+1)}, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that, if $s \geq n+1$ and $x = (x_1, ..., x_n) \in R^{(s)}$, the distribution of the random vector

$$\Theta_x = (\Theta_x(0), ..., \Theta_x(k)) \tag{6.1}$$

is Dirichlet with parameter $U(x)$; moreover if $y \in S$ and $y \neq x$ the random vector Θ_y is independent of Θ_x.

The collection of random variables Θ's can be associated in a natural way to the elements of a tree of partitions of the space $E^\infty = E \times E \times E....$

Assign to E the sigma-field consisting of all its subsets and, for $n = 1, 2, ...$, set π_n to be the coordinate function defined for $e = (c_1, c_2, ...,) \in E^\infty$ by

$$\pi_n(e) = c_n.$$

Finally endow E^∞ with the least sigma-field $\mathcal{F}$ with respect to which all coordinate functions are measurable. We now build a tree of finite measurable partitions of E^∞ by letting $\nabla_0 = \{E^\infty\}$ and, for $n \geq 1$, ∇_n to be the finite measurable partition of E^∞ generated by the coordinate functions $\pi_1, ..., \pi_n$; that is, ∇_n is the collection of sets

$$B = \{e \in E^\infty : \pi(e) = c_1, ..., \pi_n(e) = c_n\}$$

for $c_1, ..., c_n \in E$. Note that for all $n \geq 1$, ∇_n has $(k+1)^n$ elements and is finer than ∇_{n-1}; furthermore the sigma-field generated by $\bigcup_n \nabla_n$ is $\mathcal{F}$.

For $1 \leq k \leq n$, $c_1, ..., c_n \in E$ and $B = \{e \in E^\infty : \pi(e) = c_1, ..., \pi_n(e) = c_n\}$, let us call *predecessor of B at level k* the set

$$ps(B, k) = \{e \in E^\infty : \pi_1(e) = c_1, ..., \pi_k(e) = c_k\} \in \nabla_k.$$

Then, for all $n \geq 0$ and $B \in \nabla_n$ there are exactly $k+1$ sets $C_0, ..., C_k \in \nabla_{n+1}$ such that $ps(C_i, n) = B$ for $i = 0, ..., k$. A tree of partition $\{\nabla_n\}$ with such a property is said to be *isodivided*.

As a last notation, for $A \in \mathcal{F}$ and $n \geq 1$, we set

$$\mathcal{E}_n(A) = \{B \in \nabla_n : A \bigcap B \neq \emptyset\}.$$

We now introduce a sequence of random variables indexed by the elements of the tree of partitions $\{\nabla_n\}$ by setting $V_{0,E^\infty} = 1$ and, for $n \geq 1$, $c_1, ..., c_n \in E$ and $B = \{e \in E^\infty : \pi(e) = c_1, ..., \pi_n(e) = c_n\} \in \nabla_n$,

$$V_{n,B} = \Theta_{(c_1,...,c_{n-1})}(c_n)$$

where $(c_1, ..., c_{n-1}) = \emptyset$ for $n = 1$.

6.2 **Theorem.** *The sequence of random variables $\{V_{n,B} : n \geq 0, B \in \nabla_n\}$ induces a probability distribution $\mathcal{T}$ on the space $\mathbf{P} = \mathbf{P}(E^\infty)$ of all probability distributions defined on $(E^\infty, \mathcal{F})$ endowed with the sigma-field $\mathcal{P}$ generated by the topology of weak convergence. $\mathcal{T}$ is tailfree with respect to the tree of partition $\{\nabla_n\}$ and is a Polya tree.*

Proof. In order to prove that the sequence $\{V_{n,B} : n \geq 0, B \in \nabla_n\}$ induces a tailfree process on $(\mathbf{P}, \mathcal{P})$, it is enough to check the consistency conditions $(T_1) - (T_5)$ of Cifarelli, Muliere and Secchi (1999).

Conditions $(T_1) - (T_3)$ and (T_5) are easily verified and follow from the properties of the random variables Θ's by means of which the $V_{n,B}$'s have been defined. In fact, for $n \geq 0$ and $B \in \nabla_n$, $V_{n,B} \in [0, 1]$ whereas $V_{0,E^\infty} = 1$ by definition: these are precisely conditions (T_1) and (T_2). Condition (T_3) is satisfied if, for $n \geq 0$ and $B \in \nabla_n$,

$$\sum_{C \in \mathcal{E}_{n+1}(B)} V_{n+1,C} = 1. \tag{6.3}$$

In fact, either $n = 0$ or $B = \{e \in E^\infty : \pi_1(e) = c_1, ..., \pi_n(e) = c_n\}$ with $c_1, ..., c_n \in E$. In both cases the random vector whose components are the variables $V_{n+1,C}$ appearing on the left side of (6.3) has Dirichlet distribution; hence (6.3) is true. The fact that the collections of variables

$$\{V_{1,B} : B \in \nabla_1\}, \{V_{2,B} : B \in \nabla_2\}, ...$$

are independent follows from independence of the random vectors Θ_x, for $x \in S$: this proves (T_5).

A little more care is needed for proving condition (T_4) which holds if, for all sequences $\{B_n\}$ decreasing to the empty set and such that, for each $n \geq 0$, B_n is union of sets in ∇_n,

$$\lim_{n \to \infty} \sum_{B \in \mathcal{E}_n(B_n)} \prod_{i=1}^{n} E[V_{i,ps(B,i)}] = 0.$$

For this purpose, we define on E the probability measure β_1 such that

$$\beta_1(\{c\}) = \frac{n_\emptyset(c)}{\sum_{j=0}^{k} n_\emptyset(j)}$$

for $c \in E$ whereas, for $n \geq 2$, we define iteratively on E^n the probability measure β_n such that

$$\beta_n(\{(c_1, ..., c_n)\}) = \beta_{n-1}(\{(c_1, ..., c_{n-1})\}) \frac{n_{(c_1,...,c_{n-1})}(c_n)}{\sum_{j=0}^{k} n_{(c_1,...,c_{n-1})}(j)}$$

for $c_1, ..., c_n \in E$. From Kolmogorov's extension theorem it follows that the sequence of probability measures $\{\beta_n\}$ uniquely determines a probability β on $(E^\infty, \mathcal{F})$. Notice that, for $n \geq$

$1, c_1, ..., c_n \in E$ and $B = \{e \in E^\infty : \pi_1(e) = c_1, ..., \pi_n(e) = c_n\} \in \nabla_n$,

$$\begin{aligned}\prod_{i=1}^{n} E[V_{i,ps(B,i)}] &= E[\Theta_\emptyset((c_1))\Theta_{(c_1)}((c_1, c_2)) \cdots \Theta_{(c_1,...,c_n)}((c_1, ..., c_n))] \\ &= \beta_n(\{(c_1, ..., c_n)\}) \\ &= \beta(B).\end{aligned}$$

Therefore

$$\lim_{n\to\infty} \sum_{B\in\mathcal{E}_n(B_n)} \prod_{i=1}^{n} E[V_{i,ps(B,i)}] = \lim_{n\to\infty} \sum_{B\in\mathcal{E}_n(B_n)} \beta(B) = \lim_{n\to\infty} \beta(B_n) = 0.$$

This shows that (T_4) also holds for the collection of random variables $\{V_{n,B} : n \geq 0, B \in \nabla_n\}$: hence this collection induces a probability distribution $\mathcal{T}$ on $(\mathbf{P}, \mathcal{P})$ which is tailfree with respect to $\{\nabla_n\}$.

Finally and since the tree of partitions $\{\nabla_n\}$ constructed above is isodivided, in order to show that $\mathcal{T}$ is in fact a Polya tree probability distribution on $(\mathbf{P}, \mathcal{P})$, it is enough to recall that, for $n \geq 1$, $B \in \nabla_n$ and $\mathcal{E}_{n+1}(B) = \{C_0, ..., C_k\}$, the random vector $(V_{n+1,C_0}, ..., V_{n+1,C_k})$ has Dirichlet distribution as follows from (6.1). ◇

Following Mauldin, Sudderth and Williams (1992), it is now worth noticing that by mapping measurably the space E^∞ into another measurable space $(\mathbf{X}, \mathcal{X})$ we can carry over the Polya tree probability distribution $\mathcal{T}$ introduced on $\mathbf{P}(E^\infty)$ by means of recurrent urn processes to the space $\mathbf{P}(\mathbf{X})$. For instance, let $\mathbf{X} = I$ be the unit interval endowed with its Borel sigma-field and set, for all $e = (c_1, c_2, ...) \in E^\infty$,

$$\psi(e) = \sum_{n=1}^{\infty} \frac{c_n}{(k+1)^n}.$$

The mapping ψ is continuous and induces a continuous mapping from $\mathbf{P}(E^\infty)$ to $\mathbf{P}(I)$ which associates to any probability distribution $p \in \mathbf{P}(E^\infty)$ the probability distribution $p\psi^{-1} \in \mathbf{P}(I)$ induced by the random variable ψ defined on $(E^\infty, \mathcal{F}, p)$; more precisely, for all Borel subset B of I, set

$$p\psi^{-1}(B) = p(\psi^{-1}(B)).$$

Now, if P is a random element of $\mathbf{P}(E^\infty)$ with Polya tree distribution $\mathcal{T}$, then $P\psi^{-1}$ is a random element of $\mathbf{P}(I)$ with Polya tree distribution with respect to the tree of finite measurable partitions of I obtained by mapping the elements of $\{\nabla_n\}$ with ψ.

7 Neutral to the right priors

The aim of the last two sections has been to show that reinforced urn processes are adaptable enough to generate two classes of random probability distributions widely used in Bayesian nonparametrics. The fact that beta-Stacy and Polya trees priors can be constructed by means of RUPs confirms the versatility of the latter class of partially exchangeable processes. There is, however, a second reason for going through the details of the previous constructions; and this is that it is now easy to extend them with the purpose of generating the more general classes of neutral to the right and of tailfree priors insofar as in the definition of a RUP we substitute Polya urns with more general random generators of infinite exchangeable sequences of colors. Hence urns will disappear from our following constructions, but their heritage will be manifest through the inspiring examples of the previous two sections.

In this section we show how to construct neutral to the right priors on a countable state space S which, for simplicity, will be taken to be $\{0, 1, 2, ...\}$. Recall that a random probability distribution F on S is said to be neutral to the right if, for all $k \geq 1$, the random mass F_k assigned by F to the set $\{0, ..., k\}$ is given by

$$F_k = 1 - \prod_{j=0}^{k}(1 - V_j)$$

where $V_0, V_1, V_2, ...$ is a sequence of independent random variables with values in (0,1) and such that

$$\lim_{n\to\infty} \prod_{j=0}^{n} E[1 - V_j] = 0. \tag{7.1}$$

The beta-Stacy process of Section 5 is obtained when $V_0, V_1, V_2, ...$ have Beta distribution.

Our aim is to generate an infinite exchangeable sequence $\{T_n\}$ of random variables with values in S whose de Finetti measure is F by constructing a partially exchangeable process $\{X_n\}$ on S in which $\{T_n\}$ is embedded. As in Section 5, we introduce a set $E = \{w, b\}$ of two colors, white and black. Set $\{Y_n(0)\}$ to be an infinite sequence of black colors; for all $j \geq 1$, let $\{Y_n(j)\}$ be an infinite exchangeable sequence of colors of E whose de Finetti measure has density $(V_j, 1 - V_j)$. Define a law of motion q on S by setting

$$q(x, b) = x + 1 \text{ and } q(x, w) = 0$$

for all $x \in S$.

Set $X_0 = 0$ and, for all $n \geq 1$, if $X_0 = x_0, \ldots, X_{n-1} = x_{n-1}$, let

$$X_n = q(x_{n-1}, Y_{t(x_{n-1})+1}(x_{n-1})).$$

where, as in Section 4, $t(x_{n-1})$ counts the number of appearances of state z_{n-1} made in the finite sequence $(0, \ldots, z_{n-1})$. Hence the process $\{X_n\}$ starts in 0 and moves to state $X_1 = 1$. The next state X_2 is 0 or 2 according to the first color generated by the exchangeable sequence $\{Y_n(1)\}$; if $X_2 = 0$, then $X_3 = 1$ and X_4 is 0 or 2 according to the second color generated by

$\{Y_n(1)\}$. If $X_2 = 2$ then X_3 is 0 or 4 according to the first color generated by $\{Y_n(2)\}$. And so on.

For all $j \geq 1$ and $n, m \geq 0$, let us define

$$\lambda_j(n,m) = E[V_j^n(1-V_j)^m].$$

Then it is not difficult to check that, for all admissible sequences $(0, \ldots, x_n)$ of elements of S,

$$P[X_0 = 0, \ldots, X_n = x_n] = \prod_{j=1}^{\infty} \lambda_j(t(j,0), t(j,j+1)) \tag{7.2}$$

where, as in Section 4, for all $j \in S$, $t(j,0)$ counts the number of transitions from state j to state 0 and $t(j,j+1)$ counts the number of transitions from j to the following state $j+1$ in the sequence $0, \ldots, x_n$. This shows that the process $\{X_n\}$ is partially exchangeable. With an argument analogous to that of Lemma 5.1 one may prove that $\{X_n\}$ is recurrent when (7.1) holds; in this case $\{X_n\}$ is a mixture of Markov chains.

For all $n \geq 1$, let $T_n \in S$ be the last coordinate of the nth 0-block of the process $\{X_n\}$ when it is recurrent. Then $\{T_n\}$ is exchangeable. The next theorem describes the de Finetti measure of the sequence $\{T_n\}$; it can also be considered as a characterization of a neutral to the right process via a mixture of Markov chains.

7.3 **Theorem.** *The de Finetti measure for the exchangeable sequence $\{T_n\}$ is the neutral to the right prior F.*

Proof. As in the proof of Theorem 5.6, we aim to show that, for all $n \geq 1$ and all finite sequences $(k_1, \ldots, k_n)$ of strictly positive elements of S,

$$P[T_1 = k_1, \ldots, T_n = k_n] = E\left[\prod_{j=1}^{n} P_{k_j}\right],$$

where, for all $j \geq 1$, P_j is the random variable allocating the random mass to $j \in S$ defined by $P_j = F_j - F_{j-1} = V_j \prod_{i=1}^{j-1}(1-V_i)$.

Equation (7.2) implies that

$$P[T_1 = k_1, \ldots, T_n = k_n] = \prod_{j=1}^{\infty} \lambda(t(j,0), t(j,j+1)) \tag{7.4}$$

where, for all $x, y \in S$, $t(x,y)$ counts the transitions from x to y made in the admissible sequence

$$(0, 1, \ldots, k_1, 0, 1, \ldots, k_2, \ldots, 0, 1, \ldots, k_n, 0).$$

However,

$$\prod_{j=1}^{\infty} \lambda_j(t(j,0), t(j,j+1)) = \prod_{j=1}^{\infty} E\left[V_j^{t(j,0)}(1-V_j)^{t(j,j+1)}\right].$$

From here the proof proceeds analogously to that of Theorem 5.6. ◇

8 Tailfree priors

Having seen how to construct neutral to the right priors on countable spaces by means of a partially exchangeable scheme which is a generalization of a RUP, we now turn to the construction of random distributions which are tailfree with respect to a isodivided tree of partitions along the lines followed in Section 6 for constructing Polya trees.

Let $E = \{0, ..., k\}$ be a finite, nonempty set of colors and let S be the set of all finite sequences of elements of E including the empty sequence $\emptyset$. For all $x \in S$, let $\{Y_n(x)\}$ be an infinite exchangeable sequence of elements of E and denote with Θ_x the density of the de Finetti measure of the sequence: hence Θ_x is a random vector of $[0,1]^{k+1}$ with components which sum to 1. We assume that sequences of Y's corresponding to different elements of S are independent.

Finally, consider again the same law of motion that we used for constructing Polya trees in Section 6; therefore, given $s \geq 1$, let q_s be defined on $S \times E$ by setting, for all $n \geq 1, c_1, ..., c_n, c \in E$,

$$q_s(\emptyset, c) = (c)$$

and

$$q_s((c_1, ..., c_n), c) = \begin{cases} (c_1, ..., c_n, c) & \text{if } 1 \leq n < s, \\ \emptyset & \text{if } n \geq s. \end{cases}$$

We are now ready for the recursive definition of a partially exchangeable process $\{X_n^{(s)}\}$ on the space S. Set $X_0^{(s)} = \emptyset$. For $n \geq 1$, if $X_0^{(s)} = x_0, ..., X_{n-1}^{(s)} = x_{n-1}$, set

$$X_n^{(s)} = q_s(x_{n-1}, Y_{t(x_{n-1})+1}(x_{n-1}))$$

where $t(x_{n-1})$ counts the number of appearances of x_{n-1} within the sequence of states

$$x_0, ..., x_{n-1}.$$

Therefore $X_1^{(s)} = (c_1)$ where c_1 is the first color generated by the sequence $\{Y_n(\emptyset)\}$, that is the realization of $Y_1(\emptyset)$, $X_2^{(s)} = (c_1, c_2)$ where c_2 is the first color generated by $\{Y_n((c_1))\}$, and so on for the first s stages. The next state $X_{s+1}^{(s)} = \emptyset$, then $X_{s+2}^{(s)} = (c'_1)$ where c'_1 is the second color generated by $\{Y_n(\emptyset)\}$, $X_{s+3}^{(s)} = (c'_1, c'_2)$ where c'_2 is the first color generated by $\{Y_n((c'_1))\}$ if $c'_1 \neq c_1$ whereas c'_2 is the second color generated by $\{Y_n((c_1))\}$ when $c'_1 = c_1$. And so on.

The sequence $\{X_n^{(s)}\}$ is partially exchangeable; to prove this compute $P[X_0 = x_0, ..., X_n = x_n]$, for $n \geq 1$ and $x_0, x_1, ..., x_n \in S$, by first conditioning to the random variables $\Theta_{x_0}, ..., \Theta_{x_n}$ and notice that this probability depends on $x_0, ..., x_n$ only through the number of transitions $t(x_i, x_{i+1})$, for $i = 0, ..., n-1$. Moreover, $\{X_n^{(s)}\}$ is trivially recurrent and therefore is a mixture of Markov chains.

Let us now consider the space $E^\infty = E \times E \times E \times ....$ endowed with the least sigma-field $\mathcal{F}$ with respect to which all the coordinate functions π_n, for $n = 1, 2, ...$, are measurable. As before, $\{\nabla_n\}$ is the isodivided tree of measurable partitions such that $\nabla_0 = E^\infty$, whereas, for $n \geq 1$, ∇_n is the finite partition of E^∞ generated by the coordinate function π_n. To each set belonging to a partition of the tree $\{\nabla_n\}$ we associate a random variable by setting $V_{0,E^\infty} = 1$ and, for $n \geq 1, c_1, ..., c_n \in E$ and $B = \{e \in E^\infty : \pi_1(e) = c_1, ..., \pi_n(e) = c_n\} \in \nabla_n$,

$$V_{n,B} = \Theta_{(c_1,...,c_{n-1})}(c_n),$$

the c_nth coordinate of the vector $\Theta_{(c_1,...,c_{n-1})}$, where we assume $(c_1, ..., c_{n-1}) = \emptyset$ for $n = 1$.

Analogously to Theorem 6.2, the next result proves that the sequence of random variables constructed above induces a tailfree probability distribution on the space of all probability distribution defined on $(E^\infty, \mathcal{F})$; as we observed at the end of Section 5, by mapping measurably E^∞ into another measurable space $(\mathbf{X}, \mathcal{X})$ one can carry over this tailfree prior to the space of all probability distributions defined on $(\mathbf{X}, \mathcal{X})$.

8.1 **Theorem.** *The sequence of random variables $\{V_{n,B} : n \geq 0, B \in \nabla_n\}$ induces a probability distribution $\mathcal{T}$ on the space $\mathbf{P}$ of all probability distributions defined on $(E^\infty, \mathcal{F})$ endowed with the sigma-field $\mathcal{P}$ generated by the topology of weak convergence. $\mathcal{T}$ is tailfree with respect to the isodivided tree of partitions $\{\nabla_n\}$.*

Proof. As in the proof of Theorem 6.2, we need to check the consistency conditions $(T_1) - (T_5)$ of Cifarelli, Muliere and Secchi (1999). As before, this is easily done for conditions $(T_1), (T_2), (T_3)$ and (T_5). For proving condition (T_4) we define on the space E, endowed with the sigma-field of all its subsets, the probability measure β_1 such that

$$\beta_1(\{c\}) = E[\Theta_\emptyset(c)]$$

where $\Theta_\emptyset(c)$ is the cth coordinate of the vector $\Theta_\emptyset$. For $n \geq 2$ we define iteratively on the space E^n, endowed with the sigma-field of all its subsets, the probability measure β_n such that

$$\beta_n(\{(c_1, ..., c_n)\}) = \beta_{n-1}(\{(c_1, ..., c_{n-1})\})E[\Theta_{(c_1,...,c_{n-1})}(c_n)]$$

for $c_1, ..., c_n \in E$. The sequence of probability measures $\{\beta_n\}$ determines uniquely a probability β on $(E^\infty, \mathcal{F})$ such that $\beta(B) = \beta_n(B)$ for all $n \geq 0$ and $B \in \nabla_n$. Now follow the final lines of the proof of Theorem 6.2 to show that, for all sequences of sets $\{B_n\}$ decreasing to the empty set and such that, for each $n \geq 0$, B_n is union of sets in ∇_n,

$$\lim_{n\to\infty} \sum_{B \in \mathcal{E}_n(B_n)} \prod_{i=1}^{n} E[V_{i,ps(B,i)}] = \lim_{n\to\infty} \beta(B_n) = 0.$$

This shows that condition (T_4) holds and concludes the proof of the theorem. ◇

References

[1] BLACKWELL, D. and MACQUEEN, J.B. (1973). Ferguson distributions via Polya urn schemes. *Annals of Statistics*, **1**, 353–355.

[2] CIFARELLI, D.M., MULIERE, P. and SECCHI, P. (1999). Prior processes for Bayesian nonparametrics. *Technical Report n. 377/P*, Dipartimento di Matematica, Politecnico di Milano.

[3] DIACONIS, P. and FREEDMAN, D. (1980). De Finetti's Theorem for Markov chains. *Annals of Probability*, **8**, 115–130.

[4] EGGENBERGER, F. and POLYA, G. (1923). Über die Statistik verketteter Vorgänge. *Zeitschrift für Angewandte Mathemathik und Mechanik*, **3**, 279–289.

[5] FELLER, W. (1957). *An Introduction to Probability Theory and its Applications.* John Wiley & Sons, New York.

[6] FREEDMAN, D.A. (1965). Bernard Friedman's urn. *Annals of Mathematical Statistics*, **36**, 956–970.

[7] FRIEDMAN, B. (1949). A simple urn model. *Communications on Pure and Applied Mathematics*, **2**, 59–70.

[8] HILL, B.M., LANE, D. and SUDDERTH, W.D. (1987). Exchangeable urn processes. *Annals of Probability*, **15**, 1592–1596.

[9] MAULDIN, R.D., SUDDERTH, W.D. and WILLIAMS, S.C. (1992). Polya trees and random distributions. *Ann. Statist.*, **20**, 1203-1221.

[10] MULIERE, P., SECCHI, P. and Walker, S. (2000). Urn schemes and reinforced random walks. *Stochastic processes and their applications*, **88**, 58–79.

[11] WALKER, S. and MULIERE, P. (1997). Beta-Stacy processes and a generalization of the Polya urn scheme. *Annals of Statistics*, **25**, 1762–1780.