CrossMark

Adv. Appl. Probab. **54,** 849–879 (2022)
*doi:*10.1017/apr.2021.56

THE RESCALED PÓLYA URN: LOCAL REINFORCEMENT AND CHI-SQUARED GOODNESS-OF-FIT TEST

GIACOMO ALETTI,* *Università degli Studi di Milano*
IRENE CRIMALDI,** *IMT School for Advanced Studies Lucca*

Abstract

Motivated by recent studies of big samples, this work aims to construct a parametric model which is characterized by the following features: (i) a 'local' reinforcement, i.e. a reinforcement mechanism mainly based on the last observations, (ii) a random persistent fluctuation of the predictive mean, and (iii) a long-term almost sure convergence of the empirical mean to a deterministic limit, together with a chi-squared goodness-of-fit result for the limit probabilities. This triple purpose is achieved by the introduction of a new variant of the Eggenberger–Pólya urn, which we call the rescaled Pólya urn. We provide a complete asymptotic characterization of this model, pointing out that, for a certain choice of the parameters, it has properties different from the ones typically exhibited by the other urn models in the literature. Therefore, beyond the possible statistical application, this work could be interesting for those who are concerned with stochastic processes with reinforcement.

Keywords: Empirical mean; central limit theorem; chi-squared test; compact Markov chain; Pólya urn; predictive mean; preferential attachment; reinforcement learning; reinforced stochastic process; urn model

2020 Mathematics Subject Classification: Primary 60F05; 62F03
Secondary 60J05; 62F05

1. Introduction: framework and motivation

The well-known Pearson chi-squared test of goodness of fit is a statistical test applied to categorical data to establish whether an observed frequency distribution differs from a theoretical probability distribution. In this test the observations are always assumed to be independent and identically distributed (i.i.d.). Under this hypothesis, in a multinomial sample of size N, the chi-squared statistics

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = N \sum_{i=1}^{k} \frac{(\widehat{p}_i - p_i)^2}{p_i} \tag{1}$$

(where k is the number of possible values, and O_i, E_i, $\widehat{p}_i = O_i/N$, and $p_i = E_i/N$ are the observed and expected absolute and relative frequencies, respectively) is proportional to N,

Received 12 November 2020; revision received 25 September 2021.
* Postal address: ADAMSS Center, Università degli Studi di Milano, Milan, Italy.
** Postal address: IMT School for Advanced Studies Lucca, Lucca, Italy.

© The Author(s), 2022. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

which is multiplied by the chi-squared distance between the observed and expected probabilities. Therefore, the goodness-of-fit test based on this statistics is highly sensitive to the sample size N (see, for instance, [8, 36]): the larger N, the more significant a small value of the chi-squared distance. More precisely, the value of the chi-squared distance has to be compared with the 'critical' value $\chi^2_{1-\theta}(k-1)/N$, where $\chi^2_{1-\theta}(k-1)$ denotes the quantile of order $1-\theta$ of the chi-squared distribution $\chi^2(k-1)$ with $k-1$ degrees of freedom. Hence, it is clear that the larger N, the easier the rejection of H_0. As a consequence, in the context of 'big data' (e.g. [8, 11]), where one often works with correlated noised data, suitable generative models and related chi-squared goodness-of-fit tests are needed.

Different types of correlation have been taken into account and different techniques have been developed to control the performance of the goodness-of-fit test based on (1) (see, among others, [8, 14, 26, 29, 43, 44, 46, 52], where some form of correlation is introduced in the sample, and variants of the chi-squared statistics are proposed and analyzed mainly by means of simulations). Our approach differs from the one adopted in the previously quoted papers. Indeed, our starting point is that a natural way to get a positive correlation between events of the same type is to deal with the Dirichlet-multinomial (D-M) distribution: briefly, the parameters of the multinomial (M) distribution are randomized a priori with a Dirichlet distribution, yielding an exchangeable (not independent) sequence. The variance–covariance matrix of the D-M distribution is equal to that of the multinomial distribution, multiplied by a fixed constant greater than 1: precisely, given the k parameters $\boldsymbol{b_0} = (b_{0\,1}, \dots, b_{0\,k})$ of the D-M distribution and setting $|\boldsymbol{b_0}| = \sum_{i=1}^k b_{0\,i}$, we have

$$Var_{\text{D-M}}(O_i) = N\frac{b_{0\,i}}{|\boldsymbol{b_0}|}\left(1-\frac{b_{0\,i}}{|\boldsymbol{b_0}|}\right)\frac{N+|\boldsymbol{b_0}|}{1+|\boldsymbol{b_0}|} = Var_{\text{M}}(O_i)\frac{N+|\boldsymbol{b_0}|}{1+|\boldsymbol{b_0}|},$$

$$Cov_{\text{D-M}}(O_i, O_j) = -N\frac{b_{0\,i}b_{0\,j}}{|\boldsymbol{b_0}|^2}\frac{N+|\boldsymbol{b_0}|}{1+|\boldsymbol{b_0}|} = Cov_{\text{M}}(O_i, O_j)\frac{N+|\boldsymbol{b_0}|}{1+|\boldsymbol{b_0}|}, \qquad \text{for } i \neq j.$$

Therefore, if we set $|\boldsymbol{b_0}| = \frac{1-\rho^2}{\rho^2}$, we have, for any $i, j \in \{1, \dots, k\}$,

$$Cov_{\text{D-M}}(O_i, O_j) = \left(1+(N-1)\rho^2\right) Cov_{\text{M}}(O_i, O_j), \tag{2}$$

where ρ represents a correlation parameter. Roughly speaking, the D-M model adds variance to the multinomial model by taking a mixture or by adding a positive correlation. The property (2) is fundamental for our purpose. In fact, as highlighted in [47], the two conditions (i) $\widehat{p}_i = O_i/N \to p_i$ almost surely for $N \to \infty$ and (ii) $Cov(O_i, O_j) = \lambda Cov_{\text{M}}(O_i, O_j)$ with $\lambda > 1$ imply that the statistics χ^2, defined in (1), is asymptotically distributed as $\chi^2(k-1)\lambda$ (see [47, Corollary 2]), so that the critical value for the chi-squared distance becomes $\chi^2_{1-\theta}(k-1)\lambda/N$, where λ mitigates the effect of N. As already observed, the D-M model satisfies (ii), but it is well known that it does not meet the condition (i). In this paper we present a variant of the D-M model such that the condition (i) holds true and the chi-squared statistics (1) is asymptotically distributed as $\chi^2(k-1)\lambda$ with $\lambda > 1$.

The D-M distribution may be generated by means of the standard Eggenberger–Pólya urn (see [25, 38]), a model that has been widely studied and generalized (some recent variants can be found in [5, 6, 7, 10, 12, 13, 15, 17, 18, 27, 28, 37]). This urn model with k colors works as follows. An urn contains $N_{0\,i}$ balls of color i, for $i = 1, \dots, k$. At each discrete time, a ball is drawn from the urn, and then it is again placed in the urn together with $\alpha > 0$ additional balls

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

of the same color. Therefore, if we denote by $N_{n\,i}$ the number of balls of color i in the urn at time n, we have for $n \geq 1$

$$N_{n\,i} = N_{n-1\,i} + \alpha \xi_{n\,i},$$

where $\xi_{n\,i} = 1$ if the extracted ball at time n is of color i, and $\xi_{n\,i} = 0$ otherwise. The parameter α regulates the reinforcement mechanism: the greater α, the greater the dependence of $N_{n\,i}$ on $\sum_{m=1}^{n} \xi_{m\,i}$. In addition, it is well known that the *conditional expectation* of the sequential extractions, i.e. $E[\xi_{n+1\,i}|$ 'past'], also known as the *predictive mean*, converges almost surely to a beta-distributed random variable, forcing the *empirical mean* $\bar{\xi}_{N\,i} = \sum_{n=1}^{N} \xi_{n\,i}/N$ to converge almost surely to the same limit.

In this work we exhibit an urn model that preserves the relevant aspects of the models above: a reinforcement mechanism, together with a global almost sure convergence of the empirical mean of the sequential extraction toward a fixed limit. However, differently from the previous models, for a certain choice of the parameters, the predictive mean $E[\xi_{n+1\,i}|$ 'past'] randomly fluctuates without converging almost surely, asymptotically forming a stationary ergodic process. As a consequence, since the classical martingale approach and the standard stochastic approximation require or imply the convergence of $E[\xi_{n+1\,i}|$ 'past'] (e.g. [1, 9, 37]), in order to prove asymptotic results for the new urn model introduced, we need mathematical methods that are not usual in the urn modeling literature.

Rescaled Pólya urn. We introduce a new variant of the Eggenberger–Pólya urn with k colors, which we call the rescaled Pólya (RP) urn model. In this model, the almost sure limit of the empirical mean of the draws will play the rôle of an intrinsic long-run characteristic of the process, while a local mechanism generates persistent fluctuations. More precisely, the RP urn model is characterized by the introduction of the parameter β, together with the initial parameters $(b_{0\,i})_{i=1,\dots,k}$ and $(B_{0\,i})_{i=1,\dots,k}$, next to the parameter α of the original model, so that

$$\begin{aligned} N_{n\,i} &= b_{0\,i} + B_{n\,i} && \text{with} \\ B_{n\,i} &= \beta B_{n-1\,i} + \alpha \xi_{n\,i}, && n \geq 1. \end{aligned} \tag{3}$$

Therefore, the urn initially contains $b_{0\,i} + B_{0\,i}$ balls of color i, and the parameter $\beta \geq 0$, together with $\alpha > 0$, regulates the reinforcement mechanism. More precisely, $N_{n\,i}$ is the sum of three terms:

- the term $b_{0\,i}$, which remains constant along time;
- the term $\beta B_{n-1\,i}$, which links $N_{n\,i}$ to the 'configuration' at time $n-1$, through the 'scaling' parameter β that tunes the dependence on this factor;
- the term $\alpha \xi_{n\,i}$, which links $N_{n\,i}$ to the outcome of the extraction at time n, through the parameter α that tunes the dependence on this factor.

Note that the case $\beta = 1$ corresponds to the standard Eggenberger–Pólya urn with an initial number $N_{0\,i} = b_{0\,i} + B_{0\,i}$ of balls of color i, while, when $\beta \neq 1$, the RP urn does not fall among the variants of the Eggenberger–Pólya urn discussed in [45, Section 3.2], and, as explained in detail in Section 2, it does not belong to the class of reinforced stochastic processes studied in [1, 3, 2, 20, 21, 23, 50].

The quantities $p_{0\,1}, \dots, p_{0\,k}$ defined as

$$p_{0\,i} = \frac{b_{0\,i}}{\sum_{j=1}^{k} b_{0\,j}} \tag{4}$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

can be seen as an intrinsic probability distribution on the possible values (colors) $\{1, \dots, k\}$, that remains constant over time, and that will be related to the long-term characteristic of the process, while the random variables $(B_{n\,1}, \dots, B_{n\,k})$ model random fluctuations over time, so that the probability distribution on the set of the k possible values at time n is given by

$$\psi_{n\,i} = \frac{N_{n\,i}}{\sum_{j=1}^{k} N_{n\,j}} = \frac{b_{0\,i} + B_{n\,i}}{\sum_{j=1}^{k} b_{0\,j} + \sum_{j=1}^{k} B_{n\,j}}.$$

Assuming for $B_{n\,i}$ the dynamics (3) with $\beta > 0$, the probability $\psi_{n\,i}$ increases with the number of times we have observed the value i (see Equation (13) below), and so the random variables $\xi_{n\,i}$ are generated according to a reinforcement mechanism. But, in particular, when $\beta < 1$, the reinforcement at time n associated to observation $\xi_{m\,i}$, with $m = 1, \dots, n$, increases exponentially with m (we refer again to (13) below), leaving the fluctuations to be driven by the most recent draws. We refer to this feature as 'local' reinforcement. The case $\beta = 0$ is an extreme case where $\psi_{n\,i}$ depends only on the last draw $\xi_{n\,i}$ (and not on $\xi_{m\,i}$ with $m = 1, \dots, n-1$). Hence, we are mainly interested in the case $\beta \in [0, 1)$, because in this case the RP urn exhibits the following distinctive characteristics:

(a) for each i, the process $(\psi_{n\,i})_n$ randomly fluctuates, driven by the most recent observations ('local' reinforcement), and does not converge almost surely;

(b) for each i, the empirical mean $\bar{\xi}_{N\,i} = \sum_{n=1}^{N} \xi_{n\,i}/N$, that is, the empirical frequency O_i/N, converges almost surely to the deterministic limit $p_i = p_{0\,i}$;

(c) the chi-squared statistics (1) is asymptotically distributed as $\chi^2(k-1)\lambda$ with $\lambda > 1$.

As stated before, owing to (a), the usual methods adopted in the urn literature do not work for $\beta < 1$, and so different techniques are needed for the study of the RP urn model.

We have also considered the asymptotic results for $\beta > 1$, to complete the study of the RP urn model. In this situation, the process $(\psi_{n\,i})_n$ converges exponentially fast to a random limit, and so even faster than in the classical Eggenberger–Pólya urn. Therefore, in this case, we may apply the usual martingale technique (e.g. [1, 9, 37]). Note that at $\beta = 1$ we have a phase transition: for the empirical means, we pass from a deterministic limit (the case $\beta < 1$) to a random limit (the case $\beta \geq 1$), and for the predictive means, we pass from fluctuations (the case $\beta < 1$) to almost sure convergence (the case $\beta \geq 1$).

In Figure 1 we show the properties (a) and (b) for $\beta = 0$ and $\beta \in (0, 1)$ (Figure 1(A) and Figure 1(B), respectively) compared with the classical behavior of the processes for $\beta = 1$ and $\beta > 1$ (Figure 1(C) and Figure 1(D), respectively).

Goodness-of-fit result and its statistical application. Given a sample $(\boldsymbol{\xi_1}, \dots, \boldsymbol{\xi_N})$ (where $\boldsymbol{\xi_n}$ denotes the random vector with components $\xi_{n\,i}$, $i = 1, \dots, k$) generated by an RP urn, the statistics

$$O_i = \#\{n : \xi_{n\,i} = 1\} = \sum_{n=1}^{N} \xi_{n\,i}, \qquad i = 1, \dots, k,$$

counts the number of times we have observed the value i. When $\beta \in [0, 1)$, the theorem below states the almost sure convergence of the empirical mean $\bar{\xi}_{N\,i} = O_i/N$ toward the probability $p_{0\,i}$, together with a chi-squared goodness-of-fit result for the long-term probabilities $p_{0\,1}, \dots, p_{0\,k}$. More precisely, we prove the following result.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

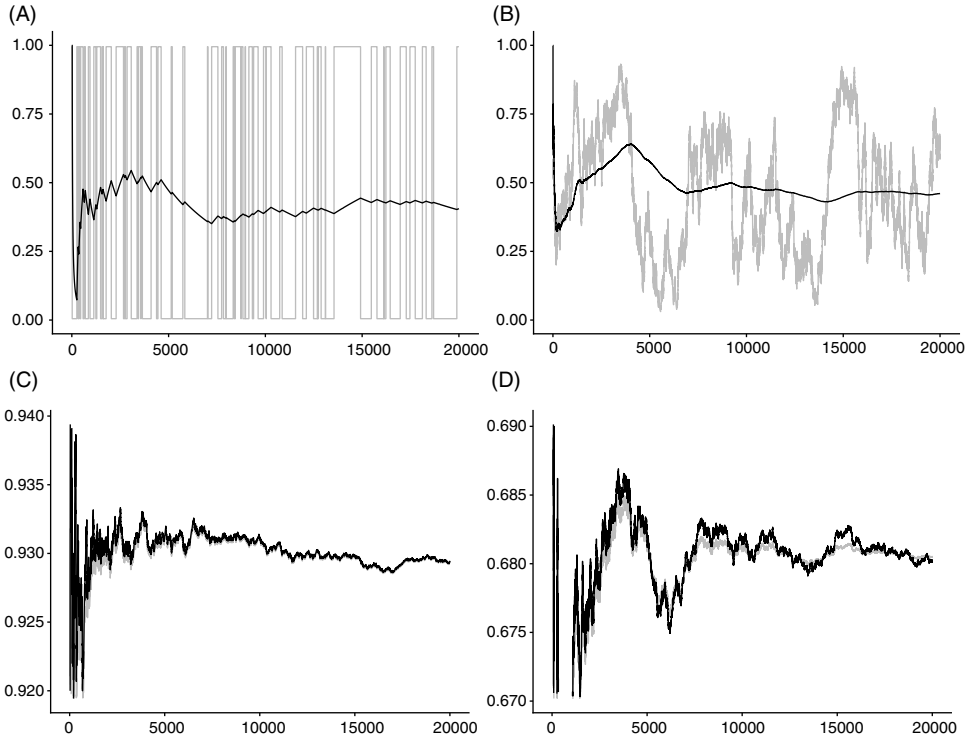

FIGURE 1. Simulations of the two processes $(\psi_{n\,1})_n$ (grey) and $(\bar{\xi}_{n\,1})_n$ (black), with $n = 1, \ldots, 20000$, $p_{0\,1} = \frac{1}{2}$, and for different values of α and β: (A) $\alpha = 199$, $\beta = 0$; (B) $\alpha = 1$, $\beta = 0.975$; (C) $\alpha = 1$, $\beta = 1$; (D) $\alpha = 0.5$, $\beta = 1.0001$. As shown, when $\beta < 1$, $(\psi_{n\,1})_n$ exhibits a persistent fluctuation, locally reinforced, and $(\bar{\xi}_{n\,1})_n$ converges to the deterministic limit $p_{0\,1}$. When $\beta \geq 1$, the y-axis is zoomed to show the random fluctuations of both the processes towards the same random limit.

Theorem 1. *Assume* $p_{0\,i} > 0$ *for all* $i = 1, \ldots, k$ *and* $\beta \in [0, 1)$. *Define the constants* γ *and* λ *as*

$$\gamma = \beta + (1-\beta)\frac{\alpha}{(1-\beta)\sum_{i=1}^{k} b_{0\,i} + \alpha} \in (\beta, 1) \qquad \textit{and} \tag{5}$$

$$\lambda = \frac{(1-\beta)^2}{(\gamma-\beta)^2 + (1-\gamma^2)}\left(1 + 2\frac{\gamma}{1-\gamma}\right) > 1. \tag{6}$$

Then $O_i/N \xrightarrow{a.s.} p_{0\,i}$ *and*

$$\sum_{i=1}^{k} \frac{(O_i - Np_{0\,i})^2}{Np_{0\,i}} \xrightarrow[N\to\infty]{d} W_* = \lambda W_0,$$

where W_0 *has distribution* $\chi^2(k-1) = \Gamma\left(\frac{(k-1)}{2}, \frac{1}{2}\right)$, *and consequently* W_* *has distribution* $\Gamma\left(\frac{k-1}{2}, \frac{1}{2\lambda}\right)$.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

From a mathematical point of view, the proof of the above result easily follows from Theorem 2, whose proof occupies the main part of the present work.

A possible application we have in mind was inspired by [11, 41] and is the following. We suppose we have a sample $\{\boldsymbol{\xi}_n : n = 1, \dots, N\}$, where the observations cannot be assumed i.i.d., but they exhibit a structure in clusters, with independence between clusters and with correlation inside each cluster. This is a usual circumstance in many applications (e.g. [16, 34, 53, 55]). More precisely, we consider the situation when inside each cluster the probability that a certain unit chooses the value i is affected by the number of units in the same cluster that have already chosen the value i, according to a reinforcement rule. For example, we can imagine that our dataset collects messages from the online social network Twitter: 'tweets' referring to different topics can be placed in different clusters. If the topics are distant from each other, we can assume independence between clusters. Inside each cluster, the tweets are temporally ordered and the associated 'sentiment' is observed to be driven by a local reinforcement mechanism: the probability of having a tweet with positive sentiment is increasing in the number of past tweets with positive sentiment, but the reinforcement is mostly driven by the most recent tweets, leading to a fluctuations of the predictive means [4]. A different clustering of the tweets can be obtained with different slots of time, sufficiently far from each other.

Another example is the following. Each cluster corresponds to an agent. The agents act independently of each other (independence between clusters). At each time-step, each agent has to choose between k brands that are related to a loyalty program: the more often the agent selects the same brand, the more loyalty points he/she gains. This fact induces the reinforcement mechanism, and it could make sense that the reinforcement is mostly driven by the most recent actions. Finally, we can have the case where clusters are associated to some products, and at each time-step a customer has to give a vote to each product on an online platform. Each cluster collects the votes for the corresponding product. If the products belong to very different categories, we can assume independence between clusters; while, if the customers can see the votes given by the previous customers, we may have a reinforcement mechanism, mainly based on the last observations.

Formally, we suppose that the N units are ordered so that we have the following L clusters of units:

$$C_\ell = \left\{ \sum_{l=1}^{\ell-1} N_l + 1, \dots, \sum_{l=1}^{\ell} N_l \right\}, \qquad \ell = 1, \dots, L.$$

Therefore, the cardinality of each cluster C_ℓ is N_ℓ. We assume that the units in different clusters are independent; that is,

$$\left[\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{N_1}\right], \dots, \left[\boldsymbol{\xi}_{\sum_{l=1}^{\ell-1} N_l+1}, \dots, \boldsymbol{\xi}_{\sum_{l=1}^{\ell} N_l}\right], \dots, \left[\boldsymbol{\xi}_{\sum_{l=1}^{L-1} N_l+1}, \dots, \boldsymbol{\xi}_N\right]$$

are L independent multidimensional random variables. Moreover, we assume that the observations inside each cluster can be modeled as an RP urn with $\beta \in [0, 1)$. We denote by $p_{0\,1}(\ell), \dots, p_{0\,k}(\ell)$ the intrinsic long-run probabilities for the cluster C_ℓ, which we assume strictly positive, and we assume the same parameter $\lambda > 1$ for each cluster (but not necessarily the same parameters α and β), so that all of the L random variables

$$Q_\ell = \sum_{i=1}^{k} \frac{\left(O_i(\ell) - N_\ell p_{0\,i}(\ell)\right)^2}{N_\ell p_{0\,i}(\ell)}, \quad \text{with } O_i(\ell) = \#\left\{n \in C_\ell : \xi_{n\,i} = 1\right\},$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

are asymptotically distributed as $\Gamma\left(\frac{k-1}{2}, \frac{1}{2\lambda}\right)$. Since $Q_1, \ldots, Q_L$ are independent because they refer to different clusters, when all the cluster sizes N_ℓ are large, we can estimate the parameter λ by means of the (asymptotic) maximum likelihood and obtain

$$\widehat{\lambda} = \frac{\sum_{\ell=1}^{L} Q_\ell}{L(k-1)} \overset{d}{\sim} \Gamma\left(\frac{L(k-1)}{2}, \frac{L(k-1)}{2\lambda}\right).$$

Note that $E[\widehat{\lambda}] = \lambda$; that is, the estimator is unbiased. Moreover, $\widehat{\lambda}/\lambda$ has asymptotic distribution $\Gamma\left(\frac{L(k-1)}{2}, \frac{L(k-1)}{2}\right)$ (that does not depend on λ), and so it can be used to construct asymptotic confidence intervals for λ. Moreover, given certain (strictly positive) intrinsic probabilities $p^*_{0\,1}(\ell), \ldots, p^*_{0\,k}(\ell)$ for each cluster C_ℓ, we can use the above procedure with $p_{0\,i}(\ell) = p^*_{0\,i}(\ell)$ for $i = 1, \ldots, k$ and $\ell = 1, \ldots, L$ to obtain an estimate $\widehat{\lambda}^*$ of λ, and then use the statistics Q_ℓ with $p_{0\,i}(\ell) = p^*_{0\,i}(\ell)$ and the corresponding asymptotic distribution $\Gamma\left(\frac{k-1}{2}, \frac{1}{2\widehat{\lambda}^*}\right)$ to perform a χ^2-test with null hypothesis

$$H_0\colon \quad p_{0\,i}(\ell) = p^*_{0\,i}(\ell) \qquad \forall i = 1, \ldots, k.$$

Regarding the probabilities $p^*_{0\,i}(\ell)$, some possibilities are as follows:

- we can take $p^*_{0\,i}(\ell) = 1/k$ for all $i = 1, \ldots, k$ if we want to test possible differences in the probabilities for the k different values;
- we can suppose that we have two different periods of time, and so two samples, say $\left\{\boldsymbol{\xi}_n^{(1)} : n = 1, \ldots, N\right\}$ and $\left\{\boldsymbol{\xi}_n^{(2)} : n = 1, \ldots, N\right\}$; take $p^*_{0\,i}(\ell) = \sum_{n \in C_\ell} \xi^{(1)}_{n\,i}/N_\ell$ for all $i = 1, \ldots, k$; and perform the test on the second sample in order to check for possible changes in the intrinsic long-run probabilities;
- we can take one of the clusters as a benchmark, say ℓ^*; set $p^*_{0\,i}(\ell) = \sum_{n \in C_{\ell^*}} \xi_{n\,i}/N_{\ell^*}$ for all $i = 1, \ldots, k$ and $\ell \neq \ell^*$; and perform the test for the other $L-1$ clusters in order to check differences with the benchmark cluster ℓ^*.

Structure of the paper. Summing up, the sequel of the paper is structured as follows. In Section 2 we set up our notation and formally define the RP urn model with parameters $\alpha > 0$ and $\beta \geq 0$. In Section 3 we provide a complete characterization of the RP urn for the three cases $\beta = 0$, $\beta \in (0, 1)$, and $\beta > 1$. (We do not deal with the case $\beta = 1$ because, as stated before, it coincides with the standard Eggenberger–Pólya urn, whose properties are well known.) In particular we show that for each i, the empirical mean of the $\xi_{n\,i}$ almost surely converges to the intrinsic probabilities $p_{0\,i}$ when $\beta \in [0, 1)$, while it almost surely converges to a random limit when $\beta > 1$. We obtain also the corresponding central limit theorems (CLTs), which, in particular for $\beta \in [0, 1)$, are the basis for the proof of Theorem 1. For completeness, we also describe the case $\alpha = 0$, which generates a sequence of independent draws. Section 4 contains the proof of Theorem 1, which gives the possibility of constructing a chi-squared test for the intrinsic long-run probabilities when the observed sample is assumed to be generated by an RP urn with $\beta \in [0, 1)$. Finally, the paper contains an appendix: in Section A.1 we state and prove a general CLT for Markov chains with a compact state space $S \subset \mathbb{R}^k$, under a certain condition which we call the 'linearity' condition, and in Section A.2 we explain a fundamental coupling technique used in the proof of the CLT for $\beta \in (0, 1)$.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

2. The rescaled Pólya urn model

In all the sequel (unless otherwise specified) we suppose given two parameters $\alpha > 0$ and $\beta \geq 0$. Given a vector $\boldsymbol{x} = (x_1, \dots, x_k)^\top \in \mathbb{R}^k$, we set $|\boldsymbol{x}| = \sum_{i=1}^k |x_i|$ and $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^\top \boldsymbol{x} = \sum_{i=1}^k |x_i|^2$. Moreover we denote by $\mathbf{1}$ and $\mathbf{0}$ the vectors with all the components equal to 1 and equal to 0, respectively, and by $\{\boldsymbol{e_1}, \dots, \boldsymbol{e_k}\}$ the canonical base of $\mathbb{R}^k$.

To formally work with the RP urn model presented in the introduction, we add here some notation. Throughout the sequel, the expression *number of balls* is not to be understood literally; all the quantities are real numbers, not necessarily integers. The urn initially contains $b_{0\,i} + B_{0\,i} > 0$ distinct balls of color i, with $i = 1, \dots, k$. We set $\boldsymbol{b_0} = (b_{0\,1}, \dots, b_{0\,k})^\top$ and $\boldsymbol{B_0} = (B_{0\,1}, \dots, B_{0\,k})^\top$. In all the sequel (unless otherwise specified) we assume $|\boldsymbol{b_0}| > 0$. Consistently with (4), we set $\boldsymbol{p_0} = \frac{\boldsymbol{b_0}}{|\boldsymbol{b_0}|}$. At each discrete time $(n+1) \geq 1$, a ball is drawn at random from the urn, producing the random vector $\boldsymbol{\xi_{n+1}} = (\xi_{n+1\,1}, \dots, \xi_{n+1\,k})^\top$ defined as

$$\xi_{n+1\,i} = \begin{cases} 1 & \text{when the ball extracted at time } n+1 \text{ is of color } i, \\ 0 & \text{otherwise,} \end{cases}$$

and the number of balls in the urn is updated as follows:

$$\boldsymbol{N_{n+1}} = \boldsymbol{b_0} + \boldsymbol{B_{n+1}} \qquad \text{with} \qquad \boldsymbol{B_{n+1}} = \beta \boldsymbol{B_n} + \alpha \boldsymbol{\xi_{n+1}}\,, \tag{7}$$

which gives (since $|\boldsymbol{\xi_{n+1}}| = 1$)

$$|\boldsymbol{B_{n+1}}| = \beta |\boldsymbol{B_n}| + \alpha. \tag{8}$$

Therefore, setting $r_n^* = |\boldsymbol{N_n}| = |\boldsymbol{b_0}| + |\boldsymbol{B_n}|$, we get

$$r_{n+1}^* = r_n^* + (\beta - 1)|\boldsymbol{B_n}| + \alpha. \tag{9}$$

Moreover, setting $\mathcal{F}_0$ equal to the trivial σ-field and $\mathcal{F}_n = \sigma(\boldsymbol{\xi_1}, \dots, \boldsymbol{\xi_n})$ for $n \geq 1$, the conditional probabilities $\boldsymbol{\psi_n} = (\psi_{n\,1}, \dots, \psi_{n\,k})^\top$ of the extraction process, also called *predictive means*, are

$$\boldsymbol{\psi_n} = E[\boldsymbol{\xi_{n+1}} | \mathcal{F}_n] = \frac{\boldsymbol{N_n}}{|\boldsymbol{N_n}|} = \frac{\boldsymbol{b_0} + \boldsymbol{B_n}}{r_n^*} \qquad \text{for } n \geq 0. \tag{10}$$

It is obvious that we have $|\boldsymbol{\psi_n}| = 1$. Finally, for the sequel, we set $\overline{\boldsymbol{\xi}}_{\boldsymbol{N}} = \sum_{n=1}^N \boldsymbol{\xi_n}/N$.

We note that, by means of (10), together with (7) and (9), we have

$$\boldsymbol{\psi_n} - \boldsymbol{\psi_{n-1}} = -\frac{(1-\beta)|\boldsymbol{b_0}|}{r_n^*}\big(\boldsymbol{\psi_{n-1}} - \boldsymbol{p_0}\big) + \frac{\alpha}{r_n^*}\big(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big). \tag{11}$$

As previously stated, the RP urn for $\beta = 1$ coincides with the well-known standard Eggenberger–Pólya urn, and so we will exclude it from the following analyses. When $\beta \neq 1$, because of the first term in the right-hand side of the above relation, the RP urn does not belong to the class of reinforced stochastic processes (RSPs) studied in [1, 3, 2, 20, 21, 23]. Generally speaking, by reinforcement in a stochastic dynamics we mean any mechanism for which the probability that a given event occurs, i.e. the predictive mean, has an increasing dependence on the number of times that the same event occurred in the past. This 'reinforcement mechanism', also known as a 'preferential attachment rule', 'rich get richer rule', or 'Matthew effect', is a key feature governing the dynamics of many biological, economic, and social systems

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

(see e.g. [45]). The RSPs are characterized by a 'strict' reinforcement mechanism such that, at each time-step, we have a strictly positive increment of the predictive mean associated to the extracted color. As an immediate consequence, the 'general' reinforcement mechanism is satisfied; that is, the predictive mean for a given color has an increasing dependence on the number of past extractions of that color. When $\beta \neq 1$, the RP urn model does not satisfy the 'strict' reinforcement mechanism, because the first term in the right-hand side of (11) is positive or negative according to the sign of $(1-\beta)$ and of $\left(\boldsymbol{\psi_{n-1}} - \boldsymbol{p_0}\right)$. However, when $\alpha,\ \beta > 0$, it satisfies the general reinforcement mechanism. Indeed, by (7), (8), (9), and (10), using $\sum_{m=0}^{n-1} x^m = (1-x^n)/(1-x)$, we have

$$r_n^* = |\boldsymbol{b_0}| + \frac{\alpha}{1-\beta} + \beta^n\left(|\boldsymbol{B_0}| - \frac{\alpha}{1-\beta}\right) \tag{12}$$

and

$$\boldsymbol{\psi_n} = \frac{\boldsymbol{b_0} + \beta^n \boldsymbol{B_0} + \alpha \sum_{m=1}^{n} \beta^{n-m} \boldsymbol{\xi_m}}{|\boldsymbol{b_0}| + \frac{\alpha}{1-\beta} + \beta^n\left(|\boldsymbol{B_0}| - \frac{\alpha}{1-\beta}\right)} = \frac{\beta^{-n}\boldsymbol{b_0} + \boldsymbol{B_0} + \alpha \sum_{m=1}^{n} \beta^{-m} \boldsymbol{\xi_m}}{\beta^{-n}\left(|\boldsymbol{b_0}| + \frac{\alpha}{1-\beta}\right) + |\boldsymbol{B_0}| - \frac{\alpha}{1-\beta}}. \tag{13}$$

In particular, for $\beta > 1$, the dependence of $\boldsymbol{\psi_n}$ on $\boldsymbol{\xi_m}$ exponentially decreases with m, because of the factor β^{-m}. For $\beta < 1$ we have the opposite behavior; that is, the dependence of $\boldsymbol{\psi_n}$ on $\boldsymbol{\xi_m}$ exponentially increases with m, because of the factor β^{n-m}, and so the main contribution is given by the most recent extractions. We refer to this phenomenon as 'local' reinforcement. The case $\beta = 0$ is an extreme case, for which $\boldsymbol{\psi_n}$ depends only on the last extraction $\boldsymbol{\xi_n}$: at each time-step $n+1 \geq 2$ we extract a ball from an urn with $b_{0\,i} + \alpha$ balls of color i, if i is the color extracted at time n, and $b_{0\,j}$ balls for each color $j \neq i$. This particular case corresponds to a version of the so-called 'memory-1 senile reinforced random walk' on a star-shaped graph introduced in [33], but the study done in that paper differs from ours. Finally, we observe that Equation (11) recalls the dynamics of an RSP with a 'forcing input' (see [1, 20, 50]), but the main difference relies on the fact that, for the RP urn, the sequence $\left(r_n^*\right)$ is such that $r_n^* \to r^* > 0$, and so $\sum_n 1/r_n^* = +\infty$ and $\sum_n 1/\left(r_n^*\right)^2 = +\infty$, when $\beta \in [0,\ 1)$, and such that $\sum_n 1/r_n^* < +\infty$ (and $\sum_n 1/\left(r_n^*\right)^2 < +\infty$) when $\beta > 1$. These facts lead to different asymptotic behavior for $(\boldsymbol{\psi_n})$. Specifically, for the RP urn with $\beta \in [0,\ 1)$, the predictive mean $\boldsymbol{\psi_n}$ randomly fluctuates and does not converge almost surely, while, for the RP urn with $\beta > 1$, the sequence $(\boldsymbol{\psi_n})$ almost surely converges to a random variable $\boldsymbol{\psi_\infty}$ and $|\boldsymbol{\psi_n} - \boldsymbol{\psi_\infty}| = O(\beta^{-n})$. By contrast, for the RSP with a 'forcing input', the almost sure convergence of $\boldsymbol{\psi_n}$ toward the forcing input (which is a constant) holds true, and the corresponding rate of convergence depends on a model parameter $\gamma \in (1/2,\ 1]$ and equals $n^{-\gamma/2}$.

3. Properties of the rescaled Pólya urn model

We study separately the three cases $\beta = 0$, $\beta \in (0,\ 1)$, and $\beta > 1$. In particular, the case $\beta = 0$ is placed at the end of this section, because it is elementary.

3.1. The case $\beta \in (0,\ 1)$

In this case, we have $\lim_n \beta^n = 0$ and $\sum_{n\geq 1} \beta^n = \beta/(1-\beta)$. Therefore, setting $r = \frac{\alpha}{1-\beta}$ and $r^* = |\boldsymbol{b_0}| + r$, we have by (8) and (9)

$$r_n^* = |\boldsymbol{b_0}| + |\boldsymbol{B_n}| = |\boldsymbol{b_0}| + r + \beta^n(|\boldsymbol{B_0}| - r) \longrightarrow r^* > 0, \tag{14}$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

and so we have that the denominator r_n^* in $\boldsymbol{\psi_n}$ (see Equation (10)) goes exponentially fast to the limit r^*. Moreover, recalling the definition of the constant γ in (5), we have $\beta < \gamma < 1$ (remember that $|\boldsymbol{b_0}| > 0$ by assumption),

$$\gamma - \beta = \frac{\alpha}{r^*}, \qquad \text{and} \qquad 1 - \gamma = \frac{(1-\beta)|\boldsymbol{b_0}|}{r^*}. \tag{15}$$

Therefore, by (14), the terms $\frac{(1-\beta)|\boldsymbol{b_0}|}{r_n^*}$ and $\frac{\alpha}{r_n^*}$ in the dynamics (11) converge exponentially fast to $(1-\gamma)$ and $(\gamma-\beta)$, respectively. Furthermore, as we will see, the fact that the constant γ is strictly smaller than 1 will play a central rôle, because it will imply the existence of a contraction of the process $\boldsymbol{\psi} = (\boldsymbol{\psi_n})_n$ in a proper metric space with (sharp) constant γ. Consequently, it is not a surprise that this constant enters naturally in the parameters of the asymptotic distribution, given in the following result.

Theorem 2. *We have* $\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \boldsymbol{p_0}$ *and*

$$\sqrt{N}\big(\overline{\boldsymbol{\xi}}_N - \boldsymbol{p_0}\big) = \frac{\sum_{n=1}^N \big(\boldsymbol{\xi_n} - \boldsymbol{p_0}\big)}{\sqrt{N}} \xrightarrow{d} \mathcal{N}\big(\boldsymbol{0}, \Sigma^2\big),$$

where

$$\Sigma^2 = \lambda\big(\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^T\big),$$

with λ *defined in* (6) *as a function of* β *and* γ.

Remark 1. Note that in Theorem 2 we do not assume $b_{0\,i} > 0$ for all i, but only $|\boldsymbol{b_0}| > 0$ (as stated in Section 2). The asymptotic result given above fails when $\boldsymbol{b_0} = \boldsymbol{0}$, and a different behavior is observed. Indeed, from (11), we have

$$\boldsymbol{\psi_n} - \boldsymbol{\psi_{n-1}} = \frac{\alpha}{r_n^*}\big(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big), \tag{16}$$

and hence $(\boldsymbol{\psi_n})$ is a martingale. Therefore the bounded process $(\boldsymbol{\psi_n})$ converges almost surely (and in mean) to a bounded random variable $\boldsymbol{\psi_\infty}$. In addition, since $r_n^* \to r = \alpha/(1-\beta)$, from (16), we obtain that the unique possible limits $\boldsymbol{\psi_\infty}$ are those for which $\boldsymbol{\xi_n} = \boldsymbol{\psi_\infty}$ eventually. Hence $\boldsymbol{\psi_\infty}$ takes values in $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_k}\}$, and since we have $E[\boldsymbol{\psi_\infty}] = E[\boldsymbol{\psi_0}] = \boldsymbol{B_0}/|\boldsymbol{B_0}|$, we get $P(\boldsymbol{\psi_\infty} = \boldsymbol{e_i}) = \frac{B_{0\,i}}{|\boldsymbol{B_0}|}$ for all $i = 1, \ldots, k$.

We will split the proof of Theorem 2 into two main steps: first, we will prove that the convergence behavior of $\overline{\boldsymbol{\xi}}_N$ does not depend on the initial constant $|\boldsymbol{B}_0|$, and then, without loss of generality, we will assume $|\boldsymbol{B}_0| = r$ and we will give the proof of the theorem under this assumption.

3.1.1. *Independence of the asymptotic properties of the empirical mean from* $|\boldsymbol{B_0}|$. We use a coupling method to prove that the convergence results stated in Theorem 2 are not affected by the value of the initial constant $|\boldsymbol{B_0}|$.

Set $\boldsymbol{\xi_n^{(1)}} = \boldsymbol{\xi_n}$ and $\boldsymbol{\psi_n^{(1)}} = \boldsymbol{\psi_n}$, which follows the dynamics (11), together with (10), starting from a certain initial point $\boldsymbol{\psi_0^{(1)}} = \boldsymbol{\psi_0}$. By (7) and the relations (15), we can write

$$\begin{aligned}
\boldsymbol{\psi_{n+1}^{(1)}} &= \frac{\boldsymbol{b_0} + \beta \boldsymbol{B_n} + \alpha \boldsymbol{\xi_{n+1}^{(1)}}}{r_{n+1}^*} = \frac{\boldsymbol{b_0} + \beta\Big(r_n^* \boldsymbol{\psi_n^{(1)}} - \boldsymbol{b_0}\Big) + \alpha \boldsymbol{\xi_{n+1}^{(1)}}}{r_{n+1}^*} \\
&= \beta \boldsymbol{\psi_n^{(1)}} \frac{r_n^*}{r_{n+1}^*} + \frac{(1-\beta)\boldsymbol{b_0}}{r_{n+1}^*} + \frac{\alpha}{r_{n+1}^*} \boldsymbol{\xi_{n+1}^{(1)}} \\
&= \beta \boldsymbol{\psi_n^{(1)}} + (\gamma - \beta)\boldsymbol{\xi_{n+1}^{(1)}} + \boldsymbol{l_{n+1}^{(1)}}\Big(\boldsymbol{\psi_n^{(1)}}, \boldsymbol{\xi_{n+1}^{(1)}}\Big) + (1-\gamma)\boldsymbol{p_0},
\end{aligned}$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

where

$$\boldsymbol{l}_{\boldsymbol{n+1}}^{(\mathbf{1})}(\boldsymbol{x},\boldsymbol{y}) = \left(\frac{r^*}{r_{n+1}^*} - 1\right)\big[(1-\gamma)\boldsymbol{p_0} + (\gamma-\beta)\boldsymbol{y}\big] + \left(\frac{r_n^*}{r_{n+1}^*} - 1\right)\beta\boldsymbol{x}.$$

Since, by (14), we have $r^*/r_{n+1}^* - 1 = O\big(\beta^{n+1}\big)$ and $r_n^*/r_{n+1}^* - 1 = O\big(\beta^{n+1}\big)$, we get $|\boldsymbol{l}_{\boldsymbol{n+1}}^{(\mathbf{1})}| = O\big(\beta^{n+1}\big)$. Now, take $\boldsymbol{\xi}^{(\mathbf{2})} = \big(\boldsymbol{\xi}_{\boldsymbol{n}}^{(\mathbf{2})}\big)_n$ and $\boldsymbol{\psi}^{(\mathbf{2})} = \big(\boldsymbol{\psi}_{\boldsymbol{n}}^{(\mathbf{2})}\big)_n$ following the same dynamics given in (10) and (11), but starting from an initial point with $|\boldsymbol{B}_{\mathbf{0}}^{(\mathbf{2})}| = r$. Therefore, we have

$$\boldsymbol{\psi}_{\boldsymbol{n+1}}^{(\mathbf{2})} = \beta\boldsymbol{\psi}_{\boldsymbol{n}}^{(\mathbf{2})} + (\gamma-\beta)\boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\mathbf{2})} + (1-\gamma)\boldsymbol{p_0}.$$

Both dynamics are of the form (A.9) with $a_0 = \beta$, $a_1 = (\gamma-\beta)$, $\boldsymbol{c} = (1-\gamma)\boldsymbol{p_0}$, $c_n^{(1)} = \beta^n$, and $c_n^{(2)} = 0$, and, by (10), the condition (A.10) holds true. Hence we can apply Theorem 10 so that there exist two stochastic processes $\widetilde{\boldsymbol{\psi}}^{(1)}$ and $\widetilde{\boldsymbol{\psi}}^{(2)}$, following the dynamics (A.11) (with the same specifications as above), together with (A.12), starting from the same initial points and such that (A.14) holds true; that is,

$$\begin{aligned} E\Big[\big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n+1}}^{(1)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n+1}}^{(2)}\big| \,\big|\, \widetilde{\boldsymbol{\psi}}_0^{(2)}, \widetilde{\boldsymbol{\psi}}_0^{(1)}\Big] &\le \gamma^{n+1}\big|\widetilde{\boldsymbol{\psi}}_0^{(1)} - \widetilde{\boldsymbol{\psi}}_0^{(2)}\big| + O\left(\sum_{j=1}^{n+1}\gamma^{n+1-j}\beta^j\right) \\ &= O\big((n+2)\max(\gamma,\beta)^{n+1}\big) = O\big((n+2)\gamma^{n+1}\big). \end{aligned}$$

Since $\gamma < 1$, if we subtract (A.11) with $\ell = 2$ from (A.11) with $\ell = 1$, we obtain that

$$\sum_{n=1}^{+\infty} E\Big[\big|\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(1)} - \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(2)}\big|\Big] < +\infty,$$

which implies $\sum_{n=0}^{+\infty}\left|\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(1)} - \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(2)}\right| < +\infty$ almost surely. Recall that $\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(1)} \in \{0,1\}$ and $\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(2)} \in \{0,1\}$, and hence

$$\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(1)} = \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(2)} \qquad \text{eventually (i.e. for sufficiently large } n). \tag{17}$$

Therefore, if we prove some asymptotic results for $\boldsymbol{\xi}^{(\mathbf{2})}$, then they hold true also for $\widetilde{\boldsymbol{\xi}}^{(2)}$ (since they have the same joint distribution); then they hold true also for $\widetilde{\boldsymbol{\xi}}^{(1)}$ (since (17)), and finally they hold true also for $\boldsymbol{\xi}^{(\mathbf{1})}$ (since they have the same joint distribution). Summing up, without loss of generality, we may prove Theorem 2 under the additional assumption $|\boldsymbol{B_0}| = r$.

3.1.2. *The case* $|\boldsymbol{B_0}| = r$. Thanks to what we have observed in the previous subsection, we here assume that $|\boldsymbol{B_0}| = r = \alpha/(1-\beta)$, which implies $|\boldsymbol{B_n}| = r$ and

$$r_n^* = r^* = |\boldsymbol{b_0}| + r \tag{18}$$

for any n. Hence, we can simplify (10) as

$$\psi_{n\,i} = P(\xi_{n+1\,i} = 1|\mathcal{F}_n) = \frac{b_{0\,i} + B_{n\,i}}{|\boldsymbol{b_0}| + r} = \frac{b_{0\,i} + B_{n\,i}}{r^*}.$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

The process $\boldsymbol{\psi} = (\boldsymbol{\psi_n})_{n\geq 0}$ is then a Markov chain with state space

$$S = \left\{ \boldsymbol{x} : x_i \in \left[\frac{b_{0\,i}}{r^*}, \frac{b_{0\,i} + r}{r^*} \right], |\boldsymbol{x}| = 1 \right\},$$

which, endowed with the distance induced by the norm $|\cdot|$, is a compact metric space.

In the sequel, according to the context, since we work with a Markov chain with state space $S \subset \mathbb{R}^k$, the notation P will be used for the following:

- a kernel $P : S \times \mathcal{B}(S) \to [0, 1]$, where $\mathcal{B}(S)$ is the Borel σ-field on S and we will use the notation $P(\boldsymbol{x}, \boldsymbol{dy})$ in the integrals;
- an operator $P : C(S) \to M(S)$, where $C(S)$ and $M(S)$ denote the spaces of the continuous and measurable functions on S, respectively, defined as

$$(Pf)(\boldsymbol{x}) = \int_S f(\boldsymbol{y}) P(\boldsymbol{x}, \boldsymbol{dy}).$$

In addition, when f is the identity map, that is, $f(\boldsymbol{y}) = \boldsymbol{y}$, we will write $(Pid)(\boldsymbol{x})$ or $(P\boldsymbol{y})(\boldsymbol{x})$.

Moreover, we set $P^0 f = f$ and $P^n f = P\big(P^{n-1} f\big)$.

By (11), together with (15) and (18), the process $\boldsymbol{\psi} = (\boldsymbol{\psi_n})_n$ follows the dynamics

$$\boldsymbol{\psi_{n+1}} = \beta \boldsymbol{\psi_n} + \boldsymbol{b_0} \frac{(1-\beta)}{r^*} + \boldsymbol{\xi_{n+1}} \frac{\alpha}{r^*} = \beta \boldsymbol{\psi_n} + (1-\gamma)\boldsymbol{p_0} + (\gamma - \beta)\boldsymbol{\xi_{n+1}}. \tag{19}$$

Therefore, given $\boldsymbol{z} = (z_1, \ldots, z_k)^T$ and setting

$$\boldsymbol{z}_{(i)} = \left(z_1, \ldots, z_i + \frac{\alpha}{r^*}, \ldots, z_k \right)^T = \big(z_1, \ldots, z_i + (\gamma - \beta), \ldots, z_k \big)^T,$$

for any $i = 1, \ldots, k$, we get

$$(Pf)(\boldsymbol{x}) = E\big[f\big(\boldsymbol{\psi_{n+1}}\big) | \boldsymbol{\psi_n} = \boldsymbol{x} \big] = \sum_{i=1}^{k} x_i f\Big(\big(\beta \boldsymbol{x} + \boldsymbol{p_0}(1-\gamma) \big)_{(i)} \Big). \tag{20}$$

In particular, from the above equality, we get

$$(Pid)(\boldsymbol{x}) - \boldsymbol{p_0} = E\big[\boldsymbol{\psi_{n+1}} - \boldsymbol{p_0} | \boldsymbol{\psi_n} = \boldsymbol{x} \big] = \gamma(\boldsymbol{x} - \boldsymbol{p_0}). \tag{21}$$

We now show that $\boldsymbol{\psi}$ is an irreducible, aperiodic, compact Markov chain (see Definitions 3 and 4).

Checking that $\boldsymbol{\psi}$ is a compact Markov chain: By Lemma 1, it is sufficient to show that P defined in (20) is weak Feller (Definition 1) and that it is a semi-contractive operator on $Lip(S)$ (Definition 2). From (20), we have immediately that the function Pf is continuous whenever f is continuous and hence P is weak Feller. In order to prove the contractive property, we start by observing that the dynamics (19) of $\boldsymbol{\psi}$ is of the form (A.9) with $a_0 = \beta$, $a_1 = (\gamma - \beta)$, $\boldsymbol{c} = \boldsymbol{p_0}(1-\gamma)$, and $\boldsymbol{l_n} \equiv \boldsymbol{0}$ for each n. Moreover, by (10), the condition (A.10) holds true. Then we let $\boldsymbol{\psi}^{(1)}$ and $\boldsymbol{\psi}^{(2)}$ be two stochastic processes following the dynamics (A.9) with the same specifications as above, together with (A.10), and starting, respectively, from the points $\boldsymbol{x}$ and $\boldsymbol{y}$.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

Applying Theorem 10, we get two stochastic processes $\widetilde{\boldsymbol{\psi}}^{(1)}$ and $\widetilde{\boldsymbol{\psi}}^{(2)}$, evolving according to (A.11), together with (A.12), starting from the same initial points $\boldsymbol{x}$ and $\boldsymbol{y}$ and such that

$$E\Big[\Big|\widetilde{\boldsymbol{\psi}}_1^{(2)} - \widetilde{\boldsymbol{\psi}}_1^{(1)}\Big|\Big] = E\Big[\Big|\widetilde{\boldsymbol{\psi}}_1^{(2)} - \widetilde{\boldsymbol{\psi}}_1^{(1)}\Big|\Big|\widetilde{\boldsymbol{\psi}}_0^{(1)} = \boldsymbol{x},\ \widetilde{\boldsymbol{\psi}}_0^{(2)} = \boldsymbol{y}\Big] \le \gamma\,|\boldsymbol{x}-\boldsymbol{y}|.$$

Therefore, if we take $f \in Lip(S)$ with

$$|f|_{Lip} = \sup_{\boldsymbol{x},\boldsymbol{y}\in S,\,\boldsymbol{x}\neq\boldsymbol{y}} \frac{|f(\boldsymbol{y}) - f(\boldsymbol{x})|}{|\boldsymbol{y}-\boldsymbol{x}|},$$

we obtain

$$\begin{aligned}|(Pf)(\boldsymbol{y}) - (Pf)(\boldsymbol{x})| &= \Big|E\Big[f\Big(\widetilde{\boldsymbol{\psi}}_1^{(2)}\Big)\Big|\widetilde{\boldsymbol{\psi}}_0^{(2)} = \boldsymbol{y}\Big] - E\Big[f\Big(\widetilde{\boldsymbol{\psi}}_1^{(1)}\Big)\Big|\widetilde{\boldsymbol{\psi}}_0^{(1)} = \boldsymbol{x}\Big]\Big| \\ &= E\Big[\ \Big|f\Big(\widetilde{\boldsymbol{\psi}}_1^{(2)}\Big) - f\Big(\widetilde{\boldsymbol{\psi}}_1^{(1)}\Big)\Big|\ \Big] \\ &\le |f|_{Lip} E\Big[\Big|\widetilde{\boldsymbol{\psi}}_1^{(2)} - \widetilde{\boldsymbol{\psi}}_1^{(1)}\Big|\Big] \le |f|_{Lip}\ \gamma\ |\boldsymbol{x}-\boldsymbol{y}|\end{aligned}$$

and so

$$|(Pf)|_{Lip} = \sup_{\boldsymbol{x},\boldsymbol{y}\in S,\,\boldsymbol{x}\neq\boldsymbol{y}} \frac{|(Pf)(\boldsymbol{y}) - (Pf)(\boldsymbol{x})|}{|\boldsymbol{y}-\boldsymbol{x}|} \le \gamma\,|f|_{Lip},$$

with $\gamma < 1$, as desired. □

Checking that $\boldsymbol{\psi}$ is irreducible and aperiodic: We prove the irreducibility and aperiodicity condition stated in Definition 4, using Theorem 7. Therefore, let us denote by π an invariant probability measure for P. Moreover, let $\boldsymbol{\psi}^{(1)}$ and $\boldsymbol{\psi}^{(2)}$ be two processes that follow the same dynamics (19) of $\boldsymbol{\psi}$, but for the first process, we set the initial distribution equal to π, while for the second process, we take any other initial distribution ν on S. Again, as above, since (19) is of the form (A.9) with $a_0 = \beta$, $a_1 = (\gamma - \beta)$, $\boldsymbol{c} = \boldsymbol{p_0}(1-\gamma)$, and $\boldsymbol{l_n} \equiv \boldsymbol{0}$ for each n, and, by (10), the condition (A.10) holds true, we can apply Theorem 5.2 and obtain two stochastic processes $\widetilde{\boldsymbol{\psi}}^{(1)}$ and $\widetilde{\boldsymbol{\psi}}^{(2)}$, evolving according to (A.11) (with the same specifications as above), together with (A.12), starting from the same initial random variables $\boldsymbol{\psi_0^{(1)}}$ (with distribution π) and $\boldsymbol{\psi_0^{(2)}}$ (with distribution ν) and such that

$$E\Big[\Big|\widetilde{\boldsymbol{\psi}}_{n+1}^{(1)} - \widetilde{\boldsymbol{\psi}}_{n+1}^{(2)}\Big|\Big|\widetilde{\boldsymbol{\psi}}_0^{(2)}, \widetilde{\boldsymbol{\psi}}_0^{(1)}\Big] \le \gamma^{n+1}\Big|\widetilde{\boldsymbol{\psi}}_0^{(1)} - \widetilde{\boldsymbol{\psi}}_0^{(2)}\Big|.$$

Hence, since $\gamma < 1$, we have $E\Big[\Big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)}\Big|\Big] \longrightarrow 0$, and, since the distribution of $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}$ is always π (by the definition of an invariant probability measure), we can conclude that $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)}$, and so $\boldsymbol{\psi_n^{(2)}}$ (because they have the same distribution), converge in distribution to π. □

Proof of the almost sure convergence: We have already proven that the Markov chain $\boldsymbol{\psi}$ has one invariant probability measure π. Furthermore, from (19) we get

$$\begin{aligned}\boldsymbol{\xi_{n+1}} - \boldsymbol{p_0} &= \frac{1}{\gamma-\beta}\big[\boldsymbol{\psi_{n+1}} - \beta\boldsymbol{\psi_n} - \boldsymbol{p_0}(1-\gamma)\big] - \boldsymbol{p_0} \\ &= \frac{1}{\gamma-\beta}\big[\big(\boldsymbol{\psi_{n+1}} - \boldsymbol{p_0}\big) - \beta\big(\boldsymbol{\psi_n} - \boldsymbol{p_0}\big)\big]. \qquad (22)\end{aligned}$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

Therefore, applying [31, Corollary 5.3 and Corollary 5.12], we obtain

$$\begin{aligned}\overline{\boldsymbol{\xi}}_N - \boldsymbol{p_0} &= \frac{1}{N}\sum_{m=1}^{N}(\boldsymbol{\xi_m} - \boldsymbol{p_0}) = \frac{1-\beta}{\gamma-\beta}\frac{1}{N}\sum_{m=1}^{N} E\big[\boldsymbol{\psi_m} - \boldsymbol{p_0}\big] + \frac{O(1)}{N} \\ &\xrightarrow{a.s.} \frac{1-\beta}{\gamma-\beta} E\big[\boldsymbol{\psi_n^{(\pi)}} - \boldsymbol{p_0}\big],\end{aligned}$$

where $\big(\boldsymbol{\psi_n^{(\pi)}}\big)_{n\geq 0}$ is a Markov chain with transition kernel P and initial distribution π, and so stationary. From (21), we have

$$E\big[\boldsymbol{\psi_n^{(\pi)}} - \boldsymbol{p_0}\big] = E\big[\boldsymbol{\psi_{n+1}^{(\pi)}} - \boldsymbol{p_0}\big] = E\big[E\big[\boldsymbol{\psi_{n+1}^{(\pi)}} - \boldsymbol{p_0} \mid \boldsymbol{\psi_n^{(\pi)}}\big]\big] = \gamma E\big[\boldsymbol{\psi_n^{(\pi)}} - \boldsymbol{p_0}\big]$$

and so, since $\gamma < 1$, we get $E\big[\boldsymbol{\psi_n^{(\pi)}} - \boldsymbol{p_0}\big] = 0$. This means that $\int_S \boldsymbol{x}\,\pi(d\boldsymbol{x}) = \boldsymbol{p_0}$ and $\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \boldsymbol{p_0}$. □

Proof of the CLT: We apply Theorem 9, taking into account that we have already proven that $\boldsymbol{\psi}$ is an irreducible and aperiodic compact Markov chain. By (22) and what we have already proven before, we have

$$\boldsymbol{\xi_{n+1}} - \boldsymbol{p_0} = \boldsymbol{f}\big(\boldsymbol{\psi_n}, \boldsymbol{\psi_{n+1}}\big) \quad \text{with} \quad \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\gamma-\beta}\big[(\boldsymbol{y} - \boldsymbol{p_0}) - \beta(\boldsymbol{x} - \boldsymbol{p_0})\big]$$

and $\boldsymbol{p_0} = \int_S \boldsymbol{x}\,\pi(d\boldsymbol{x})$. Hence $\boldsymbol{f}$ and P form a linear model as defined in Definition 5 (see also Remark 5). Indeed, we have $A_1 = -\frac{\beta}{\gamma-\beta}Id$ and $A_2 = \frac{1}{\gamma-\beta}Id$. Moreover, by (21), we have $P(id)(\boldsymbol{x}) - \boldsymbol{p_0} = \gamma(\boldsymbol{x} - \boldsymbol{p_0})$, which means $A_P = \gamma Id$. Therefore Theorem 9 holds true with

$$D_0 = (1-\gamma)^{-1}Id, \quad D_1 = -\frac{\gamma(1-\beta)}{(\gamma-\beta)(1-\gamma)}Id, \quad D_2 = \frac{(1-\beta)}{(\gamma-\beta)(1-\gamma)}Id,$$

and so, after some computations, with

$$\Sigma^2 = \frac{(1-\beta)^2(1+\gamma)}{(\gamma-\beta)^2(1-\gamma)}\Sigma_\pi^2 = \frac{(1-\beta)^2}{(\gamma-\beta)^2}\left(1 + 2\frac{\gamma}{1-\gamma}\right)\Sigma_\pi^2. \tag{23}$$

In order to conclude, we take a Markov chain $\big(\boldsymbol{\psi_n^{(\pi)}}\big)_{n\geq 0}$ with transition kernel P and initial distribution π and we set

$$\boldsymbol{\xi_{n+1}^{(\pi)}} - \boldsymbol{p_0} = f\Big(\boldsymbol{\psi_n^{(\pi)}}, \boldsymbol{\psi_{n+1}^{(\pi)}}\Big) = A_2\Big(\boldsymbol{\psi_{n+1}^{(\pi)}} - \boldsymbol{p_0}\Big) + A_1\Big(\boldsymbol{\psi_n^{(\pi)}} - \boldsymbol{p_0}\Big).$$

Then we observe that, by (A.4) and (A.5), we have

$$\begin{aligned}\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^T &= E\Big[\Big(\boldsymbol{\xi_1^{(\pi)}} - \boldsymbol{p_0}\Big)\Big(\boldsymbol{\xi_1^{(\pi)}} - \boldsymbol{p_0}\Big)^\top\Big] \\ &= A_1\Sigma_\pi^2 A_1^\top + A_2\Sigma_\pi^2 A_2^\top + A_1\Sigma_\pi^2 A_P^\top A_2^\top + A_2 A_P \Sigma_\pi^2 A_1 \\ &= \frac{(\gamma-\beta)^2 + \big(1-\gamma^2\big)}{(\gamma-\beta)^2}\Sigma_\pi^2.\end{aligned} \tag{24}$$

Finally, it is enough to combine (23) and (24). □

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

3.2. The case $\beta > 1$

In this case, $\lim_n \beta^n = +\infty$ and $\sum_{n\geq 1} \beta^{-n} = 1/(\beta - 1)$. Moreover, by (12), r_n^* increases exponentially to $+\infty$. Hence, the following results hold true.

Theorem 3. *We have*

$$\boldsymbol{\psi}_N \xrightarrow{a.s.} \boldsymbol{\psi}_\infty = \frac{\boldsymbol{B_0} + \alpha \sum_{n=1}^{+\infty} \beta^{-n} \boldsymbol{\xi_n}}{|\boldsymbol{B_0}| + \frac{\alpha}{\beta-1}}$$

and

$$|\boldsymbol{\psi}_N - \boldsymbol{\psi}_\infty| \overset{a.s.}{=} O\big(\beta^{-N}\big).$$

Moreover, $\boldsymbol{\psi}_\infty$ takes values in $\big\{\boldsymbol{x} \in [0, 1]^k : |\boldsymbol{x}| = 1\big\}$, and if $B_{0\,i} > 0$ for all $i = 1, \ldots, k$, then $P\{\psi_{\infty\,i} \in (0, 1)\} = 1$ for each i.

Note that $\boldsymbol{\psi}_\infty$ is a function of $\boldsymbol{\xi} = (\boldsymbol{\xi_n})_{n\geq 1}$ which takes values in $\big(\big\{\boldsymbol{x} \in \{0, 1\}^k : |\boldsymbol{x}| = 1\big\}\big)^\infty$.

Proof. By (13), we have

$$\boldsymbol{\psi}_N = \frac{\boldsymbol{b_0} + \boldsymbol{B}_N}{r_N^*} = \frac{\boldsymbol{b_0}\beta^{-N} + \boldsymbol{B_0} + \alpha \sum_{n=1}^{N} \beta^{-n} \boldsymbol{\xi_n}}{|\boldsymbol{b_0}|\beta^{-N} + |\boldsymbol{B_0}| + \frac{\alpha}{\beta-1}\big(1 - \beta^{-N}\big)}.$$

Hence, the almost sure convergence immediately follows because $\big|\sum_{n\geq 1} \beta^{-n} \boldsymbol{\xi_n}\big| \leq \sum_{n\geq 1} \beta^{-n} < +\infty$. Moreover, after some computations, we have

$$\boldsymbol{\psi}_N - \boldsymbol{\psi}_\infty = \frac{-\alpha\Big(|\boldsymbol{B_0}| + \frac{\alpha}{\beta-1}\Big) \sum_{n\geq N+1} \beta^{-n} \boldsymbol{\xi_n} + \beta^{-N} \boldsymbol{R}}{\Big(|\boldsymbol{B_0}| + \frac{\alpha}{\beta-1}\Big)\Big(|\boldsymbol{B_0}| + \frac{\alpha}{\beta-1} + \beta^{-N}\Big(|\boldsymbol{b_0}| - \frac{\alpha}{\beta-1}\Big)\Big)},$$

where

$$\boldsymbol{R} = \left(|\boldsymbol{B_0}| + \frac{\alpha}{\beta - 1}\right)\boldsymbol{b_0} - \left(|\boldsymbol{b_0}| - \frac{\alpha}{\beta - 1}\right)\left(\boldsymbol{B_0} + \alpha \sum_{n=1}^{+\infty} \beta^{-n} \boldsymbol{\xi_n}\right).$$

Therefore, since $|\sum_{n\geq N+1} \beta^{-n} \boldsymbol{\xi_n}| \leq \sum_{n\geq N+1} \beta^{-n} = \beta^{-N}/(\beta - 1)$ and $|\boldsymbol{R}|$ is bounded by a constant, we obtain that

$$|\boldsymbol{\psi}_N - \boldsymbol{\psi}_\infty| \overset{a.s.}{=} O\big(\beta^{-N}\big).$$

In order to conclude, it is enough to recall that, by definition, we have $\boldsymbol{\psi}_N \in [0, 1]^k$ with $|\boldsymbol{\psi}_N| = 1$ and observe that, if $B_{0,\,i} > 0$ for all i, then we have

$$0 < \frac{B_{0\,i}}{|\boldsymbol{B_0}| + \frac{\alpha}{\beta-1}} \leq \psi_{\infty\,i} = \frac{B_{0\,i} + \alpha \sum_{n=1}^{\infty} \beta^{-n} \xi_{n\,i}}{|\boldsymbol{B_0}| + \frac{\alpha}{\beta-1}} \leq \frac{B_{0\,i} + \frac{\alpha}{\beta-1}}{|\boldsymbol{B_0}| + \frac{\alpha}{\beta-1}} < 1.$$

□

Theorem 4. *We have $\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \boldsymbol{\psi}_\infty$, and*

$$\sqrt{N}\big(\overline{\boldsymbol{\xi}}_N - \boldsymbol{\psi}_N\big) \xrightarrow{s} \mathcal{N}\big(\boldsymbol{0}, \Sigma^2\big) \quad \textit{and} \quad \sqrt{N}\big(\overline{\boldsymbol{\xi}}_N - \boldsymbol{\psi}_\infty\big) \xrightarrow{s} \mathcal{N}\big(\boldsymbol{0}, \Sigma^2\big),$$

where $\Sigma^2 = \mathrm{diag}(\boldsymbol{\psi}_\infty) - \boldsymbol{\psi}_\infty \boldsymbol{\psi}_\infty^\top$ and $\xrightarrow{s}$ means stable convergence.

Note that if $B_{0\,i} > 0$ for all $i = 1, \ldots, k$, then $\Sigma_{i,j}^2 \in (0, 1)$ for each pair (i, j).

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

Stable convergence was introduced in [48]; for its definition and properties, we refer to [19, 22, 32].

Proof. The almost sure convergence of $\overline{\boldsymbol{\xi}}_N$ to $\boldsymbol{\psi}_\infty$ follows by the usual martingale arguments (see, for instance, [9, Lemma 2]) because $E\big[\boldsymbol{\xi_{n+1}}|\mathcal{F}_n\big] = \boldsymbol{\psi_n} \to \boldsymbol{\psi}_\infty$ almost surely and $\sum_{n\geq 1} E\big[\|\boldsymbol{\xi_n}\|^2\big]n^{-2} \leq \sum_{n\geq 1} n^{-2} < +\infty$.

Regarding the CLTs, we observe that, by means of (11), we can write

$$\boldsymbol{\psi_{n+1}} - \boldsymbol{\psi_n} = \frac{\boldsymbol{H}(\boldsymbol{\psi_n})}{r^*_{n+1}} + \frac{\Delta \boldsymbol{M}_{n+1}}{r^*_{n+1}}, \tag{25}$$

where $\boldsymbol{H}(\boldsymbol{x}) = (\beta - 1)|\boldsymbol{b_0}|(\boldsymbol{x} - \boldsymbol{p_0})$ and $\Delta \boldsymbol{M}_{n+1} = \alpha\big(\boldsymbol{\xi_{n+1}} - \boldsymbol{\psi_n}\big)$. Therefore, we get

$$\begin{aligned}\sqrt{N}\big(\overline{\boldsymbol{\xi}}_N - \boldsymbol{\psi}_N\big) &= \frac{1}{\sqrt{N}}\big(N\overline{\boldsymbol{\xi}}_N - N\boldsymbol{\psi}_N\big) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\big[\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}} + n\big(\boldsymbol{\psi_{n-1}} - \boldsymbol{\psi_n}\big)\big] \\ &= \sum_{n=1}^{N} \boldsymbol{Y}_{N,\boldsymbol{n}} + \boldsymbol{Q}_N,\end{aligned}$$

where

$$\boldsymbol{Y}_{N,\boldsymbol{n}} = \frac{\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}}{\sqrt{N}} = \frac{\Delta \boldsymbol{M_n}}{\alpha\sqrt{N}}$$

and

$$\boldsymbol{Q}_N = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} n\big(\boldsymbol{\psi_{n-1}} - \boldsymbol{\psi_n}\big) = -\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\frac{n}{r^*_n}\big(\boldsymbol{H}\big(\boldsymbol{\psi_{n-1}}\big) + \Delta \boldsymbol{M_n}\big).$$

Since $\sum_{n\geq 1} n/r^*_n < +\infty$ and $|\boldsymbol{H}\big(\boldsymbol{\psi_{n-1}}\big)| + |\Delta \boldsymbol{M_n}|$ is uniformly bounded by a constant, we have that $\boldsymbol{Q}_N$ converges to zero almost surely. Therefore it is enough to prove that $\sum_{n=1}^{N} \boldsymbol{Y}_{N,\boldsymbol{n}}$ converges stably to the desired Gaussian kernel. For this purpose we observe that $E\big[\boldsymbol{Y}_{N,\boldsymbol{n}}|\mathcal{F}_{n-1}\big] = \boldsymbol{0}$, and so, to prove the stable convergence, we have to check the following conditions (see [22, Corollary 7] or [19, Corollary 5.5.2]):

(c1) $E\big[\max_{1\leq n\leq N}|\boldsymbol{Y}_{N,\boldsymbol{n}}|\big] \to 0$, and

(c2) $\sum_{n=1}^{N} \boldsymbol{Y}_{N,\boldsymbol{n}}\boldsymbol{Y}_{N,\boldsymbol{n}}^\top \xrightarrow{P} \Sigma^2$.

Regarding (c1), we observe that

$$\max_{1\leq n\leq N}|\boldsymbol{Y}_{N,\boldsymbol{n}}| \leq \frac{1}{\sqrt{N}}\max_{1\leq n\leq N}|\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}| \leq \frac{1}{\sqrt{N}} \to 0.$$

In order to conclude, we have to prove the condition (c2), that is,

$$\sum_{n=1}^{N} \boldsymbol{Y}_{N,\boldsymbol{n}}\boldsymbol{Y}_{N,\boldsymbol{n}}^\top = \frac{1}{N}\sum_{n=1}^{N}\big(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big)\big(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big)^\top \xrightarrow{P} \Sigma^2.$$

The above convergence holds true even almost surely by standard martingale arguments (see, for instance, [9, Lemma 2]). Indeed, we have $\sum_{n\geq 1} E\Big[\big\|\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big\|^2\Big]/n^2 \leq \sum_{n\geq 1} n^{-2} < +\infty$ and

$$E\Big[\big(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big)\big(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big)^\top|\mathcal{F}_{n-1}\Big] = \mathrm{diag}\big(\boldsymbol{\psi_{n-1}}\big) - \boldsymbol{\psi_{n-1}}\boldsymbol{\psi_{n-1}}^\top \xrightarrow{a.s} \Sigma^2.$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

The last stable convergence follows from the equality

$$\sqrt{N}\big(\overline{\boldsymbol{\xi}}_N - \boldsymbol{\psi}_\infty\big) = \sqrt{N}\big(\overline{\boldsymbol{\xi}}_N - \boldsymbol{\psi}_N\big) + \sqrt{N}\big(\boldsymbol{\psi}_N - \boldsymbol{\psi}_\infty\big),$$

where the last term converges almost surely to zero by Theorem 3. □

Remark 2. Equation (25) implies that the bounded stochastic process $\boldsymbol{\psi} = (\boldsymbol{\psi_n})_n$ is a positive (i.e. nonnegative) almost supermartingale [49] and also a quasi-martingale [39], because $\boldsymbol{H}(\boldsymbol{\psi_n})$ is uniformly bounded by a constant and $\sum_{n\geq 1} 1/r^*_{n+1} < +\infty$.

3.3. Trivial cases

We here consider the two elementary cases $\beta = 0$ and $\alpha = 0$.

3.3.1. *The case* $\beta = 0$. In this case, by (7), (8), and (9), we have for all $i = 1, \dots, k$

$$\psi_{0\,i} = \frac{b_{0\,i} + B_{0\,i}}{|\boldsymbol{b_0}| + |\boldsymbol{B_0}|} \qquad \text{and} \qquad \psi_{n\,i} = \frac{b_{0\,i} + \alpha\xi_{n\,i}}{|\boldsymbol{b_0}| + \alpha} \qquad \text{for } n \geq 1. \tag{26}$$

We now focus on $\boldsymbol{\psi_n}$ for $n \geq 1$. The process $(\boldsymbol{\psi_n})_{n\geq 1}$ is a k-dimensional Markov chain with a finite state space $S = \{\boldsymbol{s_1}, \dots, \boldsymbol{s_k}\}$, where

$$\boldsymbol{s_i} = \frac{1}{|\boldsymbol{b_0}| + \alpha}\big(b_{0\,1}, \dots, b_{0\,i} + \alpha, \dots, b_{0\,k}\big)^\top, \qquad \text{for } i = 1, \dots, k\,,$$

and transition probability matrix

$$P = \frac{1}{|\boldsymbol{b_0}| + \alpha}\big(\mathbf{1}_k\,\boldsymbol{b_0}^\top + \alpha \mathrm{Id}_k\big) = \frac{|\boldsymbol{b_0}|}{|\boldsymbol{b_0}| + \alpha}\Big(\mathbf{1}_k\,\boldsymbol{p_0}^\top + \tfrac{\alpha}{|\boldsymbol{b_0}|}\mathrm{Id}_k\Big)\,,$$

which is irreducible and aperiodic. Now, since $\mathbf{1}_k\,\boldsymbol{p_0}^\top$ is idempotent and commutes with the identity, we have

$$\begin{aligned}
P^n &= \left(\frac{|\boldsymbol{b_0}|}{|\boldsymbol{b_0}| + \alpha}\right)^n \left(\left(\sum_{j=0}^{n-1}\binom{n}{j}\left(\tfrac{\alpha}{|\boldsymbol{b_0}|}\right)^j\right)\mathbf{1}_k\,\boldsymbol{p_0}^\top + \left(\tfrac{\alpha}{|\boldsymbol{b_0}|}\right)^n \mathrm{Id}_k\right) \\
&= \left(\frac{|\boldsymbol{b_0}|}{|\boldsymbol{b_0}| + \alpha}\right)^n \left(\left(\left(1 + \tfrac{\alpha}{|\boldsymbol{b_0}|}\right)^n - \left(\tfrac{\alpha}{|\boldsymbol{b_0}|}\right)^n\right)\mathbf{1}_k\,\boldsymbol{p_0}^\top + \left(\tfrac{\alpha}{|\boldsymbol{b_0}|}\right)^n \mathrm{Id}_k\right) \\
&= \mathbf{1}_k\,\boldsymbol{p_0}^\top + \left(\frac{\alpha}{|\boldsymbol{b_0}| + \alpha}\right)^n \Big(\mathrm{Id}_k - \mathbf{1}_k\,\boldsymbol{p_0}^\top\Big) \\
&= \mathbf{1}_k\,\boldsymbol{p_0}^\top + \gamma^n\Big(\mathrm{Id}_k - \mathbf{1}_k\,\boldsymbol{p_0}^\top\Big)\,,
\end{aligned} \tag{27}$$

where γ is the constant given in (5), which becomes equal to $\frac{\alpha}{|\boldsymbol{b_0}|+\alpha}$ for $\beta = 0$. We note that $\gamma < 1$ (since $|\boldsymbol{b_0}| > 0$ by assumption) and so $P^n \to \mathbf{1}_k\,\boldsymbol{p_0}^\top$, and the unique invariant probability measure on S is hence $\boldsymbol{\pi} = \boldsymbol{p_0}$.

Theorem 5. *We have* $\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \boldsymbol{p_0}$ *and*

$$\sqrt{N}\big(\overline{\boldsymbol{\xi}}_N - \boldsymbol{p_0}\big) = \frac{\sum_{n=1}^N \big(\boldsymbol{\xi_n} - \boldsymbol{p_0}\big)}{\sqrt{N}} \xrightarrow[N\to\infty]{d} \mathcal{N}\big(0, \Sigma^2\big),$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

where

$$\Sigma^2 = \lambda\Big(\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^\top\Big), \tag{28}$$

with λ *defined in* (6) *(taking* $\beta = 0$*).*

Proof. We observe that, by (26), we have for each $n \geq 0$

$$\{\xi_{n+1\,i} = 1\} = \{\boldsymbol{\psi_{n+1}} = \boldsymbol{s_i}\}.$$

Therefore, the strong law of large numbers for Markov chains immediately yields

$$\overline{\boldsymbol{\xi}}_N = \frac{1}{N}\sum_{n=1}^{N}\big(\mathbb{1}_{\{\boldsymbol{\psi_n}=\boldsymbol{s_1}\}}, \ldots, \mathbb{1}_{\{\boldsymbol{\psi_n}=\boldsymbol{s_k}\}}\big)^\top \xrightarrow{a.s.} \boldsymbol{p_0}.$$

Take a vector $\boldsymbol{c} = (c_1, \ldots, c_k)^T$ and define $g(\boldsymbol{x}) = \boldsymbol{c}^T(\boldsymbol{x} - \boldsymbol{p_0})$. Recall that $g(\boldsymbol{\xi}_n) = g\big(\mathbb{1}_{\{\boldsymbol{\psi_n}=\boldsymbol{s_1}\}}, \ldots, \mathbb{1}_{\{\boldsymbol{\psi_n}=\boldsymbol{s_k}\}}\big)$ and apply the CLT for uniformly ergodic Markov chains (see, for instance, [40, Theorem 17.0.1]): the sequence

$$\left(\frac{\sum_{n=1}^{N} g(\boldsymbol{\xi_n})}{\sqrt{N}}\right)$$

converges in distribution to the Gaussian distribution $\mathcal{N}\big(0, \sigma_{\boldsymbol{c}}^2\big)$, with

$$\begin{aligned}\sigma_{\boldsymbol{c}}^2 &= \mathrm{Var}\big[g\big(\boldsymbol{\xi_0^{(\pi)}}\big)\big] + 2\sum_{n\geq 1}\mathrm{Cov}\big(g\big(\boldsymbol{\xi_0^{(\pi)}}\big), g\big(\boldsymbol{\xi_n^{(\pi)}}\big)\big)\\ &= \boldsymbol{c}^T\left(\mathrm{Var}\big[\boldsymbol{\xi_0^{(\pi)}}\big] + 2\sum_{n\geq 1}\mathrm{Cov}\big(\boldsymbol{\xi_0^{(\pi)}}, \boldsymbol{\xi}_n^{(\pi)}\big)\right)\boldsymbol{c}\,,\end{aligned}$$

where

$$\boldsymbol{\xi_n^{(\pi)}} = \left(\mathbb{1}_{\left\{\boldsymbol{\psi_n^{(\pi)}}=\boldsymbol{s_1}\right\}}, \ldots, \mathbb{1}_{\left\{\boldsymbol{\psi_n^{(\pi)}}=\boldsymbol{s_k}\right\}}\right)$$

and $\big(\boldsymbol{\psi_n^{(\pi)}}\big)_{n\geq 0}$ is a Markov chain with transition matrix P and initial distribution $\boldsymbol{\pi}$, that is, $\boldsymbol{p_0}$. Now, by definition, $\boldsymbol{\xi_0^{(\pi)}}\boldsymbol{\xi_0^{(\pi)}}^T = \mathrm{diag}\big(\boldsymbol{\xi_0^{(\pi)}}\big)$, and hence $\mathrm{Var}\big[\boldsymbol{\xi_0^{(\pi)}}\big] = \mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^T$. Moreover, by means of (27),

$$E\Big[\boldsymbol{\xi_0^{(\pi)}}\boldsymbol{\xi_n^{(\pi)}}^\top\Big] = \mathrm{diag}(\boldsymbol{p_0})P^n = \boldsymbol{p_0}\boldsymbol{p_0}^\top + \gamma^n\Big(\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^\top\Big).$$

Hence, since $\gamma < 1$, we have $\sum_{n\geq 1}\gamma^n = \gamma/(1-\gamma)$ and so

$$\mathrm{Var}\Big[\boldsymbol{\xi_0^{(\pi)}}\Big] + 2\sum_{n\geq 1}\mathrm{Cov}\big(\boldsymbol{\xi_0^{(\pi)}}, \boldsymbol{\xi}_n^{(\pi)}\big) = \Big(\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^\top\Big)\left(1 + \frac{2\gamma}{1-\gamma}\right).$$

By the Cramér–Wold device, the theorem is proved with Σ^2 given in (28). □

Remark 3. Note that in Theorem 5 we do not assume $b_{0\,i} > 0$ for all i, but only $|\boldsymbol{b_0}| > 0$ (as stated in Section 2). A different behavior is observed when $\boldsymbol{b_0} = \boldsymbol{0}$. In this case, (26) gives $\boldsymbol{\psi_n} = \boldsymbol{\xi_n}$ for $n \geq 1$. Since $\psi_{n\,i} = P(\xi_{n+1\,i} = 1|\mathcal{F}_n)$, the above equality implies recursively $\boldsymbol{\psi_n} = \boldsymbol{\xi_n} = \boldsymbol{\xi_1}$ for each $n \geq 1$. In other words, the process of extractions $\boldsymbol{\xi} = (\boldsymbol{\xi_n})_{n\geq 1}$ is constant, with $P(\xi_{1\,i} = 1) = \psi_{0\,i} = B_{0\,i}/|\boldsymbol{B_0}|$.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

3.3.2. *The case* $\alpha = 0$. The model introduced above for $\alpha > 0$ also makes sense when $\alpha = 0$. For completeness, in this section we discuss this case. Recall that we are assuming $|\boldsymbol{b_0}| > 0$ and $b_{0\,i} + B_{0\,i} > 0$ (see Section 2). For the case $\beta > 1$, we here assume also $|\boldsymbol{B_0}| > 0$.

When $\alpha = 0$, the random vectors $\boldsymbol{\xi_n}$ are independent with

$$P(\xi_{n+1\,i} = 1) = \psi_{n\,i} = \frac{b_{0\,i} + \beta^n B_{0\,i}}{|\boldsymbol{b_0}| + \beta^n |\boldsymbol{B_0}|}.$$

Therefore, we have

$$\psi_{n\,i} = \frac{b_{0\,i} + B_{0\,i}}{|\boldsymbol{b_0}| + |\boldsymbol{B_0}|}$$

for all n if $\beta = 1$ (which corresponds to the classical multinomial model) and

$$\psi_{n\,i} \longrightarrow \begin{cases} \frac{b_{0\,i}}{|\boldsymbol{b_0}|} & \text{if } \beta \in [0,\,1), \\ \frac{B_{0\,i}}{|\boldsymbol{B_0}|} & \text{if } \beta > 1. \end{cases}$$

Moreover, the following result holds true.

Theorem 6. *We have*

$$\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \overline{\boldsymbol{\xi}}_\infty = \begin{cases} \frac{\boldsymbol{b_0}+\boldsymbol{B_0}}{|\boldsymbol{b_0}|+|\boldsymbol{B_0}|} & \textit{if } \beta = 1, \\ \frac{\boldsymbol{b_0}}{|\boldsymbol{b_0}|} = \boldsymbol{p_0} & \textit{if } \beta \in [0,\,1), \\ \frac{\boldsymbol{B_0}}{|\boldsymbol{B_0}|} & \textit{if } \beta > 1. \end{cases}$$

Moreover, we have

$$\sqrt{N}\left(\overline{\boldsymbol{\xi}}_N - \overline{\boldsymbol{\xi}}_\infty\right) \xrightarrow{s} \mathcal{N}\left(\boldsymbol{0},\,\Sigma^2\right),$$

where $\Sigma^2 = \mathrm{diag}(\boldsymbol{\psi}_\infty) - \boldsymbol{\psi}_\infty \boldsymbol{\psi}_\infty^\top$ *and* $\xrightarrow{s}$ *means stable convergence.*

Proof. The almost sure convergence follows from the Borel–Cantelli lemmas (see, for instance, [54, Section 12.15]). Indeed, we have the following:

- if $\sum_{n\geq 0} \psi_{n,i} < +\infty$, then $\sum_{n=1}^N \xi_{n\,i} \xrightarrow{a.s.} \xi_{\infty\,i}$, with $P(\xi_{\infty\,i} < +\infty) = 1$;
- if $\sum_{n\geq 0} \psi_{n,i} = +\infty$, then $\sum_{n=1}^N \xi_{n\,i} / \sum_{n=1}^N \psi_{n-1\,i} \xrightarrow{a.s.} 1$.

Hence, the statement of the theorem follows because

(i) if $\beta = 1$, then $\sum_{n\geq 0} \psi_{n\,i} = +\infty$ and

$$\sum_{n=1}^N \psi_{n-1\,i} \sim \frac{b_{0\,i} + B_{0\,i}}{|\boldsymbol{b_0}| + |\boldsymbol{B_0}|} N;$$

(ii) if $\beta \in [0,\,1)$ and $b_{0\,i} > 0$, then $\sum_{n\geq 0} \psi_{n\,i} = +\infty$ and

$$\sum_{n=1}^N \psi_{n-1\,i} \sim \frac{b_{0\,i}}{|\boldsymbol{b_0}|} N = p_{0\,i} N;$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

(iii) if $\beta \in [0, 1)$ and $b_{0\,i} = 0$, then

$$\sum_{n\geq 0} \psi_{n\,i} \leq \frac{B_{0\,i}}{|\boldsymbol{b_0}|} \sum_{n\geq 0} \beta^n < +\infty$$

and so $\sum_{n\geq 1} \xi_{n\,i} < +\infty$ almost surely; that is, $\xi_{n\,i} = 0$ eventually with probability one;

(iv) if $\beta > 1$ and $B_{0\,i} > 0$, then $\sum_{n\geq 0} \psi_{n\,i} = +\infty$ and $\sum_{n=1}^{N} \psi_{n-1\,i} \sim \frac{B_{0\,i}}{|\boldsymbol{B_0}|} N$;

(v) if $\beta > 1$ and $B_{0\,i} = 0$, then

$$\sum_{n\geq 0} \psi_{n\,i} \leq \frac{b_{0\,i}}{|\boldsymbol{B_0}|} \sum_{n\geq 0} \beta^{-n} < +\infty$$

and so $\sum_{n\geq 1} \xi_{n\,i} < +\infty$ almost surely; that is, $\xi_{n\,i} = 0$ eventually with probability one.

For the CLT we argue as in the proof of Theorem 4. Indeed, we set $\boldsymbol{Y_{N,n}} = \frac{\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}}{\sqrt{N}}$ so that we have

$$\sqrt{N}\left(\overline{\boldsymbol{\xi}}_N - \overline{\boldsymbol{\xi}}_\infty\right) = \sum_{n=1}^{N} \boldsymbol{Y_{N,n}} + \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(\boldsymbol{\psi_{n-1}} - \overline{\boldsymbol{\xi}}_\infty\right),$$

where the second term converges to zero because

$$\sum_{n=1}^{N} \left|\boldsymbol{\psi_{n-1}} - \overline{\boldsymbol{\xi}}_\infty\right| = \begin{cases} 0 & \text{if } \beta = 1, \\ O\left(\sum_{n=1}^{N} \beta^n\right) & \text{if } \beta \in [0, 1), \\ O\left(\sum_{n=1}^{N} \beta^{-n}\right) & \text{if } \beta > 1. \end{cases}$$

Therefore it is enough to prove that $\sum_{n=1}^{N} \boldsymbol{Y_{N,n}}$ converges stably to the desired Gaussian kernel. To this end we observe that $E[\boldsymbol{Y_{N,n}} | \mathcal{F}_{n-1}] = \boldsymbol{0}$, and so, to prove the stable convergence, we have to check the following conditions (see [22, Corollary 7] or [19, Corollary 5.5.2]):

(c1) $E\left[\max_{1\leq n\leq N} |\boldsymbol{Y_{N,n}}|\right] \to 0$, and

(c2) $\sum_{n=1}^{N} \boldsymbol{Y_{N,n}} \boldsymbol{Y_{N,n}}^\top \xrightarrow{P} \Sigma^2$.

Regarding (c1), we note that

$$\max_{1\leq n\leq N} |\boldsymbol{Y_{N,n}}| \leq \frac{1}{\sqrt{N}} \max_{1\leq n\leq N} |\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}| \leq \frac{1}{\sqrt{N}} \to 0.$$

Regarding (c2), we observe that

$$\sum_{n=1}^{N} \boldsymbol{Y_{N,n}} \boldsymbol{Y_{N,n}}^\top = \frac{1}{N} \sum_{n=1}^{N} \left(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\right)\left(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\right)^\top \xrightarrow{a.s.} \Sigma^2,$$

because (see, for instance, [9, Lemma 2]) $\sum_{n\geq 1} E\left[\left\|\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\right\|^2\right]/n^2 \leq \sum_{n\geq 1} n^{-2} < +\infty$ and

$$E\left[\left(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\right)\left(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\right)^\top | \mathcal{F}_{n-1}\right] = \operatorname{diag}\left(\boldsymbol{\psi_{n-1}}\right) - \boldsymbol{\psi_{n-1}} \boldsymbol{\psi_{n-1}}^\top \xrightarrow{a.s} \Sigma^2.$$

□

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

4. Proof of the goodness-of-fit result (Theorem 1)

The proof is based on Theorem 2 (for $0<\beta<1$) and Theorem 5 (for $\beta=0$), whose proofs are in Subsections 3.1 and 3.3.1, respectively. The almost sure convergence of O_i/N immediately follows since $O_i/N=\overline{\xi}_{N\,i}$. In order to prove the stated convergence in distribution, we mimic the classical proof for the Pearson chi-squared test based on the Sherman–Morrison formula (see [51]); however, see also [47, Corollary 2].

Proof. We start by recalling the Sherman–Morrison formula: if A is an invertible square matrix and $1-\boldsymbol{v}^\top A^{-1}\boldsymbol{u}\neq 0$, then

$$\left(A-\boldsymbol{u}\boldsymbol{v}^\top\right)^{-1}=A^{-1}+\frac{A^{-1}\boldsymbol{u}\boldsymbol{v}^\top A^{-1}}{1-\boldsymbol{v}^\top A^{-1}\boldsymbol{u}}.$$

Given the observation $\boldsymbol{\xi_n}=(\xi_{n\,1},\dots,\xi_{n\,k})^\top$, we define the 'truncated' vector $\boldsymbol{\xi_n^*}=\left(\xi^*_{n\,1},\dots,\xi^*_{n\,k-1}\right)^\top$, given by the first $k-1$ components of $\boldsymbol{\xi_n}$. Theorem 2 (for $\beta\in(0,1)$) and Theorem 5 (for $\beta=0$) give the CLT for $(\boldsymbol{\xi_n})_n$, which immediately implies

$$\sqrt{N}\left(\overline{\boldsymbol{\xi}}^*_{\boldsymbol{N}}-\boldsymbol{p}^*\right)=\frac{\sum_{n=1}^N\left(\boldsymbol{\xi_n^*}-\boldsymbol{p}^*\right)}{\sqrt{N}}\xrightarrow{d}\mathcal{N}\left(\boldsymbol{0},\,\Sigma_*^2\right),\tag{29}$$

where $\boldsymbol{p}^*$ is given by the first $k-1$ components of $\boldsymbol{p_0}$ and

$$\Sigma_*^2=\lambda\left(\operatorname{diag}(\boldsymbol{p}^*)-\boldsymbol{p}^*\boldsymbol{p}^{*^T}\right).$$

By assumption $p_{0\,i}>0$ for all $i=1,\dots,k$ and so $\operatorname{diag}(\boldsymbol{p}^*)$ is invertible with inverse $\operatorname{diag}(\boldsymbol{p}^*)^{-1}=\operatorname{diag}\left(\frac{1}{p_{0\,1}},\dots,\frac{1}{p_{0\,k-1}}\right)$; since $\left(\operatorname{diag}(\boldsymbol{p}^*)^{-1}\right)\boldsymbol{p}^*=\boldsymbol{1}\in\mathbb{R}^{k-1}$, we have

$$1-\boldsymbol{p}^{*^T}\operatorname{diag}(\boldsymbol{p}^*)^{-1}\boldsymbol{p}^*=1-\sum_{i=1}^{k-1}p_{0\,i}=\sum_{i=1}^{k}p_{0\,i}-\sum_{i=1}^{k-1}p_{0\,i}=p_{0\,k}>0.$$

Therefore we can use the Sherman–Morrison formula with $A=\operatorname{diag}(\boldsymbol{p}^*)$ and $\boldsymbol{u}=\boldsymbol{v}=\boldsymbol{p}^*$, and we obtain

$$\left(\Sigma_*^2\right)^{-1}=\frac{1}{\lambda}\left(\operatorname{diag}(\boldsymbol{p}^*)-\boldsymbol{p}^*\boldsymbol{p}^{*^T}\right)^{-1}=\frac{1}{\lambda}\left(\operatorname{diag}\left(\tfrac{1}{p_{0\,1}},\dots,\tfrac{1}{p_{0\,k-1}}\right)+\frac{1}{p_{0\,k}}\boldsymbol{1}\boldsymbol{1}^\top\right).\tag{30}$$

Now, since $\sum_{i=1}^k\left(\overline{\xi}_{N\,i}-p_{0\,i}\right)=0$, we have $\overline{\xi}_{N\,k}-p_{0\,k}=-\sum_{i=1}^{k-1}\left(\overline{\xi}_{N\,i}-p_{0\,i}\right)$, and so we get

$$\begin{aligned}\sum_{i=1}^k\frac{\left(O_i-Np_{0\,i}\right)^2}{Np_{0\,i}}&=N\sum_{i=1}^k\frac{\left(\overline{\xi}_{N\,i}-p_{0\,i}\right)^2}{p_{0\,i}}=N\left[\sum_{i=1}^{k-1}\frac{\left(\overline{\xi}_{N\,i}-p_{0\,i}\right)^2}{p_{0\,i}}+\frac{\left(\overline{\xi}_{N\,k}-p_{0\,k}\right)^2}{p_{0\,k}}\right]\\&=N\left[\sum_{i=1}^{k-1}\frac{\left(\overline{\xi}_{N\,i}-p_{0\,i}\right)^2}{p_{0\,i}}+\frac{\left(\sum_{i=1}^{k-1}\left(\overline{\xi}_{N\,i}-p_{0\,i}\right)\right)^2}{p_{0\,k}}\right]\\&=N\sum_{i_1,i_2=1}^{k-1}\left(\overline{\xi}_{N\,i_1}-p_{0\,i_1}\right)\left(\overline{\xi}_{N\,i_2}-p_{0\,i_2}\right)\left(\delta_{i_1}^{i_2}\frac{1}{p_{0\,i_1}}+\frac{1}{p_{0\,k}}\right),\end{aligned}$$

where $\delta_{i_1}^{i_2}$ is equal to 1 if $i_1=i_2$ and equal to zero otherwise. Finally, from the above equalities, recalling (29) and (30), we obtain

$$\sum_{i=1}^k\frac{\left(O_i-Np_{0\,i}\right)^2}{Np_{0\,i}}=\lambda N\left(\overline{\boldsymbol{\xi}}^*_{\boldsymbol{N}}-\boldsymbol{p}^*\right)^\top\left(\Sigma_*^2\right)^{-1}\left(\overline{\boldsymbol{\xi}}^*_{\boldsymbol{N}}-\boldsymbol{p}^*\right)\xrightarrow{d}\lambda W_0=W_*,$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

where W_0 is a random variable with distribution $\chi^2(k-1) = \Gamma((k-1)/2, 1/2)$, where $\Gamma(a, b)$ denotes the gamma distribution with density function

$$f(w) = \frac{b^a}{\Gamma(a)} w^{a-1} e^{-bw}.$$

As a consequence, W_* has distribution $\Gamma((k-1)/2, 1/(2\lambda))$. □

Appendix A

A.1. A central limit theorem for a multidimensional compact Markov chain

In this section we prove the general CLT for Markov chains, used for the proof of Theorem 2.

Let (S, d) be a *compact metric space* and denote by $C(S)$ the space of continuous real functions on S, by $Lip(S)$ the space of Lipschitz continuous real functions on S, and by $Lip(S \times S)$ the space of Lipschitz continuous real functions on $S \times S$. Moreover, we define $\|f\|_\infty = \sup_{x \in S} |f(x)|$ for each f in $C(S)$ and, for each f in $Lip(S)$,

$$|f|_{Lip} = \sup_{x,y \in S,\, x \neq y} \frac{|f(y) - f(x)|}{d(x, y)} \qquad \text{and} \qquad \|f\|_{Lip} = |f|_{Lip} + \|f\|_\infty.$$

Let $P(x, dy)$ be a *Markovian kernel* on S and set $(Pf)(x) = \int_S f(y) P(x, dy)$. We now recall some definitions and results concerning Markov chains with values in S.

Definition 1. We say that P is *weak Feller* if $(Pf)(x) = \int_S f(y) P(x, dy)$ defines a linear operator $P : C(S) \to C(S)$. A Markov chain with a weak Feller transition kernel is called a weak Feller Markov chain.

Remark 4. If P is weak Feller, then the sequence $(P^n)_{n \geq 1}$ of operators from $C(S)$ to $C(S)$ is uniformly bounded with respect to $\|\cdot\|_\infty$: indeed, we simply have

$$\begin{aligned} \|P^n f\|_\infty = \sup_{x \in S} |P^n f(x)| &= \sup_{x \in S} \left| \int_S f(y) P^n(x, dy) \right| \\ &\leq \sup_{x \in S} \left(\int_S \sup_{y \in S} |f(y)| P^n(x, dy) \right) = \sup_{y \in S} |f(y)| = \|f\|_\infty. \end{aligned}$$

Moreover, the existence of at least one invariant probability measure for P is easily shown. In fact, the set of probability measures $\mathcal{P}(S)$ on S, endowed with the topology of weak convergence, is a compact convex set. In addition, the adjoint operator of P, namely

$$P^* : \mathcal{P}(S) \to \mathcal{P}(S), \qquad (P^* \nu)(B) = \int_S \nu(dx) P(x, B),$$

is continuous on $\mathcal{P}(S)$ (since P is weak Feller). Then, the existence of an invariant probability measure π is a consequence of Brouwer's fixed-point theorem.

Definition 2. We say that P is *semi-contractive* or a *semi-contraction* on *Lip(S)* if it maps *Lip(S)* into itself and there exists a constant $\gamma < 1$ such that

$$|Pf|_{Lip} \leq \gamma |f|_{Lip}$$

for each $f \in Lip(S)$.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

We now give the definition of *compact Markov chain* (see [42, Chapter 3] for a general exposition of the theory of these processes, and [24] for the beginning of this theory).

Definition 3. We say that P is a *Doeblin–Fortet operator* if it is weak Feller and a bounded operator from $(Lip(S), \|\cdot\|_{Lip})$ into itself, and there are finite constants $n_0 \geq 1$, $\gamma < 1$, and $R \geq 0$ such that

$$|P^{n_0} f|_{Lip} \leq \gamma |f|_{Lip} + R\|f\|_\infty$$

for each $f \in Lip(S)$. A Markov chain with a Doeblin–Fortet operator on a compact set S is called *compact Markov* chain (or *process*).

Note that the Doeblin–Fortet operator, the weak Feller property, and the semi-contraction may also be defined for a non-compact state space. In general, a compact Markov process is a Doeblin–Fortet process in a compact state space. In our framework, since S is compact, the two concepts coincide and the result below follows immediately.

Lemma 1. *If P is weak Feller and a semi-contractive operator on $Lip(S)$, then P is a Doeblin–Fortet operator. In other words, a weak Feller Markov chain such that its transition kernel is semi-contractive on $Lip(S)$ is a compact Markov chain.*

Definition 4. We say that P is *irreducible* and *aperiodic* if

$$Pf = e^{i\theta} f, \quad \text{with } \theta \in \mathbb{R}, \quad f \in Lip(S) \Rightarrow e^{i\theta} = 1, \quad \text{and} \quad f = \text{constant}.$$

A Markov chain with an irreducible and aperiodic transition kernel is called an irreducible and aperiodic Markov chain.

Under the hypotheses of the theorem of Ionescu-Tulcea and Marinescu in [35], the spectral radius of P is 1, the set of eigenvalues of P of modulus 1 has only a finite number of elements, and each relative eigenspace is finite-dimensional. This theorem can always be applied to a compact Markov chain (see [42, Theorem 3.3.1]). More specifically, every compact Markov chain has d disjoint closed sets, called *ergodic sets*, contained in its compact state space S. These sets are both the support of the base of the ergodic invariant probability measures, and the support of a base of the eigenspaces related to the eigenvalues of modulus 1 (see [42, Theorem 3.4.1]). In addition, each of these ergodic sets may be subdivided into p_j closed disjoint subsets. The number p_j is the period of the jth irreducible component, and the ergodic subdivision gives the support of the eigenfunctions related to the p_j roots of 1 (see [42, Theorem 3.5.1]). Then, as also explained in [42, Section 3.6], there are no other eigenvalues of modulus 1 except 1 (aperiodicity) and no other eigenfunctions except the constant for the eigenvalue equal to 1 (irreducibility) if and only if the compact Markov chain has but one ergodic kernel, and this kernel has period 1. In other words, the following result holds true.

Theorem 7. *Let $\psi = (\psi_n)_{n\geq 0}$ be a compact Markov chain and let π an invariant probability measure with respect to its transition kernel. If $\psi = (\psi_n)_{n\geq 0}$ converges in distribution to π, regardless of its initial distribution, then π is the unique invariant probability measure and ψ is irreducible and aperiodic.*

We now note that if P is Doeblin–Fortet, irreducible, and aperiodic, then it satisfies all the conditions given in [30, Définition 0] and [30, Définition 1]. Therefore, it has a *unique* invariant probability measure π, and for any $f \in Lip(S \times S)$, there exists a unique (up to a constant) function $u_f \in Lip(S)$ such that

$$u_f(x) - Pu_f(x) = \int_S f(x, y)P(x, dy) - \int_S \int_S f(x, y)P(x, dy)\pi(dx).$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

By means of this function u_f, it is possible to define the (unique) function $f'(x, y) = f(x, y) + u_f(y) - u_f(x)$ so that we have

$$m(f) = \int_S \int_S f(x, y) P(x, dy) \pi(dx) = \int_S \int_S f'(x, y) P(x, dy) \pi(dx) = m(f').$$

In addition, we may define the quantity $\sigma^2(f) \geq 0$ as follows (see [30, Equation (6)]):

$$\begin{aligned}\sigma^2(f) &= \int_S \int_S \big[f'(x, y) - m(f')\big]^2 P(x, dy) \pi(dx) \\ &= \int_S \int_S \big[f(x, y) - m(f) + u_f(y) - u_f(x)\big]^2 P(x, dy) \pi(dx). \quad \text{(A.1)}\end{aligned}$$

Finally, we have the following convergence result.

Theorem 8. ([30, Théorème 1 and Théorème 2]) *Let $\psi = (\psi_n)_{n \geq 0}$ be an irreducible and aperiodic compact Markov chain and denote by π its unique invariant probability measure. Let $f \in Lip(S \times S)$ such that $m(f) = 0$ and $\sigma^2(f) > 0$. Then, setting $S_N(f) = \sum_{n=0}^{N-1} f(\psi_n, \psi_{n+1})$, we have*

$$\frac{S_N(f)}{\sqrt{N}} \xrightarrow[N \to \infty]{d} \mathcal{N}\big(0, \sigma^2(f)\big),$$

and

$$\sup_t \Big| P\big(S_N(f) < t\sqrt{N}\big) - \mathcal{N}\big(0, \sigma^2(f)\big)(-\infty, t) \Big| = O\big(1/\sqrt{N}\big).$$

Now let us specialize our assumptions, taking as S a compact subset of $\mathbb{R}^k$. Therefore, in the sequel we will use boldface to highlight the fact that we are working with vectors.

Definition 5. (*'Linearity' condition.*) We say that P and $\boldsymbol{f} : S \times S \to \mathbb{R}^d$ form a *linear model* if $\boldsymbol{f}$ is linear (in $\boldsymbol{x}$ and $\boldsymbol{y}$) with $m(\boldsymbol{f}) = \boldsymbol{0}$ and the function

$$(P\boldsymbol{y})(\boldsymbol{x}) = \int_S \boldsymbol{y} P(\boldsymbol{x}, \boldsymbol{dy})$$

is linear (in $\boldsymbol{x}$).

Remark 5. Denote by $\boldsymbol{p_0} = \int_S \boldsymbol{x} \pi(\boldsymbol{dx})$ the mean value under the invariant probability measure π of P. If P and $\boldsymbol{f}$ form a linear model, then there exist two matrices $A_1, A_2 \in \mathbb{R}^{d \times k}$ such that

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = A_1(\boldsymbol{x} - \boldsymbol{p_0}) + A_2(\boldsymbol{y} - \boldsymbol{p_0}) \quad \text{(A.2)}$$

and a square matrix $A_P \in \mathbb{R}^{k \times k}$ such that

$$(P(\boldsymbol{y} - \boldsymbol{p_0}))(\boldsymbol{x}) = \int_S (\boldsymbol{y} - \boldsymbol{p_0}) P(\boldsymbol{x}, \boldsymbol{dy}) = A_P(\boldsymbol{x} - \boldsymbol{p_0}). \quad \text{(A.3)}$$

Indeed, if $(P\boldsymbol{y})(\boldsymbol{x}) = A_P \boldsymbol{x} + \boldsymbol{b}$, using that π is invariant with respect to P, we obtain

$$\boldsymbol{p_0} = \int_S \boldsymbol{y} \pi(\boldsymbol{dy}) = \int_S \int_S \boldsymbol{y} P(\boldsymbol{x}, \boldsymbol{dy}) \pi(\boldsymbol{dx}) = \int_S [A_P \boldsymbol{x} + \boldsymbol{b}] \pi(\boldsymbol{dx}) = A_P \boldsymbol{p_0} + \boldsymbol{b},$$

and hence $(P(\boldsymbol{y} - \boldsymbol{p_0}))(\boldsymbol{x}) = A_P(\boldsymbol{x} - \boldsymbol{p_0})$. Moreover, if $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = A_1 \boldsymbol{x} + A_2 \boldsymbol{y} + \boldsymbol{b}$, then

$$\begin{aligned} m(A_1 \boldsymbol{x} + A_2 \boldsymbol{y} + \boldsymbol{b}) &= \int_S P(\boldsymbol{x}, \boldsymbol{dy}) \int_S A_1 \boldsymbol{x} \pi(\boldsymbol{dx}) + \int_S A_2 \boldsymbol{y} \pi(\boldsymbol{dy}) + \boldsymbol{b} \\ &= (A_1 + A_2) \boldsymbol{p_0} + \boldsymbol{b}, \end{aligned}$$

and hence, if $m(\boldsymbol{f}) = \boldsymbol{0}$, we obtain $\boldsymbol{f} = A_1(\boldsymbol{x} - \boldsymbol{p_0}) + A_2(\boldsymbol{y} - \boldsymbol{p_0})$.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

Theorem 9. *Let $\psi = (\psi_n)_{n\geq 0}$ be an irreducible and aperiodic compact Markov chain, and denote by P its transition kernel and by π its unique invariant measure. Assume that P and $\boldsymbol{f}$ form a linear model and let A_1, A_2, and A_P be defined as in (A.2) and in (A.3). Then, setting $\boldsymbol{S}_N(\boldsymbol{f}) = \sum_{n=0}^{N-1} \boldsymbol{f}(\boldsymbol{\psi_n}, \boldsymbol{\psi_{n+1}})$, we have*

$$\frac{\boldsymbol{S}_N(\boldsymbol{f})}{\sqrt{N}} \xrightarrow[N\to\infty]{d} \mathcal{N}\big(\boldsymbol{0}, \Sigma^2\big),$$

where

$$\Sigma^2 = D_1\Sigma_\pi^2 D_1^\top + D_1\Sigma_\pi^2 A_P^\top D_2^\top + D_2 A_P \Sigma_\pi^2 D_1^\top + D_2\Sigma_\pi^2 D_2^\top,$$

with

$$\Sigma_\pi^2 = \int_S (\boldsymbol{x}-\boldsymbol{p_0})(\boldsymbol{x}-\boldsymbol{p_0})^\top \pi(\boldsymbol{dx})$$

(the variance–covariance matrix under the invariant probability measure π),

$$D_1 = A_1 - D_0, \quad \text{and} \quad D_2 = A_2 + D_0,$$

where $D_0 = (A_1 + A_2A_P)(Id - A_P)^{-1}$. Moreover, for any $\boldsymbol{c} \in \mathbb{R}^k$,

$$\sup_t \Big| P\big(\boldsymbol{S}_N\big(\boldsymbol{c}^\top \boldsymbol{f}\big) < t\sqrt{N}\big) - \mathcal{N}\big(0, \boldsymbol{c}^\top \Sigma^2 \boldsymbol{c}\big)(-\infty, t)\Big| = O\big(1/\sqrt{N}\big).$$

Proof. As a consequence of Definition 2, the spectral radius of A_P must be less than one, and hence $Id - A_P$ is invertible. Therefore, we may define

$$\boldsymbol{u_f}(\boldsymbol{x}) = D_0(\boldsymbol{x}-\boldsymbol{p_0}) = (A_1 + A_2A_P)(Id - A_P)^{-1}(\boldsymbol{x}-\boldsymbol{p_0}),$$

so that we have

$$\begin{aligned}
\boldsymbol{u_f}(\boldsymbol{x}) - (P\boldsymbol{u_f})(\boldsymbol{x}) &= (A_1 + A_2A_P)(Id - A_P)^{-1}(\boldsymbol{x}-\boldsymbol{p_0}) \\
&\qquad - (A_1 + A_2A_P)(Id - A_P)^{-1}A_P(\boldsymbol{x}-\boldsymbol{p_0}) \\
&= (A_1 + A_2A_P)(\boldsymbol{x}-\boldsymbol{p_0}) \\
&= A_1(\boldsymbol{x}-\boldsymbol{p_0})\int_S P(\boldsymbol{x}, \boldsymbol{dy}) + A_2\int_S (\boldsymbol{y}-\boldsymbol{p_0})P(\boldsymbol{x}, \boldsymbol{dy}) \\
&= \int_S \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})P(\boldsymbol{x}, \boldsymbol{dy}) - \boldsymbol{0} \\
&= \int_S \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})P(\boldsymbol{x}, \boldsymbol{dy}) - \int_S\int_S \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})P(\boldsymbol{x}, \boldsymbol{dy})\pi(\boldsymbol{dx}).
\end{aligned}$$

We immediately get that the function $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{u_f}(\boldsymbol{y}) - \boldsymbol{u_f}(\boldsymbol{x})$ is linear and it may be written as $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = D_1(\boldsymbol{x}-\boldsymbol{p_0}) + D_2(\boldsymbol{y}-\boldsymbol{p_0})$. Taking into account that

$$\begin{aligned}
\int_S\int_S (\boldsymbol{y}-\boldsymbol{p_0})P(\boldsymbol{x}, \boldsymbol{dy})\pi(\boldsymbol{dx}) &= \int_S (\boldsymbol{y}-\boldsymbol{p_0})\pi(\boldsymbol{dy}) = \boldsymbol{0}, \\
\int_S\int_S (\boldsymbol{y}-\boldsymbol{p_0})(\boldsymbol{y}-\boldsymbol{p_0})^\top P(\boldsymbol{x}, \boldsymbol{dy})\pi(\boldsymbol{dx}) &= \int_S (\boldsymbol{y}-\boldsymbol{p_0})(\boldsymbol{y}-\boldsymbol{p_0})^\top \pi(\boldsymbol{dy}) = \Sigma_\pi^2,
\end{aligned} \tag{A.4}$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

and

$$\int_S \int_S (\boldsymbol{y}-\boldsymbol{p_0})(\boldsymbol{x}-\boldsymbol{p_0})^\top P(\boldsymbol{x},\boldsymbol{dy})\pi(\boldsymbol{dx}) = A_P \int_S (\boldsymbol{x}-\boldsymbol{p_0})(\boldsymbol{x}-\boldsymbol{p_0})^\top \pi(\boldsymbol{dx}) = A_P \Sigma_\pi^2, \tag{A.5}$$

we can compute the quantity

$$\begin{aligned}
&\int_S \int_S \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})^\top P(\boldsymbol{x},\boldsymbol{dy})\pi(\boldsymbol{dx}) \\
&\qquad = \int_S \int_S \big[D_1(\boldsymbol{x}-\boldsymbol{p_0}) + D_2(\boldsymbol{y}-\boldsymbol{p_0})\big] \\
&\qquad\qquad \big[D_1(\boldsymbol{x}-\boldsymbol{p_0}) + D_2(\boldsymbol{y}-\boldsymbol{p_0})\big]^\top P(\boldsymbol{x},\boldsymbol{dy})\pi(\boldsymbol{dx}) \\
&\qquad = D_1\Sigma_\pi^2 D_1^\top + D_1\Sigma_\pi^2 A_P^\top D_2^\top + D_2 A_P \Sigma_\pi^2 D_1^\top + D_2 \Sigma_\pi^2 D_2^\top \\
&\qquad = \Sigma^2.
\end{aligned}$$

By the Cramér–Wold device, the theorem is proven with Σ^2 given above if we prove that, for any $\boldsymbol{c}$,

$$\boldsymbol{c}^\top \frac{\boldsymbol{S_N(f)}}{\sqrt{N}} = \frac{\boldsymbol{S_N}\big(\boldsymbol{c}^\top \boldsymbol{f}\big)}{\sqrt{N}} \xrightarrow[N\to\infty]{d} \mathcal{N}\big(0, \boldsymbol{c}^\top \Sigma \boldsymbol{c}\big).$$

Therefore, in order to conclude, it is enough to note that the above convergence is a consequence of Theorem 8 with $f = \boldsymbol{c}^\top \boldsymbol{f}$. Indeed, by definition $f \in Lip(S \times S)$ and the function $u_f \in Lip(S)$ in (A.1) may be chosen as $u_f = \boldsymbol{c}^\top \boldsymbol{u_f}$, so that $m(f) = 0$ and $\sigma^2(f) = \boldsymbol{c}^\top \Sigma \boldsymbol{c}$. □

A.2. Coupling technique

The result proven in this subsection plays a relevant rôle in the proof of Theorem 2. Indeed, it shows that, under suitable assumptions, two stochastic processes can be 'coupled' in a suitable way, preserving their respective joint distributions.

Set $S^* = \{\boldsymbol{x} : x_i \geq 0,\ |\boldsymbol{x}| = 1\}$ (i.e. the standard (or probability) simplex in $\mathbb{R}^k$), and recall that $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_k}\}$ denotes the canonical base of $\mathbb{R}^k$. We have the following technical lemma.

Lemma 2. *There exist two measurable functions* $\boldsymbol{h^{(1)}}, \boldsymbol{h^{(2)}} : S^* \times S^* \times (0,1) \to \{\boldsymbol{e_1}, \ldots, \boldsymbol{e_k}\}$, *such that for any* $\boldsymbol{x}, \boldsymbol{y} \in S^*$,

$$\begin{aligned}
&\int_{(0,1)} \mathbb{1}_{\{\boldsymbol{h^{(1)}}(\boldsymbol{x},\boldsymbol{y},u)=\boldsymbol{e_i}\}} du = x_i, \quad \forall i = 1, \ldots, k, \\
&\int_{(0,1)} \mathbb{1}_{\{\boldsymbol{h^{(2)}}(\boldsymbol{x},\boldsymbol{y},u)=\boldsymbol{e_i}\}} du = y_i, \quad \forall i = 1, \ldots, k,
\end{aligned} \tag{A.6}$$

and

$$\int_{(0,1)} \mathbb{1}_{\{\boldsymbol{h^{(1)}}(\boldsymbol{x},\boldsymbol{y},u)\neq\boldsymbol{h^{(2)}}(\boldsymbol{x},\boldsymbol{y},u)\}} du \leq \frac{|\boldsymbol{x}-\boldsymbol{y}|}{2}. \tag{A.7}$$

As a consequence, we have

$$\int_{(0,1)} \big|\boldsymbol{h^{(1)}}(\boldsymbol{x},\boldsymbol{y},u) - \boldsymbol{h^{(2)}}(\boldsymbol{x},\boldsymbol{y},u)\big| du \leq |\boldsymbol{x}-\boldsymbol{y}|. \tag{A.8}$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

Proof. Given $\boldsymbol{x}, \boldsymbol{y} \in S^*$, define $\underline{\boldsymbol{xy}} = \boldsymbol{x} \wedge \boldsymbol{y}$ so that $\underline{xy}_i = \min(x_i, y_i)$. Set $u_0 = |\underline{\boldsymbol{xy}}| = \sum_{i=1}^k \min(x_i, y_i)$, and note that $0 \le u_0 \le 1$. Moreover, for any $i \in \{1, \dots, k\}$, set

$$A_{\underline{\boldsymbol{xy}}\, i} = \left\{ u : \sum_{j=1}^{i-1} \underline{xy}_j < u \le \sum_{j=1}^{i} \underline{xy}_j \right\},$$

$$A_{\boldsymbol{x}\, i} = \left\{ u : u_0 + \sum_{j=1}^{i-1} \left(x_j - \underline{xy}_j\right) < u \le u_0 + \sum_{j=1}^{i} \left(x_j - \underline{xy}_j\right) \right\},$$

$$A_{\boldsymbol{y}\, i} = \left\{ u : u_0 + \sum_{j=1}^{i-1} \left(y_j - \underline{xy}_j\right) < u \le u_0 + \sum_{j=1}^{i} \left(y_j - \underline{xy}_j\right) \right\},$$

and let

$$\boldsymbol{h}^{(1)}(\boldsymbol{x}, \boldsymbol{y}, u) = \boldsymbol{e_i} \qquad \text{if } u \in A_{\underline{\boldsymbol{xy}}\, i} \cup A_{\boldsymbol{x}\, i}, \quad \text{and}$$
$$\boldsymbol{h}^{(2)}(\boldsymbol{x}, \boldsymbol{y}, u) = \boldsymbol{e_i} \qquad \text{if } u \in A_{\underline{\boldsymbol{xy}}\, i} \cup A_{\boldsymbol{y}\, i}.$$

Observe that since $1 = u_0 + \sum_{i=1}^k \left(x_i - \underline{xy}_i\right) = u_0 + \sum_{i=1}^k \left(y_i - \underline{xy}_i\right)$, the equalities above uniquely define $\boldsymbol{h}^{(1)}, \boldsymbol{h}^{(2)}$ on the whole domain. Moreover, since $x_i = \underline{xy}_i + \left(x_i - \underline{xy}_i\right)$ and $y_i = \underline{xy}_i + \left(y_i - \underline{xy}_i\right)$, the two conditions collected in Equation (A.6) are satisfied.

To check (A.7), just note that $\boldsymbol{h}^{(1)}(\boldsymbol{x}, \boldsymbol{y}, u)$ is equal to $\boldsymbol{h}^{(2)}(\boldsymbol{x}, \boldsymbol{y}, u)$ on the set $\cup_i A_{\underline{\boldsymbol{xy}}\, i} = (0, u_0)$, and we have

$$\begin{aligned} 2(1 - u_0) &= \sum_{i=1}^k x_i + \sum_{i=1}^k y_i - 2\sum_{i=1}^k \underline{xy}_i = \sum_{i=1}^k (x_i + y_i - \min(x_i, y_i)) - \min(x_i, y_i) \\ &= \sum_{i=1}^k \max(x_i, y_i) - \min(x_i, y_i) = |\boldsymbol{x} - \boldsymbol{y}|. \end{aligned}$$

Finally, (A.8) follows immediately from (A.7) since $|\boldsymbol{h}^{(1)} - \boldsymbol{h}^{(2)}| \le 2$. □

Now we are ready to prove the following 'coupling result'.

Theorem 10. *Let $\boldsymbol{\psi}^{(1)} = \left(\boldsymbol{\psi}_n^{(1)}\right)_n$ and $\boldsymbol{\psi}^{(2)} = \left(\boldsymbol{\psi}_n^{(2)}\right)_n$ be two stochastic processes with values in S^* that evolve according to the following dynamics:*

$$\boldsymbol{\psi}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})} = a_0 \boldsymbol{\psi}_{\boldsymbol{n}}^{(\boldsymbol{\ell})} + a_1 \boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})} + \boldsymbol{l}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})}\left(\boldsymbol{\psi}_{\boldsymbol{n}}^{(\boldsymbol{\ell})}, \boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})}\right) + \boldsymbol{c}, \quad \ell = 1,\ 2, \tag{A.9}$$

where $a_0, a_1 \ge 0$, $\boldsymbol{c} \in \mathbb{R}^k$, $\boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})}$ are random variables taking values in $\{\boldsymbol{e_1}, \dots, \boldsymbol{e_k}\}$ and such that

$$\begin{aligned} &P\left(\boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})} = \boldsymbol{e_i} \,\middle|\, \boldsymbol{\psi}_{\boldsymbol{0}}^{(\boldsymbol{\ell})},\ \boldsymbol{\xi}_{\boldsymbol{1}}^{(\boldsymbol{\ell})}, \dots, \boldsymbol{\xi}_{\boldsymbol{n}}^{(\boldsymbol{\ell})}\right) = \\ &\qquad P\left(\boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})} = \boldsymbol{e_i} \,\middle|\, \boldsymbol{\psi}_{\boldsymbol{n}}^{(\boldsymbol{\ell})}\right) = \psi_{n\,i}^{(\ell)}, \quad \textit{for} i = 1, \dots, k, \end{aligned} \tag{A.10}$$

and $\boldsymbol{l}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})}$ are measurable functions such that $\left|\boldsymbol{l}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})}\right| = O\left(c_{n+1}^{(\ell)}\right)$. Then there exist two stochastic processes $\widetilde{\boldsymbol{\psi}}^{(\boldsymbol{\ell})} = \left(\widetilde{\boldsymbol{\psi}}_n^{(\ell)}\right)_{n \ge 0}$, $\ell = 1, 2$, evolving according to the dynamics

$$\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(\ell)} = a_0 \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(\ell)} + a_1 \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}+1}^{(\ell)} + \boldsymbol{l}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})}\left(\widetilde{\boldsymbol{\psi}}_n^{(\ell)}, \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}+1}^{(\ell)}\right) + \boldsymbol{c}, \quad \ell = 1,\ 2, \tag{A.11}$$

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

with $\widetilde{\boldsymbol{\psi}}_0^{(\ell)} = \boldsymbol{\psi}_{\mathbf{0}}^{(\boldsymbol{\ell})}$ *and*

$$P\Big(\widetilde{\boldsymbol{\xi}}_{n+1}^{(\ell)} = \boldsymbol{e_i}\Big|\boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{1})},\ \boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{2})}, \widetilde{\boldsymbol{\xi}}_1^{(1)}, \widetilde{\boldsymbol{\xi}}_1^{(2)} \ldots, \widetilde{\boldsymbol{\xi}}_n^{(1)}, \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(2)}\Big) = P\Big(\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}+1}^{(\ell)} = \boldsymbol{e_i}\Big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(\ell)}\Big) = \widetilde{\psi}_{n\,i}^{(\ell)}, \quad \textit{for} i = 1, \ldots, k, \tag{A.12}$$

and such that, for any $n \geq 0$*, we have*

$$E\Big[\Big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(2)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(1)}\Big|\Big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{m}}^{(1)}, \widetilde{\boldsymbol{\psi}}_{\boldsymbol{m}}^{(2)}, m \leq n\Big] \leq (a_0 + a_1)\big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}\big| + O\big(c_{n+1}^{(1)}\big) + O\big(c_{n+1}^{(2)}\big). \tag{A.13}$$

Remark 6. As a consequence, for each $\ell = 1,\ 2$, the two stochastic processes $\widetilde{\boldsymbol{\psi}}^{(\boldsymbol{\ell})}$ and $\widetilde{\boldsymbol{\xi}}^{(\boldsymbol{\ell})}$ have the same joint distribution of $\boldsymbol{\psi}^{(\boldsymbol{\ell})}$ and of $\boldsymbol{\xi}^{(\boldsymbol{\ell})}$, respectively. Indeed, $\widetilde{\boldsymbol{\psi}}_{\mathbf{0}}^{(\boldsymbol{\ell})} = \boldsymbol{\psi}_{\mathbf{0}}^{(\boldsymbol{\ell})}$, and by (A.9), (A.10), (A.11), and (A.12), the conditional distributions of $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n+1}}^{\boldsymbol{\ell}}$ given $\Big[\widetilde{\boldsymbol{\psi}}_{\mathbf{0}}^{(\boldsymbol{\ell})}, \ldots, \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(\boldsymbol{\ell})}\Big]$ and of $\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n+1}}^{\boldsymbol{\ell}}$ given $\Big[\widetilde{\boldsymbol{\psi}}_{\mathbf{0}}^{(\boldsymbol{\ell})}, \widetilde{\boldsymbol{\xi}}_{\mathbf{1}}^{(\boldsymbol{\ell})} \ldots, \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(\boldsymbol{\ell})}\Big]$ are the same as those of $\boldsymbol{\psi}_{\boldsymbol{n+1}}^{\boldsymbol{\ell}}$ given $\Big[\boldsymbol{\psi}_{\mathbf{0}}^{(\boldsymbol{\ell})}, \ldots, \boldsymbol{\psi}_{n}^{(\boldsymbol{\ell})}\Big]$ and of $\boldsymbol{\xi}_{\boldsymbol{n+1}}^{\boldsymbol{\ell}}$ given $\Big[\boldsymbol{\psi}_{\mathbf{0}}^{(\boldsymbol{\ell})}, \boldsymbol{\xi}_{\mathbf{1}}^{(\boldsymbol{\ell})} \ldots, \boldsymbol{\xi}_{\boldsymbol{n}}^{(\boldsymbol{\ell})}\Big]$, respectively.

Moreover, from the inequality (A.13), by recursion, we obtain

$$E\Big[\Big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(2)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(1)}\Big|\Big|\boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{1})}, \boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{2})}\Big] \leq (a_0 + a_1)^{n+1}\big|\boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{2})} - \boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{1})}\big| + O\Bigg(\sum_{j=1}^{n+1}\big(a_0 + a_1\big)^{n+1-j}\Big(c_j^{(1)} + c_j^{(2)}\Big)\Bigg). \tag{A.14}$$

Proof. We set $\widetilde{\boldsymbol{\psi}}_0^{(\ell)} = \boldsymbol{\psi}_{\mathbf{0}}^{(\boldsymbol{\ell})}$, for $\ell = 1,\ 2$, and we take a sequence $(U_n)_{n\geq 1}$ of i.i.d. (0, 1)-uniform random variables, independent of $\sigma\big(\boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{1})}, \boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{2})}\big)$. Then we take the two functions $\boldsymbol{h}^{(\mathbf{1})}, \boldsymbol{h}^{(\mathbf{2})}$ of Lemma 2, and for each ℓ and any $n \geq 0$ we recursively define

$$\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}+1}^{(\ell)} = \boldsymbol{h}^{(\boldsymbol{\ell})}\Big(\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}, \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)}, U_{n+1}\Big),$$
$$\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(\ell)} = a_0\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(\ell)} + a_1\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}+1}^{(\ell)} + \boldsymbol{l}_{\boldsymbol{n+1}}^{(\boldsymbol{\ell})}\Big(\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(\ell)}, \widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}+1}^{(\ell)}\Big) + \boldsymbol{c}.$$

Setting $\widetilde{\mathcal{F}}_n = \sigma\Big(\boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{1})}, \boldsymbol{\psi}_{\mathbf{0}}^{(\mathbf{2})}, U_1, \ldots, U_n\Big)$, we have that U_{n+1} is independent of $\widetilde{\mathcal{F}}_n$ and, by definition, $\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}}^{(\ell)}$ and $\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(\ell)}$ are $\widetilde{\mathcal{F}}_n$-measurable, for any $\ell = 1, 2$ and $n \geq 0$. Therefore, using the relation (A.6), we get, for any $\ell = 1,\ 2$, $n \geq 0$, and $i = 1, \ldots, k$,

$$P\Big(\widetilde{\boldsymbol{\xi}}_{\boldsymbol{n}+1}^{(\ell)} = \boldsymbol{e_i}\big|\widetilde{\mathcal{F}}_n\Big) = \int \mathbb{1}_{\big\{\boldsymbol{h}^{(\boldsymbol{\ell})}(\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}, \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)}, u) = \boldsymbol{e_i}\big\}}\, du = \widetilde{\psi}_{n\,i}^{(\ell)}.$$

This means that (A.11), together with (A.12), holds true. Finally, by the relation (A.8), we have

$$E\Big[\Big|\boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\mathbf{2})} - \boldsymbol{\xi}_{\boldsymbol{n+1}}^{(\mathbf{1})}\Big|\Big|\widetilde{\mathcal{F}}_n\Big] = \int_{(0,1)}\Big|\boldsymbol{h}^{(\mathbf{1})}\Big(\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}, \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)}, u\Big) - \boldsymbol{h}^{(\mathbf{2})}\Big(\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}, \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)}, u\Big)\Big|du \leq \big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)}\big|;$$

hence, subtracting (A.11) with $\ell = 2$ from the same relation with $\ell = 1$, we obtain

$$E\Big[\Big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(2)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}+1}^{(1)}\Big|\Big|\widetilde{\mathcal{F}}_n\Big] \leq a_0\big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}\big| + a_1\big|\widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(2)} - \widetilde{\boldsymbol{\psi}}_{\boldsymbol{n}}^{(1)}\big| + O\big(c_{n+1}^{(1)}\big) + O\big(c_{n+1}^{(2)}\big),$$

and so the inequality (A.13) holds true. □

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

Declaration

All the authors developed the theoretical results, performed the numerical simulations, and contributed to the final version of the manuscript.

Acknowledgements

Giacomo Aletti is a member of the Gruppo Nazionale per il Calcolo Scientifico, of the Istituto Nazionale di Alta Matematica (Italy). Irene Crimaldi is a member of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni, of the Istituto Nazionale di Alta Matematica.

Funding information

Irene Crimaldi is partially supported by the Programma di Attività Integrata (PAI), under the project 'TOol for Fighting FakEs' (TOFFE) funded by the IMT School for Advanced Studies Lucca.

Competing interests

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

References

[1] Aletti, G., Crimaldi, I. and Ghiglietti, A. (2017). Synchronization of reinforced stochastic processes with a network-based interaction. *Ann. Appl. Prob.* **27**, 3787–3844.

[2] Aletti, G., Crimaldi, I. and Ghiglietti, A. (2019). Networks of reinforced stochastic processes: asymptotics for the empirical means. *Bernoulli* **25**, 3339–3378.

[3] Aletti, G., Crimaldi, I. and Ghiglietti, A. (2020). Interacting reinforced stochastic processes: statistical inference based on the weighted empirical means. *Bernoulli* **26**, 1098–1138.

[4] Aletti, G., Crimaldi, I. and Saracco, F. (2021). A model for the Twitter sentiment curve. *PLOS ONE* **16**, 1–28.

[5] Aletti, G., Ghiglietti, A. and Paganoni, A. M. (2013). Randomly reinforced urn designs with prespecified allocations. *J. Appl. Prob.* **50**, 486–498.

[6] Aletti, G., Ghiglietti, A. and Rosenberger, W. F. (2018). Nonparametric covariate-adjusted response-adaptive design based on a functional urn model. *Ann. Statist.* **46**, 3838–3866.

[7] Aletti, G., Ghiglietti, A. and Vidyashankar, A. N. (2018). Dynamics of an adaptive randomly reinforced urn. *Bernoulli* **24**, 2204–2255.

[8] Bergh, D. (2015). Sample size and chi-squared test of fit—a comparison between a random sample approach and a chi-square value adjustment method using Swedish adolescent data. In *Pacific Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceedings*, eds Q. Zhang and H. Yang, Springer, Berlin, Heidelberg, pp. 197–211.

[9] Berti, P., Crimaldi, I., Pratelli, L. and Rigo, P. (2011). A central limit theorem and its applications to multicolor randomly reinforced urns. *J. Appl. Prob.* **48**, 527–546.

[10] Berti, P., Crimaldi, I., Pratelli, L. and Rigo, P. (2016). Asymptotics for randomly reinforced urns with random barriers. *J. Appl. Prob.* **53**, 1206–1220.

[11] Bertoni, D. *et al.* (2018). Farmland use transitions after the CAP greening: a preliminary analysis using Markov chains approach. *Land Use Policy* **79**, 789–800.

[12] Caldarelli, G., Chessa, A., Crimaldi, I. and Pammolli, F. (2013). Weighted networks as randomly reinforced urn processes. *Phys. Rev. E* **87**, 020106.

[13] Caron, F. *et al.* (2017). Generalized Pólya urn for time-varying Pitman–Yor processes. *J. Machine Learning Res.* **18**, 1–32.

[14] Chanda, K. C. (1999). Chi-squared tests of goodness-of-fit for dependent observations. In *Asymptotics, Nonparametrics, and Time Series*, CRC Press, Boca Raton, FL, pp. 743–756.

[15] Chen, M.-R. and Kuba, M. (2013). On generalized Pólya urn models. *J. Appl. Prob.* **50**, 1169–1186.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

[16] CHESSA, A., CRIMALDI, I., RICCABONI, M. AND TRAPIN, L. (2014). Cluster analysis of weighted bipartite networks: a new copula-based approach. *PLOS ONE* **9**, 1–12.
[17] COLLEVECCHIO, A., COTAR, C. AND LICALZI, M. (2013). On a preferential attachment and generalized Pólya's urn model. *Ann. Appl. Prob.* **23**, 1219–1253.
[18] CRIMALDI, I. (2016). Central limit theorems for a hypergeometric randomly reinforced urn. *J. Appl. Prob.* **53**, 899–913.
[19] CRIMALDI, I. (2016). *Introduzione alla nozione di convergenza stabile e sue varianti*. Unione Matematica Italiana, Bologna.
[20] CRIMALDI, I., DAI PRA, P., LOUIS, P.-Y. AND MINELLI, I. G. (2019). Synchronization and functional central limit theorems for interacting reinforced random walks. *Stoch. Process. Appl.* **129**, 70–101.
[21] CRIMALDI, I., DAI PRA, P. AND MINELLI, I. G. (2016). Fluctuation theorems for synchronization of interacting Pólya's urns. *Stoch. Process. Appl.* **126**, 930–947.
[22] CRIMALDI, I., LETTA, G. AND PRATELLI, L. (2007). A strong form of stable convergence. In *Séminaire de Probabilités XL*, Springer, Berlin, Heidelberg, pp. 203–225.
[23] DAI PRA, P., LOUIS, P.-Y. AND MINELLI, I. G. (2014). Synchronization via interacting reinforcement. *J. Appl. Prob.* **51**, 556–568.
[24] DOEBLIN, W. AND FORTET, R. (1937). Sur des chanes à liaisons complètes. *Bull. Soc. Math. France* **65**, 132–148.
[25] EGGENBERGER, F. AND PÓLYA, G. (1923). Über die Statistik verketteter Vorgänge. *Z. Angew. Math. Mech.* **3**, 279–289.
[26] GASSER, T. (1975). Goodness-of-fit tests for correlated data. *Biometrika* **62**, 563–570.
[27] GHIGLIETTI, A. AND PAGANONI, A. M. (2014). Statistical properties of two-color randomly reinforced urn design targeting fixed allocations. *Electron. J. Statist.* **8**, 708–737.
[28] GHIGLIETTI, A., VIDYASHANKAR, A. N. AND ROSENBERGER, W. F. (2017). Central limit theorem for an adaptive randomly reinforced urn model. *Ann. Appl. Prob.* **27**, 2956–3003.
[29] GLESER, L. J. AND MOORE, D. S. (1983). The effect of dependence on chi-squared and empiric distribution tests of fit. *Ann. Statist.* **11**, 1100–1108.
[30] GUIVARC'H, Y. AND HARDY, J. (1988). Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov. *Ann. Inst. H. Poincaré Prob. Statist.* **24**, 73–98.
[31] HAIRER, M. Ergodic properties of Markov processes. Available at http://www.hairer.org/notes/Markov.pdf.
[32] HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.
[33] HOLMES, M. AND SAKAI, A. (2007). Senile reinforced random walks. *Stoch. Process. Appl.* **117**, 1519–1539.
[34] IEVA, F., PAGANONI, A. M., PIGOLI, D. AND VITELLI, V. (2013). Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *J. R. Statist. Soc. C [Appl. Statist.]* **62**, 401–418.
[35] IONESCU TULCEA, C. T. AND MARINESCU, G. (1950). Théorie ergodique pour des classes d'opérations non complètement continues. *Ann. Math.* **52**, 140–147.
[36] KNOKE, D., BOHRNSTEDT, G. W. AND POTTER MEE, A. (2002). *Statistics for Social Data Analysis*. F. E. Peacock Publishers, Itasca, IL.
[37] LARUELLE, S. AND PAGÉS, G. (2013). Randomized urn models revisited using stochastic approximation. *Ann. Appl. Prob.* **23**, 1409–1436.
[38] MAHMOUD, H. M. (2009). *Pólya Urn Models*. CRC Press, Boca Raton, FL.
[39] MÉTIVIER, M. (1982). *Semimartingales*. Walter de Gruyter, Berlin.
[40] MEYN, S. AND TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.
[41] MICHELETTI, A. *et al.* (2019). A weighted χ^2 test to detect the presence of a major change point in non-stationary Markov chains. Submitted.
[42] NORMAN, M. F. (1972). *Markov Processes and Learning Models*. Academic Press, New York.
[43] PAN, W. (2002). Goodness-of-fit tests for GEE with correlated binary data. *Scand. J. Statist.* **29**, 101–110.
[44] PEI, Y., TANG, M.-L. AND GUO, J. (2008). Testing the equality of two proportions for combined unilateral and bilateral data. *Commun. Statist. Simul. Comput.* **37**, 1515–1529.
[45] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Prob. Surveys* **4**, 1–79.
[46] RADLOW, R. AND ALF, E. F., Jr. (1975). An alternate multinomial assessment of the accuracy of the χ^2 test of goodness of fit. *J. Amer. Statist. Assoc.* **70**, 811–813.
[47] RAO, J. N. K. AND SCOTT, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Amer. Statist. Assoc.* **76**, 221–230.
[48] RÉNYI, A. (1963). On stable sequences of events. *Sankhyā A* **25**, 293 302.
[49] ROBBINS, H. AND SIEGMUND, D. (1971). A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing Methods in Statistics*, Academic Press, New York, pp. 233–257.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press

[50] SAHASRABUDHE, N. (2016). Synchronization and fluctuation theorems for interacting Friedman urns. *J. Appl. Prob.* **53**, 1221–1239.
[51] SHERMAN, J. AND MORRISON, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann. Math. Statist.* **21**, 124–127.
[52] TANG, M.-L., PEI, Y.-B., WONG, W.-K. AND LI, J.-L. (2012). Goodness-of-fit tests for correlated paired binary data. *Statist. Methods Med. Res.* **21**, 331–345.
[53] THARWAT, A. (2018). Independent component analysis: an introduction. *Appl. Comput. Informat.*
[54] WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge University Press.
[55] XU, D. AND TIAN, Y. (2015). A comprehensive survey of clustering algorithms. *Ann. Data Sci.* **2**, 165–193.

https://doi.org/10.1017/apr.2021.56 Published online by Cambridge University Press