

Generalized Rescaled Pólya urn and its statistical application*

Giacomo Aletti^{1,†} and Irene Crimaldi^{2,‡}

¹*ADAMSS Center, Università degli Studi di Milano, Milan, Italy*
e-mail: giacomo.aletti@unimi.it

²*IMT School for Advanced Studies, Lucca, Italy*
e-mail: irene.crimaldi@imtlucca.it

Abstract: We introduce the Generalized Rescaled Pólya (GRP) urn, that provides a generative model for a chi-squared test of goodness of fit for the long-term probabilities of clustered data, with independence between clusters and correlation, due to a reinforcement mechanism, inside each cluster. We apply the proposed test to a data set of Twitter posts about COVID-19 pandemic: in a few words, for a classical chi-squared test the data result strongly significant for the rejection of the null hypothesis (the daily long-run sentiment rate remains constant), but, taking into account the correlation among data, the introduced test leads to a different conclusion. Beside the statistical application, we point out that the GRP urn is a simple variant of the standard Eggenberger-Pólya urn, that, with suitable choices of the parameters, shows “local” reinforcement, almost sure convergence of the empirical mean to a deterministic limit and different asymptotic behaviours of the predictive mean. Moreover, the study of this model provides the opportunity to analyze stochastic approximation dynamics, that are unusual in the related literature.

MSC2020 subject classifications: Primary 60F05, 60F15, 62F03; secondary 62F05, 62L20.

Keywords and phrases: Central limit theorem, chi-squared test, Pólya urn, reinforcement learning, reinforced stochastic process, stochastic approximation, urn model.

Received June 2021.

Contents

1	Introduction	1636
2	The Generalized Rescaled Pólya (GRP) urn	1638
3	Related literature	1640
4	Main theorem: goodness of fit result	1642

arXiv: [2010.06373](https://arxiv.org/abs/2010.06373)

*Partially supported by the Italian “Programma di Attività Integrata” (PAI), project “TOol for Fighting FakeEs” (TOFFE) funded by IMT School for Advanced Studies Lucca.

[†]Member of “Gruppo Nazionale per il Calcolo Scientifico (GNCS)” of the Italian Institute “Istituto Nazionale di Alta Matematica (INdAM)”

[‡]Member of “Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA)” of the Italian Institute “Istituto Nazionale di Alta Matematica (INdAM)”.

5	Statistical application	1644
5.1	Estimation of the parameters	1645
6	COVID-19 epidemic Twitter analysis	1648
7	Asymptotic results for the empirical means	1652
8	Proof of Theorem 7.1	1653
A	Proofs and intermediate results	1658
A.1	Proof of Theorem 4.1	1658
A.2	A preliminary central limit theorem	1660
A.3	Proof of Proposition 7.1	1661
B	Case $\sum_n \epsilon_n < +\infty$	1661
C	Computations regarding the local reinforcement	1667
C.1	Case $\alpha = \beta \in (0, 1)$	1667
C.2	Case $\alpha = \beta = 1$	1668
C.3	Case $\beta < 1$ and $0 < \alpha < \beta < (1 + \alpha)/2$	1669
D	Technical results	1670
E	Some stochastic approximation results	1672
F	Stable convergence	1675
	Acknowledgments	1676
	References	1677

1. Introduction

The standard Eggenberger-Pólya urn (see [28, 40]) has been widely studied and generalized (for instance, some recent variants can be found in [6, 7, 14, 18, 21, 38, 39, 40]). In its simplest form, this model with k -colors works as follows. An urn contains N_{0i} balls of color i , for $i = 1, \dots, k$, and, at each discrete time, a ball is extracted from the urn and then it is returned inside the urn together with $\alpha > 0$ additional balls of the same color. Therefore, if we denote by N_{ni} the number of balls of color i in the urn at time n , we have

$$N_{ni} = N_{n-1i} + \alpha \xi_{ni} \quad \text{for } n \geq 1,$$

where $\xi_{ni} = 1$ if the extracted ball at time n is of color i , and $\xi_{ni} = 0$ otherwise. The parameter α regulates the reinforcement mechanism: the greater α , the greater the dependence of N_{ni} on $\sum_{h=1}^n \xi_{hi}$.

The Generalized Rescaled Pólya (GRP) urn model is characterized by the introduction of the sequence $(\beta_n)_n$ of parameters, together with the replacement of the parameter α of the original model by a sequence $(\alpha_n)_n$, so that

$$\begin{aligned} N_{ni} &= b_i + B_{ni} && \text{with} \\ B_{n+1i} &= \beta_n B_{ni} + \alpha_{n+1} \xi_{n+1i} && n \geq 0. \end{aligned}$$

Therefore, the urn initially contains $b_i + B_{0i}$ balls of color i and the parameters $\beta_n \geq 0$, together with $\alpha_n > 0$, regulate the reinforcement mechanism. More precisely, the term $\beta_n B_{ni}$ links N_{n+1i} to the ‘‘configuration’’ at time n through

the “scaling” parameter β_n , and the term $\alpha_{n+1}\xi_{n+1 i}$ links $N_{n+1 i}$ to the outcome of the extraction at time $n + 1$ through the parameter α_{n+1} .

We are going to show that, with a suitable choice of the model parameters, we have a long-term almost sure convergence of the empirical mean $\sum_{n=1}^N \xi_{n i}/N$ to the deterministic limit $p_i = b_i/\sum_{j=1}^k b_j$, and a chi-squared goodness of fit result for the long-term probabilities $\{p_1, \dots, p_k\}$. In particular, regarding the last point, we have that the chi-squared statistics

$$\chi^2 = N \sum_{i=1}^k \frac{(\widehat{p}_i - p_i)^2}{p_i} = \sum_{i=1}^k \frac{(O_i - Np_i)^2}{Np_i}, \quad (1.1)$$

where N is the size of the sample, $\widehat{p}_i = O_i/N$, with $O_i = \sum_{n=1}^N \xi_{n i}$ the number of observations equal to i in the sample, is asymptotically distributed as $\chi^2(k-1)\lambda$, with $\lambda > 1$, or $\chi^2(k-1)N^{1-2\nu}\lambda$, where $\lambda > 0$ may be smaller than 1, but ν is always strictly smaller than $1/2$. In both cases, the presence of correlation among observations mitigates the effect in (1.1) of the sample size N , that multiplies the chi-squared distance between the observed frequencies and the expected probabilities. This aspect is important for the statistical applications in the context of a “big sample”, when a small value of the chi-squared distance might be significant, and hence a correction related to the correlation between observations is desirable (see, for instance, [12, 15, 17, 31, 32, 36, 44, 48, 49, 53]). More precisely, in the first case, the observed value of the chi-squared distance has to be compared with the “critical” value $\chi_{1-\theta}^2(k-1)\lambda/N$, where $\chi_{1-\theta}^2(k-1)$ denotes the quantile of order $1 - \theta$ of the chi-squared distribution $\chi^2(k-1)$. In the second case, the critical value for the chi-squared distance becomes $\chi_{1-\theta}^2(k-1)\lambda/N^{2\nu}$, where, although the constant λ may be smaller than 1, the effect of the sample size N is mitigated by the exponent $2\nu < 1$. In other words, for this second case, the Fisher information given by the sample does not scale with the sample size N , but with rate $N^{2\nu}$.

Summing up, the GRP urn provides a theoretical framework for a chi-squared test of goodness of fit for the long-term probabilities of correlated data, generated according to a reinforcement mechanism. Specifically, we describe a possible application in the context of clustered data, with independence between clusters and correlation, due to a reinforcement mechanism, inside each cluster. In particular, we develop a suitable estimation technique for the fundamental model parameters. We then apply the proposed test to a data set of Twitter posts about COVID-19 pandemic. Given the null hypothesis that the daily long-run sentiment rate of the posts is the same for all the considered days (suitably spaced days in the period February 20th – April 20th 2020), performing a classical χ^2 test, the data result strongly significant for the rejection of the null hypothesis, while, taking into account the correlation among posts sent in the same day, the proposed test leads to a different conclusion.

The sequel of the paper is so structured. In Section 2 we set up the notation and we define the GRP urn. In Section 3 we illustrate its relationships with previous models and we discuss the connections with related literature. In particular, the object of the present work gives us the opportunity to study

Stochastic Approximation (SA) dynamics, which are infrequent in SA literature and so fill in some theoretical gaps. In Section 4 we provide the main result of this work, that is the almost sure convergence of the empirical means to the deterministic limits p_i and the goodness of fit result for the long-term probabilities p_i , together with comments and examples. In Section 5 we describe a possible statistical application of the GRP urn and the related results: a chi-squared test of goodness of fit for the long-term probabilities of clustered data, with independence between clusters and correlation, due to a reinforcement mechanism, inside each cluster. We apply the proposed test to a data set of Twitter posts about COVID-19 pandemic. In Section 7 we state two convergence results for the empirical means, which are the basis for the proof of the main theorem. All the shown theoretical results are analytically proven. The proofs are left to Section A in the Appendix, except for the proof of Theorem 7.1, which is methodologically new and emphasizes new techniques of martingale limit theory and so it is illustrated in Section 8. Finally, in the Appendix we also provide some complements, some technical lemmas and some recalls about stochastic approximation theory and about stable convergence.

2. The Generalized Rescaled Pólya (GRP) urn

In all the sequel, we suppose given two sequences of parameters $(\alpha_n)_{n \geq 1}$, with $\alpha_n > 0$ and $(\beta_n)_{n \geq 0}$ with $\beta_n \geq 0$. Given a vector $\mathbf{x} = (x_1, \dots, x_k)^\top \in \mathbb{R}^k$, we set $|\mathbf{x}| = \sum_{i=1}^k |x_i|$ and $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x} = \sum_{i=1}^k |x_i|^2$. Moreover we denote by $\mathbf{1}$ and $\mathbf{0}$ the vectors with all the components equal to 1 and equal to 0, respectively.

The urn initially contains $b_i + B_{0i} > 0$ distinct balls of color i , with $i = 1, \dots, k$. We set $\mathbf{b} = (b_1, \dots, b_k)^\top$ and $\mathbf{B}_0 = (B_{01}, \dots, B_{0k})^\top$. We assume $|\mathbf{b}| > 0$ and we set $\mathbf{p} = \frac{\mathbf{b}}{|\mathbf{b}|}$. At each discrete time $(n+1) \geq 1$, a ball is drawn at random from the urn, obtaining the random vector $\boldsymbol{\xi}_{n+1} = (\xi_{n+11}, \dots, \xi_{n+1k})^\top$ defined as

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

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

$$\mathbf{N}_{n+1} = \mathbf{b} + \mathbf{B}_{n+1} \quad \text{with} \quad \mathbf{B}_{n+1} = \beta_n \mathbf{B}_n + \alpha_{n+1} \boldsymbol{\xi}_{n+1}, \quad (2.1)$$

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

$$|\mathbf{B}_{n+1}| = \beta_n |\mathbf{B}_n| + \alpha_{n+1}.$$

Therefore, setting $r_n^* = |\mathbf{N}_n| = |\mathbf{b}| + |\mathbf{B}_n|$, we get

$$r_{n+1}^* = r_n^* + (\beta_n - 1) |\mathbf{B}_n| + \alpha_{n+1}, \quad (2.2)$$

that is

$$r_{n+1}^* - r_n^* = |\mathbf{b}|(1 - \beta_n) - r_n^*(1 - \beta_n) + \alpha_{n+1}. \quad (2.3)$$

Moreover, setting \mathcal{F}_0 equal to the trivial σ -field and $\mathcal{F}_n = \sigma(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ for $n \geq 1$, the conditional probabilities $\boldsymbol{\psi}_n = (\psi_{n1}, \dots, \psi_{nk})^\top$ of the extraction process, also called *predictive means*, are

$$\boldsymbol{\psi}_n = E[\boldsymbol{\xi}_{n+1} | \mathcal{F}_n] = \frac{\mathbf{N}_n}{|\mathbf{N}_n|} = \frac{\mathbf{b} + \mathbf{B}_n}{r_n^*} \quad n \geq 0. \tag{2.4}$$

It is obvious that we have $|\boldsymbol{\psi}_n| = 1$. Moreover, when $\beta_n > 0$ for all n , the probability ψ_{ni} results increasing with the number of times we observed the value i , that is the random variables ξ_{ni} are generated according to a reinforcement mechanism: the probability that the extraction of color i occurs has an increasing dependence on the number of extractions of color i occurred in the past (see, e.g. [47]). More precisely, we have

$$\boldsymbol{\psi}_n = \frac{\mathbf{b} + \mathbf{B}_0 \prod_{j=0}^{n-1} \beta_j + \sum_{h=1}^n \left(\alpha_h \prod_{j=h}^{n-1} \beta_j \right) \boldsymbol{\xi}_h}{|\mathbf{b}| + |\mathbf{B}_0| \prod_{j=0}^{n-1} \beta_j + \sum_{h=1}^n \left(\alpha_h \prod_{j=h}^{n-1} \beta_j \right)}. \tag{2.5}$$

The dependence of $\boldsymbol{\psi}_n$ on $\boldsymbol{\xi}_h$ depends on the factor $f(h, n) = \alpha_h \prod_{j=h}^{n-1} \beta_j$, with $1 \leq h \leq n, n \geq 0$. In the case of the standard Eggenberger-Pólya urn, that corresponds to $\alpha_n = \alpha > 0$ and $\beta_n = 1$ for all n , each observation $\boldsymbol{\xi}_h$ has the same “weight” $f(h, n) = \alpha$. Instead, if the factor $f(h, n)$ increases with h , then the main contribution is given by the most recent extractions. We refer to this phenomenon as “local” reinforcement. For instance, this is the case when (α_n) is increasing and $\beta_n = 1$ for all n . Another case is when $\alpha_n = \alpha > 0$ and $\beta_n < 1$ for all n . The case $\beta_n = 0$ for all n is an extreme case, for which $\boldsymbol{\psi}_n$ depends only on the last extraction $\boldsymbol{\xi}_n$ (recall that conventionally $\prod_{j=n}^{n-1} = 1$). For the next examples, we will show that they exhibit a broader sense local reinforcement, in the sense that the “weight” of the observations is eventually increasing with time.

By means of (2.4), together with (2.1) and (2.2), and recalling that $\mathbf{p} = \mathbf{b}/|\mathbf{b}|$, we have

$$\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n = -\frac{(1 - \beta_n)}{r_{n+1}^*} |\mathbf{b}| (\boldsymbol{\psi}_n - \mathbf{p}) + \frac{\alpha_{n+1}}{r_{n+1}^*} (\boldsymbol{\xi}_{n+1} - \boldsymbol{\psi}_n). \tag{2.6}$$

Setting $\boldsymbol{\theta}_n = \boldsymbol{\psi}_n - \mathbf{p}$ and $\Delta \mathbf{M}_{n+1} = \boldsymbol{\xi}_{n+1} - \boldsymbol{\psi}_n = \boldsymbol{\xi}_{n+1} - \mathbf{p} - \boldsymbol{\theta}_n$ and letting $\epsilon_n = |\mathbf{b}| \frac{(1 - \beta_n)}{r_{n+1}^*}$ and $\delta_n = \alpha_{n+1}/r_{n+1}^*$, from (2.6) we obtain

$$\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n = -\epsilon_n (\boldsymbol{\psi}_n - \mathbf{p}) + \delta_n \Delta \mathbf{M}_{n+1} \tag{2.7}$$

and so

$$\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}_n = -\epsilon_n \boldsymbol{\theta}_n + \delta_n \Delta \mathbf{M}_{n+1}. \tag{2.8}$$

Therefore, the asymptotic behaviour of $(\boldsymbol{\theta}_n)$ depends on the two sequences $(\epsilon_n)_n$ and $(\delta_n)_n$.

Finally, we observe that, setting $\bar{\xi}_N = \sum_{n=1}^N \xi_n / N$ and $\mu_n = \bar{\xi}_n - \mathbf{p}$, we have the equality

$$\mu_{n+1} - \mu_n = -\frac{1}{n}(\mu_n - \theta_n) + \frac{1}{n}\Delta M_{n+1}, \quad (2.9)$$

that links the behaviour of (μ_n) and the one of (θ_n) .

Different kinds of sequences $(\epsilon_n)_n$ and $(\delta_n)_n$ provide different kinds of asymptotic behaviour of θ_n and of μ_n , i.e. of the predictive mean ψ_n and of the empirical mean $\bar{\xi}_n$. In Section 4, we provide two cases in which we have a long-term almost sure convergence of the empirical mean $\sum_{n=1}^N \xi_{ni} / N$ toward the constant $p_i = b_i / |\mathbf{b}|$, together with a chi-squared goodness of fit result. We note that the quantities p_1, \dots, p_k are related to the initial composition of the urn (since b_i is the number of balls of color i that remains constant along time, while the quantity B_{ni} is updated according to the reinforcement mechanism), but it can be seen as a (typically unknown) long-term probability distribution on the possible values (colors) $\{1, \dots, k\}$.

3. Related literature

The particular case when in the GRP urn model we have $\beta_n = \beta = 0$ for all n corresponds to a version of the so-called “memory-1 senile reinforced random walk” on a star-shaped graph introduced in [34]. The case $\alpha_n = \alpha > 0$ and $\beta_n = \beta = 1$ for all n corresponds to the standard Eggenberger-Pólya urn with an initial number $N_{0i} = b_i + B_{0i}$ of balls of color i . When (α_n) is a not-constant sequence, while $\beta_n = \beta = 1$ for all n , the GRP urn coincides with the variant of the Eggenberger-Pólya urn introduced in [46] (see also [47, Sec. 3.2]). Instead, when $\beta \neq 1$, the GRP urn does not fall in any variants of the Eggenberger-Pólya urn discussed in [47, Sec. 3.2].

The case when $\alpha_n = \alpha > 0$ and $\beta_n = \beta \geq 0$ for all n corresponds to the Rescaled Pólya (RP) urn introduced and studied in [1] and applied in [5]. It is worthwhile to point out that the two cases studied in the present work do not include (and are not included in) the case studied in [1]. Moreover, the techniques employed here and in [1] are completely different: when $\beta_n = \beta \in [0, 1)$ as in [1], the jumps $\Delta\psi_n$ do not vanish and the process $\psi = (\psi_n)_n$ converges to a stationary Markov chain and so the appropriate Markov ergodic theory is employed; in this work, we have $|\Delta\psi_n| = o(1)$, so that the martingale limit theory is here exploited to achieve the asymptotic results. Obviously, the two techniques are not exchangeable or adaptable from one contest to the other one.

When (β_n) is not identically equal to 1, since the first term in the right hand of the above relation, the GRP urn does not belong to the class of Reinforced Stochastic Processes (RSPs) studied in [2, 4, 3, 23, 24, 26]. Indeed, the RSPs are characterized by a “strict” reinforcement mechanism such that $\xi_{ni} = 1$ implies $\psi_{ni} > \psi_{n-1i}$ and so, as a consequence, ψ_{ni} has an increasing dependence on the number of times we have $\xi_{hi} = 1$ for $h = 1, \dots, n$. When (β_n) is not identically equal to 1, the GRP urn does not satisfy the “strict” reinforcement mechanism,

because the first term is positive or negative according to the sign of $(1 - \beta_n)$ and of $(\psi_n - p)$. Furthermore, we observe that equation (2.6) recalls the dynamics of a RSP with a “forcing input” (see [2, 23, 51]), but the main difference relies on the fact that such a process is driven by a classical stochastic approximation dynamics, that is a dynamics of the kind (2.7) with $\epsilon_n = \delta_n$ (up to a constant) with $\sum_n \epsilon_n = +\infty$ and $\sum_n \epsilon_n^2 < +\infty$, while the GRP urn model also allows for ϵ_n and δ_n with different rates and also for

- $\sum_n \epsilon_n = +\infty$ and $\sum_n \delta_n^2 = +\infty$ or
- $\sum_n \epsilon_n < +\infty$.

Since (2.7) is the fundamental equation of the Stochastic Approximation (SA) theory, we deem it appropriate to say a few more words on the relationship of the present work with the SA literature. The case when $\delta_n = c\epsilon_n$ in (2.7) is essentially covered by the Stochastic Approximation (SA) theory (see Section E, where we refer to [29, 37, 43, 45, 56]. See also [9, 30]). The most known case is when $\sum_n \epsilon_n = +\infty$ and $\sum_n \epsilon_n^2 < +\infty$. The case $\epsilon_n \rightarrow 0$, $\sum_n \epsilon_n = +\infty$ and $\sum_n \epsilon_n^2 = +\infty$ is less usual in literature, but it is well characterized in [37]. The case when $(\epsilon_n)_n$ and $(\delta_n)_n$ in (2.7) go to zero with different rates is typically neglected in SA literature. To our best knowledge, it is taken into consideration only in [45], where the weak convergence rate of the sequence (ψ_N) toward a certain point ψ^* is established under suitable assumptions, given the event $\{\psi_N \rightarrow \psi^*\}$. No result is given for the empirical mean $\bar{\xi}_N$, which instead is the focus of the present paper (see Theorem 4.1 below, whose proof is based on Theorem 7.1). More precisely, the assumptions on ϵ_n and δ_n in the following Theorem 7.1 imply assumption (A1.3) in [45] and so Theorem 1 in that paper provides the weak convergence rate of the sequence $(\psi_N - \psi^*)$ given the event $\{\psi_N \rightarrow \psi^*\}$. However, this result is not useful for our scope because of two reasons: first, we need convergence results for the empirical mean $\bar{\xi}_N$, not for the predictive mean ψ_N ; second, in one case included in Theorem 7.1 (see Section 7 for more details), it seems to us not immediate to check the convergence of the predictive means and so we develop another technique that does not ask for this convergence (see Section 8). Hence, the contribution of Theorem 7.1 to the SA literature is that, for a dynamics of the type (2.7) with $(\epsilon_n)_n$ and $(\delta_n)_n$ going to zero with different rates, it provides the asymptotic behaviour of the empirical mean $\bar{\xi}_N$, covering a case when $\sum_n \epsilon_n = +\infty$ and $\sum_n \delta_n^2 = +\infty$ and without requiring the convergence of the empirical means ψ_N .

Finally, it is worthwhile to point out that we also analyze the case when $\sum_n \epsilon_n < +\infty$, which is also excluded in SA literature and so it could be relevant in that field. Specifically, we prove almost sure convergence of the predictive means and of the empirical means toward a random variable and we give a central limit theorem in the sense of stable convergence. However, even if interesting from a theoretical point of view, we collect these results in Section B, because they are not related to the chi-squared test of goodness of fit.

The following statistical application of the GRP urn was inspired by [1, 15, 42]. However, those papers only deal with the case when the statistics (1.1) is asymptotically distributed as $\chi^2(k-1)\lambda$, with $\lambda > 1$, while we also face the

case when the statistics (1.1) is asymptotically distributed as $\chi^2(k-1)N^{1-2v}\lambda$, illustrating a suitable estimation procedure for the fundamental parameters $\eta = 1 - 2v$ and λ . To the best of our knowledge, this is the first work presenting a model that provides a theoretical framework for a such chi-squared test of goodness of fit.

4. Main theorem: goodness of fit result

Given a sample (ξ_1, \dots, ξ_N) generated by a GRP urn, for each $i = 1, \dots, k$, the statistics

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

counts the number of times we observed the value i . (Note that this random variable depends on the sample size N , but we simply use the symbol O_i , instead of O_{Ni} .) The theorem below states, under suitable assumptions, the almost sure convergence of the empirical mean $\hat{p}_i = O_i/N = \sum_{n=1}^N \xi_{ni}/N$ toward the probability p_i , together with a chi-squared goodness of fit test for the long-term probabilities p_1, \dots, p_k . More precisely, we prove the following result:

Theorem 4.1. *Assume $p_i > 0$ for all $i = 1, \dots, k$ and suppose to be in one of the following cases:*

- a) $\epsilon_n = (n+1)^{-\epsilon}$ and $\delta_n = c\epsilon_n$, with $\epsilon \in (0, 1]$ and $c > 0$, or
- b) $\epsilon_n = (n+1)^{-\epsilon}$, $\delta_n \sim c(n+1)^{-\delta}$, with $\epsilon \in (0, 1)$, $\delta \in (\epsilon/2, \epsilon)$ and $c > 0$.

Define the constants v and λ as

$$v = \begin{cases} 1/2 & \text{in case a)} \\ 1/2 - (\epsilon - \delta) < 1/2 & \text{in case b)} \end{cases}$$

and

$$\lambda = \begin{cases} (c+1)^2 & \text{in case a) with } \epsilon \in (0, 1), \\ (c+1)^2 + c^2 = [2c(c+1) + 1] & \text{in case a) with } \epsilon = 1, \\ \frac{c^2}{1+2(\epsilon-\delta)} & \text{in case b)}. \end{cases} \quad (4.1)$$

Then $\hat{p}_i = O_i/N \xrightarrow{a.s.} p_i$ and

$$\frac{1}{N^{1-2v}} \sum_{i=1}^k \frac{(O_i - Np_i)^2}{Np_i} = N^{2v} \sum_{i=1}^k \frac{(\hat{p}_i - p_i)^2}{p_i} \xrightarrow[N \rightarrow \infty]{d} W_* = \lambda W_0$$

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

We note that λ is a constant greater than 1 in case a); while, in case b), it is a strictly positive quantity. Moreover, in case b), we have $0 < (\epsilon - \delta) < \epsilon/2 < 1/2$

and so $(1 - 2v) = 2(\epsilon - \delta) \in (0, 1)$. As a consequence, we have $N^{1-2v}\lambda > 1$ for N large enough.

In the next two examples we show that it is possible to construct suitable sequences $(\alpha_n)_n$ and $(\beta_n)_n$ of the model such that the corresponding sequences $(\epsilon_n)_n$ and $(\delta_n)_n$ converge to zero with the same rate or with different rates and satisfy the assumptions a) or b) of the above theorem, respectively.

Example 4.1. (Case $\epsilon_n = (n + 1)^{-\epsilon}$ and $\delta_n = c\epsilon_n$, with $\epsilon > 0$ and $c > 0$)

Take $\alpha_{n+1} = c|\mathbf{b}|(1 - \beta_n)$, with $\beta_n \in [0, 1)$ and $c > 0$, that implies $\delta_n = \frac{\alpha_{n+1}}{r_{n+1}^*} = c \frac{|\mathbf{b}|(1-\beta_n)}{r_{n+1}^*} = c\epsilon_n$. Set $r_n^* = (1 + c)|\mathbf{b}|(1 - t_n)$ so that from (2.3) we obtain $t_{n+1} = \beta_n t_n$. Hence, we have

$$t_{n+1} = t_0 \prod_{k=0}^n \beta_k = \frac{c|\mathbf{b}| - |\mathbf{B}_0|}{(1 + c)|\mathbf{b}|} \prod_{k=0}^n \beta_k$$

and so

$$r_{n+1}^* = (1 + c)|\mathbf{b}| + (|\mathbf{B}_0| - c|\mathbf{b}|) \prod_{k=0}^n \beta_k.$$

Therefore, setting $\beta^* = \prod_{k=0}^\infty \beta_k \in [0, 1)$, we get $r_n^* \rightarrow r^* = (1 + c)|\mathbf{b}| + (|\mathbf{B}_0| - c|\mathbf{b}|)\beta^* > 0$. If we choose $|\mathbf{B}_0| = c|\mathbf{b}|$, then $r_n^* = r^* = (1 + c)|\mathbf{b}|$ for each n and so, setting $\beta_n = 1 - (1 + c)(1 + n)^{-\epsilon}$ with $\epsilon > 0$, we obtain $\epsilon_n = (1 + n)^{-\epsilon}$ and $\delta_n = c\epsilon_n$. Taking $\epsilon \in (0, 1]$, we have that ϵ_n and δ_n satisfy assumption a) of Theorem 4.1. Moreover, we have $\alpha_n = c|\mathbf{b}|(1 + c)n^{-\epsilon}$ and $1 - \beta_n = (1 + c)(1 + n)^{-\epsilon}$ and so, for the behaviour of the factor $f(h, n) = \alpha_h \prod_{j=h}^{n-1} \beta_j$ in (2.5), we refer to Section C.

Example 4.2. (Case $\epsilon_n = (n + 1)^{-\epsilon}$ and $\delta_n \sim c(n + 1)^{-\delta}$, with $0 < \delta < \epsilon < 1$ and $c > 0$)

Take $0 < \delta < \epsilon < 1$ and set $\gamma = \epsilon - \delta > 0$, $r_n^* = n^\gamma$ and $(1 - \beta_n) = |\mathbf{b}|^{-1}(1 + n)^{-\delta}$. We immediately have

$$\epsilon_n = |\mathbf{b}| \frac{(1 - \beta_n)}{r_{n+1}^*} = (1 + n)^{-\delta - \gamma} = (n + 1)^{-\epsilon}$$

and (2.3) yields $\alpha_{n+1} = (n + 1)^\gamma - n^\gamma [1 - |\mathbf{b}|^{-1}(1 + n)^{-\delta}] - (1 + n)^{-\delta}$, so that

$$\begin{aligned} \delta_n &= \frac{\alpha_{n+1}}{r_{n+1}^*} = \frac{\alpha_{n+1}}{(n + 1)^\gamma} = 1 - \left(1 - \frac{1}{n + 1}\right)^\gamma [1 - |\mathbf{b}|^{-1}(1 + n)^{-\delta}] - (1 + n)^{-\delta - \gamma} \\ &= 1 - \left(1 - \gamma(n + 1)^{-1} + O(n^{-2})\right) [1 - |\mathbf{b}|^{-1}(1 + n)^{-\delta}] - (1 + n)^{-\epsilon} \\ &= |\mathbf{b}|^{-1}(1 + n)^{-\delta} \left(1 + \gamma|\mathbf{b}|(n + 1)^{-1 + \delta} - |\mathbf{b}|(1 + n)^{-\epsilon + \delta} - \gamma(n + 1)^{-1} + O(n^{-2 + \delta})\right). \end{aligned}$$

Setting $c = |\mathbf{b}|^{-1} > 0$, we obtain $\epsilon_n = (n + 1)^{-\epsilon}$ and $\delta_n \sim c(n + 1)^{-\delta}$. Taking $\delta \in (\epsilon/2, \epsilon)$, we have that ϵ_n and δ_n satisfy assumption b) of Theorem 4.1.

Moreover, we have $\alpha_n = cn^{-(2\delta-\epsilon)}(1+\gamma c^{-1}n^{-1+\delta}-c^{-1}n^{-\epsilon+\delta}-\gamma n^{-1}+O(n^{-2+\delta}))$ and $(1-\beta_n) = c(1+n)^{-\delta}$, with $0 < 2\delta - \epsilon < \delta < (1 + 2\delta - \epsilon)/2$, and so, for the behaviour of the factor $f(h, n) = \alpha_h \prod_{j=h}^{n-1} \beta_j$ in (2.5), we refer to Section C.

5. Statistical application

In a big sample the observations typically can not be assumed independent and identically distributed, but they exhibit a structure in clusters, with independence between clusters and with correlation inside each cluster [15, 20, 35, 42, 54, 55]. The model and the related results presented in [1] and in the present paper may be useful in the situation when inside each cluster the probability that a certain observation takes the value i is increasing with the number of previous observations in the same cluster that are equal to the value i , hence according to a reinforcement rule. Formally, given a “big” sample $\{\xi_n : n = 1, \dots, N\}$, we suppose that the N observations are ordered so that we have the following L clusters of observations:

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

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

$$[\xi_1, \dots, \xi_{N_1}], \dots, [\xi_{\sum_{i=1}^{\ell-1} N_i+1}, \dots, \xi_{\sum_{i=1}^{\ell} N_i}], \dots, [\xi_{\sum_{i=1}^{L-1} N_i+1}, \dots, \xi_N]$$

are L independent multidimensional random variables. Moreover, we assume that the observations inside each cluster can be modelled as a GRP urn satisfying case a) or case b) of Theorem 4.1. Given certain (strictly positive) probabilities $p_1^*(\ell), \dots, p_k^*(\ell)$ for each cluster C_ℓ , we firstly want to estimate the model parameters and then perform a test with null hypothesis

$$H_0 : p_i(\ell) = p_i^*(\ell) \quad \forall i = 1, \dots, k$$

based on the statistics

$$Q_\ell = \frac{1}{N_\ell^{2(\epsilon-\delta)}} \sum_{i=1}^k \frac{(O_i(\ell) - N_\ell p_i^*(\ell))^2}{N_\ell p_i^*(\ell)}, \quad \text{with } O_i(\ell) = \#\{n \in C_\ell : \xi_{ni} = 1\}, \tag{5.1}$$

and its corresponding asymptotic distribution $\Gamma(\frac{k-1}{2}, \frac{1}{2\lambda})$, where λ is given in (4.1). Note that we can perform the above test for a certain cluster ℓ , or we can consider all the clusters together using the aggregate statistics $\sum_{\ell=1}^L Q_\ell$ and its corresponding distribution $\Gamma(\frac{L(k-1)}{2}, \frac{1}{2\lambda})$.

Regarding the probabilities $p_i^*(\ell)$, some possibilities are:

- we can take $p_i^*(\ell) = 1/k$ for all $i = 1, \dots, k$ 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 $\{\xi_n^{(1)} : n = 1, \dots, N\}$ and $\{\xi_n^{(2)} : n = 1, \dots, N\}$, take $p_i^*(\ell) = \sum_{n \in C_\ell} \xi_{ni}^{(1)} / N_\ell$ for all $i = 1, \dots, k$, and perform the test on the second sample in order to check possible changes in the long-run probabilities;
- we can take one of the clusters as benchmark, say ℓ^* , set $p_i^*(\ell) = \sum_{n \in C_{\ell^*}} \xi_{ni} / N_{\ell^*}$ for all $i = 1, \dots, k$ and $\ell \neq \ell^*$, and perform the test for the other $L - 1$ clusters in order to check differences with the benchmark cluster ℓ^* .

Finally, if we want to test possible differences among the clusters, then we can take $p_i^*(\ell) = p_i^* = \sum_{n=1}^N \xi_{ni} / N$ for all $\ell = 1, \dots, L$ and perform the test using the aggregate statistics $\sum_{\ell=1}^L Q_\ell$ with asymptotic distribution $\Gamma(\frac{(L-1)(k-1)}{2}, \frac{1}{2\lambda})$.

5.1. Estimation of the parameters

The model parameters are ϵ, δ and c . However, as we have seen, the fundamental quantities are $\eta = 2(\epsilon - \delta)$ and λ given in (4.1). Moreover, recall that in case a), we have $\eta = 0$ and $\lambda > 1$ and, in case b), we have $\eta \in (0, 1)$ and $\lambda > 0$. Therefore, according the considered model, the pair (η, λ) belongs to $S = \{0\} \times (1, +\infty) \cup (0, 1) \times (0, +\infty)$. In order to estimate the pair $(\eta, \lambda) \in S$, we define

$$T_\ell = N_\ell^\eta Q_\ell = \sum_{i=1}^k \frac{(O_i(\ell) - N_\ell p_i^*(\ell))^2}{N_\ell p_i^*(\ell)}.$$

Given the observed values t_1, \dots, t_L , the log-likelihood function of Q_ℓ reads

$$\begin{aligned} \ln(\mathcal{L}(\eta, \lambda)) &= \ln \mathcal{L}(\eta, \lambda; t_1, \dots, t_L) \\ &= -\frac{k-1}{2} L \ln(\lambda) - \frac{k-1}{2} \eta \sum_{\ell=1}^L \ln(N_\ell) - \frac{1}{2\lambda} \sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta} + R_1, \end{aligned}$$

where R_1 is a remainder term that does not depend on (η, λ) . Now, we look for the maximum likelihood estimator of the two parameters (η, λ) .

We immediately observe that, when all the clusters have the same cardinality, that is all the N_ℓ are equal to a certain N_0 , then we cannot hope to estimate η and λ , separately. Indeed, the log-likelihood function becomes

$$\begin{aligned} \ln(\mathcal{L}(\eta, \lambda)) &= \ln \mathcal{L}(\eta, \lambda; t_1, \dots, t_L) \\ &= -\frac{k-1}{2} L \left[\ln(\lambda) + \eta \ln(N_0) \right] - \frac{1}{2\lambda N_0^\eta} \sum_{\ell=1}^L t_\ell + R_1 = f(\lambda N_0^\eta). \end{aligned}$$

This fact implies that it possible to estimate only the parameter (λN_0^η) as $\widehat{\lambda N_0^\eta} = \sum_{\ell=1}^L t_\ell / (k-1)L$.

From now on, we assume that at least two clusters have different cardinality, that is at least a pair of cardinalities N_ℓ are different. We have to find (if they

exist!) the maximum points of the function $(\eta, \lambda) \mapsto \ln(\mathcal{L}(\eta, \lambda))$ on the set S , which is not closed nor limited. First of all, we note that $\ln(\mathcal{L}(\eta, \lambda)) \rightarrow -\infty$ for $\lambda \rightarrow +\infty$ and $\lambda \rightarrow 0$. Thus, the log-likelihood function has maximum value on the closure \bar{S} of S and its maximum points are stationary points belonging to $(0, 1) \times (0, +\infty)$ or they belong to $\{0, 1\} \times (0, +\infty)$. For detecting the points of the first type, we compute the gradient of the log-likelihood function, obtaining

$$\nabla_{(\eta, \lambda)} \ln \mathcal{L} = \begin{pmatrix} -\frac{k-1}{2} \sum_{\ell=1}^L \ln(N_\ell) + \frac{1}{2\lambda} \sum_{\ell=1}^L \frac{t_\ell \ln(N_\ell)}{N_\ell^\eta} \\ -\frac{k-1}{2\lambda} L + \frac{1}{2\lambda^2} \sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta} \end{pmatrix}.$$

Hence, the stationary points (η, λ) of the log-likelihood function are solutions of the system

$$\begin{cases} \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta} \ln(N_\ell)}{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta}} = \frac{\sum_{\ell=1}^L \ln(N_\ell)}{L} \\ \lambda = \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta}}{L(k-1)}. \end{cases}$$

In particular, we get that the stationary points are of the form $(\eta, \lambda(\eta))$, with

$$\lambda(\eta) = \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta}}{L(k-1)}. \quad (5.2)$$

In order to find the maximum points on the border, that is belonging to $\{0, 1\} \times (0, +\infty)$, we observe that, fixed any η , the function

$$\lambda \mapsto -\frac{k-1}{2} L \ln(\lambda) - \frac{1}{2\lambda} \sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta} + R_2,$$

where R_2 is a remainder term not depending on λ , takes its maximum value at the point $\lambda(\eta)$ defined in (5.2).

Summing up, the problem of detecting the maximum points of the log-likelihood function on \bar{S} reduces to the study of the maximum points on $[0, 1]$ of the function

$$\eta \mapsto \ln(\mathcal{L}(\eta, \lambda(\eta))) = -\frac{k-1}{2} L \ln \left(\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta} \right) - \frac{k-1}{2} \eta \sum_{\ell=1}^L \ln(N_\ell) + R_3, \quad (5.3)$$

where R_3 is a remainder term not depending on η . To this purpose, we note that we have

$$d \frac{\ln(\mathcal{L}(\eta, \lambda(\eta)))}{d\eta} = \frac{k-1}{2} L \left[\frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta} \ln(N_\ell)}{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta}} - \frac{\sum_{\ell=1}^L \ln(N_\ell)}{L} \right] = \frac{(k-1)L}{2} g(\eta),$$

where

$$g(x) = \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^x} \ln(N_\ell)}{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^x}} - \frac{\sum_{\ell=1}^L \ln(N_\ell)}{L}.$$

Setting

$$p(x, \ell) = \frac{\frac{t_\ell}{N_\ell^x}}{\sum_{l=1}^L \frac{t_l}{N_l^x}}$$

and denoting by $E_x[\cdot]$ and by $E_u[\cdot]$ the mean value with respect to the discrete probability distribution $\{p(x, \ell) : \ell = 1, \dots, L\}$ on $\{N_1, \dots, N_L\}$ and with respect to the uniform discrete distribution on $\{N_1, \dots, N_L\}$ respectively, the above function g can be written as

$$g(x) = \sum_{\ell=1}^L p(x, \ell) \ln(N_\ell) - \frac{\sum_{\ell=1}^L \ln(N_\ell)}{L} = E_x[\ln(N)] - E_u[\ln(N)].$$

Moreover, we have

$$\begin{aligned} g'(x) &= \frac{\left(-\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^x} \ln^2(N_\ell)\right) \left(\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^x}\right) + \left(\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^x} \ln(N_\ell)\right)^2}{\left(\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^x}\right)^2} \\ &= -\sum_{\ell=1}^L p(x, \ell) \ln^2(N_\ell) + \left(\sum_{\ell=1}^L p(x, \ell) \ln(N_\ell)\right)^2 = -Var_x[\ln(N)], \end{aligned}$$

where $Var_x[\cdot]$ denotes the variance with respect to the discrete probability distribution $\{p(x, \ell) : \ell = 1, \dots, L\}$ on $\{N_1, \dots, N_L\}$. Since, we are assuming that at least two N_ℓ are different, we have $Var_x[\ln(N)] > 0$ and so the function g is strictly decreasing. Finally, we observe that we have

$$Cov_u(\ln(N), T) = \frac{\sum_{\ell=1}^L t_\ell \ln(N_\ell)}{L} - \frac{\sum_{\ell=1}^L t_\ell}{L} \frac{\sum_{\ell=1}^L \ln(N_\ell)}{L} = g(0) \frac{\sum_{\ell=1}^L t_\ell}{L}$$

and

$$Cov_u(\ln(N), \frac{T}{N}) = \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell} \ln(N_\ell)}{L} - \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell}}{L} \frac{\sum_{\ell=1}^L \ln(N_\ell)}{L} = g(1) \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell}}{L},$$

where $Cov_u(\cdot, \cdot)$ denotes the covariance with respect to the discrete joint distribution concentrated on the diagonal and such that $P\{N = N_\ell, T = t_\ell\} = 1/L$ with $\ell = 1, \dots, L$. Hence, we distinguish the following cases.

First case: $Cov_u(\ln(N), T) \leq 0$

We are in the case when $g(0) \leq 0$ and so the function (5.3) is strictly decreasing for $\eta > 0$. Thus, its maximum value on $[0, 1]$ is assumed at $\hat{\eta} = 0$. Consequently, we have $\hat{\lambda} = \lambda(0) = \frac{\sum_{\ell=1}^L t_\ell}{L(k-1)}$. Recall that we need $(0, \hat{\lambda}) \in S$ and so $\hat{\lambda} > 1$. If the model fits well the data, this is a consequence. Indeed, $\hat{\lambda}$ is an unbiased estimator: $\hat{\lambda} \stackrel{d}{\sim} \Gamma(L(k-1)/2, 1/(2\lambda))$ and so $E[\hat{\lambda}] = \lambda > 1$. A value $\hat{\lambda} \leq 1$ means a bad fit of the consider model to the data (the smaller the value of λ , the worse the fitting). Note that in the threshold case ($\hat{\eta} = 0, \hat{\lambda} = 1$), the corresponding test statistics (5.1) and its distribution coincide with the classical ones used for independent observations.

Second case: $Cov_u(\ln(N), T) > 0$ and $Cov_u(\ln(N), \frac{T}{N}) < 0$

We are in the case when $g(0) > 0$ and $g(1) < 0$. Hence, the function (5.3) has a unique stationary point $\hat{\eta} \in (0, 1)$, which is the maximum point. Consequently, we have $\hat{\lambda} = \lambda(\hat{\eta}) = \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell^\eta}}{L(k-1)} > 0$. The point $(\hat{\eta}, \hat{\lambda})$ belongs to S .

Third case: $Cov_u(\ln(N), \frac{T}{N}) \geq 0$

We are in the case when $g(1) \geq 0$ and so the function (5.3) is strictly increasing on $[0, 1]$. Hence, its maximum point is at $\hat{\eta} = 1$, and, accordingly, we have $\hat{\lambda} = \lambda(1) = \frac{\sum_{\ell=1}^L \frac{t_\ell}{N_\ell}}{L(k-1)}$. However, the point $(1, \hat{\lambda})$ does not belong to S and so, in this case, we conclude that we have a bad fit of the model to the data. Note that, if the considered model fits well the data, then we have $T/N \stackrel{d}{\sim} \lambda e^{(\eta-1)\ln(N)} \chi^2(k-1)$ with $\eta < 1$ and, consequently, we expect $Cov_u(\ln(N), \frac{T}{N}) < 0$. Moreover, a value $\eta \geq 1$ in the statistics (5.1) means a central limit theorem of the type $N^{(1-\eta)/2}(\bar{\xi}_N - \mathbf{p}) \stackrel{d}{\sim} \mathcal{N}(0, CT)$ with $(1-\eta)/2 \leq 0$. This is impossible since $(\bar{\xi}_N - \mathbf{p})$ is bounded.

6. COVID-19 epidemic Twitter analysis

We illustrate the application of the above statistical methodology to a data set containing posts on the on-line social network Twitter about the COVID-19 epidemic. More precisely, the data set covers the period from February 20th (h. 11pm) to April 20th (h. 10pm) 2020, including tweets in Italian language. More details on the keywords used for the query can be found in [16]. For every message, the relative sentiment has been calculated using the *polyglot* python module developed in [19]. This module provides a numerical value v for the sentiment and we have fixed a threshold $T = 0.35$ so that we have classified as a tweet with positive sentiment those with $v > T$ and as a tweet with negative sentiment those with $v < -T$. We have discarded tweets with a value $v \in [-T, T]$.

We are in the case $k = 2$ and the random variables $\xi_n = \xi_{n1}$ take the value 1 when the sentiment of the post n is positive and the value 0 when the sentiment of the post n is negative. We have partitioned the data so that each set P_d collect the messages of the single day d , for $d = 1$ (February 20st), \dots , 61(April 20th) and then, in order to obtain independent clusters, we have set $C_\ell = P_{1+3(\ell-1)}$, for $\ell = 1, \dots, 21 = L$. (We have tested the independence of the timed sequence $\{Q_\ell : \ell = 1, \dots, 21\}$ with a Ljung–Box test and we give the results in Table 2.) Therefore N_ℓ is the total number of tweets posted during the day $1 + 3(\ell - 1)$ and $N = \sum_{\ell=1}^L N_\ell = 699\,450$ is the sample size.

Inside each cluster the sentiment associated to each message is driven by a reinforcement mechanism, that can be modelled by means of a GRP urn:

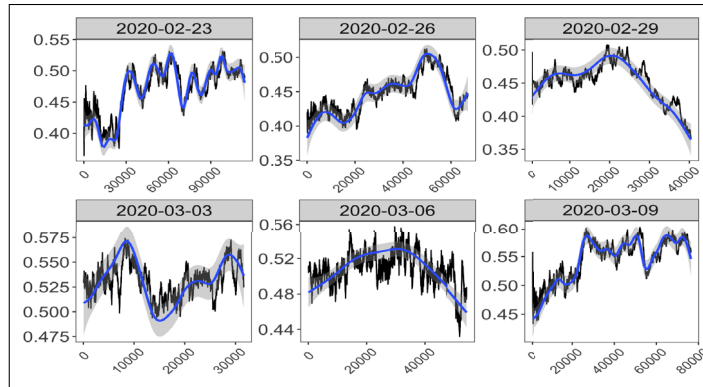


FIG 1. In each panel, the blue line is the cubic spline smoothing of the time series of the observed tweets ξ_{n+1} , together with the default confidence interval (gray), the black line represents the time series of the predictive means ψ_n given by the GRP urn model, where the parameters are estimated by maximum likelihood. Different panels refer to different clusters (we show the plots only for clusters $\ell = 2, \dots, 7$, but similar results have been obtained also for the other clusters).

the probability to have a tweet with positive sentiment is increasing with the number of past tweets with positive sentiment and the reinforcement is mostly driven by the most recent tweets, in the sense explained in Section 2 (conversely, a particle interacting approach to study opinion formation may be found in [8, 10, 11]). In Fig. 1, for clusters C_ℓ with $2 \leq \ell \leq 7$, we show the good fitting of the GRP urn model to the data (we obtain similar results also for the other clusters). Moreover, note that the main effect of the GRP urn model is the presence of “local fashions”, resulting in unexpected excursions of ψ_n around the long-run probabilities \mathbf{p} . In order to point out that the considered data set exhibits this characteristic, for each ℓ , we have computed the daily sentiment rate $\hat{p}(\ell)$, then, according to this probability, we have generated an independent sequence (ξ'_n) of Bernoulli variables, finally we have used the same smoothing procedure (i.e. classical cubic spline given in R package) to get an estimate of $\psi_n = \psi_{n-1}$, for both the real and the simulated independent data. In Fig. 2 the daily curves clearly show different behaviors in the two cases, highlighting a local reinforcement among tweets.

Our purpose is to test the null hypothesis $H_0 : \mathbf{p}(\ell) = \mathbf{p}$ for any ℓ (i.e., no differences among clusters). Therefore, taking $p_1^*(\ell) = p^* = \sum_{n=1}^N \xi_n / N$ for each ℓ , we have firstly estimated the model parameters and then we have performed the chi-squared test based on the aggregate statistics $\sum_{\ell=1}^L Q_\ell$ and its corresponding asymptotic distribution $\Gamma(\frac{(L-1)(k-1)}{2}, \frac{1}{2\lambda})$. The estimated values are $\hat{\eta} = 0.4363572$ and $\hat{\lambda} = 2.728098$ (in Fig. 3 we plot the function (5.3)). The corresponding asymptotic confidence intervals with level 0.90, computed using the asymptotic normality (with the asymptotic covariance matrix equal to the inverse of the Fisher information matrix) of the maximum likelihood estimator

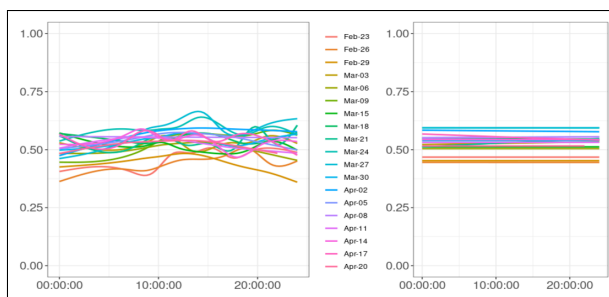


FIG 2. Smoothed daily estimate of ψ_n for the Twitter data set (left) and for the simulated independent data (right). The daily mean rate $\hat{p}(\ell)$ is the same for both the left and the right panel. x-axis: daily time. y-axis: cubic spline smoothing of the observed data ξ_n and of the simulated independent data ξ'_n .

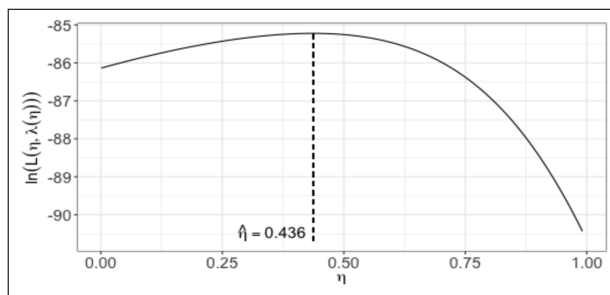


FIG 3. Plot of the function (5.3). Its maximum point gives the estimated value of the model parameter η .

and the theoretical bounds on the model parameters, are $I_\eta = (0.125; 0.747)$ and $I_\lambda = (0; 11.218)$. Moreover, since $\sum_{\ell=1}^L Q_\ell/\lambda$ is asymptotically distributed as $\Gamma((L-1)(k-1)/2, 1/2) = \chi^2((L-1)(k-1))$, we can also give the asymptotic confidence interval for λ based on that distribution: that is, $I_\lambda = (1.737; 5.028)$.

The contingency table and the associated statistics for testing H_0 is given in Table 1. The obtained χ^2 -statistics for a usual χ^2 -test is 5507.803, which is significant at any level. Under the proposed GRP urn model and the null hypothesis, the aggregate statistics $\sum_{\ell=1}^L Q_\ell$ has (asymptotic) distribution $\Gamma(\frac{L-1}{2}, \frac{1}{2\lambda})$ and the corresponding p -value associated to the data is equal to 0.4579297. The null hypothesis that the daily long-run sentiment rate of the posts is the same for all the considered days is therefore strongly rejected with a classical χ^2 test, while the same hypothesis is not rejected (at any meaningful level) if we take into account the correlation given by the reinforcement mechanism in the GRP urn model.

Note that we are not dealing with a multiple testing, since we consider a single test based on the aggregate statistics and its asymptotic distribution.

TABLE 1

Contingency table associated to COVID-Twitter data: Obs_+ (Obs_-) are the number of posts with positive (negative) sentiment posted in the day ℓ reported in the first column (DataTime); Exp_+ (Exp_-) corresponds to $N_\ell p^*$ (resp. $N_\ell(1 - p^*)$), where $N_\ell = Obs_+ + Obs_-$; χ_+^2 (χ_-^2) is the quantity $(Obs_+ - Exp_+)^2/Exp_+$ (resp. $(Obs_- - Exp_-)^2/Exp_-$); $\chi_+^{2(c)}$ ($\chi_-^{2(c)}$) is the quantity $\chi_+^2/N_\ell^{\hat{\eta}}$ (resp. $\chi_-^2/N_\ell^{\hat{\eta}}$). The statistics Q_ℓ corresponds to $\chi_+^{2(c)} + \chi_-^{2(c)}$.

Date	Obs ₊	Obs ₋	Exp ₊	Exp ₋	χ_+^2	χ_-^2	$\chi_+^{2(c)}$	$\chi_-^{2(c)}$
2020-02-20	25	43	35.11	32.89	2.91	3.11	0.46	0.49
2020-02-23	53564	60476	58886.18	55153.82	481.02	513.58	2.99	3.19
2020-02-26	29831	37175	34599.51	32406.49	657.20	701.67	5.15	5.50
2020-02-29	18220	22184	20863.18	19540.82	334.87	357.53	3.27	3.49
2020-03-03	16801	14834	16335.18	15299.82	13.28	14.18	0.14	0.15
2020-03-06	27906	27030	28366.99	26569.01	7.49	8.00	0.06	0.07
2020-03-09	41650	34769	39460.04	36958.96	121.54	129.76	0.90	0.96
2020-03-12	255	156	212.23	198.77	8.62	9.20	0.62	0.67
2020-03-15	14193	13562	14331.69	13423.31	1.34	1.43	0.02	0.02
2020-03-18	12064	10089	11439.02	10713.98	34.15	36.46	0.43	0.46
2020-03-21	11571	10026	11151.92	10445.08	15.75	16.81	0.20	0.22
2020-03-24	13339	9172	11623.88	10887.12	253.07	270.20	3.19	3.41
2020-03-27	14798	10039	12824.94	12012.06	303.55	324.09	3.67	3.92
2020-03-30	12689	10651	12051.94	11288.06	33.67	35.95	0.42	0.45
2020-04-02	12714	9300	11367.24	10646.76	159.56	170.36	2.03	2.17
2020-04-05	13373	10815	12489.82	11698.18	62.45	66.68	0.76	0.82
2020-04-08	14889	11987	13877.81	12998.19	73.68	78.67	0.86	0.92
2020-04-11	12153	10777	11840.23	11089.77	8.26	8.82	0.10	0.11
2020-04-14	13406	11430	12824.42	12011.58	26.37	28.16	0.32	0.34
2020-04-17	13977	11371	13088.80	12259.20	60.27	64.35	0.72	0.77
2020-04-20	13753	12393	13500.86	12645.14	4.71	5.03	0.06	0.06

TABLE 2

Summary of Ljung-Box test for autocorrelation of $\{Q_\ell : \ell = 1, \dots, 21\}$, with different numbers of autocorrelation lags being tested. Df: number of lags under investigation; χ^2 : Ljung-Box test statistics, which is distributed as a χ^2 distribution with Df degrees of freedom under the null hypothesis of independence; p-value: p-value of the Ljung-Box test. The strong emotional involvement of the considered period had a “mixing effect” that cancelled possible significant autocorrelation during different 3-delayed days.

Df	1	2	3	4	5	6	7	8	9	10
χ^2	3.454	3.624	4.209	4.640	5.065	7.103	8.660	8.812	10.360	12.852
p-value	0.063	0.163	0.240	0.326	0.408	0.311	0.278	0.358	0.322	0.232

Anyway, it is also possible to frame the problem in a multiple testing setting (i.e. a test for each ℓ). In that case, fixed a certain significance level, we have to compare the single statistics Q_ℓ with the quantile of the distribution $\Gamma(\frac{1}{2}, \frac{1}{2\lambda})$ associated to $(1 - \frac{level}{21})$. For example, taking the level equal to 0.05 (and so $(1 - \frac{level}{21}) = 0.997619$), we have to compare each Q_ℓ (see Table 1 or Fig. 4) with the value 25.18. As a consequence, for each ℓ , the null hypothesis is not rejected. Given the same level, if we consider the usual χ^2 -test in a multiple testing setting, we have to compare the standard χ^2 -statistics (see Table 1) with the value 9.23 and so, for almost all ℓ , the null hypothesis is strongly rejected.

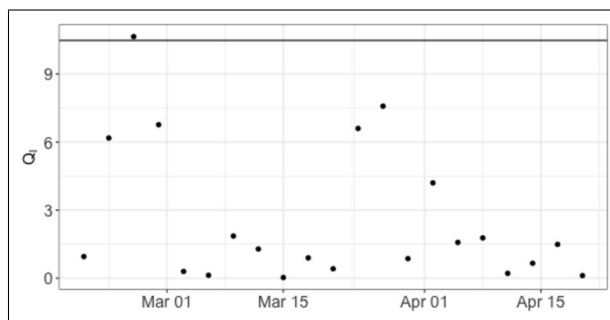


FIG 4. Plot of the Q_ℓ -series. The black line corresponds to the value 10.50. In a multiple testing setting, taking the level equal to 0.05 (and so $(1 - \frac{\text{level}}{21}) = 0.997619$), we have to compare each Q_ℓ with the value 25.18.

7. Asymptotic results for the empirical means

Theorem 4.1 is a consequence of the following Proposition 7.1 and Theorem 7.1 for the empirical means $\bar{\xi}_N$. In the sequel, we will use the symbol \xrightarrow{s} in order to denote the stable convergence: in a probability space (Ω, \mathcal{A}, P) , a sequence (Y_n) of random variables converges stably to a kernel K if, for each $H \in \mathcal{A}$ with $P(H) > 0$, it converges in distribution under the probability measure $P(\cdot|H)$ toward the mixture of probability measures $E[K(\cdot)|H]$ (for more details and a brief review on stable convergence, see Section F).

Leveraging the Stochastic Approximation results collected in Section E, we prove in Section A.3 the following result:

Proposition 7.1. Take $\epsilon_n = (n+1)^{-\epsilon}$ and $\delta_n = c\epsilon_n$, with $\epsilon \in (0, 1]$ and $c > 0$, and set $\Gamma = \text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top$. Then $\bar{\xi}_N \xrightarrow{a.s.} \mathbf{p}$ and

$$\sqrt{N}(\bar{\xi}_N - \mathbf{p}) \xrightarrow{s} \mathcal{N}(\mathbf{0}, \lambda\Gamma),$$

with $\lambda = (c+1)^2$ when $0 < \epsilon < 1$ and $\lambda = (c+1)^2 + c^2 = 2c(c+1) + 1$ when $\epsilon = 1$.

For the case when $(\epsilon_n)_n$ and $(\delta_n)_n$ in (2.7) go to zero with different rates, we prove the following theorem (the proof is illustrated in Section 8):

Theorem 7.1. Take $\epsilon_n = (n+1)^{-\epsilon}$ and $\delta_n \sim c(n+1)^{-\delta}$, with $\epsilon \in (0, 1)$, $\delta \in (\epsilon/2, \epsilon)$ and $c > 0$. Then $\bar{\xi}_N \xrightarrow{a.s.} \mathbf{p}$ and

$$N^{1/2-(\epsilon-\delta)}(\bar{\xi}_N - \mathbf{p}) \xrightarrow{s} \mathcal{N}\left(\mathbf{0}, \frac{c^2}{1+2(\epsilon-\delta)}\Gamma\right),$$

with $\Gamma = \text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top$.

In the framework of the above theorem, we can distinguish the following two cases:

- 1) $\epsilon \in (1/2, 1)$ and $\delta \in (1/2, \epsilon)$ or
- 2) $\epsilon \in (0, 1)$ and $\delta \in (\epsilon/2, \min\{\epsilon, 1/2\}] \setminus \{\epsilon\}$.

In case 1), we have $\sum_n \epsilon_n = +\infty$, $\sum_n \epsilon_n^2 < +\infty$ and $\sum_n \delta_n^2 < +\infty$ and so the typical asymptotic behaviour of the predictive mean of an urn process, that is its almost sure convergence. In case 2), we have $\sum_n \epsilon_n = +\infty$ and $\sum_n \delta_n^2 = +\infty$ (while the series $\sum_n \epsilon_n^2$ may be convergent or divergent) and it seems to us not immediate to check the convergence of the predictive means. Therefore, for the proof of Theorem 7.1 in this last case, we will employ a different technique, which is based on the L^2 -estimate of Lemma 8.1 for the predictive mean ψ_N and the almost sure convergence of the corresponding empirical mean $\bar{\psi}_{N-1}$.

Remark 7.1. It could be a bit surprising to see that the almost sure limit $\mathbf{p} = \frac{\mathbf{b}}{|\mathbf{b}|}$ of the empirical means $\boldsymbol{\xi}_N$ does not depend on the whole initial composition (i.e. $\mathbf{b} + \mathbf{B}_0$) of the urn, but only on \mathbf{b} . However, it is not so strange since b_i is the number of balls of color i that remains constant along time, while the quantity B_{ni} is scaled by β_n and updated according to the outcome of the extraction $n + 1$ (see equation (2.1)). Therefore, in particular, the quantities B_{0i} are multiplied by the factor $\prod_{j=0}^{n-1} \beta_j$ (see (2.5)), which converges to zero. Indeed, we have

$$\prod_{j=0}^{n-1} \beta_j = \prod_{j=0}^{n-1} \left(1 - \frac{\epsilon_n r_{n+1}^*}{|\mathbf{b}|}\right) \leq \prod_{j=0}^{n-1} (1 - \epsilon_n)$$

and the last term converges to zero when $\sum_{n=0}^{+\infty} \epsilon_n = +\infty$.

8. Proof of Theorem 7.1

For all the sequel, we set $\bar{\psi}_{N-1} = \sum_{n=1}^N \psi_{n-1}/N$ and $\bar{\boldsymbol{\theta}}_{N-1} = \sum_{n=1}^N \boldsymbol{\theta}_{n-1}/N$. To the proof of Theorem 7.1, we premise some intermediate results.

Lemma 8.1. *Under the same assumptions of Theorem 7.1, we have $E[\|\boldsymbol{\theta}_n\|^2] = O(n^{\epsilon-2\delta}) \rightarrow 0$.*

Proof. We observe that, starting from (2.7), we get

$$\begin{aligned} \|\boldsymbol{\theta}_{n+1}\|^2 &= \boldsymbol{\theta}_{n+1}^\top \boldsymbol{\theta}_{n+1} = (1 - \epsilon_n)^2 \|\boldsymbol{\theta}_n\|^2 + \delta_n^2 \|\Delta \mathbf{M}_{n+1}\|^2 \\ &\quad + 2(1 - \epsilon_n) \delta_n \boldsymbol{\theta}_n^\top \Delta \mathbf{M}_{n+1} \end{aligned}$$

and so

$$E[\|\boldsymbol{\theta}_{n+1}\|^2 | \mathcal{F}_n] = (1 - \epsilon_n)^2 \|\boldsymbol{\theta}_n\|^2 + \delta_n^2 E[\|\Delta \mathbf{M}_{n+1}\|^2 | \mathcal{F}_n], \tag{8.1}$$

where $\Delta \mathbf{M}_{n+1} = \boldsymbol{\xi}_{n+1} - \boldsymbol{\psi}_n$. Hence, setting $x_n = E[\|\boldsymbol{\theta}_n\|^2]$, we get

$$\begin{aligned} x_{n+1} &= (1 - 2\epsilon_n)x_n + \epsilon_n^2 x_n + \delta_n^2 E[\|\Delta \mathbf{M}_{n+1}\|^2] \\ &= (1 - 2\epsilon_n)x_n + \epsilon_n \left(\epsilon_n x_n + \frac{\delta_n^2}{\epsilon_n} E[\|\Delta \mathbf{M}_{n+1}\|^2] \right) \\ &= (1 - 2\epsilon_n)x_n + 2\epsilon_n \zeta_n, \end{aligned}$$

with $0 \leq \zeta_n = \left(\epsilon_n x_n + \frac{\delta_n^2}{\epsilon_n} E[\|\Delta M_{n+1}\|^2] \right) / 2$. Applying Lemma D.3 (with $\gamma_n = 2\epsilon_n$), we find that $\limsup_n x_n \leq \limsup_n \zeta_n$. On the other hand, since $(\Delta M_{n+1})_n$ is uniformly bounded (and so uniformly bounded in L^2) and $\epsilon_n^2 / \delta_n^2 \sim c^{-2} n^{-2(\epsilon-\delta)} \rightarrow 0$, we have $\zeta_n = O(\epsilon_n + \delta_n^2 \epsilon_n^{-1}) = O(\delta_n^2 / \epsilon_n)$ and so $x_n = O(\delta_n^2 / \epsilon_n)$. We can conclude recalling that $\delta_n^2 / \epsilon_n \sim c^2 n^{\epsilon-2\delta}$. \square

Lemma 8.2. *Under the same assumptions of Theorem 7.1, we have*

$$\bar{\theta}_{N-1} = \frac{1}{N} \sum_{n=1}^N \theta_{n-1} = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\delta_n}{\epsilon_n} \Delta M_{n+1} + \mathbf{R}_N, \tag{8.2}$$

where $\mathbf{R}_N \xrightarrow{a.s.} \mathbf{0}$ and $N^v E[|\mathbf{R}_N|] \rightarrow 0$ with $v = 1/2 - (\epsilon - \delta) \in (0, 1/2)$.

Proof. By (2.8), we have

$$\theta_n = -\frac{1}{\epsilon_n} (\theta_{n+1} - \theta_n) + \frac{\delta_n}{\epsilon_n} \Delta M_{n+1}.$$

Therefore, we can write

$$\begin{aligned} \sum_{n=0}^{N-1} \theta_n &= -\sum_{n=0}^{N-1} \frac{1}{\epsilon_n} (\theta_{n+1} - \theta_n) + \sum_{n=0}^{N-1} \frac{\delta_n}{\epsilon_n} \Delta M_{n+1} \\ &= -\left(\frac{\theta_N}{\epsilon_{N-1}} - \frac{\theta_0}{\epsilon_0} \right) - \sum_{n=1}^{N-1} \left(\frac{1}{\epsilon_{n-1}} - \frac{1}{\epsilon_n} \right) \theta_n + \sum_{n=0}^{N-1} \frac{\delta_n}{\epsilon_n} \Delta M_{n+1}, \end{aligned}$$

where the second equality is due to the Abel transformation for a series, which is the analogue of the integration by parts:

$$\begin{aligned} \sum_{n=k_1}^{k_2} f_n (g_{n+1} - g_n) &= (f_{k_2} g_{k_2+1} - f_{k_1} g_{k_1}) - \sum_{n=k_1+1}^{k_2} g_n (f_n - f_{n-1}) \\ &= (f_{k_2} g_{k_2+1} - f_{k_1} g_{k_1}) + \sum_{n=k_1+1}^{k_2} g_n (f_{n-1} - f_n). \end{aligned}$$

It follows the decomposition (8.2) with

$$\mathbf{R}_N = -\frac{1}{N} \left(\frac{\theta_N}{\epsilon_{N-1}} - \frac{\theta_0}{\epsilon_0} \right) - \frac{1}{N} \sum_{n=1}^{N-1} \left(\frac{1}{\epsilon_{n-1}} - \frac{1}{\epsilon_n} \right) \theta_n. \tag{8.3}$$

Since $|\theta_n| = O(1)$, we have

$$|\mathbf{R}_N| = O(N^{-1} \epsilon_{N-1}^{-1}) + O\left(N^{-1} \sum_{n=1}^{N-1} |\epsilon_{n-1}^{-1} - \epsilon_n^{-1}| \right)$$

Note that $\sum_{n=1}^{N-1} |\epsilon_{n-1}^{-1} - \epsilon_n^{-1}| = \epsilon_0^{-1} - \epsilon_{N-1}^{-1}$ when (ϵ_n) is decreasing and so the last term in the above expression is $O(N^{-1} \epsilon_{N-1}^{-1})$. Therefore, since $\epsilon < 1$ by assumption, we have $|\mathbf{R}_N| = O(N^{-(1-\epsilon)}) \rightarrow 0$.

Regarding the last statement of the lemma, we observe that, from what we have proven before, we obtain $N^v E[|\mathbf{R}_N|] = O(N^{v-(1-\epsilon)}) = O(N^{\delta-1/2}) \rightarrow 0$ when $\delta < 1/2$. However, in the considered cases 1) and 2), we might have $\delta \geq 1/2$. Therefore, we need other arguments in order to prove the last statement. To this purpose, we observe that, by Lemma 8.1, we have $E[|\boldsymbol{\theta}_n|] = O(n^{\epsilon/2-\delta})$ and so, by (8.3), we have

$$\begin{aligned} N^v E[|\mathbf{R}_N|] &= O(N^{-(1-v)} N^{3\epsilon/2-\delta}) + O\left(\frac{1}{N^{1-v}} \sum_{n=1}^{N-1} |\epsilon_{n-1}^{-1} - \epsilon_n^{-1}| n^{\epsilon/2-\delta}\right) \\ &= O(N^{-(1-\epsilon)/2}) + O\left(\frac{1}{N^{1-v}} \sum_{n=1}^{N-1} |\epsilon_{n-1}^{-1} - \epsilon_n^{-1}| n^{\epsilon/2-\delta}\right). \end{aligned}$$

Moreover, we have

$$\begin{aligned} \sum_{n=1}^{N-1} |\epsilon_{n-1}^{-1} - \epsilon_n^{-1}| n^{\epsilon/2-\delta} &= \sum_{n=1}^{N-1} [n^\epsilon - (n-1)^\epsilon] n^{\epsilon/2-\delta} \\ &= O\left(\sum_{n=1}^{N-1} n^{\epsilon-1+\epsilon/2-\delta}\right) = O(N^{3\epsilon/2-\delta}) = o(N^{1-v}), \end{aligned}$$

because $v = 1/2 - (\epsilon - \delta)$ and $\epsilon < 1$. Summing up, we have $N^v E[|\mathbf{R}_N|] = O(N^{-(1-\epsilon)/2}) + o(1) \rightarrow 0$. \square

Lemma 8.3. *Under the same assumptions of Theorem 7.1, we have $\bar{\boldsymbol{\theta}}_{N-1} \xrightarrow{a.s.} \mathbf{0}$, that is $\bar{\boldsymbol{\psi}}_{N-1} \xrightarrow{a.s.} \mathbf{p}$. In particular, when $\epsilon \in (1/2, 1)$ and $\delta \in (1/2, \epsilon)$, we have $\boldsymbol{\theta}_N \xrightarrow{a.s.} \mathbf{0}$, that is $\boldsymbol{\psi}_N \xrightarrow{a.s.} \mathbf{p}$.*

Proof. Let us distinguish the following two cases:

- 1) $\epsilon \in (1/2, 1)$ and $\delta \in (1/2, \epsilon)$ or
- 2) $\epsilon \in (0, 1)$ and $\delta \in (\epsilon/2, \min\{\epsilon, 1/2\}] \setminus \{\epsilon\}$.

For the case 1), we observe that, by (8.1), we have

$$E[\|\boldsymbol{\theta}_{n+1}\|^2 | \mathcal{F}_n] \leq (1 + \epsilon_n^2) \|\boldsymbol{\theta}_n\|^2 + \delta_n^2 E[\|\Delta \mathbf{M}_{n+1}\|^2 | \mathcal{F}_n],$$

where $\Delta \mathbf{M}_{n+1} = \boldsymbol{\xi}_{n+1} - \boldsymbol{\psi}_n$. Therefore, since $(\Delta \mathbf{M}_{n+1})_n$ (and so $(E[\|\Delta \mathbf{M}_{n+1}\|^2 | \mathcal{F}_n])_n$) is uniformly bounded and, in case 1), we have $\sum_n \epsilon_n^2 < +\infty$ and $\sum_n \delta_n^2 < +\infty$, the sequence $(\|\boldsymbol{\theta}_n\|^2)_n$ is a bounded non-negative almost supermartingale (see [50] for the definition of non-negative almost supermartingale). As a consequence, it converges almost surely to a certain random variable (again, see [50] for this convergence result). This limit random variable is necessarily equal to $\mathbf{0}$ because, by Lemma 8.1, we have $E[\|\boldsymbol{\theta}_n\|^2] = O(n^{\epsilon-2\delta}) \rightarrow 0$. Hence, we have the almost sure convergence of $\boldsymbol{\theta}_N$ to $\mathbf{0}$ and, consequently, the almost sure convergence of $\bar{\boldsymbol{\theta}}_{N-1}$ to $\mathbf{0}$ follows by Lemma D.2 and Remark D.1 (with $c_n = n$ and $v_{N,n} = n/N$), because $E[\boldsymbol{\theta}_{n-1} | \mathcal{F}_{n-2}] = (1 - \epsilon_{n-2}) \boldsymbol{\theta}_{n-2} \rightarrow \mathbf{0}$ almost surely.

For the case 2), we use Lemma 8.2, that gives the decomposition (8.2), with $\mathbf{R}_N \xrightarrow{a.s.} \mathbf{0}$. Indeed, by this decomposition, it is enough to prove that the term $\sum_{n=0}^{N-1} \frac{\delta_n}{\epsilon_n} \Delta \mathbf{M}_{n+1} / N$ converges almost surely to $\mathbf{0}$. To this purpose, we observe that, if we set

$$\mathbf{L}_n = \sum_{j=1}^n \frac{1}{j} \frac{\delta_{j-1}}{\epsilon_{j-1}} \Delta \mathbf{M}_j,$$

then (\mathbf{L}_n) is a square integrable martingale. Indeed, we have

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} \frac{\delta_{n-1}^2}{\epsilon_{n-1}^2} E[\|\Delta \mathbf{M}_n\|^2] = O\left(\sum_{n=1}^{+\infty} \frac{1}{n^{1+2v}}\right) < +\infty.$$

Therefore, (\mathbf{L}_n) converges almost surely, that is we have $\sum_n \frac{1}{n} \frac{\delta_{n-1}}{\epsilon_{n-1}} \Delta \mathbf{M}_n < +\infty$ almost surely. Applying Lemma D.1 (with $v_{N,n} = n/N$), we find

$$\frac{1}{N} \sum_{n=0}^{N-1} \frac{\delta_n}{\epsilon_n} \Delta \mathbf{M}_{n+1} = \sum_{n=1}^N v_{N,n} \frac{1}{n} \frac{\delta_{n-1}}{\epsilon_{n-1}} \Delta \mathbf{M}_n \xrightarrow{a.s.} \mathbf{0}$$

and so $\bar{\boldsymbol{\theta}}_{N-1} \xrightarrow{a.s.} \mathbf{0}$. □

Proof of Theorem 7.1. Set $v = 1/2 - (\epsilon - \delta) \in (0, 1/2)$ and $\lambda = c^2/[2(1 - v)] = c^2/[1 + 2(\epsilon - \delta)]$. Moreover, let us distinguish the following two cases:

- 1) $\epsilon \in (1/2, 1)$ and $\delta \in (1/2, \epsilon)$ or
- 2) $\epsilon \in (0, 1)$ and $\delta \in (\epsilon/2, \min\{\epsilon, 1/2\}] \setminus \{\epsilon\}$.

Almost sure convergence: In case 1), by Lemma 8.3, $\boldsymbol{\psi}_N$ converges almost surely to \mathbf{p} . Therefore, the almost sure convergence of $\bar{\boldsymbol{\xi}}_N$ to \mathbf{p} follows by Lemma D.2 and Remark D.1 (with $c_n = n$ and $v_{N,n} = n/N$), because $E[\boldsymbol{\xi}_{n+1} | \mathcal{F}_n] = \boldsymbol{\psi}_n \rightarrow \mathbf{p}$ almost surely and $\sum_n E[\|\boldsymbol{\xi}_n\|^2] n^{-2} \leq \sum_n n^{-2} < +\infty$.

In case 2), we use a different argument. Take $\gamma \in [0, v)$ and set

$$\mathbf{L}_n = \sum_{j=1}^n \frac{1}{j^{1-\gamma}} \frac{\delta_{j-1}}{\epsilon_{j-1}} \Delta \mathbf{M}_j.$$

Then (\mathbf{L}_n) is a square integrable martingale, because we have

$$\sum_{n=1}^{+\infty} \frac{1}{n^{2-2\gamma}} \frac{\delta_{n-1}^2}{\epsilon_{n-1}^2} E[\|\Delta \mathbf{M}_n\|^2] = O\left(\sum_{n=1}^{+\infty} \frac{1}{n^{1+2v-2\gamma}}\right) < +\infty.$$

Therefore, (\mathbf{L}_n) converges almost surely, that is we have $\sum_n \frac{1}{n^{1-\gamma}} \frac{\delta_{n-1}}{\epsilon_{n-1}} \Delta \mathbf{M}_n < +\infty$ almost surely.

By Lemma D.1 (with $v_{N,n} = (n/N)^{1-\gamma} \epsilon_{n-1} / \delta_{n-1} \sim n^{1-\gamma-\epsilon+\delta} / N^{1-\gamma}$), we get

$$\frac{1}{N^{1-\gamma}} \sum_{n=0}^{N-1} \Delta \mathbf{M}_{n+1} = \sum_{n=1}^N v_{N,n} \frac{1}{n^{1-\gamma}} \frac{\delta_{n-1}}{\epsilon_{n-1}} \Delta \mathbf{M}_n \xrightarrow{a.s.} \mathbf{0}.$$

Therefore, we have

$$N^\gamma (\bar{\xi}_N - \bar{\psi}_{N-1}) = \frac{1}{N^{1-\gamma}} \sum_{n=0}^{N-1} \Delta M_{n+1} \xrightarrow{a.s.} \mathbf{0},$$

that is $(\bar{\xi}_N - \bar{\psi}_{N-1}) = o(N^{-\gamma})$ for each $\gamma \in [0, \nu)$. Recalling Lemma 8.3, we obtain in particular that $\bar{\xi}_N$ converges almost surely to \mathbf{p} .

Second order asymptotic behaviour: We have

$$N^\nu (\bar{\xi}_N - \mathbf{p}) = N^\nu \bar{\mu}_N = N^{\nu-1/2} \sqrt{N} (\bar{\mu}_N - \bar{\theta}_{N-1}) + N^\nu \bar{\theta}_{N-1}. \quad (8.4)$$

Moreover, by Lemma 8.1, we have

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} E[|\theta_n|] &= O(N^{-1} \sum_{n=1}^N n^{\epsilon/2-\delta}) = O(N^{-1-\delta+\epsilon/2+1}) = O(N^{\epsilon/2-\delta}) \rightarrow 0, \\ \frac{1}{N} \sum_{n=0}^{N-1} E[|\theta_n|^2] &= O(N^{-1} \sum_{n=1}^N n^{\epsilon-2\delta}) = O(N^{-1-2\delta+\epsilon+1}) = O(N^{\epsilon-2\delta}) \rightarrow 0, \end{aligned}$$

and so Theorem A.1 holds true with $V = \Gamma$ (see Remark A.1). Therefore, the first term in the right side of (8.4) converges in probability to $\mathbf{0}$ because $\nu < 1/2$. Hence, if we prove that

$$N^\nu \bar{\theta}_{N-1} \xrightarrow{s} \mathcal{N}(\mathbf{0}, \lambda\Gamma), \quad (8.5)$$

then the proof is concluded.

In order to prove (8.5), we observe that, by decomposition (8.2) in Lemma 8.2, we have

$$N^\nu \bar{\theta}_{N-1} = \sum_{n=1}^N \mathbf{Y}_{N,n} + N^\nu \mathbf{R}_N,$$

where $\mathbf{Y}_{N,n} = \frac{1}{N^{1-\nu}} \frac{\delta_{n-1}}{\epsilon_{n-1}} \Delta M_n$ and $N^\nu \mathbf{R}_N$ converges in probability to $\mathbf{0}$ (because $N^\nu E[|\mathbf{R}_N|] \rightarrow 0$). Therefore, it is enough to prove that the term $\sum_{n=1}^N \mathbf{Y}_{N,n}$ stably converges to the Gaussian kernel $\mathcal{N}(0, \lambda\Gamma)$, with $\lambda = c^2/[2(1-\nu)] = c^2/[1+2(\epsilon-\delta)]$. To this purpose, we observe that $E[\mathbf{Y}_{N,n}|\mathcal{F}_{n-1}] = \mathbf{0}$ and so $\sum_{n=1}^N \mathbf{Y}_{N,n}$ converges stably to $\mathcal{N}(\mathbf{0}, \lambda\Gamma)$ if the conditions (c1) and (c2) of Theorem F.1, with $V = \lambda\Gamma$, hold true. Regarding (c1), we note that $\delta_{n-1}/\epsilon_{n-1} \sim cn^{\epsilon-\delta} = cn^{1/2-\nu}$ and so we have

$$\begin{aligned} \max_{1 \leq n \leq N} |\mathbf{Y}_{N,n}| &\leq N^{-(1-\nu)} \max_{1 \leq n \leq N} \frac{\delta_{n-1}}{\epsilon_{n-1}} |\xi_n - \psi_{n-1}| \\ &\leq N^{-(1-\nu)} \max_{1 \leq n \leq N} \frac{\delta_{n-1}}{\epsilon_{n-1}} = O(N^{-1/2}) \rightarrow 0. \end{aligned}$$

Condition (c2) means

$$\frac{1}{N^{2(1-\nu)}} \sum_{n=1}^N \frac{\delta_{n-1}^2}{\epsilon_{n-1}^2} (\xi_n - \psi_{n-1})(\xi_n - \psi_{n-1})^\top \xrightarrow{P} \lambda\Gamma. \quad (8.6)$$

We note that $N^{-2(1-v)} \sum_{n=1}^N \delta_{n-1}^2 / \epsilon_{n-1}^2 \rightarrow \lambda$, because $\delta_{n-1}^2 / \epsilon_{n-1}^2 \sim c^2 n^{1-2v}$, and

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

Therefore, in case 1), condition (8.6) immediately follows by the almost sure convergence of $\boldsymbol{\psi}_n$ to \boldsymbol{p} . It is enough to apply Lemma D.2 and Remark D.1 with $c_n = n$ and $v_{N,n} = n\delta_{n-1}^2 / (N^{2(1-v)} \epsilon_{n-1}^2) \sim c^2 n^{1+2(\epsilon-\delta)} / N^{2-2v} = c^2 (n/N)^{2(1-v)}$. In case 2), we apply again Lemma D.2 with the above c_n and $v_{N,n}$, but we note that $\boldsymbol{\psi}_n = \boldsymbol{\theta}_n + \boldsymbol{p}$ and so condition (D.1) in Lemma D.2, with $V = \lambda\Gamma$, is equivalent to

$$\frac{1}{N^{2-2v}} \sum_{n=0}^{N-1} \frac{\delta_n^2}{\epsilon_n^2} \boldsymbol{\theta}_n \xrightarrow{P} \mathbf{0} \quad \text{and} \quad \frac{1}{N^{2-2v}} \sum_{n=0}^{N-1} \frac{\delta_n^2}{\epsilon_n^2} \boldsymbol{\theta}_n \boldsymbol{\theta}_n^\top \xrightarrow{P} 0_{k \times k}.$$

These two convergences hold true because, by Lemma 8.1, we have

$$\begin{aligned} \frac{1}{N^{2-2v}} \sum_{n=0}^{N-1} \frac{\delta_n^2}{\epsilon_n^2} E[|\boldsymbol{\theta}_n|] &= O(N^{-2+2v} \sum_{n=1}^N n^{-2\delta+2\epsilon-\delta+\epsilon/2}) \\ &= O(N^{-2+2v-3\delta+5\epsilon/2+1}) = O(N^{-\delta+\epsilon/2}) \rightarrow 0, \\ \frac{1}{N^{2-2v}} \sum_{n=0}^{N-1} \frac{\delta_n^2}{\epsilon_n^2} E[|\boldsymbol{\theta}_n|^2] &= O(N^{-2+2v} \sum_{n=1}^N n^{-2\delta+2\epsilon-2\delta+\epsilon}) \\ &= O(N^{-2+2v-4\delta+3\epsilon+1}) = O(N^{-2\delta+\epsilon}) \rightarrow 0. \end{aligned}$$

Therefore, in both cases 1) and 2), conditions c1) and c2) of Theorem F.1 are satisfied and so $\sum_{n=1}^N \mathbf{Y}_{N,n}$ stably converges to the Gaussian kernel $\mathcal{N}(0, \lambda\Gamma)$. \square

Appendix A: Proofs and intermediate results

We here collect some proofs omitted in the main text of the paper.

A.1. Proof of Theorem 4.1

The proof is based on Proposition 7.1 (for case a)) and Theorem 7.1 (for case b)). The almost sure convergence of O_i/N immediately follows since $O_i/N = \bar{\xi}_{Ni}$. In order to prove the stated convergence in distribution, we mimic the classical proof for the Pearson chi-squared test based on the Sherman Morrison formula (see [52]), but see also [49, Corollary 2].

We start recalling the Sherman Morrison formula: if A is an invertible square matrix and we have $1 - \mathbf{v}^\top A^{-1} \mathbf{u} \neq 0$, then

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

Given the observation $\boldsymbol{\xi}_n = (\xi_{n1}, \dots, \xi_{nk})^\top$, we define the “truncated” vector $\boldsymbol{\xi}_n^* = (\xi_{n1}^*, \dots, \xi_{n, k-1}^*)^\top$, given by the first $k - 1$ components of $\boldsymbol{\xi}_n$. Proposition 7.1 (for case a)) and Theorem 7.1 (for case b)) give the second order asymptotic behaviour of $(\boldsymbol{\xi}_n)$, that immediately implies

$$N^v (\bar{\boldsymbol{\xi}}_N^* - \mathbf{p}^*) = \frac{\sum_{n=1}^N (\boldsymbol{\xi}_n^* - \mathbf{p}^*)}{N^{1-v}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Gamma_*), \tag{A.1}$$

where \mathbf{p}^* is given by the first $k - 1$ components of \mathbf{p} and $\Gamma_* = \lambda(\text{diag}(\mathbf{p}^*) - \mathbf{p}^* \mathbf{p}^{*T})$. By assumption $p_i > 0$ for all $i = 1, \dots, k$ and so $\text{diag}(\mathbf{p}^*)$ is invertible with inverse $\text{diag}(\mathbf{p}^*)^{-1} = \text{diag}(\frac{1}{p_1}, \dots, \frac{1}{p_{k-1}})$ and, since $(\text{diag}(\mathbf{p}^*)^{-1})\mathbf{p}^* = \mathbf{1} \in \mathbb{R}^{k-1}$, we have

$$1 - \mathbf{p}^{*T} \text{diag}(\mathbf{p}^*)^{-1} \mathbf{p}^* = 1 - \sum_{i=1}^{k-1} p_i = \sum_{i=1}^k p_i - \sum_{i=1}^{k-1} p_i = p_k > 0.$$

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

$$(\Gamma_*)^{-1} = \frac{1}{\lambda} (\text{diag}(\mathbf{p}^*) - \mathbf{p}^* \mathbf{p}^{*T})^{-1} = \frac{1}{\lambda} \left(\text{diag}(\frac{1}{p_1}, \dots, \frac{1}{p_{k-1}}) + \frac{1}{p_k} \mathbf{1}\mathbf{1}^\top \right). \tag{A.2}$$

Now, since $\sum_{i=1}^k (\bar{\xi}_{Ni} - p_i) = 0$, then $\bar{\xi}_{Nk} - p_k = -\sum_{i=1}^{k-1} (\bar{\xi}_{Ni} - p_i)$ and so we get

$$\begin{aligned} \sum_{i=1}^k \frac{(O_i - Np_i)^2}{Np_i} &= N \sum_{i=1}^k \frac{(\bar{\xi}_{Ni} - p_i)^2}{p_i} = N \left[\sum_{i=1}^{k-1} \frac{(\bar{\xi}_{Ni} - p_i)^2}{p_i} + \frac{(\bar{\xi}_{Nk} - p_k)^2}{p_k} \right] \\ &= N \left[\sum_{i=1}^{k-1} \frac{(\bar{\xi}_{Ni} - p_i)^2}{p_i} + \frac{(\sum_{i=1}^{k-1} (\bar{\xi}_{Ni} - p_i))^2}{p_k} \right] \\ &= N \sum_{i_1, i_2=1}^{k-1} (\bar{\xi}_{Ni_1} - p_{i_1})(\bar{\xi}_{Ni_2} - p_{i_2}) \left(I_{i_1, i_2} \frac{1}{p_{i_1}} + \frac{1}{p_k} \right), \end{aligned}$$

where $I_{i_1 i_2}$ is equal to 1 if $i_1 = i_2$ and equal to zero otherwise. Finally, from the above equalities, recalling (A.1) and (A.2), we obtain

$$\frac{1}{N^{1-2v}} \sum_{i=1}^k \frac{(O_i - Np_i)^2}{Np_i} = \lambda N^{2v} (\bar{\boldsymbol{\xi}}_N^* - \mathbf{p}^*)^\top (\Gamma_*)^{-1} (\bar{\boldsymbol{\xi}}_N^* - \mathbf{p}^*) \xrightarrow{d} \lambda W_0 = W_*,$$

where $1 - 2e \geq 0$ and 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 $\Gamma((k - 1)/2, 1/(2\lambda))$.

A.2. A preliminary central limit theorem

The following preliminary central limit theorem is useful for the proofs of the other central limit theorems stated in the main part and in Section B.

Theorem A.1. *If*

$$\frac{1}{N} \sum_{n=1}^N \text{diag}(\boldsymbol{\psi}_{n-1}) - \boldsymbol{\psi}_{n-1}\boldsymbol{\psi}_{n-1}^\top \xrightarrow{P} V, \tag{A.3}$$

where V is a random variable with values in the space of positive semidefinite $k \times k$ -matrices, then

$$\sqrt{N} (\bar{\boldsymbol{\mu}}_N - \bar{\boldsymbol{\theta}}_{N-1}) = \sqrt{N} (\bar{\boldsymbol{\xi}}_N - \bar{\boldsymbol{\psi}}_{N-1}) \xrightarrow{s} \mathcal{N}(\mathbf{0}, V).$$

Proof. We can write

$$\begin{aligned} \sqrt{N} (\bar{\boldsymbol{\xi}}_N - \bar{\boldsymbol{\psi}}_{N-1}) &= \frac{1}{\sqrt{N}} N (\bar{\boldsymbol{\xi}}_N - \bar{\boldsymbol{\psi}}_{N-1}) = \frac{1}{\sqrt{N}} \sum_{n=1}^N (\boldsymbol{\xi}_n - \boldsymbol{\psi}_{n-1}) \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^N \Delta M_n = \sum_{n=1}^N Y_{N,n}, \end{aligned}$$

with $\mathbf{Y}_{N,n} = N^{-1/2} \Delta M_n$. For the convergence of $\sum_{n=1}^N \mathbf{Y}_{N,n}$, we observe that $E[\mathbf{Y}_{N,k} | \mathcal{F}_{k-1}] = \mathbf{0}$ and so, by Theorem F.1, it converges stably to $\mathcal{N}(\mathbf{0}, V)$ if the conditions (c1) and (c2) hold true. Regarding (c1), we note that $\max_{1 \leq n \leq N} |\mathbf{Y}_{N,n}| \leq \frac{1}{\sqrt{N}} \max_{1 \leq n \leq N} |\boldsymbol{\xi}_n - \boldsymbol{\psi}_{n-1}| = O(1/\sqrt{N}) \rightarrow 0$. Condition (c2) means

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

The above convergence holds true by Assumption (A.3) and Lemma D.2 (with $c_n = n$ and $v_{N,n} = n/N$). Indeed, we have $\sum_{n \geq 1} E[|\boldsymbol{\xi}_n - \boldsymbol{\psi}_{n-1}|^2] / n^2 \leq \sum_{n \geq 1} n^{-2} < +\infty$ and

$$E[(\boldsymbol{\xi}_n - \boldsymbol{\psi}_{n-1})(\boldsymbol{\xi}_n - \boldsymbol{\psi}_{n-1})^\top | \mathcal{F}_{n-1}] = \text{diag}(\boldsymbol{\psi}_{n-1}) - \boldsymbol{\psi}_{n-1}\boldsymbol{\psi}_{n-1}^\top. \quad \square$$

Remark A.1. Recalling that $\boldsymbol{\psi}_n = \boldsymbol{\theta}_n + \mathbf{p}$, the convergence (A.3) with $V = \Gamma = \text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top$, means

$$\bar{\boldsymbol{\theta}}_{N-1} = \frac{1}{N} \sum_{n=1}^N \boldsymbol{\theta}_{n-1} \xrightarrow{P} \mathbf{0} \quad \text{and} \quad \frac{1}{N} \sum_{n=1}^N \boldsymbol{\theta}_{n-1}\boldsymbol{\theta}_{n-1}^\top \xrightarrow{P} 0_{k \times k},$$

where $0_{k \times k}$ is the null matrix with dimension $k \times k$.

A.3. Proof of Proposition 7.1

By Lemma D.2 (with $c_n = n$ and $v_{N,n} = n/N$), Remark D.1 and Theorem E.1, we immediately get $\bar{\xi}_N \rightarrow \mathbf{p}$ almost surely. Indeed, we have $E[\xi_{n+1} | \mathcal{F}_n] = \psi_n \rightarrow \mathbf{p}$ almost surely and $\sum_{n \geq 1} E[\|\xi_n\|^2] n^{-2} \leq \sum_{n \geq 1} n^{-2} < +\infty$.

Regarding the central limit theorem for $\bar{\xi}_N$, we have to distinguish the two cases $1/2 < \epsilon \leq 1$ or $0 < \epsilon \leq 1/2$. In the first case, the result follows from Theorem E.3, because (2.9) and the fact that $E[\Delta M_{n+1} \Delta M_{n+1}^\top | \mathcal{F}_n] = \text{diag}(\psi_{n-1}) - \psi_{n-1} \psi_{n-1}^\top \rightarrow \Gamma$ almost surely; while for the second case the result follows from Theorem A.1. Indeed, we have

$$\begin{aligned} \sqrt{N} (\bar{\xi}_N - \mathbf{p}) &= \sqrt{N} (\bar{\xi}_N - \bar{\psi}_{N-1}) + \sqrt{N} (\bar{\psi}_{N-1} - \mathbf{p}) \\ &= (c + 1) \sqrt{N} (\bar{\xi}_N - \bar{\psi}_{N-1}) - \sqrt{N} \mathbf{D}_N, \end{aligned}$$

where $\mathbf{D}_N = c (\bar{\xi}_N - \bar{\psi}_{N-1}) - (\bar{\psi}_{N-1} - \mathbf{p})$. By Theorem A.1, the term $(c + 1) \sqrt{N} (\bar{\xi}_N - \bar{\psi}_{N-1})$ stably converges to $\mathcal{N}(0, (c + 1)^2 \Gamma)$ (note that assumption (A.3) is satisfied with $V = \Gamma$, because $\psi_n \rightarrow \mathbf{p}$ almost surely). Therefore, in order to conclude, it is enough to show that $\sqrt{N} \mathbf{D}_N$ converges in probability to $\mathbf{0}$. To this purpose, we observe that, by (2.7) with $\delta_n = c\epsilon_n$, we have

$$\psi_n - \psi_{n-1} = \epsilon_{n-1} [c(\xi_n - \psi_{n-1}) - (\psi_{n-1} - \mathbf{p})]$$

and so

$$\mathbf{D}_N = \frac{1}{N} \sum_{n=1}^N \frac{\psi_n - \psi_{n-1}}{\epsilon_{n-1}}.$$

Moreover, we note that $\sum_{n=1}^{+\infty} (\psi_n - \psi_{n-1}) = \lim_N \psi_N - \psi_0 = \mathbf{p} - \psi_0 < +\infty$ and, by Lemma D.1 (with $v_{N,n} = \epsilon_{N-1}/\epsilon_{n-1}$), we get

$$\epsilon_{N-1} \sum_{n=1}^N \frac{\psi_n - \psi_{n-1}}{\epsilon_{n-1}} \xrightarrow{a.s.} \mathbf{0}.$$

For $\epsilon \leq 1/2$, this fact implies

$$\sqrt{N} \mathbf{D}_N = \frac{1}{\sqrt{N} \epsilon_{N-1}} \epsilon_{N-1} \sum_{n=1}^{N-1} \frac{\psi_n - \psi_{n-1}}{\epsilon_{n-1}} \xrightarrow{a.s.} \mathbf{0}.$$

The proof is thus concluded. □

Appendix B: Case $\sum_n \epsilon_n < +\infty$

In this section we provide some results regarding the case $\sum_n \epsilon_n < +\infty$, even if, as we will see, this case is not interesting for the chi-squared test of goodness of fit. Indeed, as shown in the following result, the empirical mean almost surely converges to a random variable, which does not coincide almost surely with a deterministic vector.

Theorem B.1. *If $\sum_{n=0}^{+\infty} \epsilon_n < +\infty$, then $\bar{\xi}_N \xrightarrow{a.s.} \psi_\infty$, where ψ_∞ is a random variable, which is not almost surely equal to a deterministic vector, that is $P(\psi_\infty \neq \mathbf{q}) > 0$ for all $\mathbf{q} \in \mathbb{R}^k$.*

Proof. When $\sum_{n=0}^{+\infty} \epsilon_n < +\infty$, the sequence (ψ_n) is a (bounded) non-negative almost supermartingale (see [50]) because, by (M2.7), we have

$$E[\psi_{n+1}|\mathcal{F}_n] = \psi_n(1 - \epsilon_n) + \epsilon_n \mathbf{p} \leq \psi_n + \epsilon_n \mathbf{p}.$$

As a consequence, it converges almost surely (and in L^p with $p \geq 1$) to a certain random variable ψ_∞ . An alternative proof of this fact follows from quasi-martingale theory [41]: indeed, since $\sum_n E[|E[\psi_{n+1}|\mathcal{F}_n] - \psi_n|] = O(\sum_n \epsilon_n) < +\infty$, the stochastic process (ψ_n) is a non-negative quasi-martingale and so it converges almost surely (and in L^p with $p \geq 1$) to a certain random variable ψ_∞ .

The almost sure convergence of $\bar{\xi}_n$ to ψ_∞ follows by Lemma D.2 and Remark D.1 (with $c_n = n$ and $v_{N,n} = n/N$), because $E[\xi_{n+1}|\mathcal{F}_n] = \psi_n \rightarrow \psi_\infty$ almost surely and $\sum_{n \geq 1} E[\|\xi_n\|^2]n^{-2} \leq \sum_{n \geq 1} n^{-2} < +\infty$.

In order to show that ψ_∞ is not almost surely equal to a deterministic vector, we set

$$y_n = E[\|\psi_n - \mathbf{p}\|^2] - \|E[\psi_n - \mathbf{p}]\|^2 = \sum_{i=1}^k Var[\psi_{ni} - p_i]$$

and observe that, starting from (M2.7), we get

$$\psi_{n+1} - \mathbf{p} = (1 - \epsilon_n)(\psi_n - \mathbf{p}) + \delta_n \Delta \mathbf{M}_{n+1}$$

and so

$$\|E[\psi_n - \mathbf{p}]\|^2 = E[\psi_n - \mathbf{p}]^\top E[\psi_n - \mathbf{p}] = (1 - \epsilon_n)^2 \|E[\psi_n - \mathbf{p}]\|^2$$

and

$$\begin{aligned} E[\|\psi_{n+1} - \mathbf{p}\|^2] &= E[(\psi_{n+1} - \mathbf{p})^\top (\psi_{n+1} - \mathbf{p})] \\ &= (1 - \epsilon_n)^2 E[\|\psi_n - \mathbf{p}\|^2] + \delta_n^2 E[\|\Delta \mathbf{M}_{n+1}\|^2]. \end{aligned}$$

Hence, we obtain

$$y_{n+1} = (1 - \epsilon_n)^2 y_n + \delta_n^2 E[\|\Delta \mathbf{M}_{n+1}\|^2] = (1 - 2\epsilon_n)y_n + \tilde{\zeta}_n \tag{B.1}$$

with $\tilde{\zeta}_n = \epsilon_n^2 y_n + \delta_n^2 E[\|\Delta \mathbf{M}_{n+1}\|^2] \geq 0$. It follows that, given \tilde{n} such that $\epsilon_n < 1/2$ for $n \geq \tilde{n}$, we have $y_N \geq y_{\tilde{n}} \prod_{n=\tilde{n}}^{N-1} (1 - 2\epsilon_n)$ for each $N \geq \tilde{n}$ and so

$$\begin{aligned} E[\|\psi_\infty - \mathbf{p}\|^2] - \|E[\psi_\infty - \mathbf{p}]\|^2 &= y_\infty = \lim_{N \rightarrow +\infty} y_N \\ &\geq y_{\tilde{n}} \prod_{n=\tilde{n}}^{+\infty} (1 - 2\epsilon_n) = y_{\tilde{n}} \exp\left(\sum_{n=\tilde{n}}^{+\infty} \ln(1 - 2\epsilon_n)\right). \end{aligned}$$

The above exponential is strictly greater than 0 because $\sum_{n=\tilde{n}}^{+\infty} \ln(1 - 2\epsilon_n) \sim -2 \sum_{n=\tilde{n}}^{+\infty} \epsilon_n > -\infty$. Therefore, if $y_{\tilde{n}} > 0$, then we have $y_\infty > 0$. This means that $\psi_\infty - \mathbf{p}$, and consequently ψ_∞ , is not almost surely equal to a deterministic vector, that is $P(\psi_\infty \neq \mathbf{q}) > 0$ for all $\mathbf{q} \in \mathbb{R}^k$. If $y_{\tilde{n}} = 0$, that is if $\psi_{\tilde{n}}$ is almost surely equal to a deterministic vector $\tilde{\psi}$, then, by (B.1), we get

$$y_{\tilde{n}+1} = \delta_n^2 E[\|\Delta \mathbf{M}_{n+1}\|^2] = \delta_n^2 E[\|\xi_{\tilde{n}+1} - \tilde{\psi}\|^2] > 0,$$

because $\delta_n > 0$ for each n and $\tilde{\psi}$ is different from a vector of the canonical base of \mathbb{R}^k by means of the assumption $b_i + B_{0i} > 0$ and equality (M2.4). It follows that we can repeat the above argument replacing \tilde{n} by $\tilde{n} + 1$ and conclude that ψ_∞ is not almost surely equal to a deterministic vector. \square

As a consequence of the above theorem, if we aim at having the almost sure convergence of $\bar{\xi}_N$ to a deterministic vector, we have to avoid the case $\sum_{n=0}^{+\infty} \epsilon_n < +\infty$. However, for the sake of completeness, we provide a second-order convergence result also in this case. First, we note that Theorem A.1 still holds true with $V = \text{diag}(\psi_\infty) - \psi_\infty \psi_\infty^\top$. Indeed, assumption (A.3) is satisfied by Lemma D.2 and Remark D.1 (with $c_n = n$ and $v_{N,n} = n/N$), because of the almost sure convergence of ψ_n to ψ_∞ . Moreover, we have the following theorem:

Theorem B.2. *Suppose to be in one of the following two cases:*

- a) $\sum_{n=1}^N n\epsilon_{n-1} = o(\sqrt{N})$ and $\sum_{n=1}^N n\delta_{n-1} = o(\sqrt{N})$;
- b) $\epsilon_n = (n+1)^{-\epsilon}$ and $\delta_n \sim c(n+1)^{-\delta}$ with $c > 0$, $\delta \in (1/2, 1)$ and $\epsilon > \delta + 1/2$ ($\epsilon = +\infty$ included, that means $\epsilon_n = 0$ for all n).

Set $v = 1/2$ and $\lambda = 1$ in case a) and $v = \delta - 1/2 \in (0, 1/2)$ and $\lambda = c^2/[2(1-v)] = c^2/(3-2\delta)$ in case b). Then, we have

$$N^v (\bar{\xi}_N - \psi_N) \xrightarrow{s} \mathcal{N}(0, \lambda\Gamma),$$

where $\Gamma = \text{diag}(\psi_\infty) - \psi_\infty \psi_\infty^\top$.

When $(\psi_N - \psi_\infty) = o_P(N^{-v})$, we also have

$$N^v (\bar{\xi}_N - \psi_\infty) \xrightarrow{s} \mathcal{N}(0, \lambda\Gamma).$$

Note that case a) covers the case $\epsilon_n = (n+1)^{-\epsilon}$ and $\delta_n \sim c(n+1)^{-\delta}$ with $c > 0$ and $\min\{\epsilon, \delta\} > 3/2$.

The case $\epsilon_n = 0$ (that is $\beta_n = 1$) for all n corresponds to the case considered in [46], but in that paper the author studies only the limit ψ_∞ and he does not provide second-order convergence results.

Proof. We have

$$\begin{aligned} N^v (\bar{\xi}_N - \psi_N) &= \frac{1}{N^{1-v}} (N\bar{\xi}_N - N\psi_N) \\ &= \frac{1}{N^{1-v}} \sum_{n=1}^N [\xi_n - \psi_{n-1} + n(\psi_{n-1} - \psi_n)] \end{aligned}$$

and, by (2.7),

$$\begin{aligned} &= \frac{1}{N^{1-v}} \sum_{n=1}^N (\xi_n - \psi_{n-1}) + \frac{1}{N^{1-v}} \sum_{n=1}^N n \epsilon_{n-1} (\psi_{n-1} - p) \\ &\quad - \frac{1}{N^{1-v}} \sum_{n=1}^N n \delta_{n-1} \Delta M_n \\ &= \frac{1}{N^{1/2-v}} \sum_{n=1}^N Y_{N,n} + Q_N + \sum_{n=1}^N Z_{N,n}, \end{aligned}$$

where

$$Y_{N,n} = \frac{\xi_n - \psi_{n-1}}{\sqrt{N}} = \frac{\Delta M_n}{\sqrt{N}}, \quad Z_{N,n} = -\frac{n \delta_{n-1} (\xi_n - \psi_{n-1})}{N^{1-v}} = \frac{n \delta_{n-1} \Delta M_n}{N^{1-v}}$$

and

$$Q_N = \frac{1}{N^{1-v}} \sum_{n=1}^N n \epsilon_{n-1} (\psi_{n-1} - p).$$

In both cases a) and b), we have $\sum_{n=1}^N n \epsilon_{n-1} = o(N^{1-v})$ and so Q_N converges almost surely to $\mathbf{0}$. Moreover, by Theorem A.1, $\sum_{n=1}^N Y_{N,n}$ stable converges to $\mathcal{N}(\mathbf{0}, V)$ with $V = \Gamma = \text{diag}(\psi_\infty) - \psi_\infty \psi_\infty^\top$. Therefore it is enough to study the convergence of $\sum_{n=1}^N Z_{N,n}$. To this purpose, we observe that, if we are in case a), then $\sum_{n=1}^N Z_{N,n}$ converges almost surely to $\mathbf{0}$ and so

$$\sqrt{N} (\bar{\xi}_N - \psi_N) \xrightarrow{s} \mathcal{N}(0, \Gamma).$$

Otherwise, if we are in case b), we observe that $E[Z_{N,n} | \mathcal{F}_{n-1}] = \mathbf{0}$ and so $\sum_{n=1}^N Z_{N,n}$ converges stably to $\mathcal{N}(\mathbf{0}, \lambda \Gamma)$ if the conditions (c1) and (c2) of Theorem F.1, with $V = \lambda \Gamma$, hold true. Regarding (c1), we observe that $\max_{1 \leq n \leq N} |Z_{N,n}| \leq \frac{1}{N^{1-v}} \max_{1 \leq n \leq N} n \delta_{n-1} |\xi_n - \psi_{n-1}| = O(1/\sqrt{N})$. Regarding condition (c2), that is

$$\sum_{n=1}^N Z_{N,n} Z_{N,n}^\top = \frac{1}{N^{2(1-v)}} \sum_{n=1}^N n^2 \delta_{n-1}^2 (\xi_n - \psi_{n-1})(\xi_n - \psi_{n-1})^\top \xrightarrow{P} \frac{c^2}{2(1-v)} \Gamma,$$

we observe that it holds true even almost surely, because $\frac{1}{N^{2(1-v)}} \sum_{n=1}^N n^2 \delta_{n-1}^2 \rightarrow c^2/[2(1-v)] = c^2/(3-2\delta)$ and

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

(see Lemma D.2 and Remark D.1 with $c_n = n$ and $v_{N,n} = n^3 \delta_{n-1}^2 / N^{2(1-v)} \sim c^2(n/N)^{3-2\delta}$). Therefore, we have

$$N^v (\bar{\xi}_N - \psi_N) \xrightarrow{s} \mathcal{N}(0, c^2(3-2\delta)^{-1} \Gamma).$$

Finally, we observe that

$$N^v (\bar{\xi}_N - \psi_\infty) = N^v (\bar{\xi}_N - \psi_N) + N^v (\psi_N - \psi_\infty).$$

Therefore, when $(\psi_N - \psi_\infty) = o_P(N^{-v})$, we have

$$N^v (\bar{\xi}_N - \psi_\infty) \xrightarrow{s} \mathcal{N}(0, \lambda\Gamma). \quad \square$$

An example of the case a) of Theorem B.2 with $(\psi_N - \psi_\infty) = o_P(N^{-v})$ is the RP urn with $\alpha_n = \alpha > 0$ and $\beta_n = \beta > 1$ (see [1]). Indeed, in this case, we have $\epsilon_n \sim c_\epsilon \beta^{-n}$ and $\delta_n \sim c_\delta \beta^{-n}$, where $c_\epsilon > 0$ and $c_\delta > 0$ are suitable constants, and $(\psi_N - \psi_\infty) = O(\beta^{-N})$. We conclude this section with other two examples regarding the case $\epsilon_n = 0$ (that is $\beta_n = 1$) for all n .

Example B.1 (Case $\epsilon_n = 0$ and $\delta_n \sim c(n+1)^{-\delta}$ with $c > 0$ and $\delta > 3/2$). If $\epsilon_n = 0$ for all n , then we have $r_n^* = |\mathbf{b}| + |\mathbf{B}_0| + \sum_{h=1}^n \alpha_h$. Therefore, if we take $\alpha_n = n^{-\delta}$, with $\delta > 3/2$, then r_n^* converges to the constant $r^* = |\mathbf{b}| + |\mathbf{B}_0| + \sum_{h=1}^{+\infty} h^{-\delta}$ and $\delta_n = \alpha_{n+1}/r_{n+1}^* \sim c\alpha_{n+1} = c(n+1)^{-\delta}$, with $c = 1/r^*$. Moreover, since $\delta > 3/2$, assumption a) of Theorem B.2 is satisfied. We also observe that $\sum_n \delta_n^2 < +\infty$ and so $\psi_{\infty i}$ is not concentrated on $\{0, 1\}$ and has no atoms in $(0, 1)$ (see [46, Th. 2 and Th. 3]). More precisely, we have

$$\psi_\infty = \frac{\mathbf{b} + \mathbf{B}_0 + \sum_{n=1}^{+\infty} \alpha_n \xi_n}{|\mathbf{b}| + |\mathbf{B}_0| + \sum_{n=1}^{+\infty} \alpha_n}$$

and so

$$\begin{aligned} & \psi_N - \psi_\infty \\ &= \frac{(\mathbf{b} + \mathbf{B}_0 + \sum_{n=1}^N \alpha_n \xi_n) \sum_{n \geq N+1} \alpha_n - (|\mathbf{b}| + |\mathbf{B}_0| + \sum_{n=1}^N \alpha_n) \sum_{n \geq N+1} \alpha_n \xi_n}{(|\mathbf{b}| + |\mathbf{B}_0| + \sum_{n=1}^N \alpha_n)(|\mathbf{b}| + |\mathbf{B}_0| + \sum_{n=1}^{+\infty} \alpha_n)} \\ &= O\left(\sum_{n \geq N+1} \alpha_n\right) = O(N^{1-\delta}). \end{aligned}$$

Since $\delta > 3/2$, we get $(\psi_N - \psi_\infty) = o(N^{-1/2})$. This fact can also be obtained as a consequence of Theorem B.3 below. Indeed, this theorem states that the rate of convergence of ψ_N to ψ_∞ is $N^{-(\delta-1/2)}$.

Note that, since $\beta_n = 1$ for all n , the factor $f(h, n)$ in (M2.5) coincides with α_h and so, in this case, it is decreasing.

Example B.2 (Case $\epsilon_n = 0$ and $\delta_n \sim c(n+1)^{-\delta}$ with $c > 0$ and $\delta \in (1/2, 1)$). As in the previous example, since $\epsilon_n = 0$ for all n , we have $r_n^* = |\mathbf{b}| + |\mathbf{B}_0| + \sum_{h=1}^n \alpha_h$. Let us set $A_n = \sum_{h=1}^n \alpha_h = \exp(bn^\alpha)$ with $b > 0$ and $\alpha \in (0, 1/2)$, which brings

to $r_n^* \sim A_n \uparrow +\infty$ and $\alpha_n = \exp(bn^\alpha) - \exp(b(n-1)^\alpha)$ and

$$\begin{aligned} \delta_{n-1} &= \frac{\alpha_n}{|\mathbf{b}| + |\mathbf{B}_0| + A_n} \sim 1 - \frac{\sum_{h=1}^{n-1} \alpha_h}{\sum_{h=1}^n \alpha_h} \\ &= 1 - \exp[b((n-1)^\alpha - n^\alpha)] \\ &= bn^\alpha (1 - (1 - n^{-1})^\alpha) + O(n^{2\alpha}(1 - (1 - n^{-1})^\alpha)^2) \\ &= bn^\alpha (\alpha n^{-1} + O(n^{-2})) + O(n^{-(2-2\alpha)}) \\ &= b\alpha n^{-(1-\alpha)} + O(n^{-(2-\alpha)}) + O(n^{-(2-2\alpha)}) = b\alpha n^{-(1-\alpha)} + O(n^{-2(1-\alpha)}), \end{aligned}$$

so that $\delta = (1 - \alpha) \in (1/2, 1)$ and $c = b\alpha > 0$. Hence, we have $\delta_n \sim c(n+1)^{-\delta}$ and assumption b) of Theorem B.2 is satisfied. We also observe that $\sum_n \delta_n^2 < +\infty$ and so $\psi_{\infty i}$ is not concentrated on $\{0, 1\}$ and has no atoms in $(0, 1)$ (see [46, Th. 2 and Th. 3]). Moreover, by Theorem B.3 below, we get that $N^v(\psi_N - \psi_\infty) \rightarrow \mathcal{N}(0, c^2(2v)^{-1}\Gamma)$, where $v = \delta - 1/2$. Hence, applying Theorem F.3, we obtain

$$N^v(\bar{\xi}_N - \psi_\infty) \xrightarrow{s} \mathcal{N}(0, c^2[2v(1-v)]^{-1}\Gamma).$$

Finally, note that, as before, since $\beta_n = 1$ for all n , the factor $f(h, n)$ in (M2.5) coincides with α_h and so, in this case, $\ell(h) = \ln(f(h, n)) = \ln(\alpha_h) \sim \ln(\delta_{h-1}) + bh^\alpha \sim bh^\alpha - b\alpha(1 - \alpha)\ln(h)$. Hence, there exists h^* such that $h \mapsto \ell(h)$ is increasing for $h \geq h^*$. Since $\max_{h \leq h^*} \ell(h) \leq C$, for a suitable constant C , the contributions of the observations until h^* are eventually smaller than those with $h \geq h^*$, that are increasing with h .

Theorem B.3. For $\epsilon_n = 0$ for all n and $\delta_n \sim c(n+1)^{-\delta}$ with $c > 0$ and $1/2 < \delta \leq 1$, we have

$$N^{\delta-1/2}(\psi_N - \psi_\infty) \rightarrow \mathcal{N}(0, c^2(2\delta-1)^{-1}\Gamma) \text{ stably in the strong sense w.r.t. } \mathcal{F},$$

where $\Gamma = \text{diag}(\psi_\infty) - \psi_\infty \psi_\infty^\top$.

Proof. We want to apply Theorem F.2. To this purpose, we recall that, when $\epsilon_n = 0$ for all n , the process (ψ_n) is a martingale with respect to \mathcal{F} . Moreover, it converges almost surely and in mean to ψ_∞ . Therefore, in order to conclude, it is enough to check conditions (c1) and (c2) of Theorem F.2. Regarding the first condition, we note that

$$\begin{aligned} N^{\delta-1/2} \sup_{n \geq N} |\psi_n - \psi_{n+1}| &= N^{\delta-1/2} \sup_{n \geq N} \delta_n |\Delta \mathbf{M}_{n+1}| \\ &= O(N^{\delta-1/2-\delta}) = O(N^{-1/2}) \rightarrow 0. \end{aligned}$$

Finally, regarding the second condition, we observe that

$$N^{2\delta-1} \sum_{n \geq N} (\psi_n - \psi_{n+1})(\psi_n - \psi_{n+1})^\top$$

$$\sim N^{2\delta-1} c^2 \sum_{n \geq N} (n+1)^{-2\delta} (\Delta M_{n+1})(\Delta M_{n+1})^\top \xrightarrow{a.s.} \frac{c^2}{(2\delta-1)} \Gamma,$$

where the almost sure convergence follows from [24, Lemma 4.1] and the fact that

$$E[(\Delta M_{n+1})(\Delta M_{n+1})^\top | \mathcal{F}_n] = E[(\xi_{n+1} - \psi_n)(\xi_{n+1} - \psi_n)^\top | \mathcal{F}_n] \xrightarrow{a.s.} \Gamma. \quad \square$$

Appendix C: Computations regarding the local reinforcement

Suppose $\alpha_n \sim an^{-\alpha}$ for $n \geq 1$ and $(1 - \beta_n) \sim b(n+1)^{-\beta}$ for $n \geq 0$. In the following subsections we study the behaviour of the factor $f(h, n) = \alpha_h \prod_{j=h}^{n-1} \beta_j$ in some particular cases that cover the cases of the two examples in Section M4. Specifically, for all the considered cases, we set $\ell(h, n) = \ln(\alpha_h \prod_{j=h}^{n-1} \beta_j) = \ln(\alpha_h) + \sum_{j=h}^{n-1} \ln(\beta_j)$ for $n \geq h$, so that

$$\ell(h_*, n) - \ell(h, n) = \ln(\alpha_{h_*}) - \ln(\alpha_h) - \sum_{j=h}^{h_*-1} \ln(\beta_j), \quad h_* \geq h, \quad (C.1)$$

and we prove that there exists h_* such that $\max_{h \leq h_*} \ell(h, n) \leq \ell(h_*, n)$ and $h \mapsto \ell(h, n)$ is increasing for $h \geq h_*$. This means that the weights $f(h, n)$ of the observations until h_* are smaller than those with $h \geq h_*$ and the contribution of the observation for $h \geq h_*$ is increasing with h .

C.1. Case $\alpha = \beta \in (0, 1)$

Suppose $\alpha_n = an^{-\alpha}$ and $1 - \beta_n = b(n+1)^{-\alpha}$, with $a, b > 0$ and $\alpha \in (0, 1)$. For $n \geq h$, we have

$$\begin{aligned} \ell(h+1, n) - \ell(h, n) &= \ln(a(h+1)^{-\alpha}) - \ln(ah^{-\alpha}) - \ln(1 - b(h+1)^{-\alpha}) \\ &= -\alpha \ln\left(1 + \frac{1}{h}\right) - \ln\left(1 - \frac{b}{(h+1)^\alpha}\right) = -\frac{\alpha}{h} + \frac{b}{(h+1)^\alpha}. \end{aligned}$$

Since $\alpha < 1$, there exists h_0 such that the function $h \mapsto \ell(h, n)$ is monotonically increasing for $h \geq h_0$. Now, fix $\eta > 0$ and let j_0 such that $j \geq j_0$ implies $\ln(\beta_j) \leq -\frac{bj^{-\alpha}}{1+\eta}$. Then take $h_* \geq \max(h_0, j_0) + 1$ and $h \leq h_0 - 1$. For h_* large enough, we get by (C.1)

$$\begin{aligned} \ell(h_*, n) - \ell(h, n) &= \ln(ah_*^{-\alpha}) - \ln(ah^{-\alpha}) - \sum_{j=h}^{h_*-1} \ln(\beta_j) \\ &\geq \ln(h_*^{-\alpha}) + \sum_{j=\max(h_0, j_0)}^{h_*-1} \frac{bj^{-\alpha}}{1+\eta} \end{aligned}$$

$$\begin{aligned}
&\geq -\alpha \ln(h_*) + C_1 + \frac{b}{1+\eta} \int_{\max(h_0, j_0)}^{h_*-1} x^{-\alpha} dx \\
&= -\alpha \ln(h_*) + C_1 + \frac{b}{(1+\eta)(1-\alpha)} [(h_*-1)^{1-\alpha} \\
&\quad - \max(h_0, j_0)^{1-\alpha}] \\
&= C_2 - \alpha \ln(h_*) + \frac{b}{(1+\eta)(1-\alpha)} (h_*-1)^{1-\alpha} \geq 0.
\end{aligned}$$

Therefore, taking h^* large enough, we have $\max_{h \leq h_*} \ell(h, n) = \max_{h \leq h_0-1} \ell(h, n) \vee \max_{h_0 \leq h \leq h_*} \ell(h, n) \leq \ell(h_*, n)$.

C.2. Case $\alpha = \beta = 1$

Suppose $\alpha_n = an^{-1}$ and $1 - \beta_n = b(n+1)^{-1}$, with $a > 0$ and $b > 1$. For $n \geq h$, we have

$$\begin{aligned}
\ell(h+1, n) - \ell(h, n) &= \ln(a(h+1)^{-1}) - \ln(ah^{-1}) - \ln(1 - b(h+1)^{-1}) \\
&= -\ln\left(1 + \frac{1}{h}\right) - \ln\left(1 - \frac{b}{(h+1)}\right) = \frac{b-1}{h+1} + o(h^{-1}).
\end{aligned}$$

Since $b > 1$, we can argue as in the previous subsection. Therefore, there exists h_0 such that the function $h \mapsto \ell(h, n)$ is monotonically increasing for $h \geq h_0$. Now, fix $\eta = (b-1)/(b+1) > 0$ and let j_0 such that $j \geq j_0$ implies $\ln(\beta_j) \leq -\frac{bj^{-1}}{1+\eta}$. Then take $h_* \geq \max(h_0, j_0) + 1$ and $h \leq h_0 - 1$. For h_* large enough, we get by (C.1)

$$\begin{aligned}
\ell(h_*, n) - \ell(h, n) &= \ln(ah_*^{-1}) - \ln(ah^{-1}) - \sum_{j=h}^{h_*-1} \ln(\beta_j) \\
&\geq \ln(h_*^{-1}) + \sum_{j=\max(h_0, j_0)}^{h_*-1} \frac{bj^{-1}}{1+\eta} \\
&\geq -\ln(h_*) + C_1 + \frac{b}{1+\eta} \int_{\max(h_0, j_0)}^{h_*-1} x^{-1} dx \\
&= -\ln(h_*) + C_1 + \frac{b}{(1+\eta)} [\ln(h_*-1) - \ln(\max(h_0, j_0))] \\
&= C_2 + \frac{b-1-\eta}{(1+\eta)} \ln(h_*) - O(1/h_*) \\
&= C_2 + \frac{b(b-1)}{2b} \ln(h_*) - O(1/h_*) \geq 0.
\end{aligned}$$

Therefore, taking h^* large enough, we have $\max_{h \leq h_*} \ell(h, n) = \max_{h \leq h_0-1} \ell(h, n) \vee \max_{h_0 \leq h \leq h_*} \ell(h, n) \leq \ell(h_*, n)$.

C.3. Case $\beta < 1$ and $0 < \alpha < \beta < (1 + \alpha)/2$

Suppose

$$\alpha_n = an^{-\alpha} \left(1 + \frac{c_1}{n^{1-\beta}} + \frac{c_2}{n^{\beta-\alpha}} + \frac{c_3}{n} + O(1/n^{2-\beta}) \right)$$

and $1 - \beta_n = b(n+1)^{-\beta}$, with $a, b > 0, 0 < \alpha < \beta < (1 + \alpha)/2$ and $c_1, c_2, c_3 \in \mathbb{R}$. Set $\gamma = \beta - \alpha \in (0, 1/2)$. For $n \geq h$, we have

$$\begin{aligned} \ell(h+1, n) - \ell(h, n) &= \ln(a(h+1)^{-\alpha}) - \ln(ah^{-\alpha}) - \ln(1 - b(h+1)^{-\beta}) \\ &\quad + \ln\left(1 + c_1/(h+1)^{1-\beta} + c_2/(h+1)^\gamma + c_3/(h+1) + O(1/h^{2-\beta})\right) \\ &\quad - \ln\left(1 + c_1/h^{1-\beta} + c_2/h^\gamma + c_3/h + O(1/h^{2-\beta})\right). \end{aligned} \tag{C.2}$$

Now, we aim at obtaining a series expansion with a reminder term of the type $o(1/h^\beta)$. Since $\beta < 1$, the first three terms of the right-hand side of the above equation give

$$\begin{aligned} &\ln(a(h+1)^{-\alpha}) - \ln(ah^{-\alpha}) - \ln(1 - b(h+1)^{-\beta}) \\ &= -\alpha \ln\left(1 + \frac{1}{h}\right) - \ln\left(1 - \frac{b}{(h+1)^\beta}\right) = \frac{b}{(h+1)^\beta} + o(h^{-\beta}). \end{aligned}$$

We deal now with the last two terms of (C.2). We recall that

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \dots + (-1)^{j-1} \frac{x^j}{j} + o(x^j),$$

and therefore, since $2 - \beta = 1 + 1 - \beta > 1 > \beta$ and $j(1 - \beta) > \beta$ and $j\gamma = j(\beta - \alpha) > \beta$ for j large enough, there are only a finite number J_0 of terms with an order $\tau_j \leq \beta$. In other words, we can write

$$\begin{aligned} &\ln\left(1 + c_1/(h+1)^{1-\beta} + c_2/(h+1)^\gamma + c_3/(h+1) + O(1/n^{2-\beta})\right) \\ &\quad - \ln\left(1 + c_1/h^{1-\beta} + c_2/h^\gamma + c_3/h + O(1/n^{2-\beta})\right) \\ &= \sum_{j=1}^{J_0} C_j (h+1)^{-\tau_j} - \sum_{j=1}^{J_0} C_j h^{-\tau_j} + o(1/h^\beta) \\ &= \sum_{j=1}^{J_0} C_j h^{-\tau_j} [(1+h^{-1})^{-\tau_j} - 1] + o(1/h^\beta) \\ &= \sum_{j=1}^{J_0} C_j h^{-\tau_j} (\tau_j h^{-1} + o(1/h)) + o(1/h^\beta) = o(1/h^\beta). \end{aligned}$$

Summing up, we have

$$\ell(h+1, n) - \ell(h, n) = \frac{b}{(h+1)^\beta} + o(h^{-\beta}).$$

Then there exists h_0 such that the function $h \mapsto \ell(h, n)$ is monotonically increasing for $h \geq h_0$. Now, fix $\eta > 0$ and let j_0 such that $j \geq j_0$ implies $\ln(\beta_j) \leq -\frac{bj^{-\beta}}{1+\eta}$. Then take $h_* \geq \max(h_0, j_0) + 1$ and $h \leq h_0 - 1$. Since $\beta < (1 + \alpha)/2$, we have $\alpha_n = an^{-\alpha}(1 + O(1/n^\gamma))$ and so, for h_* large enough, we get by (C.1)

$$\begin{aligned} \ell(h_*, n) - \ell(h, n) &= \ln(ah_*^{-\alpha}) - \ln(ah^{-\alpha}) + \ln(1 + O(h_*^{-\gamma})) + C_1 - \sum_{j=h}^{h_*-1} \ln(\beta_j) \\ &\geq \ln(h_*^{-\alpha}) + \ln(1 + O(h_*^{-\gamma})) + C_1 + \sum_{j=\max(h_0, j_0)}^{h_*-1} \frac{bj^{-\beta}}{1+\eta} \\ &\geq -\alpha \ln(h_*) + O(h_*^{-\gamma}) + C_2 + \frac{b}{1+\eta} \int_{\max(h_0, j_0)}^{h_*-1} x^{-\beta} dx \\ &= -\alpha \ln(h_*) + O(h_*^{-\gamma}) + C_2 \\ &\quad + \frac{b}{(1+\eta)(1-\beta)} [(h_* - 1)^{1-\beta} - \max(h_0, j_0)^{1-\beta}] \\ &= C_3 + O(h_*^{-\gamma}) - \alpha \ln(h_*) + \frac{b}{(1+\eta)(1-\beta)} (h_* - 1)^{1-\beta} \geq 0. \end{aligned}$$

Therefore, taking h^* large enough, we have $\max_{h \leq h^*} \ell(h, n) = \max_{h \leq h_0-1} \ell(h, n) \vee \max_{h_0 \leq h \leq h^*} \ell(h, n) \leq \ell(h^*, n)$.

Appendix D: Technical results

We recall the generalized Kronecker lemma [3, Corollary A.1]:

Lemma D.1 (Generalized Kronecker Lemma). *Let $\{v_{N,n} : 1 \leq n \leq N\}$ and $(z_n)_n$ be respectively a triangular array and a sequence of complex numbers such that $v_{N,n} \neq 0$ and*

$$\lim_N v_{N,n} = 0, \quad \lim_n v_{n,n} \text{ exists finite}, \quad \sum_{n=1}^N |v_{N,n} - v_{N,n-1}| = O(1)$$

and $\sum_n z_n$ is convergent. Then $\lim_N \sum_{n=1}^N v_{N,n} z_n = 0$.

The above corollary is useful to get the following result for complex random variables, which slightly extends the version provided in [3, Lemma A.2]:

Lemma D.2. *Let $\mathcal{H} = (\mathcal{H}_n)_n$ be a filtration and $(Y_n)_n$ a \mathcal{H} -adapted sequence of complex random variables. Moreover, let $(c_n)_n$ be a sequence of strictly positive real numbers such that $\sum_n E[|Y_n|^2] / c_n^2 < +\infty$ and let $\{v_{N,n}, 1 \leq n \leq N\}$ be a triangular array of complex numbers such that $v_{N,n} \neq 0$ and*

$$\lim_N v_{N,n} = 0, \quad \lim_n v_{n,n} \text{ exists finite}, \quad \sum_{n=1}^N |v_{N,n} - v_{N,n-1}| = O(1).$$

Suppose that

$$\sum_{n=1}^N v_{N,n} \frac{E[Y_n | \mathcal{H}_{n-1}]}{c_n} \xrightarrow{P} V, \tag{D.1}$$

where V is a suitable random variable. Then $\sum_{n=1}^N v_{N,n} Y_n / c_n \xrightarrow{P} V$.

If the convergence in (D.1) is almost sure, then also the convergence of $\sum_{n=1}^N v_{N,n} Y_n / c_n$ toward V is almost sure.

Proof. Consider the martingale $(M_n)_n$ defined by

$$M_n = \sum_{j=1}^n \frac{Y_j - E[Y_j | \mathcal{H}_{j-1}]}{c_j}.$$

It is bounded in L^2 since $\sum_n \frac{E[|Y_n|^2]}{c_n^2} < +\infty$ by assumption and so it is almost surely convergent, that means

$$\sum_n \frac{Y_n(\omega) - E[Y_n | \mathcal{H}_{n-1}](\omega)}{c_n} < +\infty$$

for $\omega \in B$ with $P(B) = 1$. Therefore, fixing $\omega \in B$ and setting $z_n = \frac{Y_n(\omega) - E[Y_n | \mathcal{H}_{n-1}](\omega)}{c_n}$, by Lemma D.1, we get

$$\lim_N \sum_{n=1}^N v_{N,n} \frac{Y_n(\omega) - E[Y_n | \mathcal{H}_{n-1}](\omega)}{c_n} = 0,$$

that is

$$\sum_{n=1}^N v_{N,n} \frac{Y_n - E[Y_n | \mathcal{H}_{n-1}]}{c_n} \xrightarrow{a.s.} 0.$$

In order to conclude, it is enough to observe that

$$\sum_{n=1}^N v_{N,n} \frac{Y_n}{c_n} = \sum_{n=1}^N v_{N,n} \frac{Y_n - E[Y_n | \mathcal{H}_{n-1}]}{c_n} + \sum_{n=1}^N v_{N,n} \frac{E[Y_n | \mathcal{H}_{n-1}]}{c_n}$$

and use assumption (D.1). □

Remark D.1. If we have $\sum_{n=1}^N \frac{|v_{N,n}|}{c_n} = O(1)$, $\lim_N \sum_{n=1}^N \frac{v_{N,n}}{c_n} = \lambda \in \mathbb{C}$ and $E[Y_n | \mathcal{H}_{n-1}] \xrightarrow{a.s.} Y$, then (D.1) is satisfied with almost sure convergence and $V = \lambda Y$. Indeed, if we denote by A an event such that $P(A) = 1$ and $\lim_n E[Y_n | \mathcal{H}_{n-1}](\omega) = Y(\omega)$ for each $\omega \in A$, then we can fix $\omega \in A$, set $w_n = E[Y_n | \mathcal{H}_{n-1}](\omega)$ and $w = Y(\omega)$, and apply the generalized Toeplitz lemma [3, Lemma A.1] (with $z_{N,n} = v_{N,n}/(c_n \lambda)$ and $s = 1$ when $\lambda \neq 0$ and with $z_{N,n} = v_{N,n}/c_n$ and $s = 0$ when $\lambda = 0$) in order to get $\sum_{n=1}^N v_{N,n} \frac{w_n}{c_n} \rightarrow \lambda Y$ almost surely.

The proof of the following lemma can be found in [27]. We here rewrite the proof only for the reader's convenience.

Lemma D.3 ([27, Lemma 18]). *Let x_n, ζ_n, γ_n be non-negative sequences such that $\gamma_n \rightarrow 0$, $\sum_n \gamma_n = +\infty$ and*

$$x_n \leq (1 - \gamma_n)x_{n-1} + \gamma_n \zeta_n.$$

Then $\limsup_n x_n \leq \limsup_n \zeta_n$.

Proof. Take $L > \limsup_n \zeta_n$ and n^* large enough so that $\zeta_n < L$ and $\gamma_n \leq 1$ when $n \geq n^*$. Then, using that $(x + y)^+ \leq x^+ + y^+$, we have for $n \geq n^*$

$$\begin{aligned} (x_n - L)^+ &\leq ((1 - \gamma_n)(x_{n-1} - L) + \gamma_n(\zeta_n - L))^+ \\ &\leq (1 - \gamma_n)(x_{n-1} - L)^+ + \gamma_n(\zeta_n - L)^+ \\ &\leq (1 - \gamma_n)(x_{n-1} - L)^+. \end{aligned}$$

Since $\sum_n \gamma_n = +\infty$, the above inequality implies that $\lim_n (x_n - L)^+ = 0$. This is enough to conclude, because we can choose L arbitrarily close to $\limsup_n \zeta_n$. \square

Appendix E: Some stochastic approximation results

Consider a stochastic process (θ_n) taking values in $\Theta = [-1, 1]^k$, adapted to a filtration $\mathcal{F} = (\mathcal{F}_n)_n$ and following the dynamics

$$\theta_{n+1} = (1 - \epsilon_n)\theta_n + c\epsilon_n \Delta M_{n+1}, \quad (\text{E.1})$$

where $c > 0$, $(\Delta M_{n+1})_n$ is a uniformly bounded martingale difference sequence with respect to \mathcal{F} and $\epsilon_n = (n + 1)^{-\epsilon}$ with $\epsilon \in (0, 1]$ so that $\epsilon_n \rightarrow 0$ and $\sum_n \epsilon_n = +\infty$. Setting $\widetilde{\Delta M}_{n+1} = c\Delta M_{n+1}$, equation (E.1) becomes

$$\theta_{n+1} = (1 - \epsilon_n)\theta_n + \epsilon_n \widetilde{\Delta M}_{n+1}.$$

Then:

Theorem E.1. *In the above setting, we have $\theta_N \xrightarrow{a.s.} \mathbf{0}$.*

Proof. We have the following two cases:

- $\epsilon \in (1/2, 1]$ so that $\sum_n \epsilon_n^2 < +\infty$ or
- $\epsilon \in (0, 1/2]$ so that $\sum_n \epsilon_n^2 = +\infty$.

For the first case, we refer to [37, Cap. 5, Th. 2.1]. For the second case, we refer to [37, Cap. 5, Th. 3.1]). In this case, since (θ_n) and $(\widetilde{\Delta M}_n)$ are uniformly bounded, the key assumption to be verified in order to apply [37, Cap. 5, Th. 3.1] is the “rate of change” condition (see [37, p. 137]), that is

$$\limsup_N \sup_{t \in [0, 1]} |M^0(N + t) - M^0(N)| = 0, \quad a.s.$$

where $M^0(t) = \sum_{j=0}^{m(t)-1} \epsilon_j \widetilde{\Delta M}_{j+1}$ and $m(t) = \inf\{n: t < t_{n+1} = \sum_{j=0}^n \epsilon_j\}$ (see [37, p. 122]). Since $(\widetilde{\Delta M}_n)$ is uniformly bounded, the above condition is satisfied when the following simpler conditions are satisfied (see [37, pp. 139-141]):

- (i) For each $u > 0$ $\sum_n e^{-u/\epsilon_n} < +\infty$;
- (ii) For some $T < +\infty$, there exists a constant $c(T) < +\infty$ such that $\sup_{n \leq j \leq m(t_n+T)} \frac{\epsilon_j}{\epsilon_n} \leq c(T)$.

When $\epsilon_n = (1 + n)^{-\epsilon}$, condition (i) is obviously verified, because we have $\lim_n n^2/e^{u(1+n)^{-\epsilon}} = 0$. Finally, condition (ii) is always satisfied when ϵ_n is decreasing, as it is in the case $\epsilon_n = (1 + n)^{-\epsilon}$. Indeed, we simply have

$$\sup_{n \leq j \leq m(t_n+T)} \epsilon_j/\epsilon_n = \epsilon_n/\epsilon_n = 1. \quad \square$$

Theorem E.2. *In the above setting, if we have $E[\Delta M_{n+1} \Delta M_{n+1}^\top | \mathcal{F}_n] \xrightarrow{a.s.} \Gamma$ with Γ a symmetric positive definite matrix, then we have*

$$\frac{1}{\sqrt{\epsilon_N}} \boldsymbol{\theta}_N \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma),$$

where $\Sigma = c^2\Gamma/2$ when $\epsilon \in (0, 1)$ and $\Sigma = c^2\Gamma$ when $\epsilon = 1$.

Proof. We have $\boldsymbol{\theta}_N \xrightarrow{a.s.} \mathbf{0}$ and $\mathbf{0}$ belongs to the interior part of Θ . Moreover, we have

$$E[\widetilde{\Delta M}_{n+1} \widetilde{\Delta M}_{n+1}^\top | \mathcal{F}_n] \xrightarrow{a.s.} c^2\Gamma.$$

For the case $\epsilon \in (1/2, 1]$, we refer to [29, Th. 2.1] (with $h = Id$, $U_* = c^2\Gamma$ and $\gamma_* = 1$) and [45, Th. 1] (with $H = -Id$, $\gamma_n = \sigma_n = \epsilon_n$ and so $\gamma_0 = 1$ and $\beta = \epsilon$). For the case $\epsilon \in (0, 1/2]$, we refer to [37, cap.10, Th. 2.1] (with $A = -Id$). The key assumption for applying this theorem is $\boldsymbol{\theta}_n/\sqrt{\epsilon_n}$ tight. On the other hand, in the considered setting, this last condition is satisfied because of [37, Th. 4.1]. Note that the limit distribution corresponds to the stationary distribution of the diffusion

$$dU_t = (-Id + c(\epsilon)) U_t dt + c\Gamma^{1/2} dW_t,$$

where $W = (W_t)_t$ is a standard Wiener process and

$$c(\epsilon) = \begin{cases} 0 & \text{for } \epsilon < 1 \\ 1/2 & \text{for } \epsilon = 1. \end{cases}$$

Therefore the limit covariance matrix is determined by solving the associated Lyapunov's equation [45], that, in the considered case, simply is

$$2(-Id + c(\epsilon)Id) \Sigma = -c^2\Gamma. \quad \square$$

Theorem E.3. *In the above setting, let $(\boldsymbol{\mu}_n)$ be another stochastic process taking values in $\Theta = [-1, 1]^k$, adapted to a filtration \mathcal{F} and following the dynamics*

$$\boldsymbol{\mu}_{n+1} - \boldsymbol{\mu}_n = -\frac{1}{n}(\boldsymbol{\mu}_n - \boldsymbol{\theta}_n) + \frac{1}{n} \Delta M_{n+1}.$$

Suppose that $E[\Delta\mathbf{M}_{n+1}\Delta\mathbf{M}_{n+1}^\top | \mathcal{F}_n] \xrightarrow{a.s.} \Gamma$. If $\epsilon \in (1/2, 1)$, then we have

$$\begin{pmatrix} \sqrt{N}\boldsymbol{\mu}_N \\ \epsilon_N^{-1/2}\boldsymbol{\theta}_N \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} (c+1)^2\Gamma & \mathbf{0} \\ \mathbf{0} & \frac{c^2}{2}\Gamma \end{pmatrix}\right).$$

If $\epsilon = 1$, then we have

$$\begin{pmatrix} \sqrt{N}\boldsymbol{\mu}_N \\ \epsilon_N^{-1/2}\boldsymbol{\theta}_N \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} [(c+1)^2 + c^2]\Gamma & c(c+1)\Gamma \\ c(c+1)\Gamma & c^2\Gamma \end{pmatrix}\right).$$

Proof. The dynamics for the pair $(\boldsymbol{\mu}_n, \boldsymbol{\theta}_n)_n$ is

$$\begin{cases} \boldsymbol{\mu}_{n+1} - \boldsymbol{\mu}_n &= -\frac{1}{n}(\boldsymbol{\mu}_n - \boldsymbol{\theta}_n) + \frac{1}{n}\Delta\mathbf{M}_{n+1} \\ \boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}_n &= -\epsilon_n\boldsymbol{\theta}_n + c\epsilon_n\Delta\mathbf{M}_{n+1} = -\epsilon_n\boldsymbol{\theta}_n + \epsilon_n\Delta\widetilde{\mathbf{M}}_{n+1}. \end{cases}$$

with $E[\Delta\mathbf{M}_{n+1}\Delta\mathbf{M}_{n+1}^\top | \mathcal{F}_n] \xrightarrow{a.s.} \Gamma$. Therefore, when $1/2 < \epsilon < 1$, the statement follows from [43] (with $Q_{11} = Q_{22} = -Id$, $Q_{12} = Id$, $Q_{21} = \mathbf{0}$, $b = \beta_0 = 1$, $a = \epsilon$, $\Gamma_{11} = \Gamma$, $\Gamma_{22} = c^2\Gamma$ and $\Gamma_{12} = \Gamma_{21} = c\Gamma$). In particular, the two blocks of the limit covariance matrix, say Σ_μ and Σ_θ , are determined solving the equations

$$(H + \frac{1}{2}Id)\Sigma_\mu + \Sigma_\mu(H^\top + \frac{1}{2}Id) = -\Gamma_\mu,$$

where $H = Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} = -Id + \mathbf{0}$ and $\Gamma_\mu = \Gamma_{11} + Q_{12}Q_{22}^{-1}\Gamma_{22}(Q_{22}^{-1})^\top Q_{12}^\top - \Gamma_{12}(Q_{22}^{-1})^\top Q_{12}^\top - Q_{12}Q_{22}^{-1}\Gamma_{21} = \Gamma + c^2\Gamma + c\Gamma + c\Gamma = (c+1)^2\Gamma$, and

$$Q_{22}\Sigma_\theta + \Sigma_\theta Q_{22}^\top = -\Gamma_{22}.$$

When $\epsilon = 1$, we can conclude by [45] or [56] taking $\mathbf{X}_n = (\boldsymbol{\mu}_n, \boldsymbol{\theta}_n)^\top$. Indeed, in this case the covariance matrix is given by

$$(H + \frac{1}{2}Id)\Sigma + \Sigma(H^\top + \frac{1}{2}Id) = -\widetilde{\Gamma},$$

where

$$H = \begin{pmatrix} -Id & Id \\ \mathbf{0} & -Id \end{pmatrix} \quad \text{and} \quad \widetilde{\Gamma} = \begin{pmatrix} \Gamma & c\Gamma \\ c\Gamma & c^2\Gamma \end{pmatrix}.$$

Therefore, if we split Σ in blocks, say Σ_μ , Σ_θ and $\Sigma_{\mu\theta}$, we find the system

$$\begin{aligned} -\Sigma_\mu + 2\Sigma_{\mu\theta} &= -\Gamma \\ -\Sigma_{\mu\theta} + \Sigma_\theta &= -c\Gamma \\ -\Sigma_\theta &= -c^2\Gamma \end{aligned}$$

and so the proof is concluded by solving this system. □

Appendix F: Stable convergence

This brief section contains some basic definitions and results concerning stable convergence. For more details, we refer the reader to [22, 25, 33] and the references therein.

Let (Ω, \mathcal{A}, P) be a probability space, and let S be a Polish space, endowed with its Borel σ -field. A *kernel* on S , or a random probability measure on S , is a collection $K = \{K(\omega) : \omega \in \Omega\}$ of probability measures on the Borel σ -field of S such that, for each bounded Borel real function f on S , the map

$$\omega \mapsto Kf(\omega) = \int f(x) K(\omega)(dx)$$

is \mathcal{A} -measurable. Given a sub- σ -field \mathcal{H} of \mathcal{A} , a kernel K is said \mathcal{H} -measurable if all the above random variables Kf are \mathcal{H} -measurable. A probability measure ν can be identified with a constant kernel $K(\omega) = \nu$ for each ω .

On (Ω, \mathcal{A}, P) , let $(Y_n)_n$ be a sequence of S -valued random variables, let \mathcal{H} be a sub- σ -field of \mathcal{A} , and let K be a \mathcal{H} -measurable kernel on S . Then, we say that Y_n converges \mathcal{H} -stably to K , and we write $Y_n \rightarrow K$ \mathcal{H} -stably, if

$$P(Y_n \in \cdot | H) \xrightarrow{weakly} E[K(\cdot) | H] \quad \text{for all } H \in \mathcal{H} \text{ with } P(H) > 0,$$

where $K(\cdot)$ denotes the random variable defined, for each Borel set B of S , as $\omega \mapsto KI_B(\omega) = K(\omega)(B)$. In the case when $\mathcal{H} = \mathcal{A}$, we simply say that Y_n converges *stably* to K and we write $Y_n \rightarrow K$ stably. Clearly, if $Y_n \rightarrow K$ \mathcal{H} -stably, then Y_n converges in distribution to the probability distribution $E[K(\cdot)]$. The \mathcal{H} -stable convergence of Y_n to K can be stated in terms of the following convergence of conditional expectations:

$$E[f(Y_n) | \mathcal{H}] \xrightarrow{\sigma(L^1, L^\infty)} Kf \tag{F.1}$$

for each bounded continuous real function f on S . In [25] the notion of \mathcal{H} -stable convergence is firstly generalized in a natural way replacing in (F.1) the single sub- σ -field \mathcal{H} by a collection $\mathcal{G} = (\mathcal{G}_n)$ (called conditioning system) of sub- σ -fields of \mathcal{A} and then it is strengthened by substituting the convergence in $\sigma(L^1, L^\infty)$ by the one in probability (i.e. in L^1 , since f is bounded). Hence, according to [25], we say that Y_n converges to K *stably in the strong sense*, with respect to $\mathcal{G} = (\mathcal{G}_n)$, if

$$E[f(Y_n) | \mathcal{G}_n] \xrightarrow{P} Kf$$

for each bounded continuous real function f on S .

We now conclude this section recalling some convergence results that we apply in our proofs.

From [33, Th. 3.2] (see also [25, Th. 5 and Cor. 7] or [22, Th. 5.5.1 and Cor. 5.5.2]), we get:

Theorem F.1. *Given a filtration $\mathcal{F} = (\mathcal{F}_n)_n$, let $(\mathbf{Y}_{N,n})_{N,n}$ be a triangular array of random variables with values in \mathbb{R}^k such that $Y_{N,n}$ is \mathcal{F}_n -measurable and $E[\mathbf{Y}_{N,n} | \mathcal{F}_{n-1}] = \mathbf{0}$. Suppose that the following two conditions are satisfied:*

- (c1) $E[\max_{1 \leq n \leq N} |\mathbf{Y}_{N,n}|] \rightarrow 0$ and
 (c2) $\sum_{n=1}^N \mathbf{Y}_{N,n} \mathbf{Y}_{N,n}^\top \xrightarrow{P} V$, where V is a random variable with values in the space of positive semidefinite $k \times k$ -matrices.

Then $\sum_{n=1}^N \mathbf{Y}_{N,n}$ converges stably to the Gaussian kernel $\mathcal{N}(\mathbf{0}, V)$.

From [25, Th. 5, Cor. 7, Rem. 4] or [22, Th. 5.5.1, Cor. 5.5.2, Rem. 5.5.2]), we obtain:

Theorem F.2. Let (\mathbf{L}_n) be a \mathbb{R}^k -valued martingale with respect to the filtration $\mathcal{F} = (\mathcal{F}_n)$. Suppose that $\mathbf{L}_n \xrightarrow{a.s., L^1} \mathbf{L}$ for some \mathbb{R}^k -valued random variable \mathbf{L} and

- (c1) $n^v E[\sup_{j \geq n} |\mathbf{L}_{j-1} - \mathbf{L}_j|] \rightarrow 0$ and
 (c2) $n^{2v} \sum_{j \geq n} (\mathbf{L}_{j-1} - \mathbf{L}_j)(\mathbf{L}_{j-1} - \mathbf{L}_j)^\top \xrightarrow{P} V$, where V is a random variable with values in the space of positive semidefinite $k \times k$ -matrices.

Then

$$n^v (\mathbf{L}_n - \mathbf{L}) \rightarrow \mathcal{N}(0, V) \quad \text{stably in strong sense w.r.t. } \mathcal{F}.$$

Indeed, following [25, Example 6], it is enough to observe that $\mathbf{L}_n - \mathbf{L}$ can be written as $\mathbf{L}_n - \mathbf{L} = \sum_{j \geq n} (\mathbf{L}_j - \mathbf{L}_{j+1})$.

Finally, the following result combines together a stable convergence and a stable convergence in the strong sense [13, Lemma 1].

Theorem F.3. Suppose that C_n and D_n are S -valued random variables, that M and N are kernels on S , and that $\mathcal{G} = (\mathcal{G}_n)_n$ is an (increasing) filtration satisfying for all n

$$\sigma(C_n) \subseteq \mathcal{G}_n \quad \text{and} \quad \sigma(D_n) \subseteq \sigma(\bigcup_n \mathcal{G}_n).$$

If C_n stably converges to M and D_n converges to N stably in the strong sense, with respect to \mathcal{G} , then

$$[C_n, D_n] \rightarrow M \otimes N \quad \text{stably.}$$

(Here, $M \otimes N$ is the kernel on $S \times S$ such that $(M \otimes N)(\omega) = M(\omega) \otimes N(\omega)$ for all ω .)

This last result contains as a special case the fact that stable convergence and convergence in probability combine well: that is, if C_n stably converges to M and D_n converges in probability to a random variable D , then (C_n, D_n) stably converges to $M \otimes \delta_D$, where δ_D denotes the Dirac kernel concentrated in D .

Acknowledgments

We are grateful to the Referees for their useful comments.

References

- [1] ALETTI, G. and CRIMALDI, I. (2022). The Rescaled Pólya Urn: local reinforcement and chi-squared goodness of fit test. *Advances in Applied Probability* **54**.
- [2] ALETTI, G., CRIMALDI, I. and GHIGLIETTI, A. (2017). Synchronization of reinforced stochastic processes with a network-based interaction. *Ann. Appl. Probab.* **27** 3787–3844. [MR3737938](#)
- [3] ALETTI, G., CRIMALDI, I. and GHIGLIETTI, A. (2019). Networks of reinforced stochastic processes: asymptotics for the empirical means. *Bernoulli* **25** 3339–3378. [MR4010957](#)
- [4] ALETTI, G., CRIMALDI, I. and GHIGLIETTI, A. (2020). Interacting Reinforced Stochastic Processes: Statistical Inference based on the Weighted Empirical Means. *Bernoulli* **26** 1098–1138. [MR4058362](#)
- [5] ALETTI, G., CRIMALDI, I. and SARACCO, F. (2021). A model for the Twitter sentiment curve. *PLOS ONE* **16** 1–28.
- [6] ALETTI, G., GHIGLIETTI, A. and ROSENBERGER, W. F. (2018). Nonparametric covariate-adjusted response-adaptive design based on a functional urn model. *Ann. Statist.* **46** 3838–3866. [MR3852670](#)
- [7] ALETTI, G., GHIGLIETTI, A. and VIDYASHANKAR, A. N. (2018). Dynamics of an adaptive randomly reinforced urn. *Bernoulli* **24** 2204–2255. [MR3757528](#)
- [8] ALETTI, G., NALDI, G. and TOSCANI, G. (2007). First-order continuous models of opinion formation. *SIAM J. Appl. Math.* **67** 837–853. [MR2300313](#)
- [9] BENAÏM, M. (1999). *Dynamics of stochastic approximation algorithms. In Séminaire de Probabilités XXXIII, Lecture Notes in Mathematics* **1709**. Springer, Berlin, Heidelberg. [MR1767993](#)
- [10] BENFENATI, A. and COSCIA, V. and (2013). Nonlinear microscale interactions in the kinetic theory of active particles. *Appl. Math. Lett.* **26** 979–983. [MR3078979](#)
- [11] BENFENATI, A. and COSCIA, V. and (2014). Modeling opinion formation in the kinetic theory of active particles I: spontaneous trend. *Ann. Univ. Ferrara Sez. VII Sci. Mat.* **60** 35–53. [MR3208786](#)
- [12] BERGH, D. (2015). Sample Size and Chi-Squared Test of Fit— A Comparison Between a Random Sample Approach and a Chi-Square Value Adjustment Method Using Swedish Adolescent Data. In *Pacific Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceedings* (Q. ZHANG and H. YANG, eds.) 197–211. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [13] BERTI, P., CRIMALDI, I., PRATELLI, L. and RIGO, P. (2011). A central limit theorem and its applications to multicolor randomly reinforced urns. *J. Appl. Probab.* **48** 527–546. [MR2840314](#)
- [14] BERTI, P., CRIMALDI, I., PRATELLI, L. and RIGO, P. (2016). Asymptotics for randomly reinforced urns with random barriers. *J. Appl. Probab.* **53** 1206–1220. [MR3581252](#)

- [15] BERTONI, D., ALETTI, G., FERRANDI, G., MICHELETTI, A., CAVICCHIOLI, D. and PRETOLANI, R. (2018). Farmland Use Transitions After the CAP Greening: a Preliminary Analysis Using Markov Chains Approach. *Land Use Policy* **79** 789–800.
- [16] CALDARELLI, G., DE NICOLA, R., PETROCCHI, M., PRATELLI, M. and SARACCO, F. (2020). Analysis of online misinformation during the peak of the COVID-19 pandemics in Italy. *arXiv: 2010.01913*.
- [17] CHANDA, K. C. (1999). Chi-squared tests of goodness-of-fit for dependent observations. In *Asymptotics, Non-Parametrics and Time Series, Statist. Textbooks Monogr.*, **158** 743–756. Dekker. [MR1725275](#)
- [18] CHEN, M.-R. and KUBA, M. (2013). On generalized Pólya urn models. *J. Appl. Probab.* **50** 1169–1186. [MR3161380](#)
- [19] CHEN, Y. and SKIENA, S. (2014). Building sentiment lexicons for all major languages. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)* 383–389.
- [20] CHESSA, A., CRIMALDI, I., RICCABONI, M. and TRAPIN, L. (2014). Cluster Analysis of Weighted Bipartite Networks: A New Copula-Based Approach. *PLOS ONE* **9** 1–12.
- [21] COLLEVECCHIO, A., COTAR, C. and LICALZI, M. (2013). On a preferential attachment and generalized Pólya’s urn model. *Ann. Appl. Probab.* **23** 1219–1253. [MR3076683](#)
- [22] CRIMALDI, I. (2016). *Introduzione alla nozione di convergenza stabile e sue varianti (Introduction to the notion of stable convergence and its variants)* **57**. Unione Matematica Italiana, Monograf s.r.l., Bologna, Italy. Book written in Italian.
- [23] CRIMALDI, I., DAI PRA, P., LOUIS, P.-Y. and MINELLI, I. G. (2019). Synchronization and functional central limit theorems for interacting reinforced random walks. *Stochastic Processes and their Applications* **129** 70–101. [MR3906991](#)
- [24] CRIMALDI, I., DAI PRA, P. and MINELLI, I. G. (2016). Fluctuation theorems for synchronization of interacting Pólya’s urns. *Stochastic Process. Appl.* **126** 930–947. [MR3452818](#)
- [25] CRIMALDI, I., LETTA, G. and PRATELLI, L. (2007). *A Strong Form of Stable Convergence* **1899** 203–225. Springer. [MR2409006](#)
- [26] DAI PRA, P., LOUIS, P.-Y. and MINELLI, I. G. (2014). Synchronization via interacting reinforcement. *J. Appl. Probab.* **51** 556–568. [MR3217785](#)
- [27] DELYON, B. (2000). Stochastic approximation with decreasing gain: Convergence and asymptotic theory. *Technical report*.
- [28] EGGENBERGER, F. and PÓLYA, G. (1923). Über die Statistik verketteter Vorgänge. *ZAMM – Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* **3** 279–289.
- [29] FORT, G. (2015). Central limit theorems for stochastic approximation with controlled Markov chain dynamics. *ESAIM: PS* **19** 60–80. [MR3374869](#)

- [30] GANGOPADHYAY, U. and MAULIK, K. (2019). Stochastic approximation with random step sizes and urn models with random replacement matrices having finite mean. *The Annals of Applied Probability* **29** 2033–2066. [MR3984252](#)
- [31] GASSER, T. (1975). Goodness-of-fit tests for correlated data. *Biometrika* **62** 563–570. [MR0403125](#)
- [32] GLESER, L. J. and MOORE, D. S. (1983). The Effect of Dependence on Chi-Squared and Empiric Distribution Tests of Fit. *The Annals of Statistics* **11** 1100–1108. [MR0720256](#)
- [33] HALL, P. and HEYDE, C. C. (1980). *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics. [MR624435](#) (83a:60001)
- [34] HOLMES, M. and SAKAI, A. (2007). Senile reinforced random walks. *Stochastic Processes and their Applications* **117** 1519–1539. [MR2353038](#)
- [35] IEVA, F., PAGANONI, A. M., PIGOLI, D. and VITELLI, V. (2013). Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **62** 401–418. [MR3060623](#)
- [36] KNOKE, D., BOHRNSTEDT, G. W. and POTTER MEE, A. (2002). *Statistics for Social Data Analysis*. F.E.Peacock Publishers.
- [37] KUSHNER, H. J. and YIN, G. G. (2003). *Stochastic approximation and recursive algorithms and applications*, second ed. *Applications of Mathematics (New York)* **35**. Springer-Verlag, New York Stochastic Modelling and Applied Probability. [MR1993642](#)
- [38] LARUELLE, S. and PAGÉS, G. (2013). Randomized urn models revisited using stochastic approximation. *Ann. Appl. Proba.* **23** 1409–1436. [MR3098437](#)
- [39] LASMAR, N., MAILLER, C. and SELMI, O. (2018). Multiple drawing multi-colour urns by stochastic approximation. *J. Appl. Probab.* **55** 254–281. [MR3780393](#)
- [40] MAHMOUD, H. M. (2009). *Pólya urn models. Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. [MR2435823](#)
- [41] MÉTIVIER, M. (1982). *Semimartingales*. Walter de Gruyter and Co., Berlin. [MR0688144](#)
- [42] MICHELETTI, A., ALETTI, G., FERRANDI, G., BERTONI, D., CAVICCHIOLI, D. and PRETOLANI, R. (2020). A weighted χ^2 test to detect the presence of a major change point in non-stationary Markov chains. *Stat. Methods Appl.* **29** 899–912. [MR4174691](#)
- [43] MOKKADEM, A. and PELLETIER, M. (2006). Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *Ann. Appl. Probab.* **16** 1671–1702. [MR2260078](#)
- [44] PAN, W. (2002). Goodness-of-fit tests for GEE with correlated binary data. *Scand. J. Statist.* **29** 101–110. [MR1894384](#)
- [45] PELLETIER, M. (1998). Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.* **8** 10–44. [MR1620405](#)

- [46] PEMANTLE, R. (1990). A time-dependent version of Pólya's urn. *J. Theor. Probab.* **3** 627–637. [MR1067672](#)
- [47] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probab. Surveys* **4** 1–79. [MR2282181](#)
- [48] RADLOW, R. and ALF JR., E. F. (1975). An Alternate Multinomial Assessment of the Accuracy of the χ^2 Test of Goodness of Fit. *Journal of the American Statistical Association* **70** 811–813.
- [49] RAO, J. N. K. and SCOTT, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Amer. Statist. Assoc.* **76** 221–230. [MR624328](#)
- [50] ROBBINS, H. and SIEGMUND, D. (1971). A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing Methods in Statistics* 233–257. Academic Press. [MR0343355](#)
- [51] SAHASRABUDHE, N. (2016). Synchronization and fluctuation theorems for interacting Friedman urns. *J. Appl. Probab.* **53** 1221–1239. [MR3581253](#)
- [52] SHERMAN, J. and MORRISON, W. J. (1950). Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix. *Ann. Math. Statist.* **21** 124–127. [MR0035118](#)
- [53] TANG, M.-L., PEI, Y.-B., WONG, W.-K. and LI, J.-L. (2012). Goodness-of-fit tests for correlated paired binary data. *Stat. Methods Med. Res.* **21** 331–345. [MR3190613](#)
- [54] THARWAT, A. (2018). Independent component analysis: An introduction. *Applied Computing and Informatics*.
- [55] XU, D. and TIAN, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science* **2** 165–193.
- [56] ZHANG, L.-X. (2016). Central limit theorems of a recursive stochastic algorithm with applications to adaptive designs. *Ann. Appl. Probab.* **26** 3630–3658. [MR3582813](#)