Periodica Mathematica Hungarica Vol. 56 (2), 2008, pp. 183–211
DOI: 10.1007/s10998-008-6183-3

A CENTRAL LIMIT THEOREM FOR NORMALIZED PRODUCTS OF RANDOM MATRICES

Rolando Cavazos-Cadena[1] and Daniel Hernández-Hernández[2]

[Communicated by István Berkes]

[1]Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro
Buenavista, Saltillo COAH, 25315, México
E-mail: rcavazos@uaaan.mx

[2]Centro de Investigación en Matemáticas, Apartado Postal 402
Guanajuato, GTO, 36000, México
E-mail: dher@cimat.mx

(*Received February 19, 2007; Accepted November 27, 2007*)

Abstract

This note concerns the asymptotic behavior of a Markov process obtained from normalized products of independent and identically distributed random matrices. The weak convergence of this process is proved, as well as the law of large numbers and the central limit theorem.

1. Introduction

Motivated by the study of ergodic properties of dynamical systems, the analysis of the asymptotic behavior of products of random matrices can be traced back, at least, to the early sixties. Fundamental results were obtained in Furstenberg and Kesten (1960), Furstenberg (1963) and Oseledec (1968). The first of these papers considered a process $\{M_n\}$ taking values in the space of $k \times k$ real matrices endowed with an appropriate norm $\|\cdot\|$ and, assuming that the M_ns are independent and identically distributed (iid), the authors studied the grow rate of the products $M_n \cdots M_1$ given by

$$\lim_{n\to\infty} \|M_n M_{n-1} \cdots M_1\|^{1/n}.$$

Mathematics subject classification number: 60F05, 60B10; 60B05.
Key words and phrases: random products, Birkhoff's distance, delayed process, invariant measure, law of large numbers, divisibility of a distribution, uniform differentiability of a family of characteristic functions.
This work was supported by the PSF Organization under Grant No. 2005-7–02, and by the Consejo Nacional de Ciencia y Tecnología under Grants 25357 and 61423.

0031-5303/2008/$20.00

Akadémiai Kiadó, Budapest
Springer, Dordrecht

It was proved that such a limit exists, and the law of large numbers as well as the central limit theorem were established; later on, Fustenberg and Kifer (1983) studied the asymptotic behavior of the vector norms

$$\|M_n M_{n-1} \cdots M_1 \mathbf{x}\|^{1/n}, \quad \mathbf{x} \in \mathbb{R}^k.$$

In the fundamental papers mentioned above, the main issue was to study the asymptotic growth of the products $M_n \cdots M_1$, and the theory of Lyapunov exponents was developed from logarithmic transformation of the products. In recent years, applications of products of random matrices in statistical physics, chaotic dynamical systems, filtering and Schrodinger operators has motivated a deep study of this theory; see, for instance, Cristiani, Paladin and Valpiani (1993), Atar and Zeitouni (1997), Bougerol and Lacroix (1985) and Carmona and Lacroix (1990).

On the other hand, the assertion in Furstenberg and Kifer (1983) that 'there are simple questions that are unanswered', can be completed requiring also simple answers to simple questions. This work deals with some of those problems proving the central limit theorem and the ergodic theorem for the Markov process explained below using basic tools of probability and linear algebra. More precisely, the present note is concerned with the process

$$M_n M_{n-1} \cdots M_1 \mathbf{x}, \tag{1.1}$$

when the matrix valued random variables $M_1,\ M_2, \ldots$ are iid and the vector $\mathbf{x} \in \mathbb{R}^k$ has nonnegative components. However, the focus of the paper is not on the asymptotic growth already studied in the literature, but on the limit properties of the probability vector X_n obtained by projecting (1.1) over the probability simplex in $\mathbb{R}^k$. The process $\{X_n\}$ obtained in this way is a Markov chain and, under the mild (communication) conditions in Assumption 2.1 below, *the main objective of the paper* is to prove that $\{X_n\}$ converges weakly and, more remarkably, that the law of large numbers and the central limit theorem hold; see Theorems 2.1–2.3 in Section 2. Even though some limit theorems have been obtained previously by Hennion (1997), see also Mukherjea (1987), the techniques used in this paper are completely different with simple arguments based on fundamental theorems of probability theory and linear algebra.

The motivation to study the asymptotic behavior of the process $\{X_n\}$ just described, comes from the analysis of partially observable Markov chains $\{S_t\}$ evolving over the finite set $\{1, 2, \ldots, k\}$ and endowed with a risk-sensitive performance index. In such models, the true state of the underlying chain is not directly observed, but at each time t the analyzer gets a signal Y_t and knows the probabilities $Q_{y\,x}$ of the true state $S_t = x$ when the observed signal is $Y_t = y$. The vector X_t represents the beliefs about the true unobservable state at time t given the available information up to time t, and the transfer matrix M_t is determined by the transition matrix of the chain $\{S_t\}$, the matrix $[Q_{y\,x}]$ relating observable signals and unobservable states, and the costs incurred while the system is running. In this context, the specific assumptions made below on the matrices M_n are satisfied when the matrix $[Q_{y\,x}]$ has positive entries, and the unobservable chain $\{S_t\}$ is communicating; see Fleming and Hernández-Hernández (1999) and Cavazos-Cadena and Hernández-Hernández (2004) for details. The arguments used to establish the main conclusions of the paper combine algebraic and probabilistic ideas. The main algebraic tool is the contraction property of positive matrices with respect to Birkhoff's distance (Seneta, 1980, Cavazos-Cadena, 2003). On the other hand, the probabilistic part is based on the introduction of 'delayed' processes which are probabilistic replicas of $\{X_n\}$.

The organization of the paper is as follows: In Section 2 the structural Assumptions are precisely formulated, the Markov process $\{X_n\}$ is introduced, and the main results are stated in the form of Theorems 2.1–2.3. Next, the basic technical preliminaries involving Birkhoff's distance and the delayed processes are given in Section 3, and these results are used in Sections 4 and 5 to prove the weak convergence result stated in Theorem 2.1, and the strong law of large numbers in Theorem 2.2, respectively.

On the other hand, the proof of the central limit theorem stated as Theorem 2.3 is substantially more technical, and the necessary preliminary results, concerning the summations $S_n = \sum_{t=1}^{n}(f(X_n) - E[f(X_n)])$ where f is a Lipchitz continuous function, are presented in Sections 6–9. In Section 6 it is shown that $E[S_n^4] = O(n^2)$, and in Section 7 it is proved that $\{E[S_n^2]/n\}$ is a convergent sequence. Next, these results are used to establish two fundamental properties of the family $\mathcal{W}$ of weak limits of the sequence $\{S_n/\sqrt{n}\}$ of normalized averages. Section 8 concerns a uniform differentiability property of the characteristic functions of the members of $\mathcal{W}$, and in Section 9 the following divisibility property is established: Given a positive integer m, each $\nu \in \mathcal{W}$ is the distribution of a normalized mean of m iid random variables whose common distribution also belongs to $\mathcal{W}$.

After these preliminaries, the central limit theorem is finally proved in Section 10.

Notation. Throughout the remainder $\mathbb{N}$ denotes the set of all positive integers. Given $k \in \mathbb{N}$, $\mathbb{R}^k$ is the k-dimensional Euclidean space of *column* vectors with real components. If $\mathcal{S}$ is a metric space, $\mathcal{B}(\mathcal{S})$ denotes the class of Borel subsets of $\mathcal{S}$, whereas $\mathbb{P}(\mathcal{S})$ stands for the space of all probability measures on $\mathcal{B}(\mathcal{S})$. On the other hand, if $f\colon \mathcal{S} \to \mathbb{R}$ is a given function, $\|f\|$ denotes the corresponding supremum norm, i.e., $\|f\| := \sup_{x\in\mathcal{S}} |f(x)|$.

2. Random probability vectors and main results

Throughout the remainder $\{M_n\}$ is a sequence of $k \times k$ random matrices defined on a probability space $(\Omega, \mathcal{F}, P)$, and the following conditions are supposed to hold.

ASSUMPTION 2.1.
(i) $M_1, M_2, \ldots$, are independent and identically distributed (iid);
(ii) all the entries of each matrix M_i are nonnegative;
(iii) there exists an integer N as well as $B_0, B_1 > 0$ such that

$$P\Big[B_1 \geq (M_1 M_2 \cdots M_N)_{i\,j} \geq B_0, \quad i, j = 1, 2, \ldots, k\Big] = 1.$$

Notice that under this assumption the following statements are always valid: For different positive integers $n_1, n_2, \ldots, n_N$

$$B_1 \geq (M_{n_1} M_{n_2} \cdots M_{n_N})_{i\,j} \geq B_0, \quad P\text{-a.s.} \tag{2.1}$$

and for each $n \in \mathbb{N}$ and $j = 1, 2, \ldots, k$

$$\sum_{i=1}^{k} (M_n)_{i\,j} > 0, \quad P\text{-a.s.}. \tag{2.2}$$

Next, let $\mathcal{P}_k \subset \mathbb{R}^k$ be the space of k-dimensional probability vectors, i.e.,

$$\mathcal{P}_k = \Big\{\mathbf{x} = (x_1, \ldots, x_k)' \,|\, x_i \geq 0,\ i = 1, 2, \ldots, k,\ \sum_{i=1}^{k} x_i = 1\Big\},$$

and denote by $\mathbb{1}$ the vector in $\mathbb{R}^k$ which has all its components equal to 1. Using this notation, (2.2) can be written as

$$(\mathbb{1}' M_n)_j > 0 \quad P\text{-a.s. for all } j = 1, 2, \ldots, k, \quad n \in \mathbb{N}. \tag{2.3}$$

DEFINITION 2.1. Let $\mathbf{x} \in \mathcal{P}_k$ be arbitrary but fixed. The sequence $\{X_n\}$ of random probability vectors is recursively defined by $X_0 := \mathbf{x}$ and

$$X_n := \frac{1}{\mathbb{1}' M_n X_{n-1}} M_n X_{n-1}, \quad n \in \mathbb{N}; \tag{2.4}$$

by (2.3), the right-hand side of this equality is well-defined P-almost surely.

REMARK 2.1. From (2.4) it is not difficult to see that

$$X_n = \frac{1}{\mathbb{1}' M_n \cdots M_1 X_0} M_n \cdots M_1 X_0, \quad n \in \mathbb{N}, \tag{2.5}$$

so that if $X_0 = \mathbf{e}_i$, the i-th vector in the canonical basis of $\mathbb{R}^k$, then X_n is the i-th column of the product $M_n \cdots M_1$ normalized in such a way that its components add up to 1.

From Assumption 2.1 and Definition 2.1 it follows that $\{X_n\}$ is a $\mathcal{P}_k$-valued Markov process, and the main objective of this note is to study its asymptotic behavior. For each $n \in \mathbb{N}$, let $\mu_n \in \mathbb{P}(\mathcal{P}_k)$ be the distribution of X_n, i.e.,

$$\mu_n(A) := P[X_n \in A], \quad A \in \mathcal{B}(\mathcal{P}_k). \tag{2.6}$$

The main results of this work are stated in the following three theorems.

THEOREM 2.1. *There exists $\mu \in \mathbb{P}(\mathcal{P}_k)$ such that $\mu_n \xrightarrow{\text{w}} \mu$. Moreover, μ does not depend on the initial state $X_0 = \mathbf{x}$.*

It is not difficult to see that $\mu \in \mathbb{P}(\mathcal{P}_k)$ in this theorem is the unique invariant distribution of the Markov process $\{X_n\}$, and Theorem 2.1 is analogous to the classical result for discrete-time Markov chains evolving on a finite state space: If such a chain is communicating and aperiodic, then the rows the n-th power of the transition matrix T converge pointwise to the unique invariant distribution of T.

Next, for each $\mathbf{y} \in \mathcal{P}_k$, let $\delta_{\mathbf{y}} \in \mathbb{P}(\mathcal{P}_k)$ be the Dirac measure at $\mathbf{y}$, i.e., for $A \in \mathcal{B}(\mathcal{P}_k)$,

$$\delta_{\mathbf{y}}(A) := \begin{cases} 1, & \text{if } \mathbf{y} \in A, \\ 0, & \text{otherwise,} \end{cases}$$

and for $n \in \mathbb{N}$ define the empirical measure $\tilde{\mu}_n \in \mathbb{P}(\mathcal{P}_k)$ associated to $X_1, X_2, \ldots, X_n$ as follows:

$$\tilde{\mu}_n(A) := \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j}(A), \quad A \in \mathcal{B}(\mathcal{P}_k). \tag{2.7}$$

The following is the ergodic theorem for $\{X_n\}$.

THEOREM 2.2. *Let $\mu \in \mathbb{P}(\mathcal{P}_k)$ be as in Theorem 2.1. In this case, $\tilde{\mu}_n \xrightarrow{\text{w}} \mu$ with probability 1. More precisely, if $f\colon \mathcal{P}_k \to \mathbb{R}$ is a continuous function, then as $n \to \infty$*

$$\int_{\mathcal{P}_k} f(\mathbf{x}) \tilde{\mu}_n(d\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} f(X_i) \to \int_{\mathcal{P}_k} f(\mathbf{x}) \mu(d\mathbf{x}) =: \mu^*(f), \quad P\text{-a.s.}. \tag{2.8}$$

The next result is a central limit theorem, and provides conditions under which the (appropriately) normalized deviations of the time average $\int_{\mathcal{P}_k} f(\mathbf{x}) \tilde{\mu}_n(d\mathbf{x})$ around the limit value $\mu^*(f)$ in (2.8) approximate a normal distribution.

THEOREM 2.3. *Let $\mathcal{P}_k^+$ be the set of all of $\mathbf{x} = (x_1, \ldots, x_k)' \in \mathcal{P}_k$ with $x_i > 0$ for all $i = 1, 2, \ldots, k$, and assume that $f\colon \mathcal{P}_k^+ \to \mathbb{R}$ is Lipschitz continuous, i.e., for some constant L_f*

$$|f(\mathbf{x}) - f(\mathbf{y})| \le L_f \|\mathbf{x} - \mathbf{y}\|, \quad \mathbf{x}, \mathbf{y} \in \mathcal{P}_k^+, \tag{2.9}$$

where $\|\mathbf{v}\|$ denotes the Euclidean norm of $\mathbf{v} \in \mathbb{R}^k$. In this context, the following assertions (i) *and* (ii) *hold:*

(i) $\lim_{n\to\infty} \frac{1}{n} \mathrm{Var}\left(\sum_{i=1}^n f(X_i)\right) =: v < \infty$ *exists and, moreover, v does not depend on the the initial state $X_0 = \mathbf{x}$.*

(ii) *As $n \to \infty$*

$$\sqrt{n}\left[\frac{\sum_{i=1}^n f(X_i)}{n} - \mu^*(f)\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0, v), \tag{2.10}$$

where $\mathcal{N}(0, v)$ is the normal distribution with mean 0 and variance v, and following the usual convention, if v is null, $\mathcal{N}(0, v)$ is interpreted as the unit of mass at zero.

Before going into the details of the proofs, let $\mathbf{w} \in \mathbb{R}^k$ be arbitrary, and notice that application of Theorem 2.3 to the function $f(\mathbf{x}) = \mathbf{w}'\mathbf{x}$ leads to

$$\sqrt{n}\mathbf{w}'\left[\frac{\sum_{i=1}^n X_i}{n} - \mu^*\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0, v_{\mathbf{w}}),$$

where

$$\mu^* = \int_{\mathcal{P}_k} \mathbf{x}\mu(d\mathbf{x}) \tag{2.11}$$

is the mean of the invariant distribution μ in Theorem 2.1, and $v_{\mathbf{w}} = \lim\limits_{n\to\infty} n^{-1}\mathbf{w}'V_n\mathbf{w}$ where V_n is the variance matrix of $X_1 + \cdots + X_n$. Since this latter limit exists for each $\mathbf{w} \in \mathbb{R}^k$, it follows that $\{n^{-1}\mathbf{w}_1 V_n \mathbf{w}_2\}$ is a convergent sequence for every $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^k$, so that

$$\lim_{n\to\infty} n^{-1}\mathrm{Var}\left(X_1 + \cdots + X_n\right) =: V \tag{2.12}$$

is well-defined and does not depend on the initial state $X_0 = \mathbf{x}$. Therefore, $v_{\mathbf{w}} = \mathbf{w}'V\mathbf{w}$, and it follows that

$$\sqrt{n}\mathbf{w}'\left[\frac{\sum_{i=1}^n X_i}{n} - \mu^*\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathbf{w}'V\mathbf{w}).$$

Since this convergence holds for every $\mathbf{w} \in \mathbb{R}^k$, Cramer's theorem yields the following.

COROLLARY 2.1. *Under Assumption* 2.1,

$$\sqrt{n}\left[\frac{\sum_{i=1}^{n} X_i}{n} - \mu^*\right] \xrightarrow{\mathrm{d}} \mathcal{N}_k\left(\mathbf{0}, V\right)$$

where $\mu^* \in \mathcal{P}_k$ *and the* $k \times k$ *matrix* V *do not depend on the initial state* $X_0 = \mathbf{x}$ *and are specified in* (2.11) *and* (2.12)*, respectively.*

The proofs of Theorems 2.1 and 2.2, based on the preliminaries contained in the following section, are presented in Sections 4 and 5, respectively. On the other hand, the proof of Theorem 2.3 is substantially more technical, and will be given in Section 10 after the necessary tools are established in Sections 6–9.

3. Basic preliminaries

This section contains the fundamental technical results that will be used to establish Theorems 2.1–2.3. Firstly, the essential algebraic tool, concerning a contraction property of positive matrices, is briefly discussed.

Birkhoff's Distance. Let $\mathcal{C}_k$ be the positive cone in $\mathbb{R}^k$, i.e.,

$$\mathcal{C}_k := \{\mathbf{x} \in \mathbb{R}^k : x_i > 0,\ i = 1, 2, \ldots, k\},$$

so that the set $\mathcal{P}_k^+$ of positive probability vectors is given by

$$P_k^+ = \mathcal{C}_k \cap \mathcal{P}_k.$$

On $\mathcal{C}_k$ define the Birkhoff (pseudo) metric by

$$d_B(\mathbf{x}, \mathbf{y}) = \max_{i,j=1,2,\ldots,k} \log\left(\frac{x_i/y_i}{x_j/y_j}\right), \quad \mathbf{x}, \mathbf{y} \in \mathcal{C}_k, \tag{3.1}$$

so that

$$d_B(c\mathbf{x}, d\mathbf{y}) = d_B(\mathbf{x}, \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathcal{C}_k, \quad c, d > 0, \tag{3.2}$$

and if $\mathbf{x}, \mathbf{y} \in \mathcal{P}_k^+$, then $d_B(\mathbf{x}, \mathbf{y}) = 0$ leads to $\mathbf{x} = \mathbf{y}$. Thus, $d_B(\cdot, \cdot)$ induces a genuine metric on $\mathcal{P}_k^+$, and it is not difficult to verify that this metric is topologically equivalent to the Euclidean one. The contraction property in the following lemma, whose proof can be found in Seneta (1980), or in Cavazos-Cadena (2003), will play a central role in the subsequent development.

LEMMA 3.1. (i) *Let A be a $k \times k$ real matrix, and assume that for some constants $B_0, B_1 > 0$,*

$$B_1 \geq A_{i\,j} \geq B_0, \quad i, j = 1, 2, \ldots, k. \tag{3.3}$$

In this case there exists a constant

$$\beta_0 \equiv \beta_0(B_0, B_1) < 1$$

such that the following Birkhoff's *inequality holds:*

$$d_B(A\mathbf{x}, A\mathbf{y}) \leq \beta_0 d_B(\mathbf{x}, \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathcal{C}_k.$$

(ii) *In particular, if each matrix $A^{(1)}, A^{(2)}, \ldots, A^{(r)}$ satisfies* (3.3), *then*

$$d_B\left(\prod_{i=1}^{r} A^{(i)}\mathbf{x}, \prod_{i=1}^{r} A^{(i)}\mathbf{y}\right) \leq \beta_0^r d_B(\mathbf{x}, \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathcal{C}_k.$$

Support of μ_n. For each $a \in (0, 1/k]$, define

$$\mathbb{K}_a := \{\mathbf{x} \in \mathcal{P}_k : x_i \geq a\}, \tag{3.4}$$

which is a compact subset of $\mathcal{P}_k^+$. Observing that

$$a \leq x_i/y_i \leq 1/a, \quad i = 1, 2, \ldots, k, \quad \mathbf{x}, \mathbf{y} \in \mathbb{K}_a \tag{3.5}$$

it follows from (3.1) that

$$d_B(\mathbf{x}, \mathbf{y}) \leq \log(1/a^2) = -2\log(a), \quad \mathbf{x}, \mathbf{y} \in \mathbb{K}_a. \tag{3.6}$$

On the other hand, for $x \in [1, b]$ where $b > 1$, the mean value theorem implies that $(x-1)/\log(x) = x_0$ for some $x_0 \in [1, b]$, so that $x - 1 \leq b\log(x)$. Given $\mathbf{x}, \mathbf{y} \in \mathbb{K}_a$, select i^* such that

$$\frac{x_{i^*}}{y_{i^*}} = \max_{i=1,2,\ldots,k} \frac{x_i}{y_i} \in [1, 1/a].$$

With this notation, $x_i - y_i = y_i(x_i/y_i - 1) \leq y_i(x_{i^*}/y_{i^*} - 1) \leq (x_{i^*}/y_{i^*} - 1) \leq \log(x_{i^*}/y_{i^*})/a \leq d_n(\mathbf{x}, \mathbf{y})/a$. Interchanging the roles of $\mathbf{x}$ and $\mathbf{y}$ it follows that $|x_i - y_i| \leq d_n(\mathbf{x}, \mathbf{y})/a$, and then

$$\|\mathbf{x} - \mathbf{y}\| \leq \frac{\sqrt{k}}{a} d_B(\mathbf{x}, \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{K}_a. \tag{3.7}$$

Next, it will be shown that, for n large enough, X_n takes values in a set $\mathbb{K}_a$.

LEMMA 3.2. *For each $n \geq N$, $P[X_n \in \mathbb{K}_\alpha] = 1$, where $\alpha = B_0/kB_1$, so that μ_n is supported on $\mathbb{K}_\alpha$; see Assumption* 2.1, (2.6) *and* (3.4).

PROOF. Let $\mathbf{J}$ be the $k \times k$ matrix with all its components equal to 1. Since X_n is a probability vector, it follows that

$$\mathbf{J} X_n = \mathbb{1},$$

whereas (2.1) yields that

$$B_0 \mathbf{J} \leq M_n \cdots M_{n-N+1} \leq B_1 \mathbf{J} \quad P\text{-a.s.},$$

where the inequalities are componentwise. On the other hand, from Definition 2.1 it is not difficult to see that for $n \geq N$

$$X_n = \frac{1}{\mathbb{1}' M_n \cdots M_{n-N+1} X_{n-N}} M_n \cdots M_{n-N+1} X_{n-N},$$

and then, with probability 1,

$$B_0 \mathbb{1} = B_0 \mathbf{J} X_{n-N} \leq M_n \cdots M_{n-N+1} X_{n-N} \leq B_1 \mathbf{J} X_{n-N} = B_1 \mathbb{1}$$

so that $\mathbb{1}' M_n \cdots M_{n-N+1} X_{n-N} \leq B_1 \mathbb{1}' \mathbb{1} = k B_1$. Combining this with the last two displayed equations it follows that $X_n \geq (B_0/kB_1)\mathbb{1} = \alpha \mathbb{1}$ P-a.s. for $n \geq N$. □

Delayed Processes. For each nonnegative integer n set

$$Y_0 := X_0, \quad Y_n := M_n M_{n-1} \cdots M_1 X_0, \tag{3.8}$$

so that

$$X_n = \frac{1}{\mathbb{1}' Y_n} Y_n; \tag{3.9}$$

see (2.5). Also, for nonnegative integers n and m define

$$Y_{n,0} := X_0, \quad Y_{n,m} := M_n M_{n-1} \cdots M_{n-m+1} X_0, \quad n \geq m > 0, \tag{3.10}$$

and

$$X_{n,m} := \frac{1}{\mathbb{1}' Y_{n,m}} Y_{n,m}, \quad n \geq m; \tag{3.11}$$

notice that

$$X_{n,n} = X_n, \quad n = 0, 1, 2, \ldots. \tag{3.12}$$

The *delayed process starting at time m* is defined as $\{X_{m+t,t} : t = 0, 1, 2, 3, \ldots\}$. Using Assumption 2.1 these definitions immediately yield the following.

LEMMA 3.3. *For each nonnegative integer m the following properties are satisfied by the delayed process $\{X_{m+t,\,t} : t = 0, 1, 2, \ldots\}$ at time m:*

(i) *$\{X_{m+t,\,t} : t = 0, 1, 2, \ldots\}$ and the original process $\{X_t : t = 0, 1, 2, 3, \ldots\}$ have the same distribution.*

Consequently,

(ii) *For $n \ge m$, $X_{n,\,m}$ and X_m have the same distribution.*

(iii) *$(X_1, X_2, \ldots, X_m)$ and $\{X_{m+t,\,t} : t \in \mathbb{N}\}$ are independent.*

By convenience, the following convention is enforced throughout the remainder of the paper: For nonnegative integers n and m,

$$X_{n,\,m} := X_n, \quad n < m. \tag{3.13}$$

The following estimate of the distance between X_n and the delayed vector $X_{n,\,m}$ will be useful.

LEMMA 3.4. (i) *Let $B_0, B_1 > 0$ and the positive integer N be as in Assumption* 2.1(iii), *and let $\beta_0 = \beta_0(B_0, B_1) \in (0, 1)$ and $\alpha > 0$ be as in Lemmas* 3.1 *and* 3.2, *respectively. In this case,*

$$d_B(X_n, X_{n,\,m}) \le \beta^m K, \quad n = 0, 1, 2, \ldots, \quad m \ge 2N \quad P\text{-a.s.}, \tag{3.14}$$

where

$$\beta = \beta_0^{1/N} \in (0, 1) \quad \text{and} \quad K = \frac{-2\log(\alpha)}{\beta^{2N}}.$$

(ii) *Let $\{X_n'\}$ be as in Definition* 2.1 *with $X_0 = \mathbf{y}$. In this case*

$$d_B(X_m, X_m') \le \beta^m K, \quad m \ge 2N.$$

PROOF. (i) By (3.13), the inequality in (3.14) holds when $n < m$, so that it is sufficient to consider the case $n \ge m \ge 2N$. Let $m \ge 2N$ be fixed and write $m = (s+1)N + r$, where $0 \le r < N$ and $s \ge 1$. Notice that for $n \ge m$

$$d_B(X_n, X_{n,\,m}) = d_B(Y_n, Y_{n,\,m}) \tag{3.15}$$

by (3.2) and (3.8)–(3.11), and set

$$A_j = \prod_{i=0}^{N-1} M_{n-(j-1)N-i}, \quad j = 1, 2, \ldots, s. \tag{3.16}$$

Next, observe that (3.8) and (3.10) can be written as

$$Y_n = \prod_{j=1}^{s} A_j Z_1, \quad Y_{n,\,m} = \prod_{j=1}^{s} A_j Z_0,$$

where

$$Z_1 = M_{n-Ns}M_{n-Ns-1}\cdots M_1\mathbf{x}, \quad Z_0 = M_{n-Ns}M_{n-Ns-1}\cdots M_{n-m+1}\mathbf{x}$$

and the right-hand side of these equalities contain, at least, N factors M_t. Setting $c_i = \mathbb{1}'Z_i$ and $\hat{Z}_i = (1/c_i)Z_i$, it follows by Assumption 2.1 and Lemma 3.2 that $\hat{Z}_i \in \mathbb{K}_\alpha$ P-a.s., and in this case

$$d_B(\hat{Z}_1, \hat{Z}_0) \le -2\log(\alpha), \quad P\text{-a.s.}, \tag{3.17}$$

by (3.6), whereas

$$Y_n = c_1 \prod_{j=1}^{s} A_j \hat{Z}_1, \quad Y_{n,\,m} = c_0 \prod_{j=1}^{s} A_j \hat{Z}_0$$

and (3.2) together imply that

$$d_B\left(Y_n, Y_{n,\,m}\right) = d_B\left(\prod_{j=1}^{s} A_j \hat{Z}_1, \prod_{j=1}^{s} A_j \hat{Z}_0\right).$$

By the iid part of Assumption 2.1, and observing that each matrix Aj is the product of N factors M_i, it follows that all the entries A_j lay in $[B_0, B_1]$ P-a.s., and then

$$d_B(Y_n, Y_{n,\,m}) \le \beta_0^s d_B(\hat{Z}_1, \hat{Z}_0), \quad P\text{-a.s.},$$

by Lemma 3.1. Together with (3.15) and (3.17), this implies that

$$d_B(X_n, X_{n,\,m}) \le \beta_0^s(-2\log(\alpha)) = \beta^{N(s+1)+r}\frac{(-2\log(\alpha))}{\beta^{N+r}} \le \beta^m \frac{(-2\log(\alpha))}{\beta^{2N}}$$

and (3.14) follows, establishing part (i). The proof of part (ii) is similar.

□

4. Convergence to the invariant distribution

In this section Theorem 2.1 will be proved. To begin with, recall that $\{\mu_n\}$ is a sequence of probability measures on the Borel sets of the space $\mathcal{P}_k$, which is compact. Therefore, Prohorov's theorem yields that there exists a subsequence $\{\mu_{n_r}\}$ and $\mu \in \mathbb{P}(\mathcal{P}_k)$ such that

$$\mu_{n_r} \xrightarrow{\text{w}} \mu \quad \text{as } r \to \infty. \tag{4.1}$$

Proof of Theorem 2.1. With $\mu \in \mathbb{P}(\mathcal{P}_k)$ as above, it will be proved that the whole sequence $\{\mu_n\}$ converges weakly to μ. Recalling that μ_n is supported on the compact set $\mathbb{K}_\alpha \subset \mathcal{P}_k^+$ when $n \geq N$, by Lemma 3.2, it is sufficient to show that for every nonempty open set $A \subset \mathcal{P}_k^+$

$$\liminf_{n\to\infty} \mu_n(A) \geq \mu(A). \tag{4.2}$$

Given such a set A, for each $\varepsilon > 0$ define

$$A_\varepsilon := \{\mathbf{x} \in A : \mathbf{y} \in A \text{ if } \mathbf{y} \in \mathcal{P}_k^+ \text{ satisfies } \ d_B(\mathbf{x}, \mathbf{y}) < \varepsilon\},$$

and notice that each A_ε is open, so that

$$\liminf_{r\to\infty} \mu_{n_r}(A_\varepsilon) \geq \mu(A_\varepsilon), \tag{4.3}$$

by (4.1). Moreover,

$$A_\varepsilon \nearrow A \quad \text{as} \quad \varepsilon \searrow 0. \tag{4.4}$$

Next, let $\delta > 0$ be arbitrary but fixed satisfying $A_\delta \neq \emptyset$ and, with β and K as in Lemma 3.4, select an integer $m \geq 2N$ such that

$$\beta^m K < \delta, \tag{4.5}$$

and for $n_r \geq m$ set

$$\Omega_{n,\,n_r} := [d_B(X_n, X_{n,\,n_r}) \leq \beta^{n_r} K], \quad n \geq n_r (\geq m).$$

Then, the definition of A_δ and (4.5) together yield

$$\Omega_{n,\,n_r} \cap [X_{n,\,n_r} \in A_\delta] \subset [X_n \in A]$$

so that

$$\mu_n(A) = P[X_n \in A] \geq P\left[\Omega_{n,\,n_r} \cap [X_{n,\,n_r} \in A_\delta]\right] = P[X_{n,\,n_r} \in A_\delta], \quad n \geq n_r,$$

where the equality used that $P[\Omega_{n,\,n_r}] = 1$, by Lemma 3.4(i). Using now that $X_{n,\,n_r}$ and X_{n_r} have the same distribution, by Lemma 3.3(ii), it follows that $P[X_{n,\,n_r} \in A_\delta] = P[X_{n_r} \in A_\delta] = \mu_{n_r}(A)$, and the above displayed relation is equivalent to $\mu_n(A) \geq \mu_{n_r}(A_\delta)$ for $n \geq n_r$, so that

$$\liminf_{n\to\infty} \mu_n(A) \geq \mu_{n_r}(A_\delta).$$

Taking the limit inferior as $r \to \infty$ it follows that

$$\liminf_{n\to\infty} \mu_n(A) \geq \liminf_{r\to\infty} \mu_{n_r}(A_\delta) \geq \mu(A_\delta),$$

where the second inequality is due to (4.3) and, via (4.4), this implies (4.2). Now let $\{X_n'\}$ be the process corresponding to the initial condition $X_0 = \mathbf{y}$. Since $d_B(X_n, X_n') \le \beta^n K$ P-a.s. when $n \ge 2N$, by Lemma 3.4(ii), it follows that $d_B(X_n, X_n') \to 0$ as $n \to \infty$ P-a.s.; since X_n and X_n' belong to $\mathbb{K}_\alpha$ for $n \ge N$, by Lemma 3.2, using (3.7) it follows that $\|X_n - X_n'\| \to 0$ P-a.s., so that $X_n' = X_n + Z_n$ where $Z_n \to 0$ with probability 1, and Slutsky's theorem yields that $\{X_n'\}$ converges weakly to μ, i.e., the limit distribution μ does not depend on the initial state.

□

5. Ergodic theorem

In this section Theorem 2.2 will be established. The argument relies on Lemma 3.4 as well as on the two lemmas stated below. To begin with, let $f\colon \mathcal{P}_k \to \mathbb{R}$ be as in the statement of Theorem 2.2, and for each integer $m \ge 2N$ define

$$\tilde{S}_r = \sum_{i=1}^{m} f(X_{[(r-1)m+i],m}), \quad r = 2, 3, 4, \dots \,. \tag{5.1}$$

LEMMA 5.1.

(i) *For each* $r = 2, 3, 4, \dots$

$$E[\tilde{S}_r] = mE[f(X_m)] = m \int_{\mathcal{P}_k} f(\mathbf{x})\mu_m(d\mathbf{x}).$$

(ii) *The* $\tilde{S}_r$*'s have the same distribution.*

(iii) $\tilde{S}_{2t}$*,* $t = 1, 2, 3, \dots$ *are iid and, similarly,* $\tilde{S}_{1+2t}$*,* $t = 1, 2, 3, \dots$ *are iid.*

PROOF. (i) By Lemma 3.3(ii), $X_{t,m}$ and X_m have the same distribution for $t \ge m$, and then the expectation of each term in the summation in (5.1) is $E[f(X_m)]$; notice that $E[f(X_m)]$ exists, since the support of μ_m is contained in the compact set $\mathbb{K}_\alpha \subset \mathcal{P}_k^+$, and f is continuous on $\mathcal{P}_k$.

(ii) Recalling that $X_0 = \mathbf{x}$ is fixed, (3.8) and (3.9) show that for a certain function F

$$(X_{(m+1),m}, \dots, X_{2m,m}) = F(M_2, \dots, M_{2m})$$

and

$$(X_{((r-1)m+1),m}, \dots, X_{rm,m}) = F(M_{(r-2)m+2}, \dots, M_{rm}); \tag{5.2}$$

since the matrices M_i are iid, via (5.1) this yields that the variables $\tilde{S}_r$ are identically distributed.

(iii) Using (5.2) it follows that for $t \ge 2$,

$$(X_{((r+t-1)m+1),m}, \dots, X_{(r+t)m,m}) = F(M_{(r+t-2)m+2}, \dots, M_{(r+t)m}),$$

so that $\tilde{S}_{r+t}$ depends on the matrices M_i with $i > rm + 1$; see (5.1). Since S_r is a function of the matrices M_j with $j \le mr$, by (5.1) and (5.2), the independence of the matrices M_t immediately yields part (iii).

□

Lemma 5.2. *As* $n \to \infty$

$$\lim_{n\to\infty} \frac{1}{mn} \sum_{r=2}^{n} \tilde{S}_r = E[f(X_m)] \quad P\text{-}a.s. \tag{5.3}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} f(X_{t,\,m}) = E[f(X_m)] \quad P\text{-}a.s.; \tag{5.4}$$

see (3.11) *and* (3.13).

Proof. By Lemma 5.1, $\tilde{S}_2, \tilde{S}_4, \tilde{S}_6, \ldots$ are iid with mean $mE[f(X_m)]$, so that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} \tilde{S}_{2r} = mE[f(X_m)] \quad P\text{-a.s.};$$

similarly,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} \tilde{S}_{2r+1} = mE[f(X_m)] \quad P\text{-a.s.}.$$

These relations together yield

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=2}^{n} \tilde{S}_r = mE[f(X_m)] \quad P\text{-a.s.},$$

which is equivalent to (5.3). Observe now that $\sum_{r=2}^{n} \tilde{S}_r = \sum_{t=m+1}^{nm} f(X_{t,\,m})$, so that (5.3) yields

$$\lim_{n\to\infty} \frac{1}{nm} \sum_{t=m+1}^{nm} f(X_{t,\,m}) = E[f(X_m)] \quad P\text{-a.s.},$$

and using that $X_{t,\,m}$ belongs to the compact set $\mathbb{K}_\alpha$ with probability 1 for $t > m$, this immediately leads to (5.4).

□

Proof of Theorem 2.2. Given $\varepsilon > 0$ let $\delta > 0$ be such that

$$|f(\mathbf{x}) - f(\mathbf{y})| \le \varepsilon, \quad \text{if } \mathbf{x}, \mathbf{y} \in \mathbb{K}_\alpha \text{ and } d_B(\mathbf{x}, \mathbf{y}) \le \delta,$$

and select an integer $m \ge 2N$ satisfying $\beta^m K \le \delta$, where β and K are as in Lemma 3.4. In this case, via (3.14) it follows that $|f(X_t) - f(X_{t,\,m})| \le \varepsilon$ P-a.s., and then

$$\frac{1}{n}\left[\sum_{t=1}^{n} f(X_t) - \sum_{t=1}^{n} f(X_{t,\,m})\right] \le \varepsilon \quad P\text{-a.s.}.$$

Therefore, (5.4) implies that

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} f(X_t) \le E[f(X_m)] + \varepsilon = \int_{\mathcal{P}_k} f(\mathbf{x})\mu_m(d\mathbf{x}) + \varepsilon \quad P\text{-a.s.}$$

Using that $m \ge 2N$ is an arbitrary integer satisfying $\beta_1^m K \le \delta$, taking the limit as m goes to ∞ on the right-hand side of this inequality, the convergence $\mu_m \xrightarrow{\text{w}} \mu$ in Theorem 2.1 yields

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} f(X_t) \le \int_{\mathcal{P}_k} f(\mathbf{x})\mu(d\mathbf{x}) + \varepsilon \quad P\text{-a.s.}$$

so that

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} f(X_t) \le \int_{\mathcal{P}_k} f(\mathbf{x})\mu(d\mathbf{x}) \quad P\text{-a.s.,}$$

since $\varepsilon > 0$ is arbitrary. Observing that this inequality also holds with $-f$ instead of f, Theorem 2.2 follows.

□

6. Fourth moment of normalized means

The remainder of the paper is dedicated to prove Theorem 2.3, which establishes the asymptotic normality of the sequence $\{W_n\}$, where

$$W_n := \sqrt{n}\left[\frac{\sum_{i=1}^{n} f(X_i)}{n} - \mu^*(f)\right], \tag{6.1}$$

and $\mu^*(f)$ is as in (2.8). Instead of studying the limit behavior of $\{W_n\}$ it is convenient to analyze the sequence $\{S_n/\sqrt{n}\}$ where

$$S_n = \sum_{t=1}^{n} [f(X_t) - E[f(X_t)]], \quad n = 1, 2, 3, \ldots, \tag{6.2}$$

so that

$$E[S_n] = 0, \quad n \in \mathbb{N}. \tag{6.3}$$

The argument used to establish Theorem 2.3 has been divided into four steps, and the most basic facts are contained in this and the following section. The present objective is to show that $\{W_n\}$ and $\{S_n/\sqrt{n}\}$ have the same limit behavior, and that the fourth moment of $S_n/\sqrt{n}$ remains bounded as n goes to ∞. This latter result implies that the family of distributions of the variables $S_n/\sqrt{n}$ is tight—so that it has weak limits in $\mathbb{P}(\mathbb{R})$, by Prohorov's theorem—and is the basis to establish the uniform differentiability properties in Section 8. Before going any further, it is convenient to introduce some notation as well as a simplifying assumption that will be enforced throughout the remainder of the paper. As in the statement of Theorem 2.3, let $f\colon \mathcal{P}_k^+ \to \mathbb{R}$ be a Lipschitz continuous function. Since the distribution of X_n is concentrated on the compact set $\mathbb{K}_\alpha$ for $n \geq N$ (see Lemma 3.2), without loss of generality it will be assumed that the function f is supported on a compact set $\mathbb{K}_{\alpha'}$ with $\alpha' \leq \alpha$. In this case $\|f\|$ is finite and, moreover, (2.9) and (3.7) together yield that, for some constant L,

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L d_B(\mathbf{x}, \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathcal{P}_k^+. \tag{6.4}$$

Next, having in mind (3.8)–(3.13), for nonnegative integers t and m define

$$Z_t {:=}\, f(X_t) - E[f(X_t)], \tag{6.5}$$

and

$$Z_{t,\,m} {:=}\, f(X_{t,\,m}) - E[f(X_{t,\,m})], \tag{6.6}$$

so that

$$|Z_t|, |Z_{t,\,m}| \leq 2\|f\| < \infty, \quad \text{and} \quad E[Z_{t,\,m}] = E[Z_t] = 0; \tag{6.7}$$

notice that

$$S_n = \sum_{t=1}^{n} Z_t, \quad n = 1, 2, 3, \ldots, \tag{6.8}$$

by (6.2) and (6.5). Observe now that Lemma 3.4 and (6.4) together yield that for every $t \in \mathbb{N}$, $|f(X_t) - f(X_{t,\,m})| \leq KL\beta^m$ whenever $m \geq 2N$. Setting $C = \max\{2\|f\|, KL\}/\beta^{2N}$ it follows that

$$|f(X_t) - f(X_{t,\,m})| \leq C\beta^m \quad P\text{-a.s.} \tag{6.9}$$

and

$$|E[f(X_t)] - E[f(X_{t,\,m})]| \leq C\beta^m \tag{6.10}$$

are always valid; consequently,

$$|Z_t - Z_{t,\,m}| \leq 2C\beta^m \quad P\text{-a.s.}. \tag{6.11}$$

The next lemma yields that $\{W_n\}$ and $\{S_n/\sqrt{n}\}$ have the same asymptotic distribution.

Lemma 6.1. *As* $n \to \infty$, $W_n - \dfrac{S_n}{\sqrt{n}} \to 0$.

PROOF. Given an integer m, let $t \geq m$ be arbitrary.

Since $X_{t,m}$ and X_m have the same distribution, by Lemma 3.3(ii), (6.10) implies that $|E[f(X_t)] - E[f(X_m)]| \leq C\beta^m$ for every $t \geq m$, and observing that $E[f(X_t)] \to \int_{\mathcal{P}_k} f(\mathbf{x})\mu(d\mathbf{x}) = \mu^*(f)$ as $t \to \infty$, by Theorem 2.1, it follows that $|\mu^*(f) - E[f(X_m)]| \leq C\beta^m$; via (6.1) and (6.2) this leads to

$$\left|W_n - \frac{S_n}{\sqrt{n}}\right| = \left|\frac{\sum_{t=1}^n (E[f(X_t)] - \mu^*(f))}{\sqrt{n}}\right| \leq \frac{C\sum_{t=1}^n \beta^t}{\sqrt{n}} \leq \frac{C}{\sqrt{n}(1-\beta)},$$

and the conclusion follows.

□

The main result of the section can now be stated as follows.

THEOREM 6.1. *Let S_n be as in* (6.2). *Then*

$$\sup_n \frac{1}{n^2} E\left[S_n^4\right] =: B_4 < \infty.$$

The proof of this result relies on the following two lemmas.

LEMMA 6.2.

(i) *For each $t > m > 0$*

$$|E[Z_t Z_m]| \leq 4C\|f\|\beta^{t-m}.$$

(ii) $\sup_n \frac{1}{n} E\left[S_n^2\right] =: B_2 < \infty$.

PROOF. (i) Let the positive integers t and m be such that $t > m$, and write $Z_t = Z_{t,\,t-m} + (Z_t - Z_{t,\,t-m})$, so that $E[Z_t Z_m] = E[Z_{t,\,t-m} Z_m] + E[(Z_t - Z_{t,\,t-m})Z_m]$. Notice now that $Z_{t,\,t-m}$ depends on the matrices $M_t, \ldots, M_{m+1}$, so that it is independent of Z_m which is a function of $M_1, \ldots, M_m$; see (3.8)–(3.11), (6.5) and (6.6). Therefore, $E[Z_{t,\,t-m} Z_m] = 0$, by (6.7). Finally, (6.7) and (6.11) together imply that $E[(Z_t - Z_{t,\,t-m})Z_m] \leq 4C\|f\|\beta^{t-m}$, and the conclusion follows from this inequality and the two equations displayed above.

(ii) Notice that $E[S_n^2] = \sum_{t=1}^n E[Z_t^2] + 2\sum_{t>m} E[Z_t Z_m]$; since $E[Z_t^2] \leq 4\|f\|^2$, by (6.7), via part (i) it follows that

$$\begin{aligned}
E[S_n^2] &\leq \sum_{t=1}^n 4\|f\|^2 + 2\sum_{n \geq t > m \geq 1} 4C\|f\|\beta^{t-m} \\
&= 4n\|f\|^2 + 8C\|f\| \sum_{d=1}^{n-1} \sum_{(t,m):\, t-m=d,\, n\geq t,\, m\geq 1} \beta^d \\
&= 4n\|f\|^2 + 8C\|f\| \sum_{d=1}^{n-1} \beta^d (n-d) \leq 4n\|f\|^2 + 8C\|f\| n \sum_{d=1}^{n-1} \beta^d
\end{aligned}$$

so that $E[S_n^2] \leq n[4\|f\|^2 + 8C\|f\|/(1-\beta)]$ and the conclusion follows.

□

The following simple property will be useful.

Lemma 6.3. *Let $\varepsilon > 0$ be arbitrary. For each $a, b \in \mathbb{R}$, the following inequalities hold:*

$$(1-\varepsilon)a^2 + \Big(1 - \frac{1}{\varepsilon}\Big)b^2 \le (a+b)^2 \le (1+\varepsilon)a^2 + \Big(1 + \frac{1}{\varepsilon}\Big)b^2 \tag{6.12}$$

and

$$(a+b)^4 \le (1+\varepsilon)^2 a^4 + \Big(1 + \frac{1}{\varepsilon}\Big)^2 b^4 \tag{6.13}$$

Proof. Using that $2|ab| \le a^2 + b^2$ always holds, it follows that

$$2|ab| = 2\Big|(a\sqrt{\varepsilon})\frac{b}{\sqrt{\varepsilon}}\Big| \le a^2\varepsilon + \frac{b^2}{\varepsilon}$$

and (6.12) follows combining this with $a^2 + b^2 - 2|ab| \le (a+b)^2 \le a^2 + b^2 + 2|ab|$. Notice now that the right inequality in (6.12) leads to

$$(a+b)^4 \le \Big[(1+\varepsilon)a^2 + \Big(1 + \frac{1}{\varepsilon}\Big)b^2\Big]^2$$

and (6.13) follows applying the right inequality in (6.12) with a and b replaced by $(1+\varepsilon)a^2$ and $(1+1/\varepsilon)b^2$, respectively.

□

Proof of Theorem 6.1. Given an arbitrary integer $n \ge 2$, let $\Big[\dfrac{n}{2}\Big]$ be the integral part of $n/2$, write n as

$$n = m + m_1, \quad m = \Big[\frac{n}{2}\Big], \quad m_1 = n - \Big[\frac{n}{2}\Big], \tag{6.14}$$

and observe that

$$S_n = S_m + \sum_{r=1}^{m_1} Z_{m+r} = S_m + \tilde{S}_{m_1} + D, \tag{6.15}$$

where

$$\tilde{S}_{m_1} := \sum_{r=1}^{m_1} Z_{m+r,\, r}, \quad D := \sum_{r=1}^{m_1} [Z_{m+r} - Z_{m+r,\, r}].$$

Notice now that Lemma 3.3, (6.5), (6.6) and (6.8) together imply that

(a) S_m and $\tilde{S}_{m_1}$ are independent, and

(b) S_{m_1} and $\tilde{S}_{m_1}$ have the same distribution.

Moreover, (6.11) yields

(c) $|D| \le \sum_{r=1}^{m_1} 2C\beta^r \le 2C/(1-\beta) =: \tilde{C}$.

Next, let $\varepsilon > 0$ be arbitrary and notice that (6.13) and (6.15) yield that

$$S_n^4 \le (1+\varepsilon)^2(S_m + \tilde{S}_{m_1})^4 + (1+\varepsilon^{-1})^2 D^4,$$

and expanding $(S_m+\tilde S_{m_1})^4$ and using that $E[S_n]=E[S_{m_1}]=0$, facts (a)–(c) above lead to

$$\begin{aligned}E[S_n^4] &\le (1+\varepsilon)^2\left(E[S_m^4]+E[S_{m_1}^4]+6E[S_m^2]E[S_{m_1}^2]\right)+(1+\varepsilon^{-1})^2\tilde C^4\\ &\le (1+\varepsilon)^2\left(E[S_m^4]+E[S_{m_1}^4]+6mm_1B_2^2\right)+(1+\varepsilon^{-1})^2\tilde C^4\\ &\le (1+\varepsilon)^2\left(E[S_m^4]+E[S_{m_1}^4]+3n^2B_2^2/2\right)+(1+\varepsilon^{-1})^2\tilde C^4\end{aligned}$$

where B_2 is as in Lemma 6.2(ii) and the inequality $mm_1\le n^2/4$ was used in the last step; see (6.14). It follows that

$$\frac{E[S_n^4]}{n^2}\le(1+\varepsilon)^2\left(\frac{m^2}{n^2}\frac{E[S_m^4]}{m^2}+\frac{m_1^2}{n^2}\frac{E[S_{m_1}^4]}{m_1^2}\right)+R(\varepsilon) \tag{6.16}$$

where $R(\varepsilon):=(1+\varepsilon^{-1})^2\tilde C^4+3(1+\varepsilon)^2B_2^2/2$. Now set

$$M_k:=\sup_{n\le 2^k}\frac{E[S_n^4]}{n^2},\quad k=0,1,2,\ldots. \tag{6.17}$$

Let $k>0$ be fixed. When $n=2,3,\ldots,2^k$, the positive integers m and m_1 in (6.14) do not exceed 2^{k-1}, so that (6.16) implies that

$$\begin{aligned}\frac{E[S_n^4]}{n^2}&\le(1+\varepsilon)^2\left(\frac{m^2}{n^2}M_{k-1}+\frac{m_1^2}{n^2}M_{k-1}\right)+R(\varepsilon)\\ &=(1+\varepsilon)^2\left(1-\frac{2mm_1}{(m+m_1)^2}\right)M_{k-1}+R(\varepsilon),\quad n=2,3,\ldots 2^k.\end{aligned}$$

Also, from (6.14) it is not difficult to see that $2mm_1\ge n^2/2-1/2$, and then $(1-2mm_1/n^2)\le 1/2+1/2n^2\le 5/8$ for $n\ge 2$, so that $E[S_n^4]/n^2\le 5(1+\varepsilon)^2M_{k-1}/8+R(\varepsilon)$ in this case. It follows that

$$\frac{E[S_n^4]}{n^2}\le\frac{5(1+\varepsilon)^2}{8}M_{k-1}+R(\varepsilon)+E[S_1^4],\quad n=1,2,\ldots,2^k,$$

and then

$$M_k\le\frac{5(1+\varepsilon)^2}{8}M_{k-1}+R(\varepsilon)+E[S_1^4]\le\frac{5(1+\varepsilon)^2}{8}M_k+R(\varepsilon)+E[S_1^4].$$

Selecting $\varepsilon>0$ such that $r(\varepsilon):=5(1+\varepsilon)^2/8<1$, this yields that

$$M_k\le(R(\varepsilon)+|E[S_1^4])/(1-r(\varepsilon))<\infty;$$

since this holds for every positive integer k, the conclusion follows from (6.17). □

7. Variance convergence

As a consequence of Theorem 6.1 (or Lemma 6.2(ii)), the variance of $S_n/\sqrt{n}$ stays bounded as n increases. The objective of this section is to establish the second basic result that will be used in the proof of Theorem 2.3, namely, that the second moments of the variables $S_n/\sqrt{n}$ form a convergent sequence.

Theorem 7.1.

(i) $\lim_{n\to\infty} \dfrac{E[S_n^2]}{n} =: v$ *exists.*

Moreover,

(ii) v *does not depend on the initial state* $X_0 = \mathbf{x}$.

The proof of this theorem uses the following two lemmas. The first one is an extension of Lemma 6.2(ii).

Lemma 7.1. $\sup_{n,m:\, n>m>0} \dfrac{E[(S_n - S_m)^2]}{n-m} =: B_2' < \infty.$

Proof. Let m be a fixed positive integer, and select $n > m$. Observe that

$$S_n - S_m = \sum_{t=1}^{n-m} Z_{m+t} = S'_{n-m} + D \tag{7.1}$$

where

$$S'_{n-m} = \sum_{t=1}^{n-m} Z_{m+t,t}, \qquad D = \sum_{t=1}^{n-m} [Z_{m+t} - Z_{m+t,t}].$$

By (6.11), $|Z_{m+t} - Z_{m+t,t}| \le 2C\beta^t$, so that $|D| \le 2C/(1-\beta) = \tilde{C}$, whereas via Lemma 3.3(i) and (6.5)–(6.8), it follows that S'_{n-m} and S_{n-m} have the same distribution, so that $E[S'^2_{n-m}] = E[S^2_{n-m}]$. Thus, using Lemma 6.3 with $\varepsilon = 1$, (7.1) leads to

$$E[(S_n - S_m)^2] \le 2E[(S'_{n-m})^2] + 2E[D^2] \le 2E[S^2_{n-m}] + 2\tilde{C}^2.$$

Observing that $E[S^2_{n-m}] \le (n-m)B_2$, by Lemma 6.2(ii), it follows that $E[(S_n - S_m)^2] \le 2(n-m)B_2 + 2\tilde{C}^2$, so that $E[(S_n - S_m)^2]/(n-m) \le 2B_2 + 2\tilde{C}^2$, and the conclusion follows, since the positive integers n and m with $n > m$ are arbitrary in this argument.

□

Lemma 7.2. *Let* $\varepsilon > 0$ *be arbitrary. For positive integers* n *and* m,

$$\left|E[S^2_{nm}] - mE[S^2_n]\right| \le \varepsilon nmB_2'' + m\tilde{C}^2\left(1 + \frac{1}{\varepsilon}\right) + 2m\sqrt{n}\tilde{C}\sqrt{B_2''} \tag{7.2}$$

where

$$B_2'' = \max\{B_2, B_2'\}$$

and B_2 *and* B_2' *are as in Lemmas* 6.2(ii) *and* 7.1, *respectively, and, as before,* $\tilde{C} = 2C/(1-\beta)$.

PROOF. Set $S_0 := 0$, write $S_{nm} = \sum_{t=0}^{nm} Z_t$ as

$$S_{nm} = \sum_{j=1}^{m} \tilde{S}_j, \quad \text{where} \quad \tilde{S}_j = \sum_{i=1}^{n} Z_{(j-1)n+i} = S_{jn} - S_{(j-1)n}, \quad j = 1, 2, \ldots, m, \tag{7.3}$$

and notice that

$$E[S_{nm}^2] = \sum_{j=1}^{m} E[\tilde{S}_j^2] + 2 \sum_{m \geq t > r \geq 1} E[\tilde{S}_t \tilde{S}_r], \tag{7.4}$$

and

$$E[\tilde{S}_j^2] \leq n B_2'', \tag{7.5}$$

by Lemmas 6.2(ii) and 7.1 and the definition of B_2''. Next, observe that

$$\begin{aligned}\sum_{t=r+1}^{m} E[\tilde{S}_t \tilde{S}_r] &= E\Big[\Big(\sum_{t=r+1}^{m} \tilde{S}_t\Big)\tilde{S}_r\Big] = E\Big[\Big(\sum_{j=1}^{nm-nr} Z_{nr+j}\Big)\tilde{S}_r\Big] \\ &= E\Big[\Big(\sum_{j=1}^{nm-nr} Z_{nr+j,\, j}\Big)\tilde{S}_r\Big] + E[D\tilde{S}_r]\end{aligned}$$

where $D = \sum_{j=1}^{nm-nr} [Z_{nr+j} - Z_{nr+j,\, j}]$; by (6.11), $|D| \leq 2C/(1-\beta) = \tilde{C}$, and (7.5) yields

$$|E[D\tilde{S}_r]| \leq \tilde{C} E[|\tilde{S}_r|] \leq \tilde{C} E[\tilde{S}_r^2]^{1/2} \leq \tilde{C}\sqrt{n B_2''}.$$

On the other hand, $\sum_{j=1}^{nm-nr} Z_{nr+j,j}$ is a function of the matrices M_t with $t \geq nr+1$, whereas $\tilde{S}_r$ depends on the matrices M_t with $t \leq nr$, Thus, Assumption 2.1 and (6.7) yield that

$$E\Big[\Big(\sum_{j=1}^{nm-nr} Z_{nr+j,\, j}\Big)\tilde{S}_r\Big] = 0.$$

The last three displays together imply that

$$\Big|\sum_{t=r+1}^{m} E[\tilde{S}_t \tilde{S}_r]\Big| \leq \tilde{C}\sqrt{n B_2''},$$

so that

$$\Big|\sum_{m \geq t > r \geq 1} E[\tilde{S}_t \tilde{S}_r]\Big| = \Big|\sum_{r=1}^{m-1} \sum_{t=r+1}^{m} E[\tilde{S}_t \tilde{S}_r]\Big| \leq \sum_{r=1}^{m-1} \Big|\sum_{t=r+1}^{m} E[\tilde{S}_t \tilde{S}_r]\Big| \leq \sum_{r=1}^{m-1} \tilde{C}\sqrt{n B_2''},$$

and then

$$\Big|\sum_{m \geq t > r \geq 1} E[\tilde{S}_t \tilde{S}_r]\Big| \leq m\sqrt{n}\tilde{C}\sqrt{B_2''}. \tag{7.6}$$

Now define

$$\hat{S}_j = \sum_{i=1}^{n} Z_{(j-1)n+i,i}, \quad j = 1, 2, \ldots, m,$$

and notice the following properties (a), (b):

(a) From (3.10), (3.11) and (6.6), it follows that $\hat{S}_1 = F(M_1, \ldots, M_n)$ for a certain function F, and, moreover, this function F also satisfies that $\hat{S}_j = F(M_{(j-1)n+1}, \ldots, M_{nj})$, $j = 2, \ldots, m$. Thus, $\hat{S}_1, \ldots, \hat{S}_m$ are iid, by Assumption 2.1. Also, via (3.12) and (6.6), it follows that $\hat{S}_1 = S_n$, and then all the $\hat{S}_j$ have the same distribution as S_n.

(b) $D_j = \tilde{S}_j - \hat{S}_j = \sum_{i=1}^{n} [Z_{(j-1)n+i} - Z_{(j-1)n+i,\, i}]$ satisfies $|D_j| \le 2C/(1-\beta) = \tilde{C}$, by (6.11). Now let $\varepsilon > 0$ be arbitrary and notice that, since $\tilde{S}_j = \hat{S}_j + D_j$, Lemma 6.3 implies

$$\Big(1 - \frac{1}{\varepsilon}\Big)\tilde{C}^2 + (1-\varepsilon)\hat{S}_j^2 \le \tilde{S}_j^2 \le \Big(1 + \frac{1}{\varepsilon}\Big)\tilde{C}^2 + (1+\varepsilon)\hat{S}_j^2.$$

Taking the expectation and using property (a) above, it follows that

$$\Big(1 - \frac{1}{\varepsilon}\Big)\tilde{C}^2 + (1-\varepsilon)E[S_n^2] \le E[\tilde{S}_j^2] \le \Big(1 + \frac{1}{\varepsilon}\Big)\tilde{C}^2 + (1+\varepsilon)E[S_n^2].$$

Therefore,

$$m\Big(1 - \frac{1}{\varepsilon}\Big)\tilde{C}^2 + (1-\varepsilon)mE[S_n^2] \le \sum_{j=1}^{m} E[\tilde{S}_j^2] \le m\Big(1 + \frac{1}{\varepsilon}\Big)\tilde{C}^2 + m(1+\varepsilon)E[S_n^2],$$

and then

$$\begin{aligned}\Big|\sum_{j=1}^{m} E[\tilde{S}_j^2] - mE[S_n^2]\Big| &\le \varepsilon m E[S_n^2] + m\Big(1 + \frac{1}{\varepsilon}\Big)\tilde{C}^2 \\ &\le \varepsilon mnB_2 + m\Big(1 + \frac{1}{\varepsilon}\Big)\tilde{C}^2 \le \varepsilon mnB_2'' + m\Big(1 + \frac{1}{\varepsilon}\Big)\tilde{C}^2,\end{aligned}$$

where Lemma 6.2(ii) was used to set the second inequality. Combining this with (7.6) and (7.4), inequality (7.2) follows.

□

Proof of Theorem 7.1. (i) Let $\varepsilon > 0$ be arbitrary but fixed. In (7.2) interchange the roles of n and m to obtain

$$\big|E[S_{nm}^2] - nE[S_m^2]\big| \le \varepsilon nmB_2'' + n\tilde{C}^2\Big(1 + \frac{1}{\varepsilon}\Big) + 2n\sqrt{m}\tilde{C}\sqrt{B_2''}$$

which together with (7.2) yields that for every $n, m \in \mathbb{N}$

$$\big|mE[S_n^2] - nE[S_m^2]\big| \le \varepsilon 2nmB_2'' + (n+m)\tilde{C}^2\Big(1 + \frac{1}{\varepsilon}\Big) + 2(n\sqrt{m} + m\sqrt{n})\tilde{C}\sqrt{B_2''}$$

so that

$$\left|\frac{E[S_n^2]}{n} - \frac{E[S_m^2]}{m}\right| \leq 2\varepsilon B_2'' + \Big(\frac{1}{m} + \frac{1}{n}\Big)\tilde{C}^2\Big(1 + \frac{1}{\varepsilon}\Big) + 2\Big(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\Big)\tilde{C}\sqrt{B_2''}.$$

Consequently,

$$\limsup_{n\to\infty,\ m\to\infty}\left|\frac{E[S_n^2]}{n} - \frac{E[S_m^2]}{m}\right| \leq 2\varepsilon B_2'';$$

since $\varepsilon > 0$ is arbitrary, this yields that

$$\limsup_{n\to\infty,\ m\to\infty}\left|\frac{E[S_n^2]}{n} - \frac{E[S_m^2]}{m}\right| = 0,$$

i.e., $\{E[S_n^2]/n\}$ is a Cauchy sequence, and part (i) follows.

(ii) Let $\mathbf{y} \in \mathcal{P}_k$ be arbitrary but fixed, and let $\{X_n'\}$ be the process in Definition 2.1 corresponding to the initial state $X_0 = \mathbf{y}$. By part (i) applied to $\{X_n'\}$, $\lim_{n\to\infty} E[S_n'^2]/n = v'$ exists and is finite, where S_n' is given by (6.2) with X_n' instead of X_n, and the conclusion asserts that $v = v'$. To prove this equality, notice that Lemma 3.4(ii) and (6.4) together yield that $|f(X_n) - f(X_n')| \leq LK\beta^n$ for $n \geq 2N$ so that $|E[f(X_n)] - E[f(X_n')]| \leq LK\beta^n$ in this case. It follows that for $n \geq 2N$,

$$\begin{aligned}|S_n - S_n'| &= \Big|\sum_{t=1}^{n}\{(f(X_n) - E[f(X_n)]) - (f(X_n') - E[f(X_n)])'\}\Big| \\ &\leq \sum_{t=1}^{2N} 4\|f\| + \sum_{n=2N+1}^{n} 2KL\beta^n \leq 8N\|f\| + 2KL/(1-\beta) = \tilde{D}\end{aligned}$$

and then $|S_n| \leq |S_n'| + \tilde{D}$. From this point, an application of Lemma 6.3(i) yields that, for each $\varepsilon > 0$,

$$S_n^2 \leq (1+\varepsilon)S_n'^2 + (1+\varepsilon^{-1})\tilde{D}^2, \quad n \geq 2N,$$

and it follows that

$$v = \lim_{n\to\infty}\frac{1}{n}E[S_n^2] \leq (1+\varepsilon)\lim_{n\to\infty}\frac{1}{n}E[S_n'^2] = (1+\varepsilon)v'.$$

Since $\varepsilon > 0$ is arbitrary, this yields that $v \leq v'$. Via a similar argument it follows that $v' \leq v$, and then $v = v'$, completing the proof.

□

8. Weak limits and uniform differentiability

In this section Theorems 6.1 and 7.1 are used to study the set of weak limits of the distributions of the normalized averages $S_n/\sqrt{n}$. Before going any further, it is convenient to introduce the following notation.

DEFINITION 8.1.

(i) For each $n \in \mathbb{N}$, let $\nu_n \in \mathbb{P}(\mathbb{R})$ be the distribution of $S_n/\sqrt{n}$, that is,

$$\nu_n(A) = P[S_n/\sqrt{n} \in A], \quad A \in \mathcal{B}(\mathbb{R}).$$

(ii) The class $\mathcal{W}$ consists of all measures $\nu \in \mathbb{P}(\mathbb{R})$ such that, for some subsequence $\{\nu_{n_r}\}$,

$$\nu_{n_r} \xrightarrow{\text{w}} \nu \quad \text{as } r \to \infty. \tag{8.1}$$

Notice that, by Markov's inequality and Lemma 6.2(ii),

$$\nu_n([-a,a]^c) \le \frac{1}{a^2 n} E[S_n^2] \le \frac{B_2}{a^2}, \quad a > 0,$$

so that the family $\{\nu_n\}$ is tight, and then the class $\mathcal{W}$ is nonempty, by Prohorov's theorem. Before stating the main result of this section, it is convenient to establish the following properties of the family $\mathcal{W}$.

LEMMA 8.1. *For each $\nu \in \mathcal{W}$, the following assertions hold (see Theorems* 6.1 *and* 7.1*):*

(i) $\int_{\mathbb{R}} x^4\, \nu(dx) \le B_4$ *and* $\int_{\mathbb{R}} |x|^3\, \nu(dx) \le (B_4)^{3/4}$*;*

(ii) $\int_{\mathbb{R}} x\, \nu(dx) = 0$ *and* $\int_{\mathbb{R}} x^2\, \nu(dx) = v$.

PROOF. Let $\nu \in \mathcal{W}$ be arbitrary but fixed, and let $\{a_m\}$ be a sequence of positive numbers diverging to ∞ and satisfying $\nu(\{-a_m\}) = \nu(\{a_m\}) = 0$ for every m. Also, let the subsequence $\{\nu_{n_k}\}$ be such that (8.1) holds, and notice that for positive integers m and i

$$\lim_{k\to\infty} \int_{-a_m}^{a_m} x^i\, \nu_{n_k}(dx) = \int_{-a_m}^{a_m} x^i\, \nu(dx). \tag{8.2}$$

(i) Since $B_4 \ge \int_{\mathbb{R}} x^4\, \nu_{n_k}(dx) \ge \int_{-a_m}^{a_m} x^4\, \nu_{n_k}(dx)$, by Theorem 6.1, the above displayed relation with $i = 4$ yields that $\int_{-a_m}^{a_m} x^4\, \nu(dx) \le B_4$, so that letting m go to ∞, the monotone convergence theorem implies that $\int_{\mathbb{R}} x^4\, \nu(dx) \le B_4$. From this point, Hölder's inequality leads to $\int_{\mathbb{R}} |x|^3\, \nu(dx) \le \left(\int_{\mathbb{R}} |x|^4\, \nu(dx)\right)^{3/4} \le B_4^{3/4}$.

(ii) Via Theorem 6.1 and part (i), it follows that for every integers $i, m, k > 0$ with $i \le 4$,

$$\int_{[-a_m,\, a_m]^c} |x|^i\; \nu_{n_k}(dx) \le \frac{1}{a_m^{4-i}} \int_{[-a_m,\, a_m]^c} x^4\; \nu_{n_k}(dx) \le \frac{B_4}{a_m^{4-i}}$$

and

$$\int_{[-a_m,\,a_m]^c} |x|^i \ \nu(dx) \le \frac{1}{a_m^{4-i}} \int_{[-a_m,\,a_m]^c} x^4 \ \nu(dx) \le \frac{B_4}{a_m^{4-i}},$$

so that

$$\left|\int_{\mathbb{R}} x^i \,\nu_{n_k}(dx) - \int_{\mathbb{R}} x^i \,\nu(dx)\right| \le \left|\int_{-a_m}^{a_m} x^i \,\nu_{n_k}(dx) - \int_{-a_m}^{a_m} x^i \,\nu(dx)\right| + \frac{2B_4}{a_m^{4-i}}.$$

Taking the limit superior as k goes to ∞, (8.2) implies that

$$\limsup_{k\to\infty} \left|\int_{\mathbb{R}} x^i \,\nu_{n_k}(dx) - \int_{\mathbb{R}} x^i \,\nu(dx)\right| \le \frac{2B_4}{a_m^{4-i}};$$

since $\{a_m\}$ diverges to ∞ this yields

$$\lim_{k\to\infty} \left|\int_{\mathbb{R}} x^i \,\nu_{n_k}(dx) - \int_{\mathbb{R}} x^i \,\nu(dx)\right| = 0, \quad i = 1, 2,$$

and using that $\int_{\mathbb{R}} x\,\nu_n(dx) = 0$ for every n, by (6.3) and Definition 8.1(i), as well as

$$\lim_{n\to\infty} \int_{\mathbb{R}} x^2 \,\nu_n(dx) = v,$$

by Theorem 7.1, it follows that $\int_{\mathbb{R}} x\,\nu(dx) = 0$ and $\int_{\mathbb{R}} x^2\,\nu(dx) = v$. □

For each $\nu \in \mathbb{P}(\mathbb{R})$, let

$$\varphi_\nu(t) = \int e^{itx}\,\nu(dx), \quad t \in \mathbb{R},$$

be the characteristic function of ν, so that $\varphi_\nu(0) = 1$ and the k-th derivative of $\varphi_\nu(t)$ equals $\int_{\mathbb{R}} i^k x^k e^{itx}\,\nu(dx)$ if the k-th moment of ν is finite. By Lemma 8.1, $\varphi_\nu{}'(t) = 0$ and $\varphi_\nu{}''(t) = -v$ for each $\nu \in \mathcal{W}$ so that L'Hospital's rule implies that

$$\lim_{t\to 0} \frac{\varphi_\nu(t) - 1 + vt^2/2}{t^2/2} = 0.$$

The main result of the section is the following theorem, stating that the above convergence is uniform in $\nu \in \mathcal{W}$.

THEOREM 8.1. *For each $t \in \mathbb{R}\setminus\{0\}$, set*

$$\Delta(t) := \sup_{\nu\in\mathcal{W}} \left|\frac{\varphi_\nu(t) - 1}{t^2/2} + v\right|. \tag{8.3}$$

With this notation,

$$\lim_{t\to 0} \Delta(t) = 0.$$

Proof. First, take the second order Taylor expansion of $\cos y$ and $\sin y$ around 0 to obtain

$$\cos y = 1 - \frac{y^2}{2} + \frac{y^3}{6}\sin y_1$$

and

$$\sin y = y - \frac{y^3}{6}\cos y_2$$

where y_1 and y_2 are points between 0 and y. Therefore,

$$e^{iy} = 1 + iy - \frac{y^2}{2} + y^3 R(y),$$

where $R(y) = [\sin y_1/6 - i\cos y_2/6]$, and then $|R(y)| \le 2/6 < 1$. Thus $e^{itx} = \cos(tx) + i\sin(tx) = 1 + itx - t^2x^2/2 + t^3x^3R(tx)$, where $|R(tx)| < 1$, so that Lemma 8.1 implies

$$\varphi_\nu(t) = \int_{\mathbb{R}} e^{itx}\,\nu(dx) = 1 - vt^2/2 + t^3\tilde{R}_\nu(t), \quad \nu\in\mathcal{W}, \quad t\in\mathbb{R}, \tag{8.4}$$

where $\tilde{R}_\nu(t) = \int_{\mathbb{R}} x^3 R(tx)\,\nu(dx)$ satisfies

$$|\tilde{R}_\nu(t)| \le \int_{\mathbb{R}} |x|^3|R(tx)|\nu(dx) \le \int_{\mathbb{R}} |x|^3\nu(dx) \le B_4^{3/4}, \quad \nu\in\mathcal{W}, \quad t\in\mathbb{R}, \tag{8.5}$$

by Lemma 8.1(i). Since (8.4) yields

$$\frac{\varphi_\nu(t) - 1}{t^2/2} + v = t\tilde{R}_\nu(t), \quad t \ne 0, \quad \nu\in\mathcal{W},$$

via (8.5) it follows that $\Delta(t) \le |t|B_4^{3/4}$, so that $\Delta(t)\to 0$ as $t\to 0$. □

9. Divisibility in the family of weak limits

This section is the last step before the proof of the central limit theorem. The main objective is to show the following divisibility result in the family $\mathcal{W}$: Given $m\in\mathbb{N}$, each $\nu\in\mathcal{W}$ can be expressed as the distribution of the normalized sample mean of m iid variables whose common distribution also belongs to $\mathcal{W}$. The precise result is now stated.

Theorem 9.1. *Let $\nu\in\mathcal{W}$ and the positive integer m be arbitrary but fixed. Then there exists $\rho\in\mathcal{W}$ such that*

$$\varphi_\nu(t) = \varphi_\rho\Big(\frac{t}{\sqrt{m}}\Big)^m.$$

PROOF. Let $\{n_k\}$ be a sequence of positive integers such that $n_k \to \infty$ and (8.1) holds so that

$$\frac{S_{n_k}}{\sqrt{n_k}} \xrightarrow{\mathrm{d}} \nu. \tag{9.1}$$

Setting $q_k = [n_k/m]$ and $n'_k = mq_k$, it follows that $0 \le n_k - n'_k \le m$, $|S_{n_k} - S_{n'_k}| \le 2(n_k - n'_k)\|f\|$ and $n'_k/n_k \to 1$, so that, by Slutsky's theorem, (9.1) yields $S_{n'_k}/\sqrt{n'_k} \xrightarrow{\mathrm{d}} \nu$. Therefore, without loss of generality, assume that

$$n_k = mq_k$$

and notice that

$$S_{n_k} = \sum_{t=1}^{n_k} Z_t = \sum_{i=1}^{m}\sum_{j=1}^{q_k} Z_{(i-1)q_k+j} = S_{q_k} + \sum_{i=2}^{m}\sum_{j=1}^{q_k} Z_{(i-1)q_k+j}\,.$$

Next, define

$$S^{(i)}_{q_k} = \sum_{j=1}^{q_k} Z_{(i-1)q_k+j,\ j}, \qquad D_i = \sum_{j=1}^{q_k}[Z_{(i-1)q_k+j} - Z_{(i-1)q_k+j,\ j}], \quad i = 2, 3, \ldots, m,$$

so that

$$U_{n_k} := \frac{S_{n_k}}{\sqrt{n_k}} - \sum_{i=2}^{m} \frac{D_i}{\sqrt{n_k}} = \frac{1}{\sqrt{m}}\left[\frac{S_{q_k}}{\sqrt{q_k}} + \sum_{i=2}^{m} \frac{S^{(i)}_{q_k}}{\sqrt{q_k}}\right]. \tag{9.2}$$

Now observe the following facts:

(a) By Lemma 3.3, (6.5) and (6.6), $S_{q_k}, S^{(2)}_{q_k}, \ldots, S^{(m)}_{q_k}$ are iid.

(b) By (6.11), $|D_i| \le \sum_{j=1}^{q_k} |Z_{(i-1)m+j} - Z_{(i-1)m+j,\ j}| \le \sum_{j=1}^{q_k} 2C\beta^j \le 2C/(1-\beta)$ P-a.s., so that

$$\sum_{i=2}^{m} \frac{D_i}{\sqrt{n_k}} \to 0 \quad P\text{-a.s.}$$

Now let $t \in \mathbb{R}$ be arbitrary but fixed. Combining this latter convergence with (9.1) and (9.2) it follows that $U_{n_k} \xrightarrow{\mathrm{d}} \nu$ as $k \to \infty$, so that, for each $t \in \mathbb{R}$

$$E[e^{itU_{n_k}}] \to \varphi_\nu(t) \quad \text{as } k \to \infty.$$

On the other hand, recalling that the distribution of $S_{q_k}/\sqrt{q_k}$ is ν_{q_k}, property (a) above and (9.2) together imply that

$$E[e^{itU_{n_k}}] = \varphi_{\nu_{q_k}}\Big(\frac{t}{\sqrt{m}}\Big)^m.$$

Selecting a subsequence if necessary, it can be assumed that $\nu_{q_k} \xrightarrow{\mathrm{w}} \rho \in \mathcal{W}$ as $k \to \infty$, by Prohorov's theorem, so that

$$\varphi_{\nu_{q_k}}\Big(\frac{t}{\sqrt{m}}\Big) \to \varphi_\rho\Big(\frac{t}{\sqrt{m}}\Big).$$

The last three displays together yield that

$$\varphi_\nu(t) = \varphi_\rho\Big(\frac{t}{\sqrt{m}}\Big)^m, \quad t \in \mathbb{R},$$

completing the proof.

□

10. Proof of the central limit theorem

In this section Theorem 2.3 is finally proved. The argument relies on the uniform differentiability and divisibility results in Theorems 8.1 and 9.1, respectively.

Proof of Theorem 2.3. (i) Using (6.3), this part follows directly from Theorem 7.1.

(ii) It will be shown that each member of the family $\mathcal{W}$ in Definition 8.1 coincides with the distribution $\mathcal{N}(0, v)$, or equivalently, that if $\nu \in \mathcal{W}$, then

$$\varphi_\nu(t) = e^{-vt^2/2}$$

for each $t \in \mathbb{R}$; since this latter equality always holds for $t = 0$, it is sufficient to verify the above equation for arbitrary but fixed $t \in \mathbb{R} \setminus \{0\}$ and $\nu \in \mathcal{W}$. Given a positive integer m, use Theorem 9.1 to find a distribution $\rho_m \in \mathcal{W}$ such that

$$\varphi_\nu(t) = \varphi_{\rho_m}\Big(\frac{t}{\sqrt{m}}\Big)^m, \tag{10.1}$$

and define

$$a_m(t) := \frac{\varphi_{\rho_m}(t/\sqrt{m}) - 1}{t^2/(2m)},$$

so that $1 + a_m(t)\frac{t^2}{2m} = \varphi_{\rho_m}(t/\sqrt{m})$, and (10.1) yields

$$\varphi_\nu(t) = \Big(1 + a_m(t)\frac{t^2}{2m}\Big)^m. \tag{10.2}$$

Notice now that

$$|a_m(t) + v| = \Big|\frac{\varphi_{\rho_m}(t/\sqrt{m}) - 1}{t^2/(2m)} + v\Big| \le \Delta(t/\sqrt{m}),$$

(see (8.3)), and then $\lim_{m\to\infty} |a_m(t) + v| = 0$, by Theorem 8.1, i.e.,

$$\lim_{m\to\infty} a_m(t) = -v.$$

Combining this convergence with (10.2), it follows that

$$\varphi_\nu(t) = \lim_{m\to\infty}\Big(1 + a_m(t)\frac{t^2}{2m}\Big)^m = e^{-vt^2/2},$$

so that each $\nu \in \mathcal{W}$ is the distribution $\mathcal{N}(0, v)$.

□

Acknowledgment

The authors wish to express their sincerest gratitude to the reviewer for his thoughtful reading of the original manuscript and for his helpful comments and suggestions.

References

[1] R. Atar and O. Zeitouni, Exponential stability for nonlinear filtering, *Ann. Inst. Henri Poincaré*, **33** (1997), 697–725.
[2] P. Bougerol and J. Lacroix, *Products of random matrices with applications to Schrödinger operators*, Progress in Probability and Statistics 8, Birkhäuser, Boston, 1985.
[3] R. Carmona and J. Lacroix, *Spectral Theory of Random Schrödinger Operators*, Birkhäuser, Zurich, 1990.
[4] R. Cavazos-Cadena, An alternative derivation of Birkhoff's formula for the contraction coefficient of a positive matrix, *Linear Algebra Appl.*, **375** (2003), 291–297.
[5] R. Cavazos-Cadena and D. Hernández-Hernández, Successive approximations in partially observable controlled Markov chains with risk-sensitive average criterion, *Stochastics*, **77** (2005), 537–568.
[6] A. Crisanti, G. Paladin and A. Vulpiani, *Products of Random Matrices in Statistical Physics*, Springer-Verlag, Berlin, 1993.
[7] W.H. Fleming and D. Hernández-Hernández, Risk-sensitive control of finite state machines on an infinite horizon II, *SIAM J. Control Optim.*, **37** (1999), 1048–1069.
[8] H. Furstenberg, Noncommuting random products, *Trans. Amer. Math. Soc.*, **108** (1963), 377–428.
[9] H. Furstenberg and H. Kesten, Products of random matrices, *Ann. Math. Statist.*, **31** (1960), 457–469.
[10] H. Furstenberg and Y. Kifer, Random products and measures on projective spaces, *Israel J. Math.*, **46** (1983), 12–32.
[11] H. Hennion, Limit theorems for random products of positive random matrices, *Ann. Probab.*, **25** (1997), 1545–1587.
[12] A. Mukherjea, Convergence in distribution of products of random matrices: a semigroup approach, *Trans. Amer. Math. Soc.*, **303** (1987), 395–411.
[13] V.I. Oseledec, A multiplicative ergodic theorem, Lyapunov characteristic numbers for dynamical systems, *Trans. Moscow Math. Soc.*, **19** (1968), 197–221.
[14] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer-Verlag, Berlin – New York, 1981.