Approximating a Sequence of Observation by a Simple Process

Dinah Rosenberg*, Eilon Solan† and Nicolas Vieille‡§

March 10, 2002

Abstract

Given a sequence $(s_0, s_1, \ldots, s_N)$ of observations from a finite set S, we construct a process $(\mathbf{s}_n)$ that satisfies the following properties: (i) $(\mathbf{s}_n)$ is a piecewise Markov chain, (ii) the conditional distribution of $\mathbf{s}_n$ given $\mathbf{s}_0, \ldots, \mathbf{s}_{n-1}$ is close to the empirical transition given by the observed sequence, for most n, (iii) under $(\mathbf{s}_n)$, with high probability the empirical frequency of the realized sequence is close to the one given by the observed sequence. We generalize this result to the case that the conditional distribution of $\mathbf{s}_n$ given $\mathbf{s}_0, \ldots, \mathbf{s}_{n-1}$ is required to be in some polyhedron $V_{\mathbf{s}_{n-1}}$.

*Laboratoire d'Analyse Géométrie et Applications, Institut Galilée, Université Paris Nord, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France. e-mail: dinah@zeus.math.univ-paris13.fr

†MEDS Department, Kellogg School of Management, Northwestern University, *and* the School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. e-mail: eilons@post.tau.ac.il

‡Département Finance et Economie, HEC, 1, rue de la Libération, 78 351 Jouy-en-Josas, France. e-mail: vieille@hec.fr

§We thank Ehud Lehrer, Sean Meyn and Laurent Saloff-Coste for their suggestions and references. We acknowledge the financial support of the Arc-en-Ciel/Keshet program for 2001/2002. The research of the second author was supported by the Israel Science Foundation (grant No. 03620191).

1 Introduction

In the analysis of stochastic games with partial monitoring, a problem of a statistical nature arises, whose basic expression can be summarized as follows. One player, called the adversary, controls a S-valued process $(\mathbf{s}_n)$, where S is a finite set. In each of finitely many stages, he chooses the law $\mathbf{y}_n$ according to which the next state $\mathbf{s}_{n+1}$ is selected. A second player, called the statistician, suffers a loss $r(\mathbf{s}_n, \mathbf{y}_n)$ where r is concave in y. The statistician gets only to observe the realized sequence of states, and wishes to estimate *ex-post* his total loss $L := \sum_{n=0}^{N} r(\mathbf{s}_n, \mathbf{y}_n)$. On the basis of his information, the natural idea for the statistician is to compute, for each $s \in S$, the distribution $\widehat{y}(s) \in \Delta(S)$ that is closest to the empirical transitions out of state s, and to suggest the quantity $\widehat{L} = \sum_{s \in S} N_s r(s, \widehat{y}(s))$ as an estimate for the loss, where N_s is the number of visits to s.

For a given strategy τ of the adversary (i.e., a rule that dictates for every stage n, which $\mathbf{y}_n$ to choose on the basis of the available information), the expectation $\mathbf{E}_\tau\left[\widehat{L}\right]$ of this estimator[1] is typically higher than the expected loss $\mathbf{E}_\tau[L]$, due to the concavity of r. In other words, $\widehat{L}$ will fail to be, even approximately, an unbiased estimator of the loss. For the purpose of the game-theoretic application, it is enough to show that, with high τ-probability, $\widehat{L}$ is close to $\mathbf{E}_{\widetilde{\tau}}[L]$ for some "simple" strategy $\widetilde{\tau}$. This is the basic question we address here in Section 2.

In the actual game-theoretic motivating problem, the player's strategy choice is restricted: for each $n \in \mathbf{N}$, $\mathbf{y}_n$ has to belong to a given compact polyhedron $V(\mathbf{s}_n)$ of probability measures over S. This caveat makes the analysis in Section 3 of the corresponding problem substantially more difficult.

We now abstract from the game-theoretic framework, to introduce the relevant statistical question. Let $\sigma = (s_0, s_1, \ldots, s_N)$ be a sequence over a finite set S. For $s \in S$, the number of visits to s is defined to be $N_s^\sigma = |\{n < N, s_n = s\}|$ (not counting the last state in the sequence). The *empirical frequency of* s in σ is $\nu_s^\sigma := N_s^\sigma / N$, and the *empirical transition* (along σ) out of s is

$$q^\sigma(t \mid s) = \frac{|\{n < N, (s_n, s_{n+1}) = (s, t)\}|}{N_s^\sigma}, \text{ for } t \in S.$$

Does there exist a "simple" process $(\mathbf{s}_n)$ such that

(i) the conditional law of $\mathbf{s}_{n+1}$ given $(\mathbf{s}_0, ..., \mathbf{s}_n)$ is close to $q^\sigma(\cdot|\mathbf{s}_n)$ a.s. for most n's;

[1] The expectation is taken w.r.t. the law of the process induced by the strategy.

(ii) with high probability under $(\mathbf{s}_n)$, the empirical frequency of $s \in S$ in the first N stages is close to the frequency ν^σ_s of s along σ ?

The problem may thus be seen as that of approximating a given sequence by a simple S-valued stochastic process. The naive solution is to define $(\mathbf{s}_n)$ as a Markov chain with the empirical transitions q^σ as transition function. As the next example illustrates, this solution fails.

Example 1.1 *Let $S = \{a, b\}$, and consider the sequence $\sigma = (a, a, \ldots, a, b, b, \ldots, b, a)$ of N a's followed by N b's, and one a at the end. The empirical transition q^σ is given by*

$$q^\sigma(a \mid b) = q^\sigma(b \mid a) = 1 - q^\sigma(b \mid b) = 1 - q^\sigma(a \mid a) = 1/N.$$

Let $(\mathbf{s}_n)$ be the Markov chain with transition q^σ, starting from a. With a probability bounded away from zero, $\mathbf{s}_n = a$ for every $n \leq 2N + 1$. In particular, condition (ii) is not satisfied. More generally, one can prove that for this example no Markov chain satisfies both conditions (i) and (ii).

In the sequel we show that, provided N is sufficiently large, there exists a *piecewise Markov chain* on S with at most $|S|$ pieces that approximates σ in the sense of (i) and (ii).

A remark is in order. The statistician is here insisting on one-step transitions, by asking that the approximating process be a (piecewise) Markov chain on S. By this insistence, he potentially looses much information on the structure of the sequence. Indeed, the sequence 001100110011...will be approximated (for lack of a better term) by the Markov chain on $\{0, 1\}$ with transitions $(\frac{1}{2}, \frac{1}{2})$ in each state. Plainly, a Markov chain of order 2 would approximate perfectly the given sequence. More generally, letting the statistician choose the order of the Markov chain would allow a tradeoff between the order of the chain and the quality of the approximation. Our results indicate that piecewise Markov chains (of order 1) are sufficient to get a good appromixation, when only one-step transitions matter.

We next discuss the extension of the basic problem motivated in the introductory paragraph. For every state $s \in S$ we are given a non empty polyhedron V_s of probability distributions over S.[2] Denote $V = (V_s)_{s \in S}$. A process $(\mathbf{s}_n)$ is a V-process if for every n the conditional law of $\mathbf{s}_{n+1}$ given $(\mathbf{s}_0, ..., \mathbf{s}_n)$ is in $V_{\mathbf{s}_n}$. The problem is, given a sequence, to approximate it in the sense of (i) and (ii) by a V-process.

In general, such an approximation needs not exist. Indeed, as the following two examples show, if all V-processes are reducible, or if the sequence is not "typical", meaning that the empirical transitions are "far" from any

[2] Since S is finite, we identify a probability measure over S to a point in the unit simplex of $\mathbf{R}^S$.

V-process, such a construction is not possible. In the following two examples, V_s is a singleton for each $s \in S$, so there is a unique V-process, which is a Markov chain.

Example 1.2 (A reducible Markov chain) *Consider a problem with three states $\{a, b, c\}$. Assume V is such that for any V-process, states b and c are absorbing, whereas if the process is in state a, with equal probabilities it moves to states b and c. When the initial state is a, there are two possible sequences under the unique V-process, each is realized with probability $1/2$: $(a, b, b, b, \ldots, b)$ and $(a, c, c, c, \ldots, c)$. But if the given sequence is $(a, b, b, b, \ldots, b)$ there is no V-process that satisfies both (i) and (ii).*

Example 1.3 (A non-typical sequence) *Assume there are two states $\{a, b\}$, and $V_a = V_b = \{\frac{1}{2}a + \frac{1}{2}b\}$.*

Assume the given sequence is $(a, a, \cdots, a)$. There is no V-process that satisfies (i) and (ii), provided N is sufficiently large.

In section 3 we define the notion of typical sequences w.r.t. $V = (V_s)_{s \in S}$, prove that for every V-process, the probability that the realized sequence is typical is close to 1, and prove that for every typical sequence there is a hidden piecewise V-Markov chain with at most $|S|$ pieces that approximates the typical sequence in the sense of (i) and (ii) above.

Thus, our result states that almost any sequence that is generated by a V-process, as complex as it can be, can be approximated by a simple V-process.

2 The Basic Problem

For every finite set K, let $|K|$ be the number of elements in K, and let $\Delta(K)$ be the space of probability distributions over K. Throughout the paper we fix a finite set S.

2.1 Presentation

Let $N \in \mathbf{N}$, and let $\sigma = (s_0, s_1, \ldots, s_N)$ be a finite sequence in S of length $N+1$. For $s \in S$, let $N_s = |\{n < N | s_n = s\}|$ be the *number of visits* to s in σ (the last state of the sequence is not counted), and define the *empirical frequency* of s in σ as

$$\nu_s^\sigma = \frac{N_s}{N}.$$

The *(empirical) transitions* out of s along σ are defined by

$$q^\sigma(t \mid s) = \frac{|\{n < N, (s_n, s_{n+1}) = (s, t)\}|}{N_s}, t \in S. \tag{1}$$

$q^\sigma(t \mid s)$ is defined whenever the denominator in (1) does not vanish; that is, whenever the state s is visited by the sequence. If $N_s = 0$, we let $q^\sigma(\cdot|s)$ be arbitrary. Note that q^σ is a transition function over S.

A piecewise Markov chain is the concatenation of Markov chains. Formally,

Definition 2.1 *Let K be a positive integer. A process $(\mathbf{s}_n)_{n\leq N}$ is a* piecewise Markov chain with K pieces *if there exists a non-decreasing sequence $(n_k)_{0\leq k\leq K}$ of integers with $n_0 = 0$ and $n_K = N$, such that for each $k = 1,\ldots,K$, the process $(\mathbf{s}_n)_{n_{k-1}\leq n\leq n_k}$ is a Markov chain.*

Given a S-valued process $(\mathbf{s}_n)$, $s \in S$, and $m \in \mathbf{N}$, we denote $\overline{F}^s_m = \frac{1}{m}\left|\{0 \leq n \leq m-1 \mid \mathbf{s}_n = s\}\right|$ (resp. F^s_m) the empirical frequency of s from stage 0 up to stage $m-1$ inclusive (resp. from stage 1 up to stage m). We also denote by $\mathbf{P}$ the law of $(\mathbf{s}_n)$, and by $\mathbf{q}_n$ the conditional law of $\mathbf{s}_{n+1}$ given $(\mathbf{s}_1, ..., \mathbf{s}_n)$. Our basic theorem is the following.

Theorem 2.2 *For every $\varepsilon > 0$ $\rho \in (0, 1/2(4|S|+1)$ and $\zeta \in (0, 2\rho)$, there exists $N_0 \in \mathbf{N}$ such that the following holds. For every sequence σ of length $N \geq N_0$, there is a piecewise Markov chain $(\mathbf{s}_n)$ with $|S|$ pieces over S such that, for each $s \in S$,*

B1 *If $\nu^\sigma_s \geq \frac{1}{N^\rho}$, then $\mathbf{P}(|\overline{F}^s_N - \nu^\sigma_s| \geq \varepsilon\nu^\sigma_s) \leq \frac{1}{N^\zeta}$.*

B2 *$\|\mathbf{q}_n - q^\sigma(\cdot \mid \mathbf{s}_n)\| < \varepsilon$, a.s. for at least $N - |S|$ values of $n < N$.*

2.2 On Markov chains

In the present section we present some general results on Markov chains, that have their own interest. We first provide a result on the speed of convergence of an irreducible Markov chain to its invariant measure. Next, we collect a few observations on the expected exit time from domains of S. Let $q : S \to \Delta(S)$ be a transition rule over S. Given $s \in S$ we denote by $\mathbf{P}_{s,q}$ the law of the Markov chain (S, q) starting from s. We denote by $\mathbf{E}_{s,q}$ the corresponding expectation operator. When there is no risk of confusion, we may abbreviate $\mathbf{P}_{s,q}$ and $\mathbf{E}_{s,q}$ to $\mathbf{P}_s$ and $\mathbf{E}_s$ respectively. The hitting time of $C \subseteq S$ is denoted $T_C := \inf\{n \geq 0 : \mathbf{s}_n \in C\}$. For $t \in S$, we abbreviate $T_{\{t\}}$ to T_t and we denote by $T^+_t = \inf\{n \geq 1, \mathbf{s}_n = t\}$ the first return to t. By convention, the infimum over an empty set is $+\infty$. Finally, for $C \subset S$, $\overline{C} = S \setminus C$ denotes the complement of C in S.

2.2.1 Convergence to the invariant measure

Definition 2.3 *Given $k > 0$, $q : S \to \Delta(S)$ over S is k-*mixing *if* $\mathbf{E}_{s,q}[T_t^+] \leq k$*, for every $s, t \in S$.*

Note that every mixing transition rule is irreducible.

Theorem 2.4 *Assume that $q : S \to \Delta(S)$ is n-mixing, with invariant measure μ. Let $m \in \mathbf{N}$ and $\varepsilon \in (0, \frac{1}{4})$ be such that $\varepsilon m > 4n$. Then, for every $s, t \in S$,*

$$\mathbf{P}_t(|F_m^s - \mu_s| > \varepsilon\mu_s) < \frac{9(2n+1)}{m\varepsilon^2}. \tag{2}$$

Remark 2.5 *Inspection of the proof shows that inequality (2) holds more generally for each state $s \in S$ such that $\sup_{t\in S} \mathbf{E}_{t,q}\left[T_s^+\right] \leq n$.*

Remark 2.6 *Since $\left|\overline{F}_m^s - F_m^s\right| \leq \frac{1}{m}$, one has, under the assumptions of Theorem 2.4,*

$$\mathbf{P}_t(|\overline{F}_m^s - \mu_s| > \varepsilon\mu_s + \frac{1}{m}) < \frac{9(2n+1)}{m\varepsilon^2}. \tag{3}$$

Proof. We first deal with the case $s = t$. Denote by $T_s^{+,1} + ... + T_s^{+,p}$ the pth return time to s. Note that the variables $T_s^{+,k}$ are *iid*, and share the law of T_s^+ under $\mathbf{P}_s$. For each m, the event $|F_m^s - \mu_s| \geq \varepsilon\mu_s$ is included in the union of the two events $\left\{T_s^{+,1} + ... + T_s^{+,\lceil m\mu_s(1-\varepsilon)\rceil} \geq m\right\}$ and $\left\{T_s^{+,1} + ... + T_s^{+,\lfloor m\mu_s(1+\varepsilon)\rfloor} \leq m\right\}$. For notational convenience, set $m_\varepsilon := \lceil m\mu_s(1-\varepsilon)\rceil$ and $m^\varepsilon := \lfloor m\mu_s(1+\varepsilon)\rfloor$.

By Chebycheff inequality, since $\mathbf{E}_s\left[T_s^+\right] = \frac{1}{\mu_s}$ and the variables are independent,

$$\begin{aligned}\mathbf{P}_s(T_s^{+,1} + ... + T_s^{+,m_\varepsilon} \geq m) &= \mathbf{P}_s\left(T_s^{+,1} + ... + T_s^{+,m_\varepsilon} - \frac{m_\varepsilon}{\mu_s} \geq m - \frac{m_\varepsilon}{\mu_s}\right)\\ &\leq \frac{m_\varepsilon var_s T_s^+}{\left(m - \frac{m_\varepsilon}{\mu_s}\right)^2} \leq \frac{m_\varepsilon var_s T_s^+}{\left(m\varepsilon - \frac{1}{\mu_s}\right)^2},\end{aligned}$$

where the second inequality holds since $m - \frac{m_\varepsilon}{\mu_s} \geq m\varepsilon - \frac{1}{\mu_s}$, and

$$\begin{aligned}\mathbf{P}_{s,q}(T_s^{+,1} + ... + T_s^{+,m^\varepsilon} \leq m) &\leq \mathbf{P}_s\left(\frac{m^\varepsilon}{\mu_s} - (T_s^{+,1} + ... + T_s^{+,m^\varepsilon}) \geq \frac{m^\varepsilon}{\mu_s} - m\right)\\ &\leq \frac{m^\varepsilon var_s T_s^+}{\left(\frac{m^\varepsilon}{\mu_s} - m\right)^2} \leq \frac{m^\varepsilon var_s T_s^+}{\left(m\varepsilon - \frac{1}{\mu_s}\right)^2}.\end{aligned}$$

Hence,

$$\mathbf{P}_{s,q}(|F_m^s - \mu_s| \geq \varepsilon\mu_s) \leq \frac{(m_\varepsilon + m^\varepsilon) var_s T_s^+}{\left(m\varepsilon - \frac{1}{\mu_s}\right)^2}. \tag{4}$$

Since q is n-mixing, $\frac{1}{\mu_s} = \mathbf{E}_s\left[T_s^+\right] \leq n$. Therefore, the denominator in (4) is at least $\frac{9}{16}m^2\varepsilon^2$. On the other hand, by Aldous and Fill (2002, chapter 2, page 21, identity (22)

$$var_s T_s^+ \times \mu_s = 2\mathbf{E}_\mu T_s + 1 - \frac{1}{\mu_s}. \tag{5}$$

Since q is n-mixing, $\mathbf{E}_\mu T_s \leq \mathbf{E}_\mu T_s^+ \leq n$, hence $var_s T_s^+ \times \mu_s \leq 2n+1$. Since $m\mu_s \geq m/n > 4/\varepsilon > 1$, $m_\varepsilon + m^\varepsilon \leq 2m\mu_s + 1 \leq 3m\mu_s$, and we obtain

$$\mathbf{P}_s(|F_m^s - \mu_s| \geq \varepsilon\mu_s) \leq \frac{16 \times 3(2n+1)}{9m\varepsilon^2}.$$

This concludes the proof in the case $s = t$.

Let now $s \neq t$ in S. We estimate $\mathbf{P}_t(T_s^{+,1} + ... + T_s^{+,m_\varepsilon} \geq m)$ and $\mathbf{P}_t(T_s^{+,1} + ... + T_s^{+,m^\varepsilon} \leq m)$ in turn. Since q is n-mixing, we obtain by Markov inequality

$$\mathbf{P}_t(T_s^+ \geq \varepsilon^2 m) \leq \frac{n}{m\varepsilon^2}. \tag{6}$$

On the other hand, by following the steps of the previous computation,

$$\begin{aligned}\mathbf{P}_t(T_s^{+,1} \leq \varepsilon^2 m, T_s^{+,1} + ... + T_s^{+,m_\varepsilon} \geq m) &\leq \mathbf{P}_t(T_s^{+,2} + ... + T_s^{+,m_\varepsilon} \geq m(1-\varepsilon^2)) \\ &\leq \frac{(m_\varepsilon - 1) var_s T_s^+}{(m(1-\varepsilon^2) - \frac{m_\varepsilon - 1}{\mu_s})^2} \leq \frac{(m_\varepsilon - 1) var_s T_s^+}{(m(\varepsilon - \varepsilon^2))^2} \end{aligned} \tag{7}$$

and

$$\begin{aligned}\mathbf{P}_t(T_s^{+,1} + ... + T_s^{+,m^\varepsilon} \leq m) &\leq \mathbf{P}_t(T_s^{+,2} + ... + T_s^{+,m^\varepsilon} \leq m - 1) \\ &\leq \frac{(m^\varepsilon - 1) var_s T_s^+}{(\frac{m^\varepsilon - 1}{\mu_s} + 1 - m)^2} \leq \frac{(m^\varepsilon - 1) var_s T_s^+}{(m\varepsilon + 1 - \frac{2}{\mu_s})^2}. \end{aligned} \tag{8}$$

In both (7) and (8), the denominator is at least $\left(\frac{1}{2}m\varepsilon\right)^2$. Therefore, summation of (6), (7) and (8) yields

$$\mathbf{P}_t(|F_m^s - \mu_s| \geq \varepsilon\mu_s) \leq \frac{4 var_s T_s^+ (m_\varepsilon + m^\varepsilon - 2) + nm}{m^2\varepsilon^2}.$$

Since $m_\varepsilon + m^\varepsilon - 2 \leq 2m\mu_s$, one gets

$$\mathbf{P}_t(|F_m^s - \mu_s| \geq \varepsilon\mu_s) \leq \frac{4 \times 2(2n+1) + n}{m\varepsilon^2},$$

hence the result. ■

2.2.2 Expected exit times

We assume throughout this section that q is irreducible. We use repeatedly the following inequality

$$\mathbf{E}_u\left[T_{\overline{L}}\right] \leq \mathbf{E}_u\left[T_{\overline{L}\cup v}\right] + \mathbf{E}_v\left[T_{\overline{L}}\right] \tag{9}$$

that holds for every $L \subset S$ and every $u, v \in L$.

Proposition 2.7 *Let $C \subset S$, with $|C| > 1$. Define $\rho_1(C) = \max_{D\subset C} \min_{s\in D} \mathbf{E}_s\left[T_{\overline{D}}\right]$ and $\rho_2(C) = \max_{s\in C} \mathbf{E}_s\left[T_{\overline{C}}\right]$. One has*

$$\begin{aligned} \mathbf{E}_s\left[T_{\overline{D}}\right] &\leq |D|\,\rho_1(C) \text{ for every } D \subset C \text{ and } s \in D, \text{ and} && (10)\\ \mathbf{E}_s\left[T_{\overline{C}}\right] &\geq \rho_2(C) - (|C|-1)\rho_1(C) \text{ for every } s \in C. && (11) \end{aligned}$$

Proof. We prove (10) by induction over $|D|$. Plainly, the inequality holds for singletons. Assume that the result holds for every subset of size k. Let $D \subset C$ be of size $k+1$, and $s \in D$. By the definition of $\rho_1(C)$, there is $t \in D$, such that $\mathbf{E}_t\left[T_{\overline{D}}\right] \leq \rho_1(C)$. By (9) and the induction hypothesis for $D \setminus t$,

$$\mathbf{E}_s\left[T_{\overline{D}}\right] \leq \mathbf{E}_s[T_{\overline{D}\cup t}] + \mathbf{E}_t[T_{\overline{D}}] \leq (|D|-1)\rho_1(C) + \rho_1(C).$$

We now prove (11). Let $s \in C$ be given. For $t \neq s \in C$, one has, by (9) and (10)

$$\mathbf{E}_s\left[T_{\overline{C}}\right] \geq \ \mathbf{E}_t\left[T_{\overline{C}}\right] - \mathbf{E}_t\left[T_{\overline{C}\cup s}\right] \geq \mathbf{E}_t\left[T_{\overline{C}}\right] - (|C|-1)\rho_1(C). \tag{12}$$

The result now follows by taking the maximum over t in (12). ∎

Corollary 2.8 *Under the notations of Proposition 2.7, one has*

$$\mathbf{P}_s\left(T_{\overline{C}} < T_t\right) \leq 2\,|C|\,\frac{\rho_1(C)}{\rho_2(C) - (|C|-1)\rho_1(C)} \text{ for each } C \subset S,\ s,t \in C. \tag{13}$$

Proof. Let $C \subset S$, and $s, t \in C$ be given. We modify the Markov chain by collapsing $\overline{C}$ to a single state, still denoted $\overline{C}$, and we set $q(t|\overline{C}) = 1$, so that $\mathbf{E}_{\overline{C}}[T_t] = 1$. This modification does not affect $\mathbf{P}_s\left(T_{\overline{C}} < T_t\right)$. By Aldous and Fill (2002, Chapter 2, Corollary 10),

$$\mathbf{P}_s(T_{\overline{C}} < T_t) = \frac{\mathbf{E}_s\left[T_t\right] + \mathbf{E}_t\left[T_{\overline{C}}\right] - \mathbf{E}_s\left[T_{\overline{C}}\right]}{\mathbf{E}_{\overline{C}}\left[T_t\right] + \mathbf{E}_t\left[T_{\overline{C}}\right]}. \tag{14}$$

Since $\mathbf{E}_{\overline{C}}[T_t] = 1$, one has $\mathbf{E}_s\left[T_t\right] \leq \mathbf{E}_s\left[T_{t\cup\overline{C}}\right] + 1$. By (9), $\mathbf{E}_t\left[T_{\overline{C}}\right] - \mathbf{E}_s\left[T_{\overline{C}}\right] \leq \mathbf{E}_t\left[T_{s\cup\overline{C}}\right]$. By (10), the numerator in (14) is at most $1 + \mathbf{E}_t[T_{\overline{C}\cup s}] + \mathbf{E}_s[T_{\overline{C}\cup t}] \leq 2(|C|-1)\rho_1(C) + 1$.

On the other hand, the denominator is equal to $1 + \mathbf{E}_t\left[T_{\overline{C}}\right]$, hence, by (11), at least $\rho_2(C) - (|C|-1)\rho_1(C)$. ■

The next result deals with the transition function q^C of the Markov chain q watched on C (see Aldous and Fill (2001, Chapter 2, Section 7.1)):

$$q^C(t \mid s) = q(t \mid s) + \sum_{u \notin C} q(u \mid s)\mathbf{P}_u(T_C = T_t), \text{ for every } s, t \in C. \tag{15}$$

By Aldous and Fill, q^C is irreducible, and its invariant measure μ^C coincides with the invariant measure of q, conditioned on C.

Corollary 2.9 *For $s, t \in C$, one has $\mathbf{E}_{s,q^C}\left[T_t\right] \leq \frac{(|C|-1)\rho_1(C)}{\min_{u \in C} \mathbf{P}_{u,q}(T_t < T_{\overline{C}})}$.*

Proof. Let $t \in C$ be given. For convenience, set $\alpha := \max_{s \in C} \mathbf{E}_{s,q^C}\left[T_t\right]$. Let $s \in S$ achieve the maximum in the definition of α. By (10)

$$\begin{aligned} \alpha = \mathbf{E}_{s,q^C}\left[T_t\right] &\leq \mathbf{E}_{s,q}\left[T_{\overline{C} \cup t}\right] + \mathbf{P}_{s,q}(T_{\overline{C}} < T_t)\alpha. \\ &\leq (|C|-1)\rho_1(C) + \alpha \mathbf{P}_{s,q}(T_{\overline{C}} < T_t). \end{aligned}$$

Then, for every $s' \in C$,

$$\mathbf{E}_{s',q^C}[T_t] \leq \alpha \leq \frac{(|C|-1)\rho_1(C)}{1 - \mathbf{P}_{s,q}(T_{\overline{C}} < T_t)} \leq \frac{(|C|-1)\rho_1(C)}{\min_{u \in C} \mathbf{P}_{u,q}(T_t < T_{\overline{C}})},$$

as desired. ■

2.2.3 A structure theorem

Here we prove a structure result which states that for every finite sequence of states in S there is a partition of S such that the number of times the sequence exits a given atom of the partition is much smaller than the number of visits to any strict subset of this atom. The sequence moves around inside the atom much more quickly than from one atom to another.

For every positive integer $N \in \mathbf{N}$, every sequence $(s_0, s_1, \ldots, s_N)$ of states, and every subset $C \subset S$, define

$$R_C = |\{n < N \mid s_n \notin C, s_{n+1} \in C\}| + 1_{s_0 \in C}.$$

R_C is the number of C-runs along the sequence (see Feller (1968, II.5)). For convenience of notations, we omit the dependency of R_C on the sequence. Note that $R_{C \setminus D} \leq R_C + R_D$ for every proper subset D of C, and that $|R_C - R_{S \setminus C}| \leq 1$. Note also that $R_C \geq |\{n < N | s_n \in C, s_{n+1} \notin C\}|$

Theorem 2.10 *For every positive integer N, every sequence $(s_0, s_1, \dots, s_N)$, of states in S, and every $a > 0$, there is a partition $\mathcal{C}$ of S such that the following holds for every $C \in \mathcal{C}$.*

P1 *$R_C \leq (a+1)^{|\mathcal{C}|}$.*

P2 *For each proper subset D of C, $R_D > aR_C$.*

Proof. Observe that the trivial partition $\mathcal{C} = \{S\}$ satisfies **P1**, since $R_S = 1$.

Among all the partitions that satisfy **P1**, let $\mathcal{C}$ be one with maximal number of atoms. Denote $k = |\mathcal{C}|$. We prove that $\mathcal{C}$ satisfies **P2**. Otherwise, there is $C \in \mathcal{C}$, and there is a proper subset D of C, such that $R_D \leq aR_C$.

Consider the partition $\mathcal{C} \setminus \{C\} \cup \{D, C \setminus D\}$; that is, we further partition the set C into two sets D and $C \setminus D$. We show that this new partition, that has $k+1$ elements, satisfies **P1** as well, contradicting the maximality of $\mathcal{C}$. Indeed, $R_D \leq aR_C \leq (a+1)^{k+1}$, and $R_{C\setminus D} \leq R_C + R_D \leq R_C(a+1) \leq (a+1)^{k+1}$. ■

2.3 Proof of Theorem 2.2

To prove Theorem 2.2 it is sufficient to consider only exhaustive sequences; namely, sequences that visit all states in S (by dropping from S states that are never visited). However, as the proof of the more general Theorem 3.4 below refers to the proof of Theorem 2.2, it is more convenient not to make this assumption.

We prove Theorem 2.2 first by considering periodic and exhaustive sequences, and then by looking at a general sequence.

Let $\varepsilon > 0$ be small enough and $\rho \in (0, 1/2(4|S|+1))$ be fixed. Choose $\zeta < 2\rho$.

2.3.1 The case of periodic exhaustive sequences

We choose $N_0 \in \mathbf{N}$ be such that (N.i) $N_0^{(4|S|+1)\rho-1} \leq \varepsilon/(2^{|S|}+1)$, (N.ii) $N_0^{4\rho} \geq \max\{11|S|, 2/\varepsilon\}$, (N.iii) $N_0^{2\rho-\zeta} \geq 4 \times 19\,|S|\,/\varepsilon^2$, and (N.iv) $N_0^{\rho} \geq 4\,|S|+1$. Let $N \geq N_0$ and set $a = N^{4\rho}$.

We assume here that the sequence $\sigma = (s_0, s_1, \dots, s_N)$ is periodic and exhaustive: $s_N = s_0$ and $N_s \geq 1$ for every $s \in S$. The proof of the following lemma is left to the reader.

Lemma 2.11 *The empirical transition function q^σ is irreducible. Its invariant distribution is $\mu_s = \frac{N_s}{N}$.*

Let $\mathcal{C} = (S_1, \dots, S_K)$ be the partition of S obtained when applying Theorem 2.10 to σ and a. For $C \subset S$, we let $n_C := \sum_{s\in C} N_s$ denote the number of stages spent in C along σ. We abbreviate n_{S_k} to n_k.

Proposition 2.12 *With the notations of Proposition 2.7, one has*

$$\rho_1(S_k) \leq \max_{D\subset S_k} \frac{n_D}{R_D - 1} \leq \frac{2}{a}\rho_2(S_k) \text{ for every } k \text{ such that } |S_k| > 1.$$

Proof. For $D \subset S$, set

$$\nu_D(s) := \frac{\sum_{t\in\overline{D}} \mu_t q^\sigma(s|t)}{\sum_{t\in\overline{D}} \mu_t q^\sigma(D|t)} \text{ for } s \in C, \text{ and } K_D := \sum_{s\in C} \nu_D(s)\mathbf{E}_{s,q^\sigma}[T_{\overline{D}}]. \quad (16)$$

The numerator (resp. the denominator) in (16) is the long run frequency of transitions from $\overline{D}$ to s (resp. from $\overline{D}$ to D). Thus, $\nu_D(s)$ is the probability that the first encountered state in D is s, while K_D is the average length of a visit to D. We shall use the identity (easily derived from the ergodic theorem)

$$\sum_{s\in D} \nu_D(s)\mathbf{E}_{s,q^\sigma}[T_{\overline{D}}] = \frac{\sum_{s\in D}\mu_s}{\sum_{s\in D}\mu_s q^\sigma(\overline{D} \mid s)}. \quad (17)$$

Observe that $\sum_{s\in D}\mu_s = \frac{n_D}{N}$ and $\sum_{s\in D}\mu_s q^\sigma(\overline{D} \mid s) = \frac{|\{n<N|s_n\in D, s_{n+1}\notin D\}|}{N}$. Therefore,

$$\frac{n_D}{R_D} \leq \sum_{s\in D} \nu_D(s)\mathbf{E}_{s,q^\sigma}[T_{\overline{D}}] \leq \frac{n_D}{R_D - 1} \quad (18)$$

Fix k. By (N.ii) $a \geq 2$, hence $\frac{1}{aR_{S_k}-1} \leq \frac{2}{a} \times \frac{1}{R_{S_k}}$. Hence, for every $D \subset S_k$, by (18), the definition of $\mathcal{C}$, and (18) again,

$$\min_{s\in D} \mathbf{E}_{s,q^\sigma}[T_{\overline{D}}] \leq \frac{n_D}{R_D - 1} \leq \frac{n_{S_k}}{aR_{S_k} - 1} \leq \frac{2}{a}\frac{n_k}{R_{S_k}} \leq \frac{2}{a}\max_{s\in S_k}\mathbf{E}_{s,q^\sigma}\left[T_{\overline{S}_k}\right] = \frac{2}{a}\rho_2(S_k).$$

The result follows, by taking the supremum over $D \subset S_k$. ■

We now construct a piecewise Markov chain with K pieces. The kth piece is used for n_k stages, and its goal is to approximate the empirical transitions on S_k. In those stages, the process will remain in S_k.

Let $m_1 = 0$, and for every positive integer $k \leq K - 1$ define $m_{k+1} = n_1 + n_2 + \cdots + n_k$.

For every $k = 1, \ldots, K$ define a transition rule $q'_k : S \to \Delta(S)$ as follows. If $n_k < N^{1-\rho}$ we define $q'_k = q^\sigma$. Otherwise we define

$$q'_k(t \mid s) = \begin{cases} q_k(t \mid s) & s \in S_k, t \in S_k \\ \mu_k(t) & s \notin S_k, t \in S_k \end{cases}$$

where q_k is the transition function of the Markov chain q^σ watched on S_k (see Eq. (15)) and μ_k is the invariant measure of q_k. Let $(\mathbf{s}_n)_{0\leq n\leq N}$ be the piecewise Markov chain that starts in S_1 and follows the transition rule q'_k from stage m_k up to m_{k+1}, for each k. We will show that it satisfies the requirements of Theorem 2.2.

We first show that condition **B2** is satisfied. Fix k and let $s \in S_k$. If $n_k < N^{1-\rho}$ then $q'_k(\cdot|s) = q^\sigma(\cdot|s)$. Otherwise,

$$\|q'_k(\cdot \mid s) - q^\sigma(\cdot \mid s)\| \leq \sum_{u \notin S_k} q^\sigma(u \mid s) \leq \frac{R_{S_k}}{N_s}.$$

If $S_k = \{s\}$ is a singleton, the right hand side is bounded by $\frac{(a+1)^{|S|}}{N^{1-\rho}} < \varepsilon$, while if $|S_k| \geq 2$, the right hand side is bounded by $\frac{R_{S_k}}{aR_{S_k}} < \varepsilon$. Therefore, $\|\mathbf{q}_n - q^\sigma(\cdot|\mathbf{s}_n)\| < \varepsilon$ holds a.s. whenever $n \neq m_k$, for $k = 0, .., K-1$.

We now prove that condition **B1** is satisfied. Let k be given. If $n_k < N^{1-\rho}$, then $\nu^\sigma_s < \frac{1}{N^\rho}$ for every $s \in S_k$, hence **B1** holds for such states. If $n_k \geq N^{1-\rho}$ and $S_k = \{s\}$ is a singleton then $\overline{F}^s_N = \nu^\sigma_s$, and **B1** holds as well.

We may thus assume that $n_k \geq N^{1-\rho}$ and $|S_k| \geq 2$. We establish the claim by proving first that q_k is mixing, and by using Theorem 2.4.

Lemma 2.13 *The transition function q_k on S_k is $N^{1-3\rho}$-mixing.*

Proof. By Corollary 2.9

$$\mathbf{E}_{s,q_k}[T_t] \leq \frac{(|S_k|-1)\rho_1(S_k)}{1 - \max_{u \in S_k} \mathbf{P}_{u,q^\sigma}(T_{\overline{S}_k} < T_t)}.$$

Abbreviate $\rho_1(S_k)$ and $\rho_2(S_k)$ to ρ_1 and ρ_2 respectively. By Corollary 2.8, the denominator is at least $1 - 2|S_k| \frac{\rho_1}{\rho_2 - (|S_k|-1)\rho_1}$. Therefore,

$$\mathbf{E}_{s,q_k}[T_t] \leq (|S_k|-1)\rho_1 \times \frac{\rho_2 - (|S_k|-1)\rho_1}{\rho_2 - (3|S_k|-1)\rho_1} \leq 2|S_k|\rho_1 \tag{19}$$

where the second inequality follows by Proposition 2.12 and (N.ii).

By Proposition 2.12

$$\rho_1 \leq \max_{D \subset S_k} \frac{n_D}{R_D - 1} \leq \frac{N}{a-1} \leq \frac{N^{1-3\rho}}{2|S_k|} - 1, \tag{20}$$

since $N^\rho \geq 4|S| + 1$. The result follows by (19) and (20). ■

By Section 2.2.2, the invariant measure of q_k is $\nu^\sigma(\cdot|S_k)$, where $\nu^\sigma(t \mid S_k)) = N_t/n_k$ for $t \in S_k$. By Remark 2.5, (N.ii), (N.iii),

$$\begin{aligned}
\mathbf{P}_{s,q_k}&\left(|\overline{F}^t_{n_k} - \nu^\sigma(t|S_k)| > \varepsilon\nu^\sigma(t|S_k)\right) \\
&\leq \mathbf{P}_{s,q_k}\left(|\overline{F}^t_{n_k} - \nu^\sigma(t|S_k)| > \frac{\varepsilon}{2}\nu^\sigma(t|S_k) + \frac{1}{n_k}\right) \\
&\leq 4 \times \frac{19N^{1-3\rho}}{n_k\varepsilon^2} \leq 4 \times \frac{19N^{1-3\rho}}{\varepsilon^2 N^{1-\rho}} \leq \frac{1}{|S|} \times \frac{1}{N^\zeta}.
\end{aligned} \tag{21}$$

Since the process $(\mathbf{s}_n)$ does not visit $t \in S_k$ except in the kth phase, **B1** follows from (21) by summation over t.

2.3.2 The sequence $\sigma = (s_0, s_1, \ldots, s_N)$ is arbitrary

Choose N_0 so that, (N'.i) $N^{2\rho(4|S|+1)-1} \leq \frac{\varepsilon-2\varepsilon^2}{2^{|S|+1}}$, (N'.ii) $N_0^{8\rho} \geq \max(11|S|, 2/\left(\varepsilon - 2\varepsilon^2\right))$, (N'.iii) $N_0^{4\rho-\zeta} \geq 4 \times 19\,|S|\,/\left(\varepsilon - 2\varepsilon^2\right)^2$, (N'.iv) $N_0^{1-2\rho} \geq \frac{|S|}{\varepsilon^2}$, (N'.v) $N_0^{\rho} \geq \frac{1}{1-\varepsilon^2}$, (N'vi) $N_0^{2\rho} \geq 4\,|S| + 1$

Let $N \geq N_0$ and $\sigma = (s_0, ..., s_N)$ be an arbitrary sequence. We will add few states to σ, so as to get a periodic and exhaustive sequence. We next apply the results of **Case 1** to the new sequence, and then prove that similar estimates hold for the original sequence.

Let $S^* = \cup_{n=0}^{N}\{s_n\} \subseteq S$ be the set of states visited by σ. Consider the sequence $\sigma^* = (s_0, s_1, \ldots, s_N, s_1^*, \ldots, s_r^*, s_0)$, where $r = |S| - |S^*|$ is the number of states not visited by σ, and $S \backslash S^* = \{s_1^*, \ldots, s_r^*\}$. By construction, this new sequence is periodic and exhaustive. The length $N_* + 1$ of this sequence is $N + r + 2 < N + |S| + 2$.

One can verify that N_* satisfies (N.i-iv) with $\rho' := 2\rho$ and $\varepsilon' := \varepsilon - 2\varepsilon^2$. Therefore there is a piecewise Markov chain[3] $(\mathbf{s}_n)_{n \leq N_*}$ such that **B1** and **B2** hold w.r.t. ν^{σ^*}. Observe that each state $s_j^* \in S \setminus S^*$ constitutes a singleton in the partition $\mathcal{C}$ associated with σ^*. We assume that the last r stages are devoted to these elements of the partition, and we now check that the restriction $(\mathbf{s}_n)_{n \leq N}$ of the process to the first N stages satisfies **B1** and **B2**.

We start with **B1**. Let $s \in S$ with $\nu_s^\sigma \geq \frac{1}{N^\rho}$. By (N'.iv),

$$\left|\nu_s^\sigma - \nu_s^{\sigma^*}\right| \leq \frac{r+1}{N} \leq \nu_s^\sigma \frac{r+1}{N^{1-2\rho}} \leq \varepsilon^2 \nu_s^\sigma. \tag{22}$$

By (N'.v),

$$\nu_s^\sigma \geq \frac{1}{N^\rho} \Longrightarrow \nu_s^{\sigma^*} \geq \frac{1}{N^{2\rho}} \Rightarrow \mathbf{P}\left(\left|\overline{F}_{N_*}^s - \nu_s^{\sigma^*}\right| \geq \varepsilon' \nu_s^{\sigma^*}\right) \leq \frac{1}{N^\zeta}. \tag{23}$$

In such a case, by (N'.iv)

$$\left|\overline{F}_{N_*}^s - \overline{F}_N^s\right| \leq \frac{r+1}{N} \leq \varepsilon^2 \nu_s^\sigma.$$

hence, by (22), $\left|\overline{F}_N^s - \nu_s^\sigma\right| \leq \left|\overline{F}_{N_*}^s - \nu_s^{\sigma^*}\right| + 2\varepsilon^2 \nu_s^\sigma$. Condition **B1** follows using (23).

We now prove **B2**. By construction, except for at most $|S|$ stages,

$$\mathbf{q}_n = q^\sigma(\cdot|\mathbf{s}_n) \text{ or both } \left|\mathbf{q}_n - q^{\sigma^*}(\cdot|\mathbf{s}_n)\right| \leq \varepsilon' \text{ and } N_{\mathbf{s}_n}^* \geq N_*^{1-2\rho} \text{ hold.}$$

In the latter case, by (N'.iii), $\left|q^{\sigma^*}(\cdot|\mathbf{s}_n) - q^\sigma(\cdot|\mathbf{s}_n)\right| \leq 1/N \leq \varepsilon'$, which concludes the proof.

[3] In Case 1, we set $q_k' = q^\sigma$ whenever $n_k < N^{1-\rho}$.We still set here q_k' to be q^σ and not q^{σ^*}. This does not affect conclusions **B1** and **B2** for σ^*.

3 The General Problem

3.1 Presentation

For every state $s \in S$ let $V_s \subseteq \Delta(S)$ be a non-empty polyhedron, and set $V = (V_s)_{s \in S}$. The set V_s should be thought of as the set of conceivable transitions from s.

Throughout this section, V is fixed.

Definition 3.1 *A V-*process *is a S-valued process (X_n) such that for every $n \geq 1$, the conditional distribution of $\mathbf{s}_n$ given $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{n-1}$ is in $V_{\mathbf{s}_{n-1}}$.*

We here generalize the question addressed in the previous section. Given a sequence σ, does there exist a simple V-process that approximates σ, in the sense of Section 2. As suggested in the Introduction, there are several obstructions to the existence of such an approximation:

- The sequence σ may be completely atypical of any V-process.
- Transitions out of states that are transient under any V-process may not be approximated.

Therefore, we will assume that some transition function $b = (b_s)_s \in V$ is irreducible, and limit our analysis to sequences that are typical, in the sense defined below.

Definition 3.2 *Let $N \in \mathbf{N}$, and $\delta, \varepsilon > 0$. A sequence $(s_0, s_1, \ldots, s_N)$ is* (N, δ, ε)-typical *if there exists $v \in V$ such that $\left|1 - \frac{v(t|s)}{q(t|s)}\right| < \varepsilon$ for every $s, t \in S$ that satisfy $N_s q(t \mid s) \geq N^\delta$ or $N_s v(t \mid s) \geq N^\delta$. The set of (N, δ, ε)-typical sequences is denoted by $T^N_{\delta,\varepsilon}$.*

As we prove in the sequel, under some constraints on the parameters the probability of the typical sequences is close to 1, under any V-process.

Definition 3.3 *A process $(\mathbf{s}_n)_n$ is a (piecewise) hidden Markov chain over S if there exists an auxiliary finite set T and a (piecewise) Markov chain $(\mathbf{z}_n)$ over $S \times T$ such that $(\mathbf{s}_n)$ coincides with the marginal of $(\mathbf{z}_n)$ over S.*

Following the notations of Definition 3.3, let $p(\cdot|(s,t))$ be the transition function of $(\mathbf{z}_n)$. If the marginal over S of $p(\cdot|(s,t))$ belongs to V_s for each $(s,t) \in S \times T$, the process $(\mathbf{s}_n)$ is a V-process, due to the convexity of V_s. It is typically not a Markov chain. In such a case, we say that $(\mathbf{s}_n)$ is a hidden (piecewise) V-Markov chain.

Theorem 3.4 *Assume that b is irreducible, and set $B := \max_{s,t\in S} \mathbf{E}_{s,b}[T_t]$. Let $\psi, \eta \in (0,1)$ be given. There exist $\zeta, \rho, \delta, \varepsilon > 0$ and $N_1 \in \mathbf{N}$ such that the following holds. For every $N \geq N_1$ and every (N,δ,ε)-typical sequence $(s_0, s_1, \dots, s_N)$, there exists a hidden piecewise V-Markov chain with at most $|S|$ pieces such that*

G1 *If $\nu_s^\sigma \geq \frac{1}{N^\rho}$, then $\mathbf{P}(|\overline{F}_N^s - \nu_s^\sigma| \geq \eta\nu_s^\sigma) \leq \frac{1}{N^\zeta}$.*

G2 *Let $\mathbf{N}_0 = |\{n < N : \|\mathbf{q}_n - q^\sigma(\cdot \mid \mathbf{s}_n)\| > \eta\ \}|$. Then $\mathbf{E}[\mathbf{N}_0] \leq N^\psi B$.*

3.2 Typical sequences

Theorem 3.5 below states that most sequences are typical, provided the parameters are chosen properly. Its proof uses the following large deviation estimate for Bernouilli variables. Let $(X_n)_n$ be an infinite sequence of i.i.d. Bernouilli r.v.s with parameter p, and denote for every positive integer n, $\overline{X}_n = \sum_{k=1}^n X_i/n$. By Alon et al (2000, Corollary A.14),

$$\mathbf{P}(|\ \overline{X}_n - p\ | > \varepsilon p) \leq 2\exp(-c_\varepsilon pn),$$

where $c_\varepsilon = \min\{\varepsilon^2, -\varepsilon + (1+\varepsilon)\ln(1+\varepsilon)\}$ is independent of n and p. In particular, for every positive integer k,

$$\mathbf{P}\left(\sup_{pn\geq k} |\ \overline{X}_n - p\ | > \varepsilon p\right) \leq 2\sum_{n=\lceil k/p\rceil}^{\infty} \exp(-c_\varepsilon pn) \leq \frac{2\exp(-c_\varepsilon k)}{1-\exp(-c_\varepsilon p)}. \quad (24)$$

Observe that for every ε sufficiently small, $\varepsilon^2/3 < c_\varepsilon \leq \varepsilon^2/2$.

Theorem 3.5 *Let $\delta, \varepsilon > 0$ be given. For each $\xi \in (0, \delta/4)$, there exists $N_0 \in \mathbf{N}$ such that, for every $N \geq N_0$ and every V-process π,*

$$\mathbf{P}(T_{\delta,\varepsilon}^N) \geq 1 - \frac{1}{N^\xi}.$$

Proof. Let $\delta, \varepsilon \in (0,1)$ and $\xi \in (0, \delta/4)$ be given. For each $s \in S$, let V_s^* be the (finite) set of the extreme points of V_s. Choose $\xi' \in (\xi, \delta/4)$ and $\varepsilon' \in (0,1)$ such that $\frac{\varepsilon'}{1-\varepsilon'} \leq \varepsilon$. Let N_0 be large enough so that the following conditions are satisfied for each $N \geq N_0$: (i) $\frac{2\exp(-c_{\varepsilon'}N^{\delta/4})}{1-\exp(-c_{\varepsilon'}N^{\delta/4-1})} \leq 1/N^{\xi'}$, (ii) $N^{\xi'-\xi} \geq 3|S|\sum_{s\in S}|V_s^*|$ and (iii) $N^{\delta/2} \geq 1/\varepsilon$. Let $N \geq N_0$ and $(\mathbf{s}_n)$ be any V-process.

We first present the V-process $(\mathbf{s}_n)$ in an alternative way, by writing the conditional distribution of $\mathbf{s}_{n+1}$ given $\mathbf{s}_0, \dots, \mathbf{s}_n$ as a convex combination $\sum_{v\in V_{\mathbf{s}_n}^*} \mathbf{b}_n(v)v$ of the extreme points of $V_{\mathbf{s}_n}$ (the weights $\mathbf{b}_n(v)$ being random themselves).

Next, consider the following process $\pi' = (\mathbf{s}_n, \mathbf{v}_n)_n$. Given the past, $\mathbf{v}_n \in V^*_{\mathbf{s}_n}$ is selected according to $\mathbf{b}_n$, then $\mathbf{s}_{n+1}$ is selected according to $\mathbf{v}_n$. Plainly, the law of the sequence of states is the same under both processes. We shall deal with the latter process.

Define $\mathbf{n}_{s,v} = |\{n < N, (\mathbf{s}_n, \mathbf{v}_n) = (s, v)\}|$ to be the number of times the extreme point v was chosen at s, and $\mathbf{q}(t \mid s, v) = |\{n < N, (\mathbf{s}_n, \mathbf{v}_n, \mathbf{s}_{n+1}) = (s, v, t)\}| \ / \mathbf{n}_{s,v}$. Note that the empirical transitions out of s are given by $\mathbf{q}(t \mid s) = \frac{\sum_{v \in V^*} \mathbf{n}_{s,v}\mathbf{q}(t|s,v)}{\mathbf{n}_s}$, and define $\mathbf{v}^*_s = \frac{\sum_{v\in V^*} \mathbf{n}_{s,v} v}{\mathbf{n}_s}$. As V_s is convex, $\mathbf{v}^*_s \in V_s$. We will show that with high probability, $\mathbf{v}^*= (\mathbf{v}^*_s)$ is close to $\mathbf{q}$ in the sense of Definition 3.2.

Fix for a moment $s, t \in S$ and $v \in V^*_s$. Plainly, $\mathbf{n}_{s,v} v(t) < N^{\delta/4}$ if $v(t) < N^{\delta/4-1}$. We now assume that $v(t) \geq N^{\delta/4-1}$. Let $(X_n)_n$ be a sequence of i.i.d. Bernouilli r.v.s with parameter $v(t)$. By (24) and (i)

$$\mathbf{P}\left(\mathbf{n}_{s,v} v(t) \geq N^{\delta/4} \text{ and } \mid \mathbf{q}(t \mid s, v) - v(t) \mid > \varepsilon' v(t)\right) \leq \frac{2\exp(-c_{\varepsilon'} N^{\delta/4})}{1-\exp(-c_{\varepsilon'} N^{\delta/4-1})} \leq \frac{1}{N^{\xi'}}. \tag{25}$$

We now claim that

$$\mathbf{P}\left(\mathbf{n}_{s,v} v(t) < N^{\delta/4} \text{ and } \mathbf{n}_{s,v} q(t \mid s, v) \geq N^{\delta/2}\right) \leq 2/N^{\delta/4}. \tag{26}$$

Indeed, the left hand side in (26) is at most

$$\begin{aligned}&\mathbf{P}\left(\sup\ \left\{X_1 + \cdots + X_k, kv(t) < N^{\delta/4}\right\} \geq N^{\delta/2}\right)\\ &\leq \mathbf{P}\left(X_1 + \cdots + X_n \geq N^{\delta/2}\right), \text{ where } n = \left\lceil N^{\delta/4}/v(t)\right\rceil.\end{aligned}$$

By Markov inequality, this quantity is at most $nv(t)/N^{\delta/2} \leq 2/N^{\delta/4}$.

Eqs. (25) and (26) yield together

$$\mathbf{P}\left(\mathbf{n}_{s,v} \max\{v(t), \mathbf{q}(t \mid s, v)\} \geq N^{\delta/4} \Rightarrow \mid \mathbf{q}(t \mid s, v) - v(t) \mid \leq \varepsilon' v(t)\right) \geq 1 - \frac{3}{N^{\xi'}}. \tag{27}$$

Let T be the set of all sequences $(s_1, v_1, s_2, v_2, \ldots, s_N)$ that satisfy the implication in (27) for every $s, t \in S$ and every $v \in V^*_s$. By (ii), $\mathbf{P}(T) \geq 1 - \frac{1}{N^\xi}$. We will show that every sequence in T is (N, δ, ε)-typical.

Let us be given a sequence in T, and let $s, t \in S$ satisfy $N_s q(t \mid s) \geq N^\delta$ (the same argument is valid also in the case $N_s v(t) \geq N^\delta$). We prove that $|q(t \mid s) - v(t)| \leq \varepsilon q(t \mid s)$. We first claim that for every $v \in V^*_s$,

$$n_{s,v}|q(t \mid s, v) - v(t)| \leq \frac{\varepsilon'}{1-\varepsilon'} N_s q(t|s). \tag{28}$$

Indeed, if $n_{s,v} \max(\{q(t \mid s, v), v(t)\} \geq N^{\delta/2}$ then by (27) $v(t) \leq \frac{1}{1-\varepsilon'} q(t \mid s, v)$, and therefore

$$n_{s,v}|q(t \mid s, v) - v(t)| \leq \varepsilon' n_{s,v} v(t) \leq \frac{\varepsilon'}{1-\varepsilon'} n_{s,v} q(t \mid s, v) \leq \frac{\varepsilon'}{1-\varepsilon'} N_{ss} q(t \mid s),$$

where the last inequality holds since $N_{ss}q(t \mid s) = \sum_{v \in V_s^*} n_{s,v} q(t \mid s, v)$. If, on the other hand $n_{s,v} \max\{q(t \mid s, v), v(t)\} < N^{\delta/2}$ then

$$n_{s,v}|q(t \mid s, v) - v(t)| \leq N^{\delta/2} \leq N_s q(t \mid s)/N^{\delta/2},$$

and (28) holds by (iii).

By summing (28) over all $v \in V_s^*$ we get,

$$|q(t \mid s) - v_s^*(t)| \leq \sum_{v \in V_s^*} \frac{n_{s,v}}{N_s} |q(t \mid s, v) - v(t)| \leq \frac{\varepsilon'}{1 - \varepsilon'} q(t \mid s).$$

Thus

$$|q(t \mid s) - v_s^*(t)| \leq \varepsilon v_s^*(t),$$

as desired. ■

The requirement $\xi < \delta/4$ arises from the use of Markov inequality. A slight modification of the argument would improve the bound to $\delta/2$. It is not clear whether this latter bound is optimal.

3.3 Approximation of typical sequences

We here prove Theorem 3.4. The proof mostly follows the proof of Theorem 2.2. The main complication is the following. Each piece q_k of the Markovian approximation that was constructed in Section 2 was obtained by watching the empirical transition function q on a specific subset S_k. Characteristics of the corresponding chain q_k (invariant measure, mixing time) were then easily derived from the properties of q and of the partition of S. By constrast, each piece of the approximation is here required to be a V-process. Thus, the former choice for q_k may no longer be admissible, and one is led to choose the V-process that is closest (in a yet-to-be-defined sense) to q_k on S_k. Properties of this process are obtained from results on perturbations of Markov chains to be found in Solan and Vieille (2002).

Step 0: Fixing parameters

Let ψ and η be given. Choose ε small enough so as to satisfy the following conditions, with $L = \max\{\sum_{n=1}^{|S|-1} \binom{|S|}{n} n^{|S|}, 220\}$.

E1) $\varepsilon < \eta/55L < \eta$

E2) $\varepsilon < \dfrac{1}{8L^2 |S|^2}$

E3) $\varepsilon < \dfrac{1}{4|S| L^2 (5|S| - 1)}$

E4) $\varepsilon < 1/2^{|S|}$

Note that that E3) implies E2) and E4). Fix $\beta \in (0, \frac{1}{2}\left(\frac{A}{L}\right)^{|S|} \times \frac{\varepsilon(1-\varepsilon)}{L\times|S|^4})$ where A=1/2. Set $\alpha = \frac{1}{2\beta|S|L^2}$, $\alpha' = \frac{\alpha/2-|S|}{2|S|}$. By (E3), $\beta < 1/20|S|^2L^2$, so that $\alpha' \geq 2$.

Fix $\xi \in (0, \psi/(|S|+1))$. Choose $\delta < \min(\frac{\psi}{|S|+1}, \frac{\xi}{4}, \frac{1-\psi}{2})$, and δ', ψ' such that $\delta < \delta' < \psi' < \psi$. Set $a = N^{\xi}$.

We assume that N is large enough so as to satisfy the following conditions

DP1) $N^{\delta'} \geq N^{\delta} + 1$,

DP2) $N^{\delta} \geq \dfrac{3}{\varepsilon(1-3\varepsilon)}$,

DP3) $2 + 16L\,|S|\,(N+S+2)^{\psi'} < N^{\psi}/\,|S|$,

RZ1) $B \leq \dfrac{1}{144}\dfrac{N^{1-\delta}}{N^{3\delta}}$,

RZ2) $\eta N^{1-2\delta} \geq 1$,

RZ3) $4BL\,|S|\,N^{\delta+\psi'-1} \leq N^{-\delta} \leq 1/2$,

RZ4) $1/N^{\delta} < \varepsilon$,

RZ5) $N^{1-2\delta-\psi} \geq 2B/\varepsilon$,

A1) $\beta(a-1) \geq (N+S+2)^{\delta'}$,

A2) $a - 1 \geq \dfrac{1}{2\beta\,|S|}$,

A3) $a \geq 3$,

A4) $2a \leq N$,

A5) $(a+1)^{|S|} \leq N^{\psi} \leq N$,

A6) $\dfrac{72}{N^{1-\delta}}\left(|S|\,L\dfrac{N}{a-1} + B + 1\right) \leq N^{-\delta}$,

S1) $\dfrac{1}{2N^{\delta}}N^{1-\delta} \geq 4(B+1)$,

S2) $B(1+3\varepsilon)\dfrac{(a+1)^{|S|}}{N^{1-\rho}} \leq \dfrac{1}{2N^{\zeta}}$,

S3) $\dfrac{36(2B+3)}{N^{1-\rho}} \leq \dfrac{1}{2|S|N^{\zeta}}$, and

S4) $N^{\psi}/|S| \geq 1 + 2B(1+3\varepsilon)N^{\rho}(a+1)^{|S|}$.

Step 1: The periodized sequence

Let σ be an (N,δ,ε)-typical sequence. For every $s \in S$ choose $v(\cdot|s) \in V_s$ such that, for every t,

$$N_s \max\{q^{\sigma}(t|s), v(t|s)\} \geq N^{\delta} \Rightarrow \left|1 - \frac{v(t|s)}{q^{\sigma}(t\mid s)}\right| \leq \varepsilon. \qquad (29)$$

Let $\sigma^* = (s_0^*, \dots, s_{N_*}^*)$ be the periodic and exhaustive sequence that is generated from σ as in the proof of Theorem 2.2. Following the notations used in Section 2, we let $q := q^{\sigma^*}$ denote the empirical transitions along σ^*, and $n_C^* := \sum_{s \in C} N_s^*$ denote the number of stages spent in $C \subset S$ along σ^*. By (DP1),

$$N_s^* \max\{q(t|s), v(t|s)\} \geq N_*^{\delta'} \Rightarrow N_s \max\{q^\sigma(t|s), v(t|s)\} \geq N_*^{\delta}.$$

In that case, by (DP2), $\left|1 - \frac{N_{s \to t}}{N_{s \to t}^*}\right| \leq \frac{\varepsilon}{3}$, where $N_{s \to t}$ and $N_{s \to t}^*$ are the number of transitions from s to t along σ and σ^* respectively, and $\left|1 - \frac{N_s}{N_s^*}\right| \leq \frac{\varepsilon}{3}$. Hence $\left|1 - \frac{q(t|s)}{q^\sigma(t|s)}\right| \leq \varepsilon$ (see Lemma 15 in Solan and Vieille (2002)). Therefore,

$$N_s^* \max\{q(t|s), v(t|s)\} \geq N_*^{\delta'} \Rightarrow \left|1 - \frac{v(t|s)}{q(t \mid s)}\right| \leq 3\varepsilon. \tag{30}$$

In other words, σ^* is $(N_*, \delta', 3\varepsilon)$-typical.

Lemma 3.6 *Let $s \in S$ such that $N_s \geq N^{1-\delta}$. One has $\|v(\cdot|s) - q^\sigma(\cdot|s)\| \leq \eta$.*

Proof. Let $t \in S$ be given. If $\max\{v(t|s), q^\sigma(t|s)\} \leq \eta$, one has $|v(t|s) - q^\sigma(t|s)| \leq \eta$. Otherwise, by (RZ2),

$$N_s \max\{v(t|s), q^\sigma(t|s)\} \geq \eta N^{1-\delta} \geq N^\delta.$$

Therefore, by (29) and (E1), $|v(t|s) - q^\sigma(t|s)| \leq \varepsilon q^\sigma(t|s) < \eta$. ■

Step 2 : The approximating process.

Let $(S_1, S_2, \dots, S_K)$ be the partition of S that is given by Theorem 2.10 w.r.t. σ^* and a. We abbreviate $n_{S_k}^*$ to n_k^*. Note that every state that is not visited by σ constitutes a singleton in this partition.

Let $m_1^* = 0$, and for every $k = 0, \dots, K-1$ define $m_{k+1}^* = n_1^* + n_2^* + \cdots + n_k^*$. As in the proof of Theorem 2.2, the approximating process π has K pieces. It follows a hidden V-Markov chain p_k from stage m_k^* up to m_{k+1}^*. All auxiliary Markov chains are defined on the same set $S \times T$, with $T = S \cup \{\square\}$, where $\square$ is an additional symbol. The initial state of the process is irrelevant. Unless otherwise stated, $\mathbf{E}$ stands for the expectations wrt the law of π.

If $n_k^* < N_*^{1-\delta}$, we let $p_k := q^\sigma$. Let now k be such that $n_k^* \geq N_*^{1-\delta}$.

We define a transition function p_k over $S \times T$ as follows:

- From state $(s, \Box)$, where $s \in S_k$: $s' \in S$ is first drawn according to $v(\cdot|s)$; if $s' \in S_k$, p_k moves to $(s', \Box)$; if $s' \notin S_k, t \in S_k$ is drawn with probability $\mathbf{P}_{s',q_k}(T_{S_k} = T_t)$ and p_k moves to (s', t).

- From state (s, t), where $s \neq t$ and $t \in S_k$: $s' \in S$ is first drawn according to $b(\cdot|s)$; if $s' = t$, p_k moves to $(s', \Box)$; if $s' \neq t$, p_k moves to (s', t).

- From state (s, t), where $s \notin S_k$ and $t \in \overline{S}_k \cup \{\Box\}$: (s', t') is drawn with probability $b(s'|s) \times \mathbf{P}_{s,q_k}(T_{S_k} = T_{t'})$. If $s' = t'$, p_k moves to $(s', \Box)$. Otherwise, p_k moves to (s', t').

All other transitions from these states receive probability zero. Transitions in other states are irrelevant. Note that the marginal over S of $p_k(\cdot|(s,t))$ belongs to V_s.

We now loosely describe the behavior of the S-coordinate. Starting from S_k, this coordinate evolves according to v until exit from S_k occurs. Then, the entry state in S_k is chosen at random, and the S-coordinate evolves according to the barycenter b until that particular state is reached. The behavior resumes from the beginning. The T-coordinate of the auxiliary chain serves as an indicator of whether b or v is currently used and, if relevant, specifies which entry state in S_k has been selected.

The third item in the definition of p_k is introduced to take care of the initial stage in phase k, where the current state is inherited from the previous phase. Observe that it is used only at stage m_k^*.

Note that $\mathbf{q}_n = v(\cdot|\mathbf{s}_n)$ holds whenever $\mathbf{t}_n = \Box$ and $n \neq m_k^*$, for $k = 1, .., K$. Observe that there is an ergodic set for p_k that contains $S_k \times \{\Box\}$. Let ν_k be the invariant measure of p_k this ergodic set.

For the sake of the proof, it is convenient to introduce the auxiliary transition function q_k defined by

$$q_k(\cdot \mid s) = \left\{ \begin{array}{ll} v(\cdot|s) & s \in S_k \\ q(\cdot \mid s) & s \notin S_k \end{array} \right. . \tag{31}$$

Thus, q_k coincides with v on S_k and with q on $\overline{S}_k$.

We proceed by proving several properties of the hidden Markov chain p_k.

The case where S_k is a singleton is albeit simpler, but also has some specific features. Therefore we shall postpone it and assume first in **Steps 3-6** that $|S_k| > 1$.

Step 3: Perturbation of Markov chains: reminder

We here introduce a result due to Solan and Vieille (2002). Given $C \subseteq S$ and an irreducible transition rule q^1 over S with invariant measure μ^1, set

$$\zeta_{q^1}^C = \min_{\emptyset \subset D \subset C} \sum_{s \in D} \mu_s^1 q^1(\overline{C} \mid s).$$

This is a variation of the conductance of a Markov chain, that was originally defined by Jerrum and Sinclair (1989), and was used in the study of the rate of convergence to the invariant measure (see also Lovasz and Kannan (1999), Lovasz and Simonovits (1990)).

Definition 3.7 *Let q^1 be an irreducible transition function on S, with invariant measure μ^1, and $C \subseteq S$. Let $\beta, \varepsilon > 0$. A transition rule q^2 is (β, ε)-*close to q on C *if (i) $q^2(\cdot \mid s) = q^1(\cdot \mid s)$ for every $s \notin C$; (ii) $\left|1 - \frac{q^2(t|s)}{q^1(t|s)}\right| < \varepsilon$ for every $s, t \in C$ such that $\mu_s^1 \max\{q^1(t \mid s), q^2(t|s)\} \geq \beta \zeta_{q^1}^C$.*

Given a transition function q over S, $i = 1, 2$, we set

$$\nu_C^q(s) = \frac{\sum_{t \in \overline{C}} \mu_t q(s|t)}{\sum_{t \in \overline{C}} \mu_t q(C|t)} \text{ for } C \subset S \text{ and } s \in C, \text{ and} \tag{32}$$

$$K_C^q = \sum_{s \in C} \nu_C(s) \mathbf{E}_{s,q}[e_C] \text{ for } C \subset S.$$

The numerator (resp. the denominator) in (32) is the long run frequency of transitions from $\overline{C}$ to s (resp. from $\overline{C}$ to C). Thus, $\nu_C^q(s)$ is the probability that the first stage in C the process visits is s, while K_C^q is the average length of a visit to C.

The next result summarizes Theorems 4 and 6 in Solan and Vieille (2002). Recall that $L = \max\{\sum_{n=1}^{|S|-1} \binom{|S|}{n} n^{|S|}, 220\}$.

Proposition 3.8 *Let $\varepsilon \in (0, 1/2^{|S|})$, $A > 0$ and $\beta \in (0, \frac{1}{2}\left(\frac{A}{L}\right)^{|S|} \times \frac{\varepsilon(1-\varepsilon)}{L \times |S|^4})$. Let q^1 be an irreducible transition function on S. Assume that $|C| > 1$ and that $\mathbf{P}_{s,q^1}(T_t^+ < T_{\overline{C}}^+) \geq A$, for every $s, t \in C$. Let q^2 be (β, ε)-close to q^1 on C. Then all states of C belong to the same ergodic set E for q^2. Moreover, for every $s \in C$ and $D \subset C$,*

$$|\mu^2(s|C) - \mu^1(s|C)| < 18\varepsilon L \mu^1(s|C), \tag{33}$$

$$L^{-1} \leq \frac{\mathbf{E}_{s,q^2}\left[T_{\overline{D}}\right]}{\mathbf{E}_{s,q^1}\left[T_{\overline{D}}\right]} \leq L \text{ and } L^{-1} \leq \frac{K_D^{q_2}}{K_D^{q_1}} \leq L. \tag{34}$$

In addition, let $\chi \in (0, \beta \zeta_{q^1}^C]$ be any number such that, for every $s, t \in C$,

$$\mu_s^1 \max\{q^1(t \mid s), q^2(t|s)\} \geq \chi \Rightarrow \left|1 - \frac{q^2(t \mid s)}{q^1(t \mid s)}\right| < \varepsilon. \tag{35}$$

Then at least one of the following hold.

$$(i) L^{-1}K_C^{q_1} \leq K_C^{q_2} \leq LK_C^{q_1} \text{ or } (ii) K_C^{q_1} \geq \frac{1}{2|S|}\times\frac{\mu_C^1}{\chi} \text{ and } K_C^{q_2} \geq \frac{1}{L}\times\frac{1}{2|S|}\times\frac{\mu_C^1}{\chi}. \tag{36}$$

Step 4 : Perturbation of Markov chains: application

We here apply Proposition 3.8 to the transition functions q and q_k (defined in (31)) and to $C = S_k$.

Lemma 3.9 *If $|S_k| \geq 2$ then the transition function q_k is $(\beta, 3\varepsilon)$-close to q on S_k.*

Proof. Using (30) it now suffices to prove that $\frac{N_*^{\delta'}}{N_*} \leq \beta\zeta_q^{S_k}$.
Let C be an arbitrary non-empty subset of S_k. One has

$$\sum_{s\in C}\mu_s q(\overline{C} \mid s) = \frac{R_{\overline{C}} - 1_{s_0\in\overline{C}}}{N_*} \geq \frac{R_C - 1}{N_*} \geq \frac{a-1}{N_*}. \tag{37}$$

By taking the minimum over C, this yields $\zeta_q^{S_k} \geq \frac{a-1}{N_*}$. The result follows by (A1). ■

We denote below by $\widehat{\rho}_i$, $i = 1, 2$, the value of the mixing constant ρ_i (defined in Proposition 2.7) for the transition function q_k. We abbreviate K for K^q, and $\widehat{K}$ for K^{q_k}.

Lemma 3.10 *If $|S_k| \geq 2$ then $\widehat{\rho}_1(S_k) \leq \frac{2}{\alpha}\widehat{\rho}_2(S_k)$, where $\alpha = 1/(2\beta|S|L^2)$.*

Proof. By (34), $\widehat{\rho}_1(S_k) \leq L\rho_1(S_k)$. We argue now that $K_{S_k} \geq \frac{1}{2\beta|S|} \times \frac{\mu_{S_k}}{\zeta_q^{S_k}}$. For $C \subset S_k$,

$$\frac{K_{S_k}}{\mu_{S_k}} = \frac{1}{\sum_{s\in S_k}\mu_s q(\overline{S}_k|s)} \geq \frac{N_*}{R_{S_k}} \geq a\frac{N_*}{R_C} \geq (a-1)\frac{1}{\sum_{s\in C}\mu_s q(\overline{C}|s)},$$

where the last inequality follows by (37) and since $R_C \geq aR_{S_k} \geq a$. The claim follows by optimizing over C, using (A2). By (36) and the definition of $\zeta_q^{S_k}$,

$$\widehat{\rho}_2(S_k) \geq \widehat{K}_{S_k} \geq \frac{1}{L}\times\frac{1}{2\beta|S|}\times\frac{\mu_{S_k}}{\zeta_q^{S_k}}. \tag{38}$$

Fix $C \subset S_k$. By (38),

$$\begin{aligned}\widehat{\rho}_2(S_k) &\geq \frac{1}{L}\times\frac{1}{2\beta|S|}\times\frac{\mu_C}{\sum_{s\in C}\mu_s q(\overline{C}|s)} \geq \frac{1}{L}\times\frac{1}{2\beta|S|}\times K_C \\ &\geq \frac{1}{L^2}\times\frac{1}{2\beta|S|}\times\widehat{K}_C \geq \frac{1}{L^2}\times\frac{1}{2\beta|S|}\times\inf_{s\in C}\mathbf{E}_{s,q_k}\left[T_{\overline{C}}\right].\end{aligned}$$

The result follows by taking the maximum over C. $\blacksquare$

It is convenient to let $\widetilde{F}^s_{n_k}$ denote the frequency of visits to $(s,\square)$ during phase k, and $\mathbf{N}_0^k = \left|\left\{m_k^* \leq n < m_{k+1}^* : \|\mathbf{q}_n - q^\sigma(\cdot|\mathbf{s}_n)\| > \eta\right\}\right|$.

Step 5: $\mathbf{E}\left[\mathbf{N}_0^k\right] \leq \frac{1}{|S|} B N^\psi$.

Denote by E_1 the ergodic set for p_k that contains $S_k \times \{\square\}$. Recall that ν_k is the invariant measure of p_k on E_1.

Lemma 3.11 *If* $|S_k| \geq 2$ *then* $\nu_k(S_k \times \{\square\}) \geq 1 - \frac{2B}{\widehat{\rho}_2(S_k)}$.

Proof. We shall use the following fact. Let $\overline{q}$ be an irreducible transition function over a finite set Ω, with invariant measure $\overline{\mu}$. Let $C \subset \Omega$, and $\overline{C} = \Omega \backslash C$. One has

$$\frac{\overline{\mu}(C)}{\overline{\mu}(\overline{C})} \geq \frac{\inf_{s\in C} \mathbf{E}_{s,\overline{q}}\left[T_{\overline{C}}\right]}{\sup_{s\in \overline{C}} \mathbf{E}_{s,\overline{q}}\left[T_C\right].} \tag{39}$$

We apply this observation to p_k and E_1, with $C = S_k \times \{\square\}$. Plainly, $\mathbf{E}_{(s,t),p_k}\left[T_{S_k\times\{\square\}}\right] = \mathbf{E}_{s,b}\left[T_t\right] \leq B$ for each (s,t) with $t \neq \square, s$, while by (11), Lemma 3.10, and (E2),

$$\begin{aligned}\inf_{s\in S_k} \mathbf{E}_{(s,\square),p_k}\left[T_{\overline{S_k\times\{\square\}}}\right] &= \inf_{s\in S_k} \mathbf{E}_{s,v}\left[T_{\overline{S}_k}\right] \\ &\geq \widehat{\rho}_2(S_k) - (|S_k|-1)\widehat{\rho}_1(S_k) \geq \widehat{\rho}_2(S_k)\left(1 - 4\beta L^2(|S|-1)|S|\right) \geq \frac{1}{2}\widehat{\rho}_2(S_k).\end{aligned}$$

By (39), one gets

$$\frac{\nu_k(\overline{S_k\times\{\square\}})}{\nu_k(S_k\times\{\square\})} \leq \frac{2B}{\widehat{\rho}_2(S_k)},$$

hence $\nu_k(\overline{S_k\times\{\square\}}) \leq \frac{2B}{\widehat{\rho}_2(S_k)}$. $\blacksquare$

Lemma 3.12 *If* $|S_k| \geq 2$ *then*

$$\widehat{\rho}_2(S_k) \geq \frac{1}{2L\,|S|} \frac{n_k^*}{N_*^{\psi'}}.$$

Proof. We will use the fact that $K_{S_k} \geq n_k^*/R_{S_k}$ (see Eq. (12) in Solan and Vieille (2002)).

By (30), and since $\delta' < \psi'$, (35) holds with $\chi = N_*^{\psi'-1}$. We distinguish two cases. If $K_{S_k} \geq \frac{1}{2|S|}\frac{n_k}{N\chi}$, then by (36)

$$\widehat{\rho}_2(S_k) \geq \widehat{K}_{S_k} \geq \frac{1}{L} \times \frac{1}{2\,|S|} \times \frac{n_k}{N\chi} = \frac{1}{L} \times \frac{1}{2\,|S|} \times \frac{n_k}{N_*^{\psi'}}.$$

If on the other hand $K_{S_k} < \frac{1}{2|S|}\frac{n_k}{N_\chi}$, then

$$\widehat{\rho}_2(S_k) \geq \widehat{K}_{S_k} \geq \frac{1}{L}K_{S_k} \geq \frac{1}{L} \times \frac{n_k^*}{R_{S_k}} \geq \frac{1}{L} \times \frac{n_k^*}{(a+1)^{|S|}} \geq \frac{1}{L}\frac{n_k^*}{(a+1)^{|S|}},$$

which gives also the result by (A5). ■

In particular, by (RZ3), $\nu_k(S_k \times \{\square\}) \geq 1/2$.

Lemma 3.13 *For each $\omega \in E_1$ and $s \in S_k$, one has*

$$\mathbf{E}_{\omega,p_k}\left[T^+_{(s,\square)}\right] \leq \frac{(|S_k|-1)\widehat{\rho}_1(S_k) + 2B}{\min_{u\in S_k}\mathbf{P}_{u,v}(T^+_s < T_{\overline{S}_k})} + 1.$$

Proof. It is a simple adaptation of the proof of Corollary 2.9. We repeat it, with few modifications. Let $s \in S_k$ be given. Note that

$$\mathbf{E}_{\omega,p_k}\left[T^+_{(s,\square)}\right] \leq 1 + \sup_{\omega'\in E_1}\mathbf{E}_{\omega',p_k}\left[T_{(s,\square)}\right]. \tag{40}$$

For convenience, set $\alpha := \max_{t\in S_k}\mathbf{E}_{(t,\square),p_k}\left[T_{(s,\square)}\right]$. Let $t \in S_k$ achieve the maximum in the definition of α. By (10)

$$\begin{aligned}\alpha = \mathbf{E}_{(t,\square),p_k}\left[T_{(s,\square)}\right] &\leq \mathbf{E}_{t,v}\left[T_{S_k\cup s}\right] + \mathbf{P}_{t,v}(T_{\overline{S}_k} < T^+_s)(\alpha + B)\\ &\leq (|S_k|-1)\widehat{\rho}_1(S_k) + (\alpha+B)\mathbf{P}_{t,v}(T_{\overline{S}_k} < T^+_s)\\ &\leq (|S_k|-1)\widehat{\rho}_1(S_k) + B + \alpha \times \sup_{u\in S_k}\mathbf{P}_{u,v}(T_{\overline{S}_k} < T^+_s)\end{aligned}$$

Therefore,

$$\alpha \leq \frac{(|S_k|-1)\widehat{\rho}_1(S_k) + B}{\min_{u\in S_k}\mathbf{P}_{u,v}(T^+_s < T_{\overline{S}_k})}. \tag{41}$$

For $\omega \in E_1\backslash(S_k \times \square)$,

$$\mathbf{E}_{\omega,p_k}\left[T_{(s,\square)}\right] \leq B + \alpha. \tag{42}$$

The result follows from (40), (41) and (42). ■

Lemma 3.14 *If $|S_k| \geq 2$ then for every $\omega \in S \times T$ one has*

$$\mathbf{E}_{\omega,p_k}\left[\mathbf{N}^k_0\right] \leq \frac{1}{|S|}BN^\psi.$$

Proof. Plainly, $\mathbf{E}_{\nu_k,p_k}\left[\mathbf{N}_0^k\right] = n_k^*\nu_k(E_1\backslash(S_k\times\{\square\})) \le 2B\times\frac{n_k^*}{\widehat{\rho}_2(S_k)}$: the equality holds since ν_k is invariant; the inequality holds by Lemma 3.11. Therefore,

$$\begin{aligned}\inf_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right]\nu_k(S_k\times\{\square\}) &\le \sum_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right]\nu_k(\omega)\\ &\le \mathbf{E}_{\nu_k,p_k}\left[\mathbf{N}_0^k\right] \le 2B\times\frac{n_k^*}{\widehat{\rho}_2(S_k)}.\end{aligned}$$

Since $\nu_k(S_k\times\{\square\})\ge 1/2$, this yields

$$\inf_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right] \le 2\times 2B\times\frac{n_k^*}{\widehat{\rho}_2(S_k)}. \tag{43}$$

Next, let $\gamma = \sup_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right]$ and let $\omega_1\in S_k\times\{\square\}$ be a state that achieves the supremum. Since p_k follows b once the process leaves $S_k\times\{\square\}$, one has, for each $\omega_2\in S_k\times\{\square\}$,

$$\gamma = \mathbf{E}_{\omega_1,p_k}\left[\mathbf{N}_0^k\right] \le \mathbf{E}_{\omega_2,p_k}\left[\mathbf{N}_0^k\right] + \mathbf{P}_{\omega_1}(T_{E_1\backslash(S_k\times\{\square\})} < T_{\omega_2})(B+\gamma).$$

By Corollary 2.8 and Lemma 3.10, $\mathbf{P}_{\omega_1,p_k}(T_{E_1\backslash S_k\times\{\square\}} < T_{\omega_2}) \le \frac{2|S|}{\alpha/2-|S|} = 1/\alpha'$. Since $\alpha'\ge 2$, one gets by letting ω_2 vary

$$\gamma \le \frac{\alpha'}{\alpha'-1}\inf_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right] + \frac{B}{\alpha'-1} \le 2\inf_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right] + B \tag{44}$$

Finally, for each $\omega'\in S\times T$, by (44), (43), Lemma 3.12 and (DP3),

$$\begin{aligned}\mathbf{E}_{\omega',p_k}\left[\mathbf{N}_0^k\right] &\le E_{\omega,p_k}\left[T_{S_k\times\{\square\}}\right] + \sup_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right]\\ &\le B + B + 2\inf_{\omega\in S_k\times\{\square\}}\mathbf{E}_{\omega,p_k}\left[\mathbf{N}_0^k\right]\\ &\le 2B + 2\times 2\times 2B\times\frac{n_k^*}{\widehat{\rho}_2(S_k)}\\ &\le 2B + 16L\,|S|\,BN_*^{\psi'} \le BN^{\psi}/\,|S|\,.\end{aligned}$$

■

Step 6: estimates on $\widetilde{F}^s_{n_k}$.

The lemma below is a mixing-type result. It is very close to Lemma 2.13.

Lemma 3.15 *If $|S_k|\ge 2$ then for every $\omega\in E_1$ and every $s\in S_k$,*

$$\mathbf{E}_{\omega,p_k}\left[T^+_{(s,\square)}\right] \le 2|S|L\frac{N}{a-1} + 4B + 1$$

Proof. We repeat the proof of Lemma 2.13 with minor adjustments. By Lemma 3.13,

$$\mathbf{E}_{\omega,p_k}[T^+_{(s,\Box)}] \leq \frac{(|S_k|-1)\widehat{\rho}_1(S_k)+2B}{\min_{u\in S_k}\mathbf{P}_{u,v}(T^+_s < T_{\overline{S}_k})}+1.$$

Abbreviate $\widehat{\rho}_1(S_k)$ and $\widehat{\rho}_2(S_k)$ to $\widehat{\rho}_1$ and $\widehat{\rho}_2$ respectively. By Corollary 2.8, the denominator is at least $1-2|S_k|\frac{\widehat{\rho}_1}{\widehat{\rho}_2-(|S_k|-1)\widehat{\rho}_1}$. Therefore,

$$\mathbf{E}_{\omega,p_k}[T^+_{(s,\Box)}] \leq ((|S_k|-1)\widehat{\rho}_1+2B)\times\frac{\widehat{\rho}_2-(|S_k|-1)\widehat{\rho}_1}{\widehat{\rho}_2-(3|S_k|-1)\widehat{\rho}_1}+1 \leq 2|S_k|\widehat{\rho}_1+4B+1,$$

where the second inequality follows by Lemma 3.10 and (E3).

Eq. (34) implies that $\widehat{\rho}_1 < L\rho_1(S_k)$, and by Proposition 2.12, $\rho_1(S_k) \leq \max_{D\subset S_k}\frac{n_D}{R_D-1} \leq \frac{N}{a-1}$. The result follows. ∎

Define $\nu_k^{\Box}(s) = \nu_k((s,\Box))/\nu_k(S_k\times\{\Box\})$. This is the invariant measure of p_k conditioned on $S_k\times\{\Box\}$.

Proposition 3.16 *If* $|S_k|\geq 2$ *then* $\mathbf{P}\left(\left|\widetilde{F}^s_{n_k}-\nu_k^{\Box}(s))\right| > 2\varepsilon(1+\varepsilon)\nu_k^{\Box}(s))\right) \leq \frac{1}{2|S|}\times\frac{1}{N^\delta}$.

Proof. By Remarks 2.5, 2.6 Corollary 3.15, (A4), (A6) and (E2), for each $\omega\in E_1$,

$$\begin{aligned}\mathbf{P}_{\omega,p_k}&\left(|\overline{F}^{(s,\Box)}_{n_k}-\nu_k((s,\Box))| > \varepsilon\nu_k((s,\Box))\right)\\ &\leq 2\times 9\frac{\varepsilon^2}{n_k}\left(2\left(2|S|L\frac{N}{a-1}+4B+1\right)+1\right)\\ &\leq 2\times 72\frac{\varepsilon^2}{n_k}(|S|L\frac{N}{a-1}+B) \leq \frac{1}{2|S|}\times\frac{1}{N^\delta}. \qquad (45)\end{aligned}$$

By Lemma 3.11, (RZ3), and since $\nu_k(S_k\times\{\Box\})\geq 1/2$,

$$\left|\nu_k((s,\Box))-\nu_k^{\Box}(s)\right| \leq 2\nu_k(E_1\backslash(S_k\times\{\Box\}))\times\nu_k((s,\Box)) \leq \varepsilon\nu_k((s,\Box)).$$

Therefore

$$\left|\widetilde{F}^s_{n_k}-\nu_k((s,\Box))\right| \leq \varepsilon\nu_k((s,\Box)) \Rightarrow \left|\widetilde{F}^s_{n_k}-\nu_k^{\Box}(s))\right| \leq 2\varepsilon(1+\varepsilon)\nu_k^{\Box}(s)).$$

The result thus follows by (45). ∎

Corollary 3.17 *If* $|S_k|\geq 2$ *then*

$$\mathbf{P}\left(\left|\widetilde{F}^s_{n_k}-\nu^{\sigma^*}(s|S_k))\right| > 55\varepsilon L\nu^{\sigma^*}(s|S_k)\right) \leq \frac{1}{2|S|}\times\frac{1}{N^\delta}.$$

Proof. Recall that q_k is the invariant measure of q_k conditioned on S_k. On the other hand, the invariant measure of q conditioned on S_k is simply $\nu^{\sigma^*}(\cdot|S_k)$. By Lemma 3.9 and Proposition 3.8,

$$\left|\nu_k^{\Box}(s) - \nu^{\sigma^*}(\cdot|S_k)\right| \leq 18 \times 3\varepsilon L \nu^{\sigma^*}(\cdot|S_k),$$

The claim follows by Proposition 3.16 and since $L \geq 220$. ■

Step 7: The singleton case: $S_k = \{\overline{s}\}$.

If $N_{\overline{s}} < N^{1-\delta}$, then the first item in Theorem holds trivially. In addition, $\mathbf{q}_n = q^\sigma(\cdot|\mathbf{s}_n)$ for each $m_k^* \leq n < m_{k+1}^*$. We now assume $N_{\overline{s}} \geq N^{1-\delta}$.

By Lemma 3.6, $\mathbf{N}_0^k$ is at most $|\{n : (\mathbf{s}_n, \mathbf{t}_n) \neq (\overline{s}, \Box)\}|$. The next lemma is an analog of Lemma 3.11. Its proof is however significantly different.

Lemma 3.18 *One has* $\nu_k((\overline{s}, \Box)) \geq 1 - B(1+3\varepsilon)\frac{(a+1)^{|S|}}{N^{1-\delta}}$.

Proof. We first provide a lower bound for $v(\overline{s} \mid \overline{s})$. By Theorem 2.10

$$q(S\backslash\{\overline{s}\} \mid \overline{s}) \leq \frac{R_{S\backslash\{\overline{s}\}}}{N_{\overline{s}}^*} \leq \frac{(a+1)^{|S|}}{N^{1-\delta}}.$$

Using (30), this yields

$$v(S\backslash\{\overline{s}\} \mid \overline{s}) \leq (1+3\varepsilon)\frac{(a+1)^{|S|}}{N^{1-\delta}}.$$

Let E_1 be the ergodic set for p_k that contains $(\overline{s}, \Box)$. We apply (39) to $\overline{q} = p_k$, $\Omega = E_1$ and $C = \{(\overline{s}, \Box)\}$ to get

$$\frac{\nu_k(E_1\backslash(\overline{s}, \Box))}{\nu_k((\overline{s}, \Box))} \leq Bv(S\backslash\overline{s}|\overline{s}) \leq B(1+3\varepsilon)\frac{(a+1)^{|S|}}{N^{1-\delta}}.$$

■

The rest of the proof for the singleton case follows closely the proof for $|S_k| > 1$.

Corollary 3.19 *One has* $\mathbf{P}(\left|\widetilde{F}_{n_k}^{\overline{s}} - 1\right| > 2\varepsilon) \leq \frac{1}{2K} \times \frac{1}{N^\delta}$.

Proof. By Definition of p_k, $\sup_{t \in E_1} \mathbf{E}_{t,p_k}\left[T^+_{(\overline{s},\Box)}\right] \leq B+1$. Therefore, using Remark 2.5 to p_k, ε and $s = (\overline{s}, \Box)$,

$$\mathbf{P}\left(\left|\widetilde{F}_{n_k}^{\overline{s}} - \nu_k((\overline{s}, \Box))\right| > \varepsilon\nu_k((\overline{s}, \Box))\right) \leq 4 \times \frac{9(2B+3)}{n_k}.$$

Observe that by (S2)

$$\left|\widetilde{F}_k^{\overline{s}} - \nu_k((\overline{s},\square))\right| \le \frac{1}{N^{2\delta}} \Rightarrow \left|\widetilde{F}_k^{\overline{s}} - 1\right| \le \frac{1}{N^{\delta}}.$$

By Lemma 3.18, (S2) and (RZ4), $|\nu_k((\overline{s},\square)) - 1| \le \varepsilon$. The result follows. ■

Lemma 3.20 *One has* $\mathbf{E}\left[\mathbf{N}_0^k\right] \le \frac{BN^{\psi}}{K}$.

Proof. We follow the proof of Lemma 3.14. For each $\omega \in E_1$ one has, by Lemma 3.18 and (S4)

$$\begin{aligned}
\mathbf{E}_{\omega}\left[\mathbf{N}_0^k\right] &\le \mathbf{E}_{\omega}\left[T_{(\overline{s},\square)}\right] + \mathbf{E}_{(\overline{s},\square)}\left[\mathbf{N}_0^k\right] \\
&\le B + \mathbf{E}_{(\overline{s},\square)}\left[\mathbf{N}_0^k\right] \le B + 2\mathbf{E}_{\nu_k}\left[\mathbf{N}_0^k\right] \\
&\le B + 2n_k^*\nu_k(E_1\backslash(\overline{s},\square)) \\
&\le B + 2NB(1+3\varepsilon)\frac{(a+1)^{|S|}}{N^{1-\delta}} \\
&\le BN^{\psi}/K.
\end{aligned}$$

■

Step 8: Conclusion

We here conclude the proof of Theorem 3.4. Note first that $\mathbf{N}_0 = \sum_k \mathbf{N}_0^k$. Therefore, **G2** follows from Lemmas 3.14 and 3.20. Let now $s \in S_k$ with $\nu_s^{\sigma} \ge \frac{1}{N^{\delta}}$. Plainly, $N\overline{F}_N^s - n_k^*\widetilde{F}_{n_k}^s$ is the total number of visits to s that are not counted in $\widetilde{F}_{n_k}^s$: there are at most $\sum_k \mathbf{N}_0^k$ many of them.

Since $\mathbf{E}\left[\sum_k \mathbf{N}_0^k\right] \le BN^{\psi}$ one has by Markov inequality and (RZ5)

$$\mathbf{P}\left(\sum_k \mathbf{N}_0^k > \varepsilon\frac{N}{N^{\delta}}\right) \le \frac{BN^{\delta+\psi-1}}{\varepsilon} \le \frac{1}{2N^{\delta}}.$$

Therefore, using Corollaries 3.17 and 3.19, the probability that both inequalities $\sum_k \mathbf{N}_0^k \le \varepsilon N^{1-\delta}$ and $\left|\widetilde{F}_{n_k}^s - \nu^{\sigma^*}(s|S_k))\right| \le 55\varepsilon L\nu^{\sigma^*}(s|S_k)$ hold for every k and every $s \in S_k$, is at least $1 - \frac{1}{N^{\delta}}$. On this event, by (E1),

$$\left|\overline{F}_N^s - \nu_s^{\sigma^*}\right| \le 55\varepsilon L\nu_s^{\sigma^*} \le \eta\nu_s^{\sigma^*}.$$

This proves **G1**.

References

[1] Aldous D.J. and Fill J.A. (2002) Reversible Markov Chains and Random Walks on Graphs, Book in preparation; dratf available via homepage http://www.stat.berkeley.edu/users/aldous

[2] Alon N., Spencer J.H. and Erdos P. (2000) The Probabilistic Method, John Wiley and Sons

[3] Feller W. (1968) An Introduction to Probability Theory and its Applications, John Wiley and sons

[4] Jerrum M. and Sinclair A. (1989) Approximating the permanent. *SIAM Journal on Computing* , 18:1149–1178

[5] Lovasz L. and Kannan R. (1999) Faster mixing via average conductance. *Annual ACM Symposium on Theory of Computing* , 282–287

[6] Lovasz L. and Simonovits M. (1990) The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. *31st Annual Symposium on Foundations of Computer Science, Vol. I, II* , 346–354

[7] Solan E. and Vieille N. (2002), Perturbed Markov Chains, *preprint*