An urn model with local reinforcement: a theoretical framework for a chi-squared goodness of fit test with a big sample

Giacomo Aletti *   and   Irene Crimaldi [†]

June 27, 2019

Abstract

Motivated by recent studies of big samples, this work aims at constructing 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 fluctuation of the conditional probabilities, and (iii) a long-term convergence of the empirical mean to a deterministic limit, together with a chi-squared goodness of fit result. This triple purpose has been achieved by the introduction of a new variant of the Eggenberger-Pólya urn, that we call the "Rescaled" Pólya urn. We provide a complete asymptotic characterization of this model and we underline that, for a certain choice of the parameters, it has properties different from the ones typically exhibited from the other urn models in the literature. As a byproduct, we also provide a Central Limit Theorem for a class of linear functionals of non-Harris Markov chains, where the asymptotic covariance matrix is explicitly given in linear form, and not in the usual form of a series.

keywords: central limit theorem, chi-squared test, compact Markov chain, Pólya urn, preferential attachment, reinforcement learning, reinforced stochastic process, urn model.

1 Introduction: framework and motivation

The well-known Pearson's 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 i.i.d., that is independent and identically distributed. 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, that multiplies 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, [10, 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. In other

---

*ADAMSS Center, Università degli Studi di Milano, Milan, Italy, giacomo.aletti@unimi.it

[†]IMT School for Advanced Studies, Lucca, Italy, irene.crimaldi@imtlucca.it

words, the larger N, the higher the type I error probability. As a consequence, in the context of "big data" (e.g. [10, 13]), where one often works with correlated noised data, suitable models and variants of (1) 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, [10, 15, 26, 29, 43, 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). Instead, in [44] a positive correlation is assumed directly in the likelihood function. 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 D-M distribution is randomized a priori with a Dirichlet distribution, obtaining an exchangeable (not independent) sequence. The variance-covariance matrix of the D-M distribution is equal to the one of the Multinomial (M) distribution, multiplied by a fixed constant greater than 1: precisely, given the k parameters $\boldsymbol{b_0} = (b_{0\,1}, \ldots, b_{0\,k})$ 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 $i, j \in \{1, \ldots, k\}$

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

and ρ is known as the "intra cluster" correlation parameter. Roughly speaking, the Dirichlet-Multinomial model adds variance to the multinomial model by taking a mixture or by adding an intra-cluster positive correlation. Property (2) is fundamental for our purpose. In fact, as well highlighted in [47], the two conditions (i) $\widehat{p}_i = O_i/N \to p_i$ almost surely for $N \to \infty$ and (ii) $Cov(O_i, O_j) = \lambda Cov_{\text{D}}(O_i, O_j)$ 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 becomes $\chi^2_{1-\theta}(k-1)\lambda/N$, where λ mitigate the effect of N. As already observed, the D-M model satisfies (ii), but it is well-known that it does not meet condition (i). In this paper we give a parametric extension of the D-M model so that both of the above conditions hold true.

Urn and biased-coin sequential models

The Dirichlet-Multinomial 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, 12, 14, 16, 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 $i = 1, \ldots, k$, and, at each discrete time, a ball is drawn out from the urn and then it is put again inside the urn together with $\alpha > 0$ additional balls 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 $E[\xi_{n+1\,i}|\xi_{m\,i}, m \leq n]$ 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.

The biased coin sequential models are related to urn models in many aspects, but the approach is different. In a biased coin design $E[\xi_{n+1\,i}|\xi_{m\,i}\,m \leq n]$ is directly driven to a fixed limit through

a dynamics, which contains information on the limit that should be reached, together with the information on the past that are considered relevant. The class of biased-coin designs has recently had a large auditor (see [8, 9, 34] and the references therein). In particular, in [9] general asymptotic properties of covariate-adaptive randomized designs are studied for this class of processes. Similar results may be found with a Friedman urn (see [4, 50] for asymptotic results regarding a network of such urns). Finally, a sequential re-randomization framework for sequential design has been also proposed in [54], where the experimental units are enrolled in groups, and adaptive re-randomization is used for balancing treatment/control assignment.

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 process $E[\xi_{n+1\,i}|\xi_{m\,i},\ m \leq n]$ randomly fluctuates without converging almost surely, forming asymptotically a stationary ergodic process. As a consequence, since the classical martingale approach or the stochastic approximation require or imply the convergence of $E[\xi_{n+1\,i}|\xi_{m\,i}\,m \leq n]$ (e.g. [1, 11, 37]), not usual mathematical methods are here needed in proving asymptotic results for the introduced new urn model.

"Rescaled" Pólya urn

We introduce here a new variant of the Eggenberger-Pólya urn with k-colors, that we call the "Rescaled" Pólya urn model (RP). 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 fluctuations. More precisely, the RP is characterized by the introduction of the parameters β and $(b_{0\,i})_{i=1,\ldots,k}$ in the original model, so that

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

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 does not fall in the variants of the Eggenberger-Pólya urn discussed in [45, Section 3.2] and, as explained in details in Section 2, it does not belong to the class of Reinforced Stochastic Processes studied in [1, 2, 3, 20, 21, 23, 50].

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

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

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

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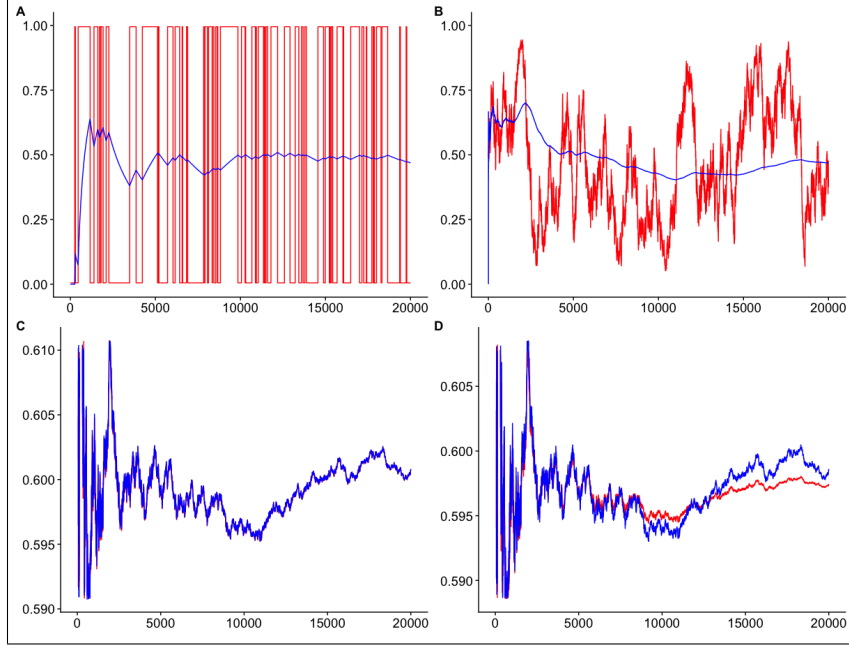$$

3

Figure 1: Simulations of the two processes $(\psi_{n\,1})_n$ (red color) and $(\bar{\xi}_{n\,1})_n$ (blue color), 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. Notably, the speed of convergence is faster for $(\psi_{n\,1})_n$ when $\beta > 1$, see Theorem 3.5.

Assuming for $B_{n\,i}$ the dynamics (3) with $\beta > 0$, the probability $\psi_{n\,i}$ results increasing with the number of times we observed the value i (see the following equation (13)) and so the random variables $\xi_{n\,i}$ are generated according to a reinforcement mechanism. But, in particular, when $\beta < 1$, the reinforce reduces exponentially, leaving the fluctuations 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 $\xi_{n\,i}$. We are mainly interested in the case $\beta \in [0, 1)$, because in this case the RP 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 do not converge almost surely;

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

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

As said before, since (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.

We have also considered the asymptotic results for $\beta > 1$, to complete the study of the RP. 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, 11, 37]).

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

4

$\beta > 1$ (Figure 1(C) and Figure 1(D), respectively).

Goodness of fit result

Given a sample $(\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_N})$ (where $\boldsymbol{\xi_n}$ denotes the random vector with components $\xi_{n\,i}$, $i = 1, \ldots, k$) generated by the RP, the statistics

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

counts the number of times we observed the value i. The theorem below shows that, when $\beta \in [0, 1)$, we can construct a chi-squared test for the intrinsic probabilities $p_{0\,1}, \ldots, p_{0\,k}$. More precisely, we will prove the following result:

Theorem 1.1. *Assume $p_{0\,i} > 0$ for all $i = 1, \ldots, k$ and $\beta \in [0, 1)$. Define the constants γ and λ as*

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

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

Then

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

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

Application to a "big sample"

A possible application we have in mind was inspired by [13, 41] and is the following. We suppose to have a "big" sample $\{\boldsymbol{\xi_n} \colon n = 1, \ldots, N\}$, where the observations can not be assumed i.i.d, because the choice of a unit is influenced by the choices of some other units, according to a reinforcement rule. More precisely, we consider the situation in which there are some clusters such that 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. Formally, suppose that the N units are ordered so that we have the following clusters of units:

$$C_1 = \{1, \ldots, N_1\}, \ldots,$$

$$C_\ell = \{\sum_{l=1}^{\ell-1} N_l + 1, \ldots, \sum_{l=1}^{\ell} N_l\}, \ldots,$$

$$C_L = \{\sum_{l=1}^{L-1} N_l + 1, \ldots, \sum_{l=1}^{L} N_l = N\}.$$

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

$$[\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_{N_1}}], \ldots, [\boldsymbol{\xi_{\sum_{l=1}^{\ell-1} N_l+1}}, \ldots, \boldsymbol{\xi_{\sum_{l=1}^{\ell} N_l}}], \ldots, [\boldsymbol{\xi_{\sum_{l=1}^{L-1} N_l+1}}, \ldots, \boldsymbol{\xi_N}].$$

are L independent multidimensional random variables. Moreover, we assume that the observations inside each cluster can be modeled as a RP with $\beta \in [0, 1)$. We denote by $p_{0\,1}(\ell), \ldots, p_{0\,k}(\ell)$ the intrinsic probabilities for the cluster C_ℓ, that we assume strictly positive, and we assume the same

5

parameter λ for each cluster (not necessarily the same parameters α and β) so that all the L random variables

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

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

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

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 not depends on λ) and so it can be used in order 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 $\ell = 1, \ldots, L$ in order 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 $\Gamma\left(\frac{k-1}{2}, \frac{1}{2\widehat{\lambda}^*}\right)$ in order to perform a χ^2-test with null hypothesis

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

Regarding the probabilities $p_{0\,i}^*(\ell)$, some possibilities are:

- 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 to have two different periods of times, and so two samples, say $\{\boldsymbol{\xi}_n^{(1)} : n = 1, \ldots, N\}$ and $\{\boldsymbol{\xi}_n^{(2)} : n = 1, \ldots, N\}$, take $p_{0\,i}^*(\ell) = \sum_{n \in C_\ell} \xi_{n\,i}^{(1)}/N_\ell$ for all $i = 1, \ldots, k$, and perform the test on the second sample in order to check possible changes in the intrinsic probabilities;

- we can take one of the clusters as 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^*.

This statistical procedure has been applied in [13], but for a sample generated by a Markov chain: specifically, each cluster is formed by the administrative data that describe, at a very detailed level, the temporal and spatial dynamics of farmland use in Lombardy, a region in North Italy. However, in that paper it is not given a general theoretical framework. Instead, here we aim at providing a general model and the related theoretical results, motivating the above statistical procedure.

CLT for linear models of Markov compact chains

The key point in proving the above result is given into two asymptotic theorems: the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT) for the empirical means given in Theorem 3.1 and Theorem 3.3, for $\beta = 0$ and $\beta \in (0,1)$, respectively. Mathematically speaking, the proof of these theorems is based on the fact that the processes $(\psi_{n\,i})_n$, with $i = 1, \ldots, k$, form a multi-dimensional Markov chain with a unique ergodic invariant limit probability distribution. For $\beta = 0$, all the quantities involved can be computed explicitly, since the Markov chain has a finite number of states. Conversely, when $\beta \in (0,1)$, the classical CLTs for Harris chains cannot be used in the present framework. It is known (see [33]) that a Markov chain with a unique invariant probability measure can either be a positive Harris recurrent chain with an absorbing set (and the invariant probability measure is non-singular), or cannot have Harris sets (and the invariant probability measure is singular). Our Markov chain falls in the second case: the systems as (22)

6

are explicitly mentioned in [33, § 3.3]). For this reason, we decide to refer to the general theory for compact Markov chains (see [42, Chapter 3]). In this framework, we provide a CLT for a general class of linear models. This byproduct result is given in Theorem A.11, and his importance is augmented by the fact that the variance-covariance matrix is explicitly calculated with linear algebra calculus, without the usual computation of the series that involves the auto-correlation function.

Structure of the paper

Summing up, the sequel of the paper is so structured. In Section 2 we set up our notation and we define the RP with parameters $\alpha > 0$ and $\beta \geq 0$. In Section 3 we provide a complete characterization of the RP for the three cases $\beta = 0$, $\beta \in [0, 1)$ and $\beta > 1$. (We do not deal with the case $\beta = 1$ because, as said 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 ξ_{ni} almost surely converges to the intrinsic probabilities p_{0i} when $\beta \in [0, 1)$; while it almost surely converges to a random limit when $\beta > 1$. We obtain also the corresponding CLTs, that, in particular for $\beta \in [0, 1)$, are the basis for the proof of Theorem 1.1. For completeness, we also describe the case $\alpha = 0$, that generates a sequence of independent draws. Section 4 contains the proof of Theorem 1.1, which gives the possibility to construct a chi-squared test for the intrinsic probabilities when the observed sample is assumed to be generated by a RP 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, that we call "linearity" condition, and in Section A.2 we explain a fundamental coupling technique used in the proof of the CLT for $\beta \in (0, 1)$.

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 $\boldsymbol{x} = (x_1, \ldots, x_k)^\top \in \mathbb{R}^k$, 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 $\boldsymbol{1}$ and $\boldsymbol{0}$ the vectors with all the components equal to 1 and equal to 0, respectively, and by $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_k}\}$ the canonical base of $\mathbb{R}^k$.

To formally work with the RP presented in the introduction, we add here some notations. As in (3), the urn initially contains a constant number of b_{0i} distinct balls of color i, with $i = 1, \ldots, k$, together with a constant number B_{0i} balls of the same color i. We set $\boldsymbol{b_0} = (b_{01}, \ldots, b_{0k})^\top$ and $\boldsymbol{B_0} = (B_{01}, \ldots, B_{0k})^\top$. In all the sequel (unless otherwise specified) we assume $|\boldsymbol{b_0}| > 0$ and $b_{0i} + B_{0i} > 0$ for each $i = 1, \ldots, k$. 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, obtaining the random vector $\boldsymbol{\xi_{n+1}} = (\xi_{n+1\,1}, \ldots, \xi_{n+1\,k})^\top$ defined as

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

and the number of balls in the urn is so updated:

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

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 $\mathcal{F}_n = \sigma(\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_n})$ for $n \geq 1$, the conditional probabilities $\boldsymbol{\psi_n} = (\psi_{n\,1}, \ldots, \psi_{n\,k})^\top$ of the extraction process are for $n \geq 0$

$$\psi_{n\,i} = P(\xi_{n+1\,i} = 1 | \mathcal{F}_n) = \frac{N_{n\,i}}{|\boldsymbol{N_n}|} = \frac{b_{0\,i} + B_{n\,i}}{r_n^*}, \qquad i = 1, \ldots k, \tag{10}$$

that is, in vectorial form,

$$\boldsymbol{\psi_n} = E[\boldsymbol{\xi_{n+1}} | \mathcal{F}_n] = \frac{\boldsymbol{N_n}}{|\boldsymbol{N_n}|} = \frac{\boldsymbol{b_0} + \boldsymbol{B_n}}{r_n^*}. \tag{11}$$

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 (11), together with (7) and (9), we have

$$\boldsymbol{\psi_n} - \boldsymbol{\psi_{n-1}} = -\frac{(1-\beta)|\boldsymbol{b_0}|}{r_n^*}(\boldsymbol{\psi_{n-1}} - \boldsymbol{p_0}) + \frac{\alpha}{r_n^*}(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}). \tag{12}$$

Since the first term in the right hand of the above relation, the RP for $\beta \neq 1$ does not belong to the class of Reinforced Stochastic Processes (RSPs) studied in [1, 2, 3, 20, 21, 23]. Generally speaking, by reinforcement in a stochastic dynamics we mean any mechanism for which the probability that a given event occurs has an increasing dependence on the number of times that the same event occurred in the past. This "reinforcement mechanism", also known as "preferential attachment rule" or "Rich get richer rule" or "Matthew effect", is a key feature governing the dynamics of many biological, economic and social systems (see, e.g. [45]). The RSPs are characterized by a "strict" reinforcement mechanism such that $\xi_{n\,i} = 1$ implies $\psi_{n\,i} > \psi_{n-1\,i}$. As a consequence, the "general" reinforcement mechanism is satisfied, in the sense that the random variable $\psi_{n\,i}$ has an increasing dependence on the number of times we have $\xi_{m\,i} = 1$ for $m = 1, \ldots, n$. When $\beta \neq 1$, the RP does not satisfy the "strict" reinforcement mechanism, because the first term is positive or negative according to the sign of $(1-\beta)$ and of $(\boldsymbol{\psi_{n-1}} - \boldsymbol{p_0})$. However, when $\alpha, \beta > 0$, it satisfies the general reinforcement mechanism. Indeed, by (7), (8), (9) and (11), using $\sum_{m=0}^{n-1} x^m = (1-x^n)/(1-x)$, we have

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

In particular, for $\beta < 1$, the dependence of $\boldsymbol{\psi_n}$ on $\boldsymbol{\xi_m}$ exponentially 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 $\boldsymbol{\xi_n}$, the last extraction. For $\beta > 1$ we have the opposite behaviour, that is the dependence of $\boldsymbol{\psi_n}$ on $\boldsymbol{\xi_m}$ exponentially decreases with m. Finally, we observe that Equation (12) recalls the dynamics of a RSP with a "forcing input" (see [1, 20, 50]), but the difference relies on the fact that for the RP the sequence (r_n^*) is such that $\sum_n 1/r_n^* = +\infty$ and $\sum_n 1/(r_n^*)^2 = +\infty$ for $\beta \in (0, 1)$ and $\sum_n 1/r_n^* < +\infty$ (and $\sum_n 1/(r_n^*)^2 < +\infty$) for $\beta > 1$. These facts lead to a different asymptotic behavior of $(\boldsymbol{\psi_n})$.

3 Properties of the "Rescaled" Pólya urn model

We study separately the three cases $\beta = 0$, $\beta \in (0, 1)$ and $\beta > 1$.

3.1 The case $\beta = 0$

In this case, by (7), (8) and (9), equation (10) becomes for all $i = 1, \ldots, k$

$$
\begin{aligned}
\psi_{0\,i} &= \frac{b_{0\,i} + B_{0\,i}}{|\boldsymbol{b_0}| + |\boldsymbol{B_0}|} &&\text{and} \\
\psi_{n\,i} &= \frac{b_{0\,i} + \alpha \xi_{n\,i}}{|\boldsymbol{b_0}| + \alpha} &&\text{for } n \geq 1.
\end{aligned}
\tag{14}
$$

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 $S = \{\boldsymbol{s_1}, \ldots, \boldsymbol{s_k}\}$, where

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

and transition probability matrix

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

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

$$
\begin{aligned}
P^n &= \Big(\frac{|\boldsymbol{b_0}|}{|\boldsymbol{b_0}| + \alpha}\Big)^n \bigg(\Big(\sum_{j=0}^{n-1} \binom{n}{j}(\tfrac{\alpha}{|\boldsymbol{b_0}|})^j\Big)\mathbf{1}_k\,\boldsymbol{p_0}^\top + (\tfrac{\alpha}{|\boldsymbol{b_0}|})^n \mathrm{Id}_k\bigg) \\
&= \Big(\frac{|\boldsymbol{b_0}|}{|\boldsymbol{b_0}| + \alpha}\Big)^n \bigg(\Big((1 + \tfrac{\alpha}{|\boldsymbol{b_0}|})^n - (\tfrac{\alpha}{|\boldsymbol{b_0}|})^n\Big)\mathbf{1}_k\,\boldsymbol{p_0}^\top + (\tfrac{\alpha}{|\boldsymbol{b_0}|})^n \mathrm{Id}_k\bigg) \\
&= \mathbf{1}_k\,\boldsymbol{p_0}^\top + \Big(\frac{\alpha}{|\boldsymbol{b_0}| + \alpha}\Big)^n \Big(\mathrm{Id}_k - \mathbf{1}_k\,\boldsymbol{p_0}^\top\Big) \\
&= \mathbf{1}_k\,\boldsymbol{p_0}^\top + \gamma^n \Big(\mathrm{Id}_k - \mathbf{1}_k\,\boldsymbol{p_0}^\top\Big)\,,
\end{aligned}
\tag{15}
$$

where γ is the constant given in (5), that 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 3.1. *We have* $\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \boldsymbol{p_0}$ *and*

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

where

$$
\Sigma^2 = \lambda\Big(\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^\top\Big),
\tag{16}
$$

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

Proof. We observe that, by (14), 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

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

9

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 $g(\boldsymbol{\xi}_n) = g\left(\mathbb{1}_{\{\psi_n = s_1\}}, \ldots, \mathbb{1}_{\{\psi_n = s_k\}}\right)$ and apply the central limit theorem 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 $\mathcal{N}(0, \sigma_{\boldsymbol{c}}^2)$, with

$$\sigma_{\boldsymbol{c}}^2 = \mathrm{Var}[g(\boldsymbol{\xi}_0^{(\boldsymbol{\pi})})] + 2\sum_{n \geq 1} \mathrm{Cov}(g(\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}), g(\boldsymbol{\xi}_n^{(\boldsymbol{\pi})}))$$

$$= \boldsymbol{c}^T\left(\mathrm{Var}[\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}] + 2\sum_{n \geq 1}\mathrm{Cov}(\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}, \boldsymbol{\xi}_n^{(\boldsymbol{\pi})})\right)\boldsymbol{c},$$

where $\boldsymbol{\xi}_n^{(\boldsymbol{\pi})} = \left(\mathbb{1}_{\{\psi_n^{(\boldsymbol{\pi})} = s_1\}}, \ldots, \mathbb{1}_{\{\psi_n^{(\boldsymbol{\pi})} = s_k\}}\right)$ and $\left(\psi_n^{(\boldsymbol{\pi})}\right)_{n \geq 0}$ is a Markov chain with transition matrix P and initial distribution $\boldsymbol{\pi}$, that is $\boldsymbol{p_0}$. Now, by definition, $\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}\boldsymbol{\xi}_0^{(\boldsymbol{\pi})T} = \mathrm{diag}(\boldsymbol{\xi}_0^{(\boldsymbol{\pi})})$, and hence $\mathrm{Var}[\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}] = \mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^T$. Moreover, by means of (15),

$$E[\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}\boldsymbol{\xi}_n^{(\boldsymbol{\pi})\top}] = \mathrm{diag}(\boldsymbol{p_0})P^n = \boldsymbol{p_0}\boldsymbol{p_0}^\top + \gamma^n\left(\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^\top\right).$$

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

$$\mathrm{Var}[\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}] + 2\sum_{n \geq 1}\mathrm{Cov}(\boldsymbol{\xi}_0^{(\boldsymbol{\pi})}, \boldsymbol{\xi}_n^{(\boldsymbol{\pi})}) = \left(\mathrm{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^\top\right)\left(1 + \frac{2\gamma}{1 - \gamma}\right).$$

By the Cramér-Wold device, the theorem is proved with Σ^2 given in (16). $\qquad\square$

Remark 3.2. Note that in Theorem 3.1 we do not assume $b_{0\,i} > 0$ for all i, but only $|\boldsymbol{b_0}| > 0$ (as said in Sec. 2). A different behavior is observed when $\boldsymbol{b_0} = \boldsymbol{0}$. In this case, (14) gives $\boldsymbol{\psi}_n = \boldsymbol{\xi}_n$ for $n \geq 1$. Since $\psi_{n\,i} = P(\xi_{n\,i} = 1|\mathcal{F}_n)$, 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 $P(\xi_{1\,i} = 1) = \psi_{0\,i} = B_{0\,i}/|\boldsymbol{B_0}|$.

3.2 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)

$$r_n^* = |\boldsymbol{b_0}| + |\boldsymbol{B_n}| = |\boldsymbol{b_0}| + r + \beta^n(|\boldsymbol{B_0}| - r) \longrightarrow r^*, \tag{17}$$

and so we have that the denominator r_n^* in $\boldsymbol{\psi}_n$ (see Eq. (10) and (11)) 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) and

$$\gamma - \beta = \frac{\alpha}{r^*} \qquad \text{and} \qquad 1 - \gamma = \frac{(1 - \beta)|\boldsymbol{b_0}|}{r^*}. \tag{18}$$

Therefore, by (17), the terms $\frac{(1-\beta)|\boldsymbol{b_0}|}{r_n^*}$ and $\frac{\alpha}{r_n^*}$ in the dynamics (12) 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 3.3. *We have* $\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \boldsymbol{p_0}$ *and*

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

10

where

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

with λ *defined in* (6).

Remark 3.4. Note that in Theorem 3.1 we do not assume $b_{0\,i} > 0$ for all i, but only $|\boldsymbol{b_0}| > 0$ (as said in Sec. 2). Again, a different behavior is observed when $\boldsymbol{b_0} = \boldsymbol{0}$. Indeed, from (12), we have

$$\boldsymbol{\psi_n} - \boldsymbol{\psi_{n-1}} = \frac{\alpha}{r_n^*}\big(\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\big), \tag{19}$$

and hence the RP is a martingale. The asymptotic result given above fails (in fact, we have $\gamma = 1$). The martingale property of the bounded process $\boldsymbol{\psi}$ implies that $\boldsymbol{\psi_n}$ converges almost surely (and in mean) to a bounded random variable $\boldsymbol{\psi_\infty}$. In addition, since $r_n^* \to \alpha/(1-\beta)$, from (19), 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 3.3 into two main steps: first, we will prove that the convergence behaviour 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.

3.2.1 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 3.3 are not affected by the value of the initial constant $|\boldsymbol{B_0}|$.

Set $\boldsymbol{\xi_n^{(1)}} = \boldsymbol{\xi_n}$ and $\boldsymbol{\psi^{(1)}} = \boldsymbol{\psi}$, that follows the dynamics (12), together with (10), starting from a certain initial point $\boldsymbol{\psi_0^{(1)}} = \boldsymbol{\psi_0}$. By (7) and relations (18), we can write

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

where

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

Since, by (17), we have $r^*/r_{n+1}^* - 1 = O(\beta^{n+1})$ and $r_n^*/r_{n+1}^* - 1 = O(\beta^{n+1})$, we get $|\boldsymbol{l_{n+1}^{(1)}}| = O(\beta^{n+1})$. Now, take $\boldsymbol{\xi^{(2)}} = (\boldsymbol{\xi_n^{(2)}})_n$ and $\boldsymbol{\psi^{(2)}} = (\boldsymbol{\psi_n^{(2)}})_n$ following the same dynamics given in (10) and (12), but starting from an initial point with $|\boldsymbol{B_0^{(2)}}| = r$. Therefore, we have

$$\boldsymbol{\psi_{n+1}^{(2)}} = \beta \boldsymbol{\psi_n^{(2)}} + (\gamma - \beta)\boldsymbol{\xi_{n+1}^{(2)}} + (1-\gamma)\boldsymbol{p_0}.$$

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), condition (A.10) holds true. Hence we can apply Theorem A.13 so that there exist two stochastic processes $\widetilde{\boldsymbol{\psi}}^{(1)}$ and $\widetilde{\boldsymbol{\psi}}^{(2)}$, following the dynamics (A.11) (with the same

11

specifications as above), together with (A.12), starting from the same initial points and such that (A.14) holds true, that is

$$E\left[|\widetilde{\psi}_{n+1}^{(1)} - \widetilde{\psi}_{n+1}^{(2)}|\,\Big|\,\widetilde{\psi}_0^{(2)},\widetilde{\psi}_0^{(1)}\right] \leq \gamma^{n+1}|\widetilde{\psi}_0^{(1)} - \widetilde{\psi}_0^{(2)}| + O\Big(\sum_{j=1}^{n+1}\gamma^{n+1-j}\beta^j\Big)$$

$$= O\left((n+2)\max(\gamma,\beta)^{n+1}\right) = O\left((n+2)\gamma^{n+1}\right).$$

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

$$\sum_{n=1}^{+\infty} E\left[|\widetilde{\xi}_n^{(1)} - \widetilde{\xi}_n^{(2)}|\right] < +\infty,$$

which implies $\sum_{n=0}^{+\infty}|\widetilde{\xi}_n^{(1)} - \widetilde{\xi}_n^{(2)}| < +\infty$ a.s., that is

$$\widetilde{\xi}_n^{(1)} = \widetilde{\xi}_n^{(2)} \qquad \text{eventually.} \tag{20}$$

Therefore, if we prove some asymptotic results for $\boldsymbol{\xi}^{(2)}$, 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 (20)), and finally they hold true also for $\boldsymbol{\xi}^{(1)}$ (since they have the same joint distribution). Summing up, without loss of generality, we may prove Theorem 3.3 under the additional assumption $|\boldsymbol{B_0}| = r$.

3.2.2 The case $|\boldsymbol{B_0}| = r$

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

$$r_n^* = r^* = |\boldsymbol{b_0}| + r \tag{21}$$

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

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

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

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

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:

- 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 space 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 $P^n f = P(P^{n-1} f)$.

By (12), together with (18) and (21), the process $\boldsymbol{\psi} = (\boldsymbol{\psi_n})_n$ follows the dynamics

$$\boldsymbol{\psi_{n+1}} = \beta \boldsymbol{\psi_n} + \boldsymbol{b_0} \frac{(1-\beta)}{r^*} + \boldsymbol{\xi_{n+1}} \frac{\alpha}{r^*} = \beta \boldsymbol{\psi_n} + (1-\gamma)\boldsymbol{p_0} + (\gamma - \beta)\boldsymbol{\xi_{n+1}}. \tag{22}$$

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

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

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

$$(Pf)(\boldsymbol{x}) = E[f(\boldsymbol{\psi_{n+1}})|\boldsymbol{\psi_n} = \boldsymbol{x}] = \sum_{i=1}^{k} x_i f\left(\left(\beta\boldsymbol{x} + \boldsymbol{p_0}(1-\gamma)\right)_{(i)}\right). \tag{23}$$

In particular, from the above equality, we get

$$(Pid)(\boldsymbol{x}) - \boldsymbol{p_0} = E[\boldsymbol{\psi_{n+1}} - \boldsymbol{p_0}|\boldsymbol{\psi_n} = \boldsymbol{x}] = \gamma(\boldsymbol{x} - \boldsymbol{p_0}). \tag{24}$$

We now show that $\boldsymbol{\psi}$ is an irreducible, aperiodic, compact Markov chain (see Def. A.4 and Def. A.6).

Check that $\boldsymbol{\psi}$ is a compact Markov chain: By Lemma A.5, it is sufficient to show that P defined in (23) is weak Feller (Definition A.1) and that it is a semi-contractive operator on $Lip(S)$ (Definition A.3). From (23), 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 (22) 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), condition (A.10) holds true. Then , 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 point $\boldsymbol{x}$ and $\boldsymbol{y}$. Then, applying Theorem A.13, 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

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

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

$$|(Pf)(\boldsymbol{y}) - (Pf)(\boldsymbol{x})| = |E[f(\widetilde{\boldsymbol{\psi}}_1^{(2)})|\widetilde{\boldsymbol{\psi}}_0^{(2)} = \boldsymbol{y}] - E[f(\widetilde{\boldsymbol{\psi}}_1^{(1)})|\widetilde{\boldsymbol{\psi}}_0^{(1)} = \boldsymbol{x}]|$$
$$= E\left[|f(\widetilde{\boldsymbol{\psi}}_1^{(2)}) - f(\widetilde{\boldsymbol{\psi}}_1^{(1)})|\right]$$
$$\leq |f|_{Lip} E\left[|\widetilde{\boldsymbol{\psi}}_1^{(2)} - \widetilde{\boldsymbol{\psi}}_1^{(1)}|\right] \leq |f|_{Lip} \, \gamma \, |\boldsymbol{x} - \boldsymbol{y}|$$

and so

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

with $\gamma < 1$, as requested. $\qquad\square$

Check that $\boldsymbol{\psi}$ is irreducible and aperiodic: We prove the irreducibility and aperiodicity condition stated in Def. A.6, using Theorem A.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 follows the same dynamics (22) 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 (22) 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), condition (A.10) holds true, then we can apply Theorem A.13 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

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

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

Proof of the almost sure convergence: We have already proven that the Markov chain $\boldsymbol{\psi}$ has one invariant probability measure π. Furthermore, from (22) we get

$$\boldsymbol{\xi_{n+1}} - \boldsymbol{p_0} = \frac{1}{\gamma - \beta}\left[\boldsymbol{\psi_{n+1}} - \beta\boldsymbol{\psi_n} - \boldsymbol{p_0}(1 - \gamma)\right] - \boldsymbol{p_0}$$

$$= \frac{1}{\gamma - \beta}\left[(\boldsymbol{\psi_{n+1}} - \boldsymbol{p_0}) - \beta(\boldsymbol{\psi_n} - \boldsymbol{p_0})\right]. \quad (25)$$

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

$$\overline{\boldsymbol{\xi}}_N - \boldsymbol{p_0} = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{\xi_n} - \boldsymbol{p_0}) \xrightarrow{a.s.} \frac{1 - \beta}{\gamma - \beta}E[\boldsymbol{\psi}_n^{(\boldsymbol{\pi})} - \boldsymbol{p_0}],$$

where $\left(\boldsymbol{\psi}_n^{(\boldsymbol{\pi})}\right)_{n \geq 0}$ is a Markov chain with transition kernel P and initial distribution π. From (24), we have

$$E[\boldsymbol{\psi}_n^{(\boldsymbol{\pi})} - \boldsymbol{p_0}] = E[\boldsymbol{\psi}_{n+1}^{(\boldsymbol{\pi})} - \boldsymbol{p_0}] = E\left[E[\boldsymbol{\psi}_{n+1}^{(\boldsymbol{\pi})} - \boldsymbol{p_0} \,|\, \boldsymbol{\psi}_n^{(\boldsymbol{\pi})}]\right] = \gamma E[\boldsymbol{\psi}_n^{(\boldsymbol{\pi})} - \boldsymbol{p_0}]$$

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

Proof of the CLT: We apply Theorem A.11, taking into account that we have already proven that $\boldsymbol{\psi}$ is an irreducible and aperiodic compact Markov chain. Since, by (25) and what we have already proven before, we have

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

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 A.9 (see also Remark A.10). Indeed, we have $A_1 = -\frac{\beta}{\gamma-\beta}Id$ and $A_2 = \frac{1}{\gamma-\beta}Id$. Moreover, by (24), we have $P(id)(\boldsymbol{x}) - \boldsymbol{p_0} = \gamma(\boldsymbol{x} - \boldsymbol{p_0})$, which means $A_P = \gamma Id$. Therefore Theorem A.11 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

$$\Sigma^2 = \frac{(1 - \beta)^2(1 + \gamma)}{(\gamma - \beta)^2(1 - \gamma)}\Sigma_\pi^2 = \frac{(1 - \beta)^2}{(\gamma - \beta)^2}\left(1 + 2\frac{\gamma}{1 - \gamma}\right)\Sigma_\pi^2. \quad (26)$$

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

$$\boldsymbol{\xi}_{n+1}^{(\boldsymbol{\pi})} - \boldsymbol{p_0} = f(\boldsymbol{\psi}_n^{(\boldsymbol{\pi})}, \boldsymbol{\psi}_{n+1}^{(\boldsymbol{\pi})}) = A_2(\boldsymbol{\psi}_{n+1}^{(\boldsymbol{\pi})} - \boldsymbol{p_0}) + A_1(\boldsymbol{\psi}_n^{(\boldsymbol{\pi})} - \boldsymbol{p_0}).$$

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

$$
\begin{aligned}
\operatorname{diag}(\boldsymbol{p_0}) - \boldsymbol{p_0 p_0}^T &= E\left[(\boldsymbol{\xi}_1^{(\boldsymbol{\pi})} - \boldsymbol{p_0})(\boldsymbol{\xi}_1^{(\boldsymbol{\pi})} - \boldsymbol{p_0})^\top\right] \\
&= A_1 \Sigma_\pi^2 A_1^\top + A_2 \Sigma_\pi^2 A_2^\top + A_1 \Sigma_\pi^2 A_P^\top A_2^\top + A_2 A_P \Sigma_\pi^2 A_1 \\
&= \frac{(\gamma - \beta)^2 + (1 - \gamma^2)}{(\gamma - \beta)^2} \Sigma_\pi^2.
\end{aligned}
\tag{27}
$$

Finally, it is enough to combine (26) and (27). $\qquad\square$

3.3 The case $\beta > 1$

In this case, $\lim_n \beta^n = +\infty$ and $\sum_{n \geq 1} \beta^{-n} = 1/(\beta - 1)$. Hence, the following results hold true:

Theorem 3.5. *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

$$|\boldsymbol{\psi_N} - \boldsymbol{\psi_\infty}| = O(\beta^{-N}).$$

Moreover $\boldsymbol{\psi_\infty}$ takes values in $\{\boldsymbol{x} \in [0,1]^k : |\boldsymbol{x}| = 1\}$ 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 $\left(\{\boldsymbol{x} \in \{0,1\}^k : |\boldsymbol{x}| = 1\}\right)^\infty$.

Proof. By (13), we have

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

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

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

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

$$|\boldsymbol{\psi_N} - \boldsymbol{\psi_\infty}| = O(\beta^{-N}).$$

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

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

Theorem 3.6. *We have $\overline{\boldsymbol{\xi}}_N \xrightarrow{a.s.} \boldsymbol{\psi_\infty}$ and*

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

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

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

The stable convergence has been introduced in [48] and, for its definition and properties, we refer to [19, 22, 32].

Proof. The almost sure convergence of $\overline{\boldsymbol{\xi}}_{\boldsymbol{N}}$ to $\boldsymbol{\psi}_{\boldsymbol{\infty}}$ follows by usual martingale arguments (see, for instance, [11, Lemma 2]) because $E[\boldsymbol{\xi}_{\boldsymbol{n+1}}|\mathcal{F}_n] = \boldsymbol{\psi}_{\boldsymbol{n}} \to \boldsymbol{\psi}_{\boldsymbol{\infty}}$ a.s. and $\sum_{n \geq 1} E[\|\boldsymbol{\xi}_{\boldsymbol{n}}\|^2] n^{-2} \leq \sum_{n \geq 1} n^{-2} < +\infty$.

Regarding the CLTs, we observe that, by means of (12), we can write

$$\boldsymbol{\psi}_{\boldsymbol{n+1}} - \boldsymbol{\psi}_{\boldsymbol{n}} = \frac{\boldsymbol{H}(\boldsymbol{\psi}_{\boldsymbol{n}})}{r_{n+1}^*} + \frac{\Delta \boldsymbol{M}_{\boldsymbol{n+1}}}{r_{n+1}^*}, \tag{28}$$

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

$$\sqrt{N}\left(\overline{\boldsymbol{\xi}}_{\boldsymbol{N}} - \boldsymbol{\psi}_{\boldsymbol{N}}\right) = \frac{1}{\sqrt{N}}\left(N\overline{\boldsymbol{\xi}}_{\boldsymbol{N}} - N\boldsymbol{\psi}_{\boldsymbol{N}}\right) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}[\boldsymbol{\xi}_{\boldsymbol{n}} - \boldsymbol{\psi}_{\boldsymbol{n-1}} + n(\boldsymbol{\psi}_{\boldsymbol{n-1}} - \boldsymbol{\psi}_{\boldsymbol{n}})]$$

$$= \sum_{n=1}^{N} \boldsymbol{Y}_{\boldsymbol{N,n}} + \boldsymbol{Q}_{\boldsymbol{N}},$$

where

$$\boldsymbol{Y}_{\boldsymbol{N,n}} = \frac{\boldsymbol{\xi}_{\boldsymbol{n}} - \boldsymbol{\psi}_{\boldsymbol{n-1}}}{\sqrt{N}} = \frac{\Delta \boldsymbol{M}_{\boldsymbol{n+1}}}{\alpha \sqrt{N}}$$

and

$$\boldsymbol{Q}_{\boldsymbol{N}} = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} n\left(\boldsymbol{\psi}_{\boldsymbol{n-1}} - \boldsymbol{\psi}_{\boldsymbol{n}}\right) = -\frac{1}{\sqrt{N}}\sum_{n=1}^{N} \frac{n}{r_n^*}(\boldsymbol{H}(\boldsymbol{\psi}_{\boldsymbol{n-1}}) + \Delta \boldsymbol{M}_{\boldsymbol{n}}).$$

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

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

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

Regarding (c1), we observe that $\max_{1 \leq n \leq N} |\boldsymbol{Y}_{\boldsymbol{N,n}}| \leq \frac{1}{\sqrt{N}} \max_{1 \leq n \leq N} |\boldsymbol{\xi}_{\boldsymbol{n}} - \boldsymbol{\psi}_{\boldsymbol{n-1}}| \leq \frac{1}{\sqrt{N}} \to 0$. In order to conclude, we have to prove condition (c2), that is

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

The above convergence holds true even almost surely by usual martingale arguments (see, for instance, [11, Lemma 2]). Indeed, we have $\sum_{n \geq 1} E[\|\boldsymbol{\xi}_{\boldsymbol{n}} - \boldsymbol{\psi}_{\boldsymbol{n-1}}\|^2]/n^2 \leq \sum_{n \geq 1} n^{-2} < +\infty$ and

$$E[(\boldsymbol{\xi}_{\boldsymbol{n}} - \boldsymbol{\psi}_{\boldsymbol{n-1}})(\boldsymbol{\xi}_{\boldsymbol{n}} - \boldsymbol{\psi}_{\boldsymbol{n-1}})^{\top}|\mathcal{F}_{n-1}] = \text{diag}(\boldsymbol{\psi}_{\boldsymbol{n-1}}) - \boldsymbol{\psi}_{\boldsymbol{n-1}}\boldsymbol{\psi}_{\boldsymbol{n-1}}^{\top} \xrightarrow{a.s} \Sigma^2.$$

The last stable convergence follows from the equality

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

where the last term converges almost surely to zero by Theorem 3.5. $\qquad \square$

Remark 3.7. Equation (28) implies that $\boldsymbol{\psi} = (\boldsymbol{\psi}_{\boldsymbol{n}})_n$ is a positive (i.e. non-negative) almost super-martingale [49] and also a bounded quasi-martingale [39] because $\boldsymbol{H}(\boldsymbol{\psi}_{\boldsymbol{n}})$ is uniformly bounded by a constant and $\sum_{n \geq 1} 1/r_{n+1}^* < +\infty$.

3.4 The case $\alpha = 0$

The model introduced above for $\alpha > 0$ makes sense also 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 Sec. 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

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

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 3.8. *We have*

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

Moreover, we have

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

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, [53, Section 12.15]). Indeed, we have:

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

(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 $\sum_{n\geq 1} \xi_{n\,i} < +\infty$ a.s., 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 $\sum_{n\geq 1} \xi_{n\,i} < +\infty$ a.s., that is $\xi_{n\,i} = 0$ eventually with probability one.

For the CLT we argue as in the proof of Theorem 3.6. Indeed, we set $\boldsymbol{Y_{N,n}} = \frac{\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}}{\sqrt{N}}$ so that we have

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

where the second term converges to zero because

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

Therefore it is enough to prove that $\sum_{n=1}^{N} \boldsymbol{Y_{N,n}}$ converges stably to the desired Gaussian kernel. To this purpose we observe that $E[\boldsymbol{Y_{N,n}}|\mathcal{F}_{n-1}] = \boldsymbol{0}$ and so, in order to prove the stable convergence, we have to check the following conditions (see [22, Cor. 7] or [19, Cor. 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 condition (c2), we observe that

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

because (see, for instance, [11, Lemma 2]) $\sum_{n \geq 1} E[\|\boldsymbol{\xi_n} - \boldsymbol{\psi_{n-1}}\|^2]/n^2 \leq \sum_{n \geq 1} n^{-2} < +\infty$ and

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

$\square$

4   Proof of the goodness of fit result (Theorem 1.1)

The proof is based on Theorem 3.1 (for $\beta = 0$) and Theorem 3.3 (for $0 < \beta < 1$), whose proofs are in Sections 3.1 and 3.2, respectively (see also [47, Corollary 2]).

Proof. We mimic here the classical proof for the Pearson chi-squared test based on Sherman Morison formula (see [51]): if A is an invertible square matrix and $1 - \boldsymbol{v}^{\top} A^{-1} \boldsymbol{u} \neq 0$, then

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

Given the observation $\boldsymbol{\xi_n} = (\xi_{n\,1}, \ldots, \xi_{n\,k})^{\top}$, we define the "truncated" vector $\boldsymbol{\xi_n^*} = (\xi_{n\,1}^*, \ldots, \xi_{n\,k-1}^*)^{\top}$, given by the first $k-1$ components of $\boldsymbol{\xi_n}$. Theorem 3.1 (for $\beta = 0$) and Theorem 3.3 (for $\beta \in (0,1)$) give the Central Limit Theorem for $(\boldsymbol{\xi_n})_n$, that immediately implies

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

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

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

By assumption $p_{0\,i} > 0$ for all $i = 1, \ldots, k$ and so $\operatorname{diag}(\boldsymbol{p}^*)$ is invertible with inverse $\operatorname{diag}(\boldsymbol{p}^*)^{-1} = \operatorname{diag}(\frac{1}{p_{0\,1}}, \ldots, \frac{1}{p_{0\,k-1}})$ and, since $(\operatorname{diag}(\boldsymbol{p}^*)^{-1})\boldsymbol{p}^* = \boldsymbol{1} \in \mathbb{R}^{k-1}$, 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 Morison formula with $A = \operatorname{diag}(\boldsymbol{p}^*)$ and $\boldsymbol{u} = \boldsymbol{v} = \boldsymbol{p}^*$, and we obtain

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

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

$$\sum_{i=1}^{k}\frac{(O_i - Np_{0\,i})^2}{Np_{0\,i}} = N\sum_{i=1}^{k}\frac{(\overline{\xi}_{N\,i} - p_{0\,i})^2}{p_{0\,i}} = N\Big[\sum_{i=1}^{k-1}\frac{(\overline{\xi}_{N\,i} - p_{0\,i})^2}{p_{0\,i}} + \frac{(\overline{\xi}_{N\,k} - p_{0\,k})^2}{p_{0\,k}}\Big]$$

$$= N\Big[\sum_{i=1}^{k-1}\frac{(\overline{\xi}_{N\,i} - p_{0\,i})^2}{p_{0\,i}} + \frac{(\sum_{i=1}^{k-1}(\overline{\xi}_{N\,i} - p_{0\,i}))^2}{p_{0\,k}}\Big]$$

$$= N\sum_{i_1,i_2=1}^{k-1}(\overline{\xi}_{N\,i_1} - p_{0\,i_1})(\overline{\xi}_{N\,i_2} - p_{0\,i_2})\Big(\delta_{i_1}^{i_2}\frac{1}{p_{0\,i_1}} + \frac{1}{p_{0\,k}}\Big),$$

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

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

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 $\Gamma((k-1)/2, 1/(2\lambda))$. $\qquad\square$

A Appendix

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

In this section we prove the general Central Limit Theorem for Markov chains, used for the proof of Theorem 3.3.

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 regarding Markov chains with values in S.

Definition A.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 said a weak Feller Markov chain.

Remark A.2. 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

$$\|P^n f\|_\infty = \sup_{x \in S}|P^n f(x)| = \sup_{x \in S}\Big|\int_S f(y)P^n(x, dy)\Big|$$

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

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 the 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 the Brouwer's fixed-point theorem.

Definition A.3. 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 $f \in Lip(S)$.

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 A.4. We say that P is a *Doeblin-Fortet operator* if it is weak Feller, 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 not-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 following result follows immediately:

Lemma A.5. *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 A.6. 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 said 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 this ergodic set may be subdivided into p_j closed disjoint subsets. The number p_j is the period of the j-th 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, § 3.6], there are not other eigenvalues of modulus 1 except 1 (aperiodicity) and not other eigenfunctions except the constant for the eigenvalue equal to 1 (irriducibility) 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 A.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, whatever is 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

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

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 (see [30, Eq. (6)])

$$\sigma^2(f) = \int_S \int_S \left[f'(x,y) - m(f')\right]^2 P(x,dy)\pi(dx)$$

$$= \int_S \int_S \left[f(x,y) - m(f) + u_f(y) - u_f(x)\right]^2 P(x,dy)\pi(dx). \quad (A.1)$$

Finally, we have the following convergence result:

Theorem A.8 ([30, Théoréme 1 and Théoréme 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*

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

and

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

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

Definition A.9 ("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

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

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

Remark A.10. Denote by $\boldsymbol{p_0} = \int_S \boldsymbol{x}\pi(d\boldsymbol{x})$ 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

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

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

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

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

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

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

Theorem A.11. *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 defined as in (A.2) and in (A.3). Then, setting $\boldsymbol{S_N}(\boldsymbol{f}) = \sum_{n=0}^{N-1} \boldsymbol{f}(\psi_n, \psi_{n+1})$, we have*

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

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),

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

where $D_0 = (A_1 + A_2 A_P)(Id - A_P)^{-1}$. Moreover, for any $\boldsymbol{c} \in \mathbb{R}^k$,

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

Proof. As a consequence of Definition A.3, the spectral radius of A_P must be less than one, and hence $Id - A_P$ is invertible. Therefore, we may define

$$\boldsymbol{u_f}(\boldsymbol{x}) = D_0(\boldsymbol{x} - \boldsymbol{p_0}) = (A_1 + A_2 A_P)(Id - A_P)^{-1}(\boldsymbol{x} - \boldsymbol{p_0}),$$

so that we have

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

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

$$\int_S\int_S (\boldsymbol{y} - \boldsymbol{p_0})P(\boldsymbol{x},\boldsymbol{dy})\pi(\boldsymbol{dx}) = \int_S (\boldsymbol{y} - \boldsymbol{p_0})\pi(\boldsymbol{dy}) = \boldsymbol{0},$$

$$\int_S \int_S (\boldsymbol{y} - \boldsymbol{p_0})(\boldsymbol{y} - \boldsymbol{p_0})^\top P(\boldsymbol{x}, d\boldsymbol{y})\pi(d\boldsymbol{x}) = \int_S (\boldsymbol{y} - \boldsymbol{p_0})(\boldsymbol{y} - \boldsymbol{p_0})^\top \pi(d\boldsymbol{y}) = \Sigma_\pi^2 \qquad \text{(A.4)}$$

and

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

we can compute the quantity

$$\int_S \int_S \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})^\top P(\boldsymbol{x}, d\boldsymbol{y})\boldsymbol{\pi_\psi}(d\boldsymbol{x})$$

$$= \int_S \int_S \left[D_1(\boldsymbol{x} - \boldsymbol{p_0}) + D_2(\boldsymbol{y} - \boldsymbol{p_0}) \right]$$

$$\left[D_1(\boldsymbol{x} - \boldsymbol{p_0}) + D_2(\boldsymbol{y} - \boldsymbol{p_0}) \right]^\top P(\boldsymbol{x}, d\boldsymbol{y})\boldsymbol{\pi_\psi}(d\boldsymbol{x})$$

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

$$= \Sigma^2.$$

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

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

Therefore, in order to conclude, it is enough to note that the above convergence is a consequence of Theorem A.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 $\sigma^2(f) = \boldsymbol{c}^\top \Sigma \boldsymbol{c}$. $\qquad \square$

A.2 Coupling technique

The result proven in this subsection plays a relevant rôle in the proof of Theorem 3.3. Indeed, it shows that, under suitable assumptions, two stochastic processes can be "coupled" in a suitable way, preserving their respective joint distributions.

Set $S^* = \{\boldsymbol{x} : x_i \geq 0, |\boldsymbol{x}| = 1\}$, that is 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 A.12. *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* $\boldsymbol{x}, \boldsymbol{y} \in S^*$

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

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

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

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

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

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

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

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

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

$$2(1 - u_0) = \sum_{i=1}^k x_i - \sum_{i=1}^k y_i - 2 \sum_{i=1}^k \underline{xy}_i = \sum_{i=1}^k (x_i + y_i - \min(x_i, y_i)) - \min(x_i, y_i)$$

$$= \sum_{i=1}^k \max(x_i, y_i) - \min(x_i, y_i) = |\boldsymbol{x} - \boldsymbol{y}|.$$

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

Now, we are ready to prove the following "coupling result":

Theorem A.13. *Let $\boldsymbol{\psi}^{(1)} = (\boldsymbol{\psi}_n^{(1)})_n$ and $\boldsymbol{\psi}^{(2)} = (\boldsymbol{\psi}_n^{(2)})_n$ be two stochastic processes with values in S^* that evolve according to the following dynamics:*

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

where $a_0, a_1 \ge 0$, $\boldsymbol{c} \in \mathbb{R}^k$, $\boldsymbol{\xi}_{n+1}^{(\ell)}$ are random variables taking values in $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$ and such that

$$P\left(\boldsymbol{\xi}_{n+1}^{(\ell)} = \boldsymbol{e}_i \,\middle|\, \boldsymbol{\psi}_0^{(1)}, \boldsymbol{\xi}_1^{(\ell)}, \dots, \boldsymbol{\xi}_n^{(\ell)} \right) =$$
$$P\left(\boldsymbol{\xi}_{n+1}^{(\ell)} = \boldsymbol{e}_i \,\middle|\, \boldsymbol{\psi}_n^{(\ell)} \right) = \psi_{n\,i}^{(\ell)}, \quad \text{for } i = 1, \dots, k, \tag{A.10}$$

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

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

with $\widetilde{\boldsymbol{\psi}}_0^{(\ell)} = \boldsymbol{\psi}_0^{(\ell)}$ and

$$P\left(\widetilde{\boldsymbol{\xi}}_{n+1}^{(\ell)} = \boldsymbol{e}_i \,\middle|\, \boldsymbol{\psi}_0^{(1)}, \boldsymbol{\psi}_0^{(2)}, \widetilde{\boldsymbol{\xi}}_1^{(1)}, \widetilde{\boldsymbol{\xi}}_1^{(2)} \dots, \widetilde{\boldsymbol{\xi}}_n^{(1)}, \widetilde{\boldsymbol{\xi}}_n^{(2)} \right) =$$
$$P\left(\widetilde{\boldsymbol{\xi}}_{n+1}^{(\ell)} = \boldsymbol{e}_i \,\middle|\, \widetilde{\boldsymbol{\psi}}_n^{(\ell)} \right) = \widetilde{\psi}_{n\,i}^{(\ell)}, \quad \text{for } i = 1, \dots, k, \tag{A.12}$$

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

$$E\left[|\widetilde{\psi}_{n+1}^{(2)} - \widetilde{\psi}_{n+1}^{(1)}|\Big|\widetilde{\psi}_m^{(1)}, \widetilde{\psi}_m^{(2)}, m \leq n\right] \leq (a_0 + a_1)|\widetilde{\psi}_n^{(2)} - \widetilde{\psi}_n^{(1)}| + O(c_{n+1}^{(1)}) + O(c_{n+1}^{(2)}). \quad \text{(A.13)}$$

Remark A.14. As a consequence, for each $\ell = 1, 2$, the two stochastic processes $\widetilde{\psi}^{(\ell)}$ and $\widetilde{\xi}^{(\ell)}$ have the same joint distribution of $\psi^{(\ell)}$ and of $\xi^{(\ell)}$, respectively. Indeed, $\widetilde{\psi}_0^{(\ell)} = \psi_0^{(\ell)}$ and, by (A.9), (A.10), (A.11) and (A.12), the conditional distributions of $\widetilde{\psi}_{n+1}^\ell$ given $[\widetilde{\psi}_0^{(\ell)}, \ldots, \widetilde{\psi}_n^{(\ell)}]$ and of $\widetilde{\xi}_{n+1}^\ell$ given $[\widetilde{\psi}_0^{(\ell)}, \widetilde{\xi}_1^{(\ell)} \ldots, \widetilde{\xi}_n^{(\ell)}]$ are the same as the one of ψ_{n+1}^ℓ given $[\psi_0^{(\ell)}, \ldots, \psi_n^{(\ell)}]$ and of ξ_{n+1}^ℓ given $[\psi_0^{(\ell)}, \xi_1^{(\ell)} \ldots, \xi_n^{(\ell)}]$, respectively.

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

$$E\left[|\widetilde{\psi}_{n+1}^{(2)} - \widetilde{\psi}_{n+1}^{(1)}|\Big|\psi_0^{(1)}, \psi_0^{(2)}\right] \leq (a_0 + a_1)^{n+1}|\psi_0^{(2)} - \psi_0^{(1)}|$$
$$+ O\Big(\sum_{j=1}^{n+1}(a_0 + a_1)^{n+1-j}(c_j^{(1)} + c_j^{(2)})\Big). \quad \text{(A.14)}$$

Proof. We set $\widetilde{\psi}_0^{(\ell)} = \psi_0^{(\ell)}$, 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 $\sigma(\psi_0^{(1)}, \psi_0^{(2)})$. Then, we take the two functions $h^{(1)}, h^{(2)}$ of Lemma A.12 and, for each ℓ and any $n \geq 0$, we recursively define

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

Setting $\widetilde{\mathcal{F}}_n = \sigma(\psi_0^{(1)}, \psi_0^{(2)}, U_1, \ldots, U_n)$, we have that U_{n+1} is independent of $\widetilde{\mathcal{F}}_n$ and, by definition, $\widetilde{\xi}_n^{(\ell)}$ and $\widetilde{\psi}_n^{(\ell)}$ are $\widetilde{\mathcal{F}}_n$-measurable, for any $\ell = 1, 2$ and $n \geq 0$. Therefore, using relation (A.6), we get for any $\ell = 1, 2$, $n \geq 0$ and $i = 1, \ldots, k$,

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

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

$$E\left[|\xi_{n+1}^{(2)} - \xi_{n+1}^{(1)}|\Big|\widetilde{\mathcal{F}}_n\right] = \int_{(0,1)} |h^{(1)}(\widetilde{\psi}_n^{(1)}, \widetilde{\psi}_n^{(2)}, u) - h^{(2)}(\widetilde{\psi}_n^{(1)}, \widetilde{\psi}_n^{(2)}, u)| du$$
$$\leq |\widetilde{\psi}_n^{(1)} - \widetilde{\psi}_n^{(2)}|$$

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

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

and so inequality (A.13) holds true. $\qquad\square$

Acknowledgments

Funding Sources

References

[1] G. Aletti, I. Crimaldi, and A. Ghiglietti. Synchronization of reinforced stochastic processes with a network-based interaction. *Ann. Appl. Probab.*, 27(6):3787–3844, 2017.

[2] G. Aletti, I. Crimaldi, and A. Ghiglietti. Interacting reinforced stochastic processes: Statistical inference based on the weighted empirical means. *arXiv: 1811.10255, submitted*, 2019.

[3] G. Aletti, I. Crimaldi, and A. Ghiglietti. Networks of reinforced stochastic processes: asymptotics for the empirical means. *Forthcoming in Bernoulli*, 2019.

[4] G. Aletti and A. Ghiglietti. Interacting generalized Friedman's urn systems. *Stochastic Process. Appl.*, 127(8):2650–2678, 2017.

[5] G. Aletti, A. Ghiglietti, and A. M. Paganoni. Randomly reinforced urn designs with prespecified allocations. *J. Appl. Probab.*, 50(2):486–498, 2013.

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

[7] G. Aletti, A. Ghiglietti, and A. N. Vidyashankar. Dynamics of an adaptive randomly reinforced urn. *Bernoulli*, 24(3):2204–2255, 2018.

[8] A. Baldi Antognini and M. Zagoraiou. Multi-objective optimal designs in comparative clinical trials with covariates: The reinforced doubly adaptive biased coin design. *Ann. Statist.*, 40(3):1315–1345, 06 2012.

[9] A. Baldi Antognini and M. Zagoraiou. Estimation accuracy under covariate-adaptive randomization procedures. *Electron. J. Statist.*, 11(1):1180–1206, 2017.

[10] D. Bergh. 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 Q. Zhang and H. Yang, editors, *Pacific Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceedings*, pages 197–211, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.

[11] P. Berti, I. Crimaldi, L. Pratelli, and P. Rigo. A central limit theorem and its applications to multicolor randomly reinforced urns. *J. Appl. Probab.*, 48(2):527–546, 2011.

[12] P. Berti, I. Crimaldi, L. Pratelli, and P. Rigo. Asymptotics for randomly reinforced urns with random barriers. *J. Appl. Probab.*, 53(4):1206–1220, 2016.

[13] D. Bertoni, G. Aletti, G. Ferrandi, A. Micheletti, D. Cavicchioli, and R. Pretolani. Farmland use transitions after the cap greening: a preliminary analysis using markov chains approach. *Land Use Policy*, 79:789 – 800, 2018.

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

[15] K. C. Chanda. Chi-squared tests of goodness-of-fit for dependent observations. In *Asymptotics, Non-Parametrics and Time Series, Statist. Textbooks Monogr.*, volume 158, pages 743–756. Dekker, 1999.

[16] M.-R. Chen and M. Kuba. On generalized pólya urn models. *J. Appl. Probab.*, 50(4):1169–1186, 12 2013.

[17] A. Collevecchio, C. Cotar, and M. LiCalzi. On a preferential attachment and generalized pólya's urn model. *Ann. Appl. Probab.*, 23(3):1219–1253, 06 2013.

[18] I. Crimaldi. Central limit theorems for a hypergeometric randomly reinforced urn. *J. Appl. Probab.*, 53(3):899–913, 2016.

[19] I. Crimaldi. *Introduzione alla nozione di convergenza stabile e sue varianti (Introduction to the notion of stable convergence and its variants)*, volume 57. Unione Matematica Italiana, Monograf s.r.l., Bologna, Italy., 2016. Book written in Italian.

[20] I. Crimaldi, P. Dai Pra, P.-Y. Louis, and I. G. Minelli. Synchronization and functional central limit theorems for interacting reinforced random walks. *Stochastic Processes and their Applications*, 129(1):70–101, 2019.

[21] I. Crimaldi, P. Dai Pra, and I. G. Minelli. Fluctuation theorems for synchronization of interacting Pólya's urns. *Stochastic Process. Appl.*, 126(3):930–947, 2016.

[22] I. Crimaldi, G. Letta, and L. Pratelli. *A Strong Form of Stable Convergence*, volume 1899, pages 203–225. 2007.

[23] P. Dai Pra, P.-Y. Louis, and I. G. Minelli. Synchronization via interacting reinforcement. *J. Appl. Probab.*, 51(2):556–568, 2014.

[24] W. Doeblin and R. Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, 65:132–148, 1937.

[25] F. Eggenberger and G. Pólya. Über die statistik verketteter vorgänge. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 3(4):279–289, 1923.

[26] T. Gasser. Goodness-of-fit tests for correlated data. *Biometrika*, 62(3):563–570, 1975.

[27] A. Ghiglietti and A. M. Paganoni. Statistical properties of two-color randomly reinforced urn design targeting fixed allocations. *Electron. J. Statist.*, 8(1):708–737, 2014.

[28] A. Ghiglietti, A. N. Vidyashankar, and W. F. Rosenberger. Central limit theorem for an adaptive randomly reinforced urn model. *Ann. Appl. Probab.*, 27(5):2956–3003, 10 2017.

[29] L. J. Gleser and D. S. Moore. The effect of dependence on chi-squared and empiric distribution tests of fit. *The Annals of Statistics*, 11(4):1100–1108, 1983.

[30] Y. Guivarc'h and J. Hardy. Théorèmes limites pour une classe de chaînes de markov et applications aux difféomorphismes d'anosov. *Annales de l'I.H.P. Probabilités et statistiques*, 24(1):73–98, 1988.

[31] M. Hairer. Ergodic Properties of Markov Processes. Online at http://www.hairer.org/notes/Markov.pdf.

[32] P. Hall and C. C. Heyde. *Martingale limit theory and its application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics.

[33] O. Hernández-Lerma and J. B. Lasserre. On the classification of Markov chains via occupation measures. *Appl. Math. (Warsaw)*, 27(4):489–498, 2000.

[34] Y. Hu and F. Hu. Asymptotic properties of covariate-adaptive randomization. *Ann. Statist.*, 40(3):1794–1815, 06 2012.

[35] C. T. Ionescu Tulcea and G. Marinescu. Théorie ergodique pour des classes d'opérations non complètement continues. *Ann. of Math. (2)*, 52:140–147, 1950.

[36] D. Knoke, G. W. Bohrnstedt, and A. Potter Mee. *Statistics for Social Data Analysis.* F.E.Peacock Publishers, 2002.

[37] S. Laruelle and G. Pagés. Randomized urn models revisited using stochastic approximation. *Ann. Appl. Proba.*, 23(4):1409–1436, 2013.

[38] H. M. Mahmoud. *Pólya urn models.* Texts in Statistical Science Series. CRC Press, Boca Raton, FL, 2009.

[39] M. Métivier. *Semimartingales.* Walter de Gruyter and Co., Berlin, 1982.

[40] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[41] A. Micheletti, G. Aletti, G. Ferrandi, D. Bertoni, D. Cavicchioli, and R. Pretolani. A weighted χ^2 test to detect the presence of a major change point in non-stationary Markov chains. Submitted for publication, 2019.

[42] M. F. Norman. *Markov processes and learning models*. Academic Press, New York-London, 1972. Mathematics in Science and Engineering, Vol. 84.

[43] W. Pan. Goodness-of-fit tests for GEE with correlated binary data. *Scand. J. Statist.*, 29(1):101–110, 2002.

[44] Y. Pei, M.-L. Tang, and J. Guo. Testing the equality of two proportions for combined unilateral and bilateral data. *Communications in statistics - Simulation and computation*, 37(8):1515–1529, 2008.

[45] R. Pemantle. A survey of random processes with reinforcement. *Probab. Surveys*, 4:1–79, 2007.

[46] R. Radlow and E. F. A. Jr. An alternate multinomial assessment of the accuracy of the χ^2 test of goodness of fit. *Journal of the American Statistical Association*, 70(352):811–813, 1975.

[47] J. N. K. Rao and A. J. Scott. 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(374):221–230, 1981.

[48] A. Rényi. On stable sequences of events. *Sankhyā Ser. A*, 25:293 302, 1963.

[49] H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing Methods in Statistics*, pages 233–257. Academic Press, 1971.

[50] N. Sahasrabudhe. Synchronization and fluctuation theorems for interacting Friedman urns. *J. Appl. Probab.*, 53(4):1221–1239, 2016.

[51] J. Sherman and W. J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann. Math. Statist.*, 21(1):124–127, 03 1950.

[52] M.-L. Tang, Y.-B. Pei, W.-K. Wong, and J.-L. Li. Goodness-of-fit tests for correlated paired binary data. *Stat. Methods Med. Res.*, 21(4):331–345, 2012.

[53] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[54] Q. Zhou, P. A. Ernst, K. L. Morgan, D. B. Rubin, and A. Zhang. Sequential rerandomization. *Biometrika*, 105(3):745–752, 06 2018.