Electronic Journal of Probability

Vol. 10 (2005), Paper no. 36, pages 1221-1235.

Journal URL
http://www.math.washington.edu/~ejpecp/

A Martingale Proof of Dobrushin's Theorem for Non-Homogeneous Markov Chains [1]

S. Sethuraman and **S.R.S. Varadhan**

Department of Mathematics, Iowa State University
Ames, IA 50011, USA
Email: sethuram@iastate.edu

and

Courant Institute, New York University
New York, NY 10012, USA
Email: varadhan@cims.nyu.edu

Abstract. In 1956, Dobrushin proved an important central limit theorem for non-homogeneous Markov chains. In this note, a shorter and different proof elucidating more the assumptions is given through martingale approximation.

Keywords and phrases: non-homogeneous Markov, contraction coefficient, central limit theorem, martingale approximation

AMS Subject Classification (2000): Primary 60J10; secondary 60F05.

Submitted to EJP on March 21, 2005. Final version accepted on September 6, 2005.

[1]Research supported in part by NSF/DMS-0071504, NSA-H982300510041 and NSF/DMS-0104343.

1 Introduction and Results

Nearly fifty years ago, R. Dobrushin proved in his thesis [2] an important central limit theorem (CLT) for Markov chains in discrete time that are not necessarily homogeneous in time. Previously, Markov, Bernstein, Sapagov, and Linnik, among others, had considered the central limit question under various sufficient conditions. Roughly, the progression of results relaxed the state space structure from 2 states to an arbitrary set of states, and also the level of asymptotic degeneracy allowed for the transition probabilities of the chain.

After Dobrushin's work, some refinements and extensions of his CLT, some of which under more stringent assumptions, were proved by Statuljavicius [16] and Sarymsakov [13]. See also Hanen [6] in this regard. A corresponding invariance principle was also proved by Gudinas [4]. More general references on non-homogeneous Markov processes can be found in Isaacson and Madsen [7], Iosifescu [8], Iosifescu and Theodorescu [9], and Winkler [18].

We now define what is meant by "degeneracy." Although there are many measures of "degeneracy," the measure which turns out to be most useful to work with is that in terms of the contraction coefficient. This coefficient has appeared in early results concerning Markov chains, however, in his thesis, Dobrushin popularized its use, and developed many of its important properties. [See Seneta [14] for some history.]

Let $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ be a Borel space, and let $\pi = \pi(x, dy)$ be a Markov transition probability on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$. Define the contraction coefficient $\delta(\pi)$ of π as

$$\begin{aligned}
\delta(\pi) &= \sup_{x_1, x_2 \in \mathbf{X}} \|\pi(x_1, \cdot) - \pi(x_2, \cdot)\|_{\mathrm{Var}} \\
&= \sup_{\substack{x_1, x_2 \in \mathbf{X} \\ A \in \mathcal{B}(\mathbf{X})}} |\pi(x_1, A) - \pi(x_2, A)| \\
&= \frac{1}{2} \sup_{\substack{x_1, x_2 \in \mathbf{X} \\ \|f\|_{L^\infty} \le 1}} \left| \int f(y)[\pi(x_1, dy) - \pi(x_2, dy)] \right|.
\end{aligned}$$

Also, define the related coefficient $\alpha(\pi) = 1 - \delta(\pi)$.

Clearly, $0 \le \delta(\pi) \le 1$, and $\delta(\pi) = 0$ if and only if $\pi(x, dy)$ does not depend on x. It makes sense to call π "non-degenerate" if $0 \le \delta(\pi) < 1$. We use the standard convention and denote by $\mu\pi$ and πu the transformations induced by π on countably additive measures and bounded measurable functions respectively,

$$(\mu\pi)(A) = \int \mu(dx)\, \pi(x, A) \qquad \text{and} \qquad (\pi u)(x) = \int \pi(x, dy)\, u(y).$$

One can see that $\delta(\pi)$ has the following properties.

$$\delta(\pi) = \sup_{\substack{x_1, x_2 \in \mathbf{X} \\ u \in \mathcal{U}}} |(\pi u)(x_1) - (\pi u)(x_2)| \tag{1.1}$$

with $\mathcal{U} = \{u : \sup_{y_1,y_2} |u(y_1) - u(y_2)| \le 1\}$. It is the operator norm of π with respect to the Banach (semi-) norm $\mathrm{Osc}(u) = \sup_{x_1,x_2} |u(x_1) - u(x_2)|$, namely the oscillation of u. In particular, for any transition probabilities π_1, π_2 we have

$$\delta(\pi_1 \, \pi_2) \le \delta(\pi_1) \, \delta(\pi_2) \tag{1.2}$$

where $\pi_1\pi_2$ is the two-step transition probability $\pi_1\pi_2(x, \cdot) = \int \pi_1(x, dy)\pi_2(y, \cdot)$.

By a non-homogeneous Markov chain of length n on state space $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ corresponding to transition operators $\{\pi_{i,i+1} = \pi_{i,i+1}(x, dy) : 1 \le i \le n-1\}$ we mean the Markov process P on the product space $(\mathbf{X}^n, \ \mathcal{B}(\mathbf{X}^n))$,

$$P[X_{i+1} \in A | X_i = x] = \pi_{i,i+1}(x, A),$$

where $\{X_i : 1 \le i \le n\}$ are the canonical projections. In particular, under the initial distribution $X_1 \sim \mu$, the distribution at time $k \ge 1$ is $\mu\pi_{1,2}\pi_{2,3} \cdots \pi_{k-1,k}$. For $i < j$ we will define

$$\pi_{i,j} = \pi_{i,i+1}\pi_{i+1,i+2} \cdots \pi_{j-1,j}.$$

We denote by $E[Z]$ and $V(Z)$ the expectation and variance of the random variable Z with respect to P.

Dobrushin's theorem concerns the fluctuations of an array of non-homogeneous Markov chains. For each $n \ge 1$, let $\{X_i^{(n)} : 1 \le i \le n\}$ be n observations of a non-homogeneous Markov chain on $\mathbf{X}$ with transition matrices $\{\pi_{i,i+1}^{(n)} = \pi_{i,i+1}^{(n)}(x, dy) : 1 \le i \le n-1\}$ and initial distribution $\mu^{(n)}$. Let also

$$\alpha_n = \min_{1 \le i \le n-1} \alpha\big(\pi_{i,i+1}^{(n)}\big).$$

In addition, let $\{f_i^{(n)} : 1 \le i \le n\}$ be real valued functions on $\mathbf{X}$. Define, for $n \ge 1$, the sum

$$S_n = \sum_{i=1}^{n} f_i^{(n)}(X_i^{(n)}).$$

Theorem 1.1 *Suppose that for some finite constants C_n,*

$$\sup_{1 \le i \le n} \sup_{x \in \mathbf{X}} |f_i^{(n)}(x)| \le C_n.$$

Then, if

$$\lim_{n \to \infty} C_n^2 \alpha_n^{-3} \left[\sum_{i=1}^{n} V\big(f_i^{(n)}(X_i^{(n)})\big) \right]^{-1} = 0, \tag{1.3}$$

we have the standard Normal convergence

$$\frac{S_n - E[S_n]}{\sqrt{V(S_n)}} \ \Rightarrow \ \mathrm{N}(0, 1). \tag{1.4}$$

Also, there is an example where the result is not true if condition (1.3) is not met.

In [2], Dobrushin also states the direct corollary which simplifies some of the assumptions.

Corollary 1.1 *When the functions are uniformly bounded, i.e.* $\sup_n C_n = C < \infty$ *and the variances are bounded below, i.e.* $V(f_i^{(n)}(X_i^{(n)})) \geq c > 0$*, for all* $1 \leq i \leq n$ *and* $n \geq 1$*, then we have the convergence (1.4) provided*

$$\lim_{n\to\infty} n^{1/3}\alpha_n = \infty.$$

We remark that in [2] (e.g. Theorems 3, 8) there are also results where the boundedness condition on $f_i^{(n)}$ is replaced by integrability conditions. As these results follow from truncation methods and Theorem 1.1 for bounded variables, we only consider Dobrushin's theorem in the bounded case.

Also, for the ease of the reader, and to be complete, we will discuss in the next section an example, given in [2] and due to Bernstein and Dobrushin, of how the convergence (1.4) may fail when the condition (1.3) is not satisfied.

We now consider Dobrushin's methods. The techniques used in [2] to prove the above results fall under the general heading of the "blocking method." The condition (1.3) ensures that well-separated blocks of observations may be approximated by independent versions with small error. Indeed, in many remarkable steps, Dobrushin exploits the Markov property and several contraction coefficient properties, which he himself derives, to deduce error bounds sufficient to apply CLT's for independent variables. However, in [2], it is difficult to see, even at the technical level, why condition (1.3) is natural.

The aim of this note is to provide a different, shorter proof of Theorem 1.1 which explains more why condition (1.3) appears in the result. The methods are through martingale approximations and martingale CLT's. These methods go back at least to Gordin [3] in the context of homogeneous processes, and have been used by others in mostly "stationary" situations (e.g. Kifer [10], Kipnis and Varadhan [11], Pinsky [12], and Wu and Woodroofe [19]). The approximation with respect to the non-homogeneous setting of Theorem 1.1 makes use of three ingredients: (1) negligibility estimates for individual components, (2) a law of large numbers (LLN) for conditional variances, and (3) lower bounds for the variance $V(S_n)$. Negligibility bounds and a LLN are well known requirements for martingale CLT's (cf. Hall-Heyde [5, ch. 3]), and in fact, as will be seen, the sufficiency of condition (1.3) is transparent in the proofs of these two components (Lemma 3.2, and Lemmas 3.3 and 3.4). The variance lower bounds which we will use (Proposition 3.2) were as well derived by Dobrushin in his proof. However, using some martingale and spectral tools, we give a more direct argument for a better estimate.

We note also, with this martingale approximation, that an invariance principle for the partial sums holds through standard martingale propositions, Hall-Heyde [5], among other results. In fact, from the martingale invariance principle, it should be possible to derive Gudynas's theorems [4] although this is not done here.

We now explain the structure of the article. In section 2, we give the Bernstein-Dobrushin example of a Markov chain with anomalous behavior. In section 3, we state a martingale

CLT and prove Theorem 1.1 assuming a lower bound on the variance $V(S_n)$. Last, in section 4, we prove this variance estimate.

2 Bernstein-Dobrushin Example

Here, we summarize the example in Dobrushin's thesis, attributed to Bernstein, which shows that condition (1.3) is sharp.

Example 2.1 Let $\mathbf{X} = \{1, 2\}$, and consider the 2×2 transition matrices on $\mathbf{X}$,

$$Q(p) = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

for $0 \leq p \leq 1$. The contraction coefficient $\delta(Q(p))$ of $Q(p)$ is $|1-2p|$. Note that $\delta(Q(p)) = \delta(Q(1-p))$. The invariant measures for all the $Q(p)$ are the same $\mu(1) = \mu(2) = \frac{1}{2}$. We will be looking at $Q(p)$ for p close to 0 or 1 and the special case of $p = \frac{1}{2}$. However, when p is small, the homogeneous chains behave very differently under $Q(p)$ and $Q(1-p)$. More specifically, when p is small there are very few switches between the two states whereas when $1-p$ is small it switches most of the time. In fact, this behavior can be made more precise (see Dobrushin [1], or from direct computation). Let $T_n = \sum_{i=1}^{n} \mathbf{1}_{\{1\}}(X_i)$ count the number of visits to state 1 in the first n steps.

Case A. Consider the homogeneous chain under $Q(p)$ with $p = \frac{1}{n}$ and initial distribution $\mu(1) = \mu(2) = \frac{1}{2}$. Then,

$$\frac{T_n}{n} \Rightarrow G \quad \text{and} \quad \lim_{n\to\infty} n^{-2}\, V(T_n) \;=\; V_A \tag{2.1}$$

where $0 < V_A < \infty$ and G is a non-degenerate distribution supported on $[0, 1]$.

Case B. Consider the homogeneous chain run under $Q(p)$ with $p = 1 - \frac{1}{n}$ and initial distribution $\mu(1) = \mu(2) = \frac{1}{2}$. Then,

$$T_n - \frac{n}{2} \Rightarrow F \quad \text{and} \quad \lim_{n\to\infty} V(T_n) \;=\; V_B \tag{2.2}$$

where $0 < V_B < \infty$ and F is a non-degenerate distribution.

Let a sequence $\alpha_n \to 0$ with $\alpha_n \geq n^{-\frac{1}{3}}$ be given . To construct the anomalous Markov chain, it will be helpful to split the time horizon $[1, 2, \ldots, n]$ into roughly $n\alpha_n$ blocks of size α_n^{-1}. We interpose a $Q(\frac{1}{2})$ between any two blocks that has the effect of making the blocks independent of each other. More precisely let $k_i^{(n)} = i[\alpha_n^{-1}]$ for $1 \leq i \leq m_n$ where $m_n = [n/[\alpha_n^{-1}]]$. Also, define $k_0^{(n)} = 0$, and $k_{m_n+1}^{(n)} = n$.

Define now, for $1 \le i \le n$,

$$\pi^{(n)}_{i,i+1} = \begin{cases} Q(\alpha_n) & \text{for } i = 1, 2, \ldots, k^{(n)}_1 - 1 \\ Q(\frac{1}{2}) & \text{for } i = k^{(n)}_1, k^{(n)}_2, \ldots, k^{(n)}_{m_n} \\ Q(1-\alpha_n) & \text{for all other } i. \end{cases}$$

Consider the non-homogeneous chain with respect to $\{\pi^{(n)}_{i,i+1} : 1 \le i \le n-1\}$ starting from equilibrium $\mu^{(n)}(0) = \mu^{(n)}(1) = \frac{1}{2}$. ¿From the definition of the chain, one observes, as $Q(\frac{1}{2})$ does not distinguish between states, that the process in time horizons $\{(k^{(n)}_i + 1, k^{(n)}_{i+1}) : 0 \le i \le m_n\}$ are mutually independent. For the first time segment 1 to $k^{(n)}_1$, the chain is in regime A, while for the other segments, the chain is in case B.

Once again, let us concentrate on the number of visits to state 1. Denote by $T^{(n)} = \sum_{i=1}^n \mathbf{1}_{\{1\}}(X^{(n)}_i)$ and $T^{(n)}(k,l) = \sum_{i=k}^l \mathbf{1}_{\{1\}}(X^{(n)}_i)$ the counts in the first n steps and in steps k to l respectively. It follows from the discussion of independence above that

$$T^{(n)} \;=\; \sum_{i=0}^{m_n} T^{(n)}(k^{(n)}_i + 1, k^{(n)}_{i+1})$$

is the sum of independent sub-counts where, additionally, the sub-counts for $1 \le i \le m_n - 1$ are identically distributed, the last sub-count perhaps being shorter. Also, as the initial distribution is invariant, we have $V(\mathbf{1}_{\{1\}}(X^{(n)}_i)) = 1/4$ for all i and n. Then, in the notation of Corollary 1.1, $C = 1$ and $c = 1/4$.

¿From (2.1), we have that

$$V(T^{(n)}(1, k^{(n)}_1)) \;\sim\; \alpha_n^{-2} V_A \quad \text{as } n \uparrow \infty.$$

Also, from (2.2) and independence of m_n sub-counts, we have that

$$V(T^{(n)}(k^{(n)}_1 + 1, n)) \;\sim\; n\alpha_n V_B \quad \text{as } n \uparrow \infty.$$

¿From these calculations, we see if $n^{1/3}\alpha_n \to \infty$, then $\alpha_n^{-2} << n\alpha_n$, and so the major contribution to $T^{(n)}$ is from $T^{(n)}(k^{(n)}_1 + 1, n)$. However, since this last count is (virtually) the sum of m_n i.i.d. sub-counts, we have that $T^{(n)}$, properly normalized, converges to $\mathrm{N}(0,1)$, as predicted by Dobrushin's Theorem 1.1.

On the other hand, if $\alpha_n = n^{-1/3}$, we have $\alpha_n^{-2} = n\alpha_n$, and count $T^{(n)}(1, k^{(n)}_1)$, independent of $T^{(n)}(k^{(n)}_1, n)$, also contributes to the sum $T^{(n)}$. After centering and scaling, then, $T^{(n)}$ approaches the convolution of a non-trivial non-normal distribution and a normal distribution, and therefore is not Gaussian.

3 Proof of Theorem 1.1

The CLT for martingale differences is a standard tool. We quote the following form of the result implied by Corollary 3.1 in Hall and Heyde [5].

Proposition 3.1 *For each $n \geq 1$, let $\{(W_i^{(n)}, \mathcal{G}_i^{(n)}) : 0 \leq i \leq n\}$ be a martingale relative to the nested family $\mathcal{G}_i^{(n)} \subset \mathcal{G}_{i+1}^{(n)}$ with $W_0^{(n)} = 0$. Let $\xi_i^{(n)} = W_i^{(n)} - W_{i-1}^{(n)}$ be their differences. Suppose that*

$$\begin{array}{rcll} \max_{1\leq i\leq n} \|\xi_i^{(n)}\|_{L^\infty} & \to & 0 & \text{and} \\ \sum_{i=1}^n E[(\xi_i^{(n)})^2|\mathcal{G}_{i-1}^{(n)}] & \to & 1 & \text{in } L^2. \end{array}$$

Then,

$$W_n^{(n)} \Rightarrow \mathrm{N}(0,1).$$

The first and second limit conditions are the so called "negligibility" assumption on the sequence, and LLN for conditional variances mentioned in the introduction.

Consider now the non-homogeneous setting of Theorem 1.1. To simplify notation, we will assume throughout that the functions $\{f_i^{(n)}\}$ are mean-zero, $E[f_i^{(n)}(X_i^{(n)})] = 0$ for $1 \leq i \leq n$ and $n \geq 1$. Define

$$Z_k^{(n)} = \sum_{i=k}^n E[f_i^{(n)}(X_i^{(n)})|X_k^{(n)}]$$

so that

$$Z_k^{(n)} = \begin{cases} f_k^{(n)}(X_k^{(n)}) + \sum_{i=k+1}^n E[f_i^{(n)}(X_i^{(n)})|X_k^{(n)}] & \text{for} \quad 1 \leq k \leq n-1 \\ f_n^{(n)}(X_n^{(n)}) & \text{for} \quad k = n. \end{cases} \tag{3.1}$$

Remark 3.1 Before going further, we remark that the sequence $\{Z_k^{(n)}\}$ can be thought of as a type of "Poisson-resolvent" sequence often seen in martingale approximations Namely, when the array $\{X_i^{(n)}\}$ is formed from the sequence $\{X_i\}$, $f_i^{(n)} = f$ for all i and n, and the chain is homogeneous, $P_n = P$ for all n, then indeed $Z_k^{(n)}$ reduces to $Z_k^{(n)} = f(X_k) + \sum_{i=1}^{n-k}(P^i f)(X_k)$ which approximates the Poisson-resolvent solution $\sum_{i=0}^\infty (P^i f)(X_k) = [(I - P)^{-1}f](X_k)$ usually used to prove the CLT in this case (cf. p. 145-6 Varadhan [17]).

Returning to the full non-homogeneous setting of Theorem 1.1, by rearranging terms in (3.1), we obtain for $1 \leq k \leq n-1$ that

$$f_k^{(n)}(X_k^{(n)}) \quad = \quad Z_k^{(n)} - E[Z_{k+1}^{(n)}|X_k^{(n)}] \tag{3.2}$$

which for $2 \leq k \leq n-1$ further equals $[Z_k^{(n)} - E[Z_k^{(n)}|X_{k-1}^{(n)}]] + [E[Z_k^{(n)}|X_{k-1}^{(n)}] - E[Z_{k+1}^{(n)}|X_k^{(n)}]]$. Then, we have the decomposition,

$$\begin{aligned} S_n &= \sum_{k=1}^n f_k^{(n)}(X_k^{(n)}) \\ &= \sum_{k=2}^n [Z_k^{(n)} - E[Z_k^{(n)}|X_{k-1}^{(n)}]] + Z_1^{(n)} \end{aligned} \tag{3.3}$$

and so in particular $V(S_n) = \sum_{k=2}^n V(Z_k^{(n)} - E[Z_k^{(n)}|X_{k-1}^{(n)}]) + V(Z_1^{(n)})$. Let us now define the scaled differences

$$\xi_k^{(n)} = \frac{1}{\sqrt{V(S_n)}}\left[Z_k^{(n)} - E[Z_k^{(n)}|X_{k-1}^{(n)}]\right] \tag{3.4}$$

and the martingale $M_k^{(n)} = \sum_{l=2}^k \xi_l^{(n)}$ with respect to $\mathcal{F}_k^{(n)} = \sigma\{X_l^{(n)} : 1 \le l \le k\}$ for $2 \le k \le n$. The plan to obtain Theorem 1.1 will now be to approximate $S_n/\sqrt{V(S_n)}$ by $M_n^{(n)}$ and use Proposition 3.1. Condition (1.3) will be a sufficent condition for "negligibility" (Lemma 3.2) and "LLN" (Lemmas 3.3 and 3.4) with regard to Proposition 3.1.

Lemma 3.1 *We have, for* $1 \le i < j \le n$,

$$\left\|\pi_{i,j} f_j^{(n)}\right\|_{L^\infty} \le 2C_n(1-\alpha_n)^{j-i} \quad \text{and} \quad \mathrm{Osc}\left(\pi_{i,j}(f_j^{(n)})^2\right) \le 2C_n^2(1-\alpha_n)^{j-i}$$

and, for $1 \le l < i < j \le n$,

$$\mathrm{Osc}\left(\pi_{l,i}(f_i^{(n)}\pi_{i,j} f_j^{(n)})\right) \le 6C_n^2\ (1-\alpha_n)^{i-l}(1-\alpha_n)^{j-i}.$$

Proof. As $\|f_j^{(n)}\|_{L_\infty} \le C_n$ its oscillation $\mathrm{Osc}(f_j^{(n)}) \le 2C_n$. From definition of $\delta(\cdot)$ (cf. (1.1)) and (1.2),

$$\mathrm{Osc}(\pi_{i,j} f_j^{(n)}) \le \mathrm{Osc}(f_j^{(n)})\ \delta(\pi_{i,j}) \le 2C_n(1-\alpha_n)^{j-i}.$$

Because $E[(\pi_{i,j} f_j^{(n)})(X_i^{(n)})] = E[f_j^{(n)}(X_j^{(n)})] = 0$, the first bound follows as

$$\|\pi_{i,j} f_j^{(n)}\|_{L_\infty} \le \mathrm{Osc}(\pi_{i,j} f_j^{(n)}) \le 2C_n(1-\alpha_n)^{j-i}.$$

The second bound is analogous. For the third bound, write

$$\begin{aligned}
\mathrm{Osc}\left(\pi_{l,i}(f_i^{(n)}\pi_{i,j} f_j^{(n)})\right) &\le (1-\alpha_n)^{i-l}\ \mathrm{Osc}\left(f_i^{(n)}\pi_{i,j} f_j^{(n)}\right) \\
&\le (1-\alpha_n)^{i-l}\Big[\mathrm{Osc}(f_i^{(n)})\ \|\pi_{i,j} f_j^{(n)}\|_{L^\infty} \\
&\qquad\qquad + \|f_i^{(n)}\|_{L^\infty}\ \mathrm{Osc}(\pi_{i,j} f_j^{(n)})\Big] \\
&\le 6C_n^2\ (1-\alpha_n)^{i-l}(1-\alpha_n)^{j-i}.
\end{aligned}$$

□

We now state a lower bound for the variance proved in the next section. For comparison, we remark that in [2] the bound $V(S_n) \ge (\alpha_n/8)\sum_{i=1}^n V(f_i^{(n)}(X_i^{(n)}))$ is given (see also section 1.2.2 [9]).

Proposition 3.2 *For $n \geq 1$,*

$$V(S_n) \ \geq \ \frac{\alpha_n}{4} \sum_{i=1}^{n} V\big(f_i^{(n)}(X_i^{(n)})\big). \tag{3.5}$$

The next estimate shows that the asymptotics of $S_n/\sqrt{V(S_n)}$ depend only on the martingale approximant $M_n^{(n)}$, and that the differences $\xi_k^{(n)}$ are negligible.

Lemma 3.2 *Under condition (1.3), we have that*

$$\lim_{n\to\infty} \sup_{1\leq k\leq n} \frac{\|Z_k^{(n)}\|_{L^\infty}}{\sqrt{V(S_n)}} = 0.$$

Proof. By Lemma 3.1,

$$\|Z_k^{(n)}\|_{L^\infty} \ \leq \ \sum_{i=k}^{n} \|E[f_i^{(n)}(X_i^{(n)})|X_k^{(n)}]\|_{L^\infty} \ \leq \ 2C_n \sum_{i=k}^{n} (1-\alpha_n)^{i-k} \ \leq \ 2C_n\alpha_n^{-1}.$$

Then, by Proposition 3.2,

$$\sup_{1\leq k\leq n} \frac{\|Z_k^{(n)}\|_{L^\infty}}{\sqrt{V(S_n)}} \ \leq \ 4C_n\left(\alpha_n^3 \sum_{i=1}^{n} V(f_i^{(n)}(X_i^{(n)}))\right)^{-1/2}$$

which in turn by (1.3) is $o(1)$. □

The next two lemmas help prove the LLN part of Proposition 3.1 for array $\{M_k^{(n)}\}$. By the oscillation of a random variable η we mean $\mathrm{Osc}(\eta) = \sup_{\omega,\omega'} |\eta(\omega) - \eta(\omega')|$.

Lemma 3.3 *Let $\{Y_l^{(n)} : 1 \leq l \leq n\}$ and $\{\mathcal{G}_l^{(n)} : 1 \leq l \leq n\}$, for $n \geq 1$, be respectively an array of non-negative variables and σ-fields such that $\sigma\{Y_1^{(n)}, \ldots, Y_l^{(n)}\} \subset \mathcal{G}_l^{(n)}$. Suppose that*

$$\lim_{n\to\infty} E\big[\sum_{l=1}^{n} Y_l^{(n)}\big] = 1 \quad \text{and} \quad \sup_{1\leq i\leq n} \|Y_i^{(n)}\|_{L^\infty} \leq \epsilon_n$$

where $\lim_{n\to\infty} \epsilon_n = 0$. In addition, assume

$$\lim_{n\to\infty} \sup_{1\leq l\leq n-1} \mathrm{Osc}\left(E\big[\sum_{j=l+1}^{n} Y_j^{(n)}|\mathcal{G}_l^{(n)}\big]\right) = 0.$$

Then,

$$\lim_{n\to\infty} \sum_{l=1}^{n} Y_l^{(n)} = 1 \quad \text{in } L^2.$$

Proof. Write

$$E\big[(\sum_{l=1}^{n} Y_l^{(n)})^2\big] = \sum_{l=1}^{n} E\big[(Y_l^{(n)})^2\big] + 2\sum_{l=1}^{n-1} E\big[Y_l^{(n)}(\sum_{j=l+1}^{n} Y_j^{(n)})\big].$$

The first sum on the right-hand side is bounded as follows. ¿From non-negativity,

$$\sum_{l=1}^{n} E\big[(Y_l^{(n)})^2\big] \ \le\ \epsilon_n \sum_{l=1}^{n} E\big[Y_l^{(n)}\big] \ =\ \epsilon_n \cdot (1 + o(1)) \ \to 0 \ \text{ as } \quad n \uparrow \infty.$$

For the second sum, write

$$\sum_{l=1}^{n-1} E\big[Y_l^{(n)}\big(\sum_{j=l+1}^{n} Y_j^{(n)}\big)\big] = \sum_{l=1}^{n-1} E\big[Y_l^{(n)} E\big[\sum_{j=l+1}^{n} Y_j^{(n)} | \mathcal{G}_l^{(n)}\big]\big].$$

¿From the oscillation assumption, we have that

$$\sup_{1\le l\le n-1} \sup_{\omega} \big|E\big[\sum_{j=l+1}^{n} Y_j^{(n)} | \mathcal{G}_l^{(n)}\big](\omega) - E\big[\sum_{j=l+1}^{n} Y_j^{(n)}\big]\big| = o(1).$$

Therefore,

$$\begin{aligned} 2\sum_{l=1}^{n-1} E\big[Y_l^{(n)}\big(\sum_{j=l+1}^{n} Y_j^{(n)}\big)\big] &= 2\sum_{l=1}^{n-1} E\big[Y_l^{(n)}\big]E\big[\sum_{j=l+1}^{n} Y_j^{(n)}\big] + o(1) \cdot \sum_{l=1}^{n-1} E\big[Y_l^{(n)}\big] \\ &= \big(\sum_{l=1}^{n} E\big[Y_l^{(n)}\big]\big)^2 - \sum_{l=1}^{n} E\big[(Y_l^{(n)})^2\big] + o(1) \\ &= 1 + o(1) \end{aligned}$$

finishing the proof. □

To apply later this result to $v_j^{(n)} = E[(\xi_j^{(n)})^2 | \mathcal{F}_{j-1}^{(n)}]$ measureable with respect to $\mathcal{G}_j^{(n)} = \mathcal{F}_{j-1}^{(n)}$ for $2 \le j \le n$ we will need the following oscillation estimate.

Lemma 3.4 *Under condition (1.3), we have*

$$\sup_{2\le l\le n-1} \mathrm{Osc}\bigg(E\big[\sum_{j=l+1}^{n} v_j^{(n)} | \mathcal{F}_{l-1}^{(n)}\big](\omega)\bigg) \ =\ o(1).$$

Proof. From the martingale property, $E[\xi_r^{(n)}\xi_s^{(n)}|\mathcal{F}_u^{(n)}] = 0$ for $r > s > u$, (3.4) and (3.2), we have

$$
\begin{aligned}
E\big[\sum_{j=l+1}^{n} v_j^{(n)}|\mathcal{F}_{l-1}^{(n)}\big] &= E\big[\sum_{j=l+1}^{n}\big(\xi_j^{(n)}\big)^2\,|\mathcal{F}_{l-1}^{(n)}\big]\\
&= E\Big[\Big(\sum_{j=l+1}^{n}\xi_j^{(n)}\Big)^2\,|X_{l-1}^{(n)}\Big]\\
&= V(S_n)^{-1}E\big[\,\big(\sum_{j=l+1}^{n} f_j^{(n)}(X_j^{(n)}) - E\big[Z_{l+1}^{(n)}|X_l^{(n)}\big]\big)^2\,|X_{l-1}^{(n)}\big]\\
&= V(S_n)^{-1}E\big[\,\big(\sum_{j=l+1}^{n} f_j^{(n)}(X_j^{(n)})\big)^2\,|X_{l-1}^{(n)}\big]\\
&\qquad -V(S_n)^{-1}E\big[E\big[Z_{l+1}^{(n)}|X_l^{(n)}\big]^2|X_{l-1}^{(n)}\big]. \qquad (3.6)
\end{aligned}
$$

By Lemma 3.2, the last term in (3.6) is bounded $\sup_{2\le l\le n-1} V(S_n)^{-1}\|Z_{l+1}^{(n)}\|_{L^\infty}^2 = o(1)$, and so its oscillation is also uniformly $o(1)$.

To estimate oscillation of the first term on right-side of (3.6), we write

$$
\begin{aligned}
&\mathrm{Osc}\Big(V(S_n)^{-1}E\big[\big(\sum_{j=l+1}^{n} f_j^{(n)}(X_j^{(n)})\big)^2|X_{l-1}^{(n)}\big]\Big) \qquad (3.7)\\
&\qquad\le\ V(S_n)^{-1}\sum_{l+1\le j,m\le n}\mathrm{Osc}\Big(E\big[f_j^{(n)}(X_j^{(n)})f_m^{(n)}(X_m^{(n)})|X_{l-1}^{(n)}\big]\Big).
\end{aligned}
$$

But, for $l+1 \le j \le m \le n$, we have from Lemma 3.1 that

$$
\mathrm{Osc}\Big(E\big[f_j^{(n)}(X_j^{(n)})f_m^{(n)}(X_m^{(n)})|X_{l-1}^{(n)}\big]\Big)\ \le\ 6C_n^2\,(1-\alpha_n)^{j-l+1}(1-\alpha_n)^{m-j}.
$$

Then, (3.7) is bounded, uniformly in l, on order $V(S_n)^{-1}C_n^2\alpha_n^{-2}$ which from Proposition 3.2 and (1.3) is $o(1)$. □

Proof of Theorem 1.1. From Lemma 3.2, we need only show that $M_n^{(n)}/\sqrt{V(S_n)} \Rightarrow$ N$(0,1)$. This will follow from martingale convergence (Proposition 3.1) as soon as we show (1) $\sup_{2\le k\le n}\|\xi_k^{(n)}\|_{L^\infty} \to 0$ and (2) $\sum_{k=2}^{n} E[(\xi_k^{(n)})^2|\mathcal{F}_{k-1}^{(n)}] \to 1$. However, (1) follows from the negligibility estimate Lemma 3.2, and (2) from LLN Lemmas 3.3 and 3.4 since "negligibility" (1) holds and $\sum_{k=2}^{n} E[(\xi_k^{(n)})^2] = 1+o(1)$ (from variance decomposition after (3.3) and Lemma 3.2). □

4 Proof of Variance Lower Bound

Let λ be a probability measure on $\mathbf{X}\times\mathbf{X}$ with marginals α and β respectively. Let $\pi(x_1, dx_2)$ and $\widehat{\pi}(x_2, dx_1)$ be the corresponding transition probabilities in the two directions so that $\alpha\pi = \beta$ and $\beta\widehat{\pi} = \alpha$.

Lemma 4.1 *Let $f(x_1)$ and $g(x_2)$ be square integrable with respect to α and β respectively. If*

$$\int f(x_1)\alpha(dx_1) = \int g(x_2)\beta(dx_2) = 0$$

then,

$$\left|\int f(x_1)g(x_2)\lambda(dx_1, dx_2)\right| \leq \sqrt{\delta(\pi)}\,\|f\|_{L_2(\alpha)}\|g\|_{L_2(\beta)}.$$

Proof. Let us construct a measure on $\mathbf{X}\times\mathbf{X}\times\mathbf{X}$ by starting with λ on $\mathbf{X}\times\mathbf{X}$ and using reversed $\widehat{\pi}(x_2, dx_3)$ to go from x_2 to x_3. The transition probability from x_1 to x_3 defined by

$$Q(x_1, A) = \int \pi(x_1, dx_2)\widehat{\pi}(x_2, A)$$

satisfies $\delta(Q) \leq \delta(\pi)\delta(\widehat{\pi}) \leq \delta(\pi)$. Moreover $\alpha Q = \alpha$ and the operator Q is self-adjoint and bounded with norm 1 on $L_2(\alpha)$. Then, if f is a bounded function with $\int f(x)\alpha(dx) = 0$ (and so $E_\alpha[Q^n f] = 0$), we have for $n \geq 1$,

$$\|Q^n f\|_{L_2(\alpha)} \leq \|Q^n f\|_{L_\infty} \leq (\delta(Q))^n \mathrm{Osc}(f). \tag{4.1}$$

Hence, as bounded functions are dense, on the subspace of functions, $M = \{f \in L_2(\alpha) : \int f(x)\alpha(dx) = 0\}$, the top of the spectrum of Q is less than $\delta(Q)$ and so $\|Q\|_{L_2(\alpha,M)} \leq \delta(Q)$. Indeed, suppose the spectral radius of Q on M is larger than $\delta(Q)+\epsilon$ for $\epsilon > 0$, and $f \in M$ is a non-trivial bounded function whose spectral decomposition is with respect to spectral values larger than $\delta(Q)+\epsilon$. Then, $\|Q^n f\|_{L^2(\alpha)} \geq \|f\|_{L^2(\alpha)}(\delta(Q)+\epsilon)^n$ which contradicts the bound (4.1) when $n \uparrow \infty$. [cf. Thm. 2.10 [15] for a proof in discrete space settings.]

Then,

$$\|\widehat{\pi} f\|^2_{L_2(\beta)} = \langle \pi\widehat{\pi} f, f\rangle_{L_2(\alpha)} = \langle Qf, f\rangle_{L_2(\alpha)} \leq \|Q\|_{L_2(\alpha,M)}\|f\|^2_{L_2(\alpha)} \leq \delta(Q)\|f\|^2_{L_2(\alpha)}.$$

Finally,

$$\left|\int f(x_1)g(x_2)\lambda(dx_1, dx_2)\right| = |\langle \widehat{\pi} f, g\rangle_{L_2(\beta)}| \leq \sqrt{\delta(\pi)}\,\|f\|_{L_2(\alpha)}\|g\|_{L_2(\beta)}.$$

□

Lemma 4.2 *Let $f(x_1)$ and $g(x_2)$ be square integrable with respect to α and β respectively. Then,*

$$E\big[\big(f(x_1)-g(x_2)\big)^2\big] \ \geq \ (1-\delta(\pi))\ V\big(f(x_1)\big)$$

as well as

$$E\big[\big(f(x_1)-g(x_2)\big)^2\big] \ \geq \ (1-\delta(\pi))\ V\big(g(x_2)\big).$$

Proof. To get lower bounds, we can assume without loss of generality that f and g have mean 0 with respect to α and β respectively. Then by Lemma 4.1

$$\begin{aligned}
E\big[\big(f(x_1)-g(x_2)\big)^2\big] &= E\big[[f(x_1)]^2\big]+E\big[[g(x_2)]^2\big]-2E\big[f(x_1)g(x_2)\big]\\
&\geq E\big[[f(x_1)]^2\big]+E\big[[g(x_2)]^2\big]-2\sqrt{\delta(\pi)}\,\|f\|_{L_2(\alpha)}\|g\|_{L_2(\beta)}\\
&\geq (1-\delta(\pi))\|f\|^2_{L_2(\alpha)}.
\end{aligned}$$

The proof of the second half is identical. □

Proof of Proposition 3.2. Applying Lemma 4.2 to the Markov pairs $\{(X_k^{(n)},X_{k+1}^{(n)}):1\leq k\leq n-1\}$ with $f(X_k^{(n)})=E[Z_{k+1}^{(n)}|X_k^{(n)}]$ and $g(X_{k+1}^{(n)})=Z_{k+1}^{(n)}$, we get

$$E\big[\big(Z_{k+1}^{(n)}-E[Z_{k+1}^{(n)}|X_k^{(n)}]\big)^2\big]\geq \alpha_n E\big[\big(Z_{k+1}^{(n)}\big)^2\big].$$

On the other hand from (3.2), for $1\leq k\leq n-1$, we have

$$\begin{aligned}
V(f_k^{(n)}(X_k^{(n)})) &\leq E\big[\big(f_k^{(n)}(X_k^{(n)})\big)^2\big]\\
&\leq 2E\big[\big(Z_k^{(n)}\big)^2\big]+2E\big[\big(E\big[Z_{k+1}^{(n)}|X_k^{(n)}\big]\big)^2\big]\\
&\leq 2E\big[\big(Z_k^{(n)}\big)^2\big]+2E\big[\big(Z_{k+1}^{(n)}\big)^2\big].
\end{aligned}$$

Summing over k, and noting $f_n^{(n)}(X_n^{(n)})=Z_n^{(n)}$ and variance decomposition after (3.3),

$$\begin{aligned}
\sum_{k=1}^{n}V(f_k^{(n)}(X_k^{(n)})) &\leq 4\sum_{k=1}^{n}E\big[\big(Z_k^{(n)}\big)^2\big]\\
&\leq \frac{4}{\alpha_n}\Bigg[\sum_{k=1}^{n-1}E\big[\big(Z_{k+1}^{(n)}-E[Z_{k+1}^{(n)}|X_k^{(n)}]\big)^2\big]+E[(Z_1^{(n)})^2]\Bigg]=\frac{4}{\alpha_n}V(S_n).
\end{aligned}$$

□

Acknowledgement. We would like to thank the referees for their comments.

References

[1] Dobrushin, R. (1953) Limit theorems for Markov chains with two states. (Russian) *Izv. Adad. Nauk SSSR* **17:4** 291-330.

[2] Dobrushin, R. (1956) Central limit theorems for non-stationary Markov chains I,II. *Theory of Probab. and its Appl.* **1** 65-80, 329-383.

[3] Gordin, M.I. (1969) The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10** 1174-1176.

[4] Gudynas, P. (1977) An invariance principle for inhomogeneous Markov chains. *Lithuanian Math. J.* **17** 184-192.

[5] Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Its Application.* Academic Press, New York.

[6] Hanen, A. (1963) Théorèmes limites pour une suite de châines de Markov. *Ann. Inst. H. Poincaré* **18** 197-301.

[7] Isaacson, D.L., and Madsen, R.W. (1976) *Markov Chains. Theory and Applications.* John Wiley and Sons, New York.

[8] Iosifescu, M. (1980) *Finite Markov Processes and Their Applications.* John Wiley and Sons, New York.

[9] Iosifescu, M., and Theodorescu, R. (1969) *Random Processes and Learning. Springer*, Berlin.

[10] Kifer, Y. (1998) Limit theorems for random transformations and processes in random environments. *Trans. Amer. Math. Soc.* **350** 1481-1518.

[11] Kipnis, C., Varadhan, S. R. S. (1986) Central limit theorem for additive functionals of reversible markov processes. *Commun. Math. Phys.* **104** 1-19.

[12] Pinsky, M. (1991) *Lectures on Random Evolution.* World Scientific, Singapore.

[13] Sarymsakov, T.A. (1961) Inhomogeneous Markov chains. *Theor. Probability Appl.* **6** 178-185.

[14] Seneta, E. (1973) On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Cambridge Philos. Soc.* **74** 507-513.

[15] Seneta, E. (1981) *Non-negative Matrices and Markov Chains.* Second Edition, Springer-Verlag, New York.

[16] Statuljavicius, V. (1969) Limit theorems for sums of random variables connected in Markov chains.(Russian) *Litovsk. Mat. Sb.* **9** 345-362; ibid. **9**, 635-672; ibid. **10** 161–169.

[17] Varadhan, S.R.S. (2001) *Probability Theory.* Courant Lecture Notes **7**, American Mathematical Society, Providence, R.I.

[18] Winkler, G. (1995) *Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction.* Applications of Mathematics **27**, Springer-Verlag, Berlin.

[19] Wu, Wei Biao, Woodroofe, M. (2004) Martingale approximations for sums of stationary processes. *Ann. Probab.* **32** 1674–1690.