NOTES ON THE MARTINGALE APPROACH TO CENTRAL LIMIT THEOREMS FOR MARKOV CHAINS

TIMO SEPPÄLÄINEN

Abstract. Notes on the Maxwell-Woodroofe martingale approach to central limit theorems for Markov chains.

1. Introduction

Suppose $\{X_k\}$ are independent and identically distributed random variables with values in some measurable space $(S, \mathcal{B})$, defined on a probability space $(\Omega, \mathcal{F}, P)$, with common distribution μ. Suppose $g : S \to \mathbb{R}$ is a measurable function such that

$$\int g\, d\mu = 0 \quad \text{and} \quad \sigma^2 = \int g^2\, d\mu < \infty.$$

Let

$$S_n = \sum_{k=0}^{n-1} g(X_k).$$

Then the basic central limit theorem says that the random variable $n^{-1/2} S_n$ converges weakly to a centered normal distribution with variance σ^2. Furthermore, define the process $Y_n(t) = n^{-1/2} S_{[nt]}$ for $t \in [0, \infty)$, with paths in the space $D[0, \infty)$. According to Donsker's invariance principle, the process Y_n converges weakly on D to the process σB where $B = \{B(t) : t \in [0, \infty)\}$ denotes standard Brownian motion.

The purpose of these notes is to develop results of this kind for the case where $\{X_k\}$ is a Markov chain. Presently we cover the approach of Maxwell-Woodroofe [1] with some small improvement from [2].

Notation. $\mathbb{Z}_+ = \{0, 1, 2, 3, \dots\}$ is the set of nonnegative integers, $\mathbb{N} = \{1, 2, 3, \dots\}$ the set of positive integers. $C, C_1, C_2, \dots$ denote constants that do not depend on the growing parameter of the context (often n) and whose precise value may change from line to line. The floor and ceiling of a real number are $[x] = \max\{n \in \mathbb{Z} : x \leq n\}$ and $\lceil x \rceil = \min\{n \in \mathbb{Z} : x \leq n\}$.

Date: November 3, 2006.

Author partially supported by National Science Foundation grant DMS-0402231.

2. A Martingale invariance principle

Let M_n be a mean 0 $\mathbb{R}^d$-valued vector martingale in L^2, with differences $Y_k = M_k - M_{k-1}$. A vector martingale simply means that each coordinate forms a real-valued martingale. Define the scaled process $\bar{M}_n(t) = n^{-1/2} M_{[nt]}$. Assume

$$\frac{1}{n}\sum_{k=1}^{n} E(\, Y_k Y_k^T \,|\, \mathcal{F}_{k-1}\,) \to \Gamma \quad \text{in probability}$$

for a symmetric, nonnegative definite $d \times d$ matrix Γ, and

$$\frac{1}{n}\sum_{k=1}^{n} E(\, |Y_k|^2 \mathbf{1}\{|Y_k| \geq \varepsilon\sqrt{n}\} \,|\, \mathcal{F}_{k-1}\,) \to 0 \quad \text{in probability.}$$

Then $\bar{M}_n$ converges weakly on the path space $D_{\mathbb{R}^d}[0,\infty)$ to a process $W = \{W(t) : t \geq 0\}$ which is a Brownian motion with diffusion matrix Γ. This last definition means that $W(0) = 0$, W has continuous paths, independent increments, and for $s < t$ the d-vector $W(t) - W(s)$ has Gaussian distribution with mean zero and covariance matrix $(t-s)\Gamma$. One can produce such a process by finding a matrix Λ such that $\Gamma = \Lambda\Lambda^T$, and by defining $W(t) = \Lambda B(t)$ where B is a d-dimensional standard Brownian motion.

3. Markov chain notation and basics

There is a measurable space $(S, \mathcal{B})$ which is the state space of the process. A transition probability p is a function $p(x, A)$ of $x \in S$ and $A \in \mathcal{B}$ and satisfies these properties:

(i) For each $A \in \mathcal{B}$, $p(x, A)$ is a measurable function of $x \in S$.
(ii) For each $x \in S$, $p(x, \cdot\,)$ is a probability measure on $(S, \mathcal{B})$.

The integral $\int f(y)\, p(x, dy)$ is a measurable function of x for any real or complex measurable function f for which the integrals are well-defined. In particular, the transition probability p defines two operators, one on the space $B(S)$ of bounded measurable functions f on $(S, \mathcal{B})$, the other on the space $M(S)$ of finite signed measures on $(S, \mathcal{B})$. Both are denoted by P, but distinguished by left and right notation:

$$Pf(x) = \int f(y)\, p(x, dy) \quad \text{and} \quad \mu P(A) = \int p(x, A)\, \mu(dx).$$

The functions and measures could also be complex-valued if desired.

The path space of the process is $\Omega = S^{\mathbb{Z}_+}$ with its product σ-algebra $\mathcal{F} = \mathcal{B}^{\otimes \mathbb{Z}_+}$. Given an initial state $x \in S$, the path measure P_x on $(\Omega, \mathcal{F})$ is the probability measure uniquely determined by the property

$$\begin{aligned} &P_x\{X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n\} \\ &\qquad = \mathbf{1}_{A_0}(x) \int p(x, dx_1)\, \mathbf{1}_{A_1}(x_1) \int p(x_1, dx_2)\, \mathbf{1}_{A_2}(x_2)\ \cdots \\ &\cdots \int p(x_{n-3}, dx_{n-2}) \mathbf{1}_{A_{n-2}}(x_{n-2}) \int p(x_{n-2}, dx_{1-1})\, \mathbf{1}_{A_{n-1}}(x_{n-1})\, p(x_{n-1}, A_n). \end{aligned} \tag{3.1}$$

The existence of P_x follows from Kolmogorov's Extension Theorem if $(S, \mathcal{B})$ is a Borel subset of a complete separable metric space with its Borel σ-algebra. However, the assumption that the transition probabilities exist allows one to do away with topological assumptions and prove the existence of P_x for an arbitrary measurable space $(S, \mathcal{B})$. This is called Tulcea's Extension Theorem. Expectation under P_x is denoted by E_x. So for a bounded measurable function G on Ω,

$$E_x[G] = \int_\Omega G(\omega)\, P_x(d\omega)$$

and this is again a measurable function of x.

Given an arbitrary probability measure μ on $(S, \mathcal{B})$, define

$$P_\mu(A) = \int P_x(A)\, \mu(dx)\,, \quad A \in \mathcal{F}.$$

Thus P_x is the special case of P_μ with $\mu = \delta_x$, point mass at x.

Let $\omega = (x_k)_{k \in \mathbb{Z}_+}$ denote a generic element of Ω. The coordinate random variables on Ω are defined by $X_k(\omega) = x_k$. The filtration $\{\mathcal{F}_k\}$ is defined by $\mathcal{F}_n = \sigma\{X_0, \dots, X_n\}$. The previous construction can now be summarized by this statement: on the probability space $(\Omega, \mathcal{F}, P_\mu)$ the process $\{X_n\}$ is a Markov chain with respect to the filtration $\{\mathcal{F}_n\}$, with initial distribution μ and with transition probability p.

This last statement has a meaning that does not depend on the particular construction. Namely, any S-valued process $\{X_n\}$ on some probability space $(\Omega, \mathcal{F}, P)$ is said to be a Markov chain with respect to the filtration $\{\mathcal{F}_n\}$, with initial distribution μ and transition probability p, if these three conditions are satisfied:

(i) Process $\{X_n\}$ is adapted to the filtration $\{\mathcal{F}_n\}$.
(ii) $P\{X_0 \in A\} = \mu(A)$ for $A \in \mathcal{B}$.
(iii) For any $n \in \mathbb{N}$ and $A \in \mathcal{B}$,

$$P(X_{n+1} \in A | \mathcal{F}_n) = p(X_n, A) \quad P\text{-almost surely.} \tag{3.2}$$

A probability measure π on $(S, \mathcal{B})$ is invariant for p if $\pi = \pi P$, and reversible for p if

$$\int g\, Pf\, d\pi = \int f\, Pg\, d\pi$$

for all $f, g \in B(S)$. Reversibility implies invariance. Invariance is equivalent to the condition that the path measure P_π is invariant under the shift mapping θ on Ω defined by $(\theta\omega)_k = \omega_{k+1}$. Invariance under θ means that $P_\pi(A) = P_\pi(\theta^{-1}A)$ for all $A \in \mathcal{F}$. Another way of saying this is that the coordinate process is stationary, which means the distributional equality

$$(X_0, X_1, X_2, \dots) \overset{d}{=} (X_1, X_2, X_3, \dots).$$

Reversibility is equivalent to the condition

$$(X_0, X_1, X_2, \dots, X_n) \overset{d}{=} (X_n, X_{n-1}, X_{n-2}, \dots X_0) \qquad \text{for all } n \in \mathbb{N}.$$

This means that the original and the time-reversed process have the same distribution.

An invariant distribution π is ergodic for p if the path measure P_π is ergodic under θ. This means that, in addition to θ-invariance, $P_\pi(A) \in \{0, 1\}$ for all θ-invariant events A. These are the members of the invariant σ-algebra

$$\mathcal{I} = \{A \in \mathcal{F} : \theta^{-1}A = A\}.$$

Equivalently, π is ergodic for p if it is an extreme point of the convex set of p-invariant probability measures on $(S, \mathcal{B})$.

In the invariant case the ergodic theorem says that

$$\frac{1}{n}\sum_{k=0}^{n-1} G \circ \theta^k \longrightarrow E_\pi[G|\mathcal{I}] \quad P_\pi\text{-almost surely}$$

for any $G \in L^1(P_\pi)$. Under the ergodicity assumption $\mathcal{I}$ is trivial under P_π, and so $E_\pi[G|\mathcal{I}] = E_\pi[G]$.

4. The martingale approach to Markov chain central limit theorems

Let $\{X_n\}$ be a Markov chain with transition p. Let h be a measurable function on the state space such that $h(X_k)$ is integrable for all k. The initial distribution is arbitrary at this point. Then there is a standard way to produce a martingale associated to h. Namely, by the Markov property

$$E\big[h(X_k) - Ph(X_{k-1})\,\big|\mathcal{F}_{k-1}\big] = 0,$$

so the process

$$M_n = \sum_{k=1}^{n} \big\{h(X_k) - Ph(X_{k-1})\big\}$$

is a mean-zero martingale.

Next we turn this into a derivation of a central limit theorem for a sum of the type

$$S_n = S_n(g) = \sum_{k=0}^{n-1} g(X_k)$$

where g is a given function on the state space S. Assume now that π is an ergodic invariant measure for the Markov transition p. Given $g \in L^2(\pi)$, suppose there exists $u \in L^2(\pi)$ such that

$$u(x) - Pu(x) = g(x) \quad \text{for } x \in S, \tag{4.1}$$

or $(I - P)u = g$ in operator short-hand. I denotes the identity operator on $L^2(\pi)$. Then it is straightforward to deduce this theorem from the martingale invariance principle.

THEOREM 4.1. *Assume* (4.1). *Then the process* $Y_n(t) = n^{-1/2} S_{[nt]}(g)$ $(t \geq 0)$ *converges weakly to* σB*, where the limiting variance is given by*

$$\sigma^2 = \int (u^2 - (Pu)^2)\, d\pi.$$

If the Poisson equation (4.1) cannot be solved for a given g, one can instead solve the approximate Poisson equation

$$(1 + \varepsilon)u_\varepsilon - Pu_\varepsilon = g \tag{4.2}$$

where $\varepsilon > 0$. This equation can always be solved by

$$u_\varepsilon = \sum_{k=1}^{\infty} \frac{P^{k-1}g}{(1+\varepsilon)^k}. \tag{4.3}$$

This series that defines u_ε converges in $L^2(\pi)$ because P is a contraction: $\|Pg\|_2 \leq \|g\|_2$. The solvability of (4.2) follows also from basic spectral considerations. Since P is a contraction on $L^2(\pi)$, its operator norm satisfies $\|P\| \leq 1$. Since $Pc = c$ for any constant function c, $\|P\| = 1$ and $\lambda = 1$ is an eigenvalue of P. Thus the spectral radius $r(P) = 1$. Consequently $1 + \varepsilon$ lies in the resolvent set of P, which is the complement of the spectrum $\sigma(P)$. This means that the inverse $(1 + \varepsilon - P)^{-1}$ exists as a bounded operator on $L^2(\pi)$, and one can define

$$u_\varepsilon = (1 + \varepsilon - P)^{-1} g.$$

Let us attempt to build on this by the same martingale ideas as before. Write

$$\begin{aligned} S_n(g) &= \sum_{k=0}^{n-1} \big((1+\varepsilon)u_\varepsilon(X_k) - Pu_\varepsilon(X_k) \big) \\ &= \sum_{k=1}^{n} \big(u_\varepsilon(X_k) - Pu_\varepsilon(X_{k-1}) \big) + \big(u_\varepsilon(X_0) - u_\varepsilon(X_n) \big) + \varepsilon S_n(u_\varepsilon). \end{aligned}$$

Rename the terms on the right above to write this as

$$S_n(g) = M_n(\varepsilon) + R_n(\varepsilon) + \varepsilon S_n(u_\varepsilon). \tag{4.4}$$

The decomposition above suggests this strategy. First let $\varepsilon \to 0$ to obtain a representation

$$S_n(g) = M_n + R_n \tag{4.5}$$

where M_n is an L^2 martingale, and then try to show that $n^{-1/2}(M_n + R_n)$ satisfies a central limit theorem. We can identify three steps that achieve this.

(a) Show that $\varepsilon S_n(u_\varepsilon) \to 0$ in L^2 as $\varepsilon \to 0$, for each n.

(b) Show that the limits $R_n = \lim_{\varepsilon\to 0} R_n(\varepsilon)$ exist in L^2, and $n^{-1/2}R_n \to 0$ in probability.

(c) Show that the limits $M_n = \lim_{\varepsilon\to 0} M_n(\varepsilon)$ exist in L^2. Since each process $\{M_n(\varepsilon) : n \in \mathbb{N}\}$ is a martingale with stationary, ergodic increments, M_n inherits these properties (exercise below).

Decomposition (4.5), items (b) and (c), and Theorem imply then that $n^{-1/2}S_n(g)$ satisfies the functional central limit theorem with limiting variance $\sigma^2 = E(M_1^2)$.

The task ahead is to investigate hypotheses under which this three-step program can be realized.

Exercise 4.2. Let $\{\mathcal{F}_n\}$ be a filtration on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that for each k, $\{M_n^{(k)} : n \geq 1\}$ is an L^2-martingale with respect to $\{\mathcal{F}_n\}$. Assume the limit $M_n = \lim_{k\to\infty} M_n^{(k)}$ exists in L^2 for each n. Then $\{M_n\}$ is also an L^2-martingale with respect to $\{\mathcal{F}_n\}$. *Hint:* Consider limits of $E[YM_n^{(k)}]$ for bounded $\mathcal{F}_{n-1}$-measurable Y.

5. Moment assumptions on the resolvent

In this section we make the following standing assumption and prove a central limit theorem for an ergodic Markov chain under this hypothesis.

Assumption 5.1. There exist constants $0 < C < \infty$ and $0 < \alpha < 1/2$ such that

$$\Big\| \sum_{k=0}^{n-1} P^k g \Big\|_{L^2(\pi)} \leq Cn^\alpha \qquad \text{for all } n \in \mathbb{N}. \tag{5.1}$$

The key point is the strict inequality $\alpha < 1/2$. When convenient we abbreviate $V_0 g = 0$ and

$$V_n g = \sum_{k=0}^{n-1} P^k g \qquad \text{for } n \geq 1.$$

We shall use $\|\cdot\|$ to denote L^2 norm without specifying the space when no confusion should arise.

Lemma 5.2. *For* $0 < \varepsilon \leq 1$, $\|u_\varepsilon\|_{L^2(\pi)} \leq C\varepsilon^{-\alpha}$.

Proof. First sum by parts:

$$u_\varepsilon = \sum_{k=0}^{\infty} \frac{V^{k+1}g - V^k g}{(1+\varepsilon)^{k+1}} = \varepsilon \sum_{k=0}^{\infty} \frac{V^k g}{(1+\varepsilon)^{k+1}}\,.$$

Note that $\log(1+\varepsilon) \geq C_1\varepsilon$ for $0 < \varepsilon \leq 1$, and bound as follows, using (5.1):

$$\begin{aligned}\|u_\varepsilon\| &\leq C\varepsilon \sum_{k=0}^{\infty} \frac{k^\alpha}{(1+\varepsilon)^{k+1}} \leq C\varepsilon \int_0^\infty (1+\varepsilon)^{-x} x^\alpha\,dx \\ &= C\varepsilon(\log(1+\varepsilon))^{-1-\alpha} \int_0^\infty e^{-y} y^\alpha\,dy \leq C\varepsilon \cdot \varepsilon^{-1-\alpha}.\end{aligned}$$

□

On the product space S^2 define the measure $\pi_2 = \pi \otimes P$ by

$$\pi_2(dx, dy) = \pi(dx)p(x, dy).$$

In other words, π_2 is the distribution of any pair (X_k, X_{k+1}) under P_π. Define the $L^2(\pi_2)$ function H_ε on S^2 by

$$H_\varepsilon(x, y) = u_\varepsilon(y) - Pu_\varepsilon(x).$$

Exercise 5.3. Use the invariance of π and the Markov property to show that for any two functions $f, h \in L^2(\pi)$,

$$\begin{aligned}&E_\pi\big[(f(X_1) - Pf(X_0) - h(X_1) + Ph(X_0))^2\big] \\ &\quad = \int_S (I+P)(h-f)\cdot(I-P)(h-f)\,d\pi.\end{aligned} \tag{5.2}$$

Lemma 5.4. *The limit* $H = \lim_{\varepsilon\to 0} H_\varepsilon$ *exists in* $L^2(\pi_2)$.

Proof. In the following calculation, first use (5.2), then Schwartz inequality, next $(I-P)u_\varepsilon = -\varepsilon u_\varepsilon + g$, then the contraction property of P as an operator on $L^2(\pi)$, and finally Lemma 5.2.

$$\begin{aligned}E_\pi\big[(H_\varepsilon - H_\delta)^2\big] &= \int_S (I+P)(u_\varepsilon - u_\delta)\cdot(I-P)(u_\varepsilon - u_\delta)\,d\pi \\ &\leq \|(I+P)(u_\varepsilon - u_\delta)\| \cdot \|(I-P)(u_\varepsilon - u_\delta)\| \\ &= \|(I+P)(u_\varepsilon - u_\delta)\| \cdot \|\varepsilon u_\varepsilon - \delta u_\delta\| \\ &\leq 2\,\|u_\varepsilon - u_\delta\| \cdot \|\varepsilon u_\varepsilon - \delta u_\delta\| \leq 2\big(\|u_\varepsilon\| + \|u_\delta\|\big)\big(\varepsilon\,\|u_\varepsilon\| + \delta\,\|u_\delta\|\big) \\ &\leq 2C(\varepsilon^{-\alpha} + \delta^{-\alpha})(\varepsilon^{1-\alpha} + \delta^{1-\alpha}).\end{aligned}$$

With $\delta_k = 2^{-k}$,

$$\begin{aligned}\sup_{\varepsilon\in[\delta_k,\delta_{k+1}]} \|H_\varepsilon - H_{\delta_k}\| &\leq C\big(2^{(k+1)\alpha} + 2^{k\alpha}\big)^{1/2}\big(2^{-k+k\alpha} + 2^{-k+k\alpha}\big)^{1/2} \\ &\leq C2^{-(\frac{1}{2}-\alpha)k}.\end{aligned}$$

The last bound is summable over k. Hence if we apply it to $\varepsilon = \delta_{k+1}$ we can conclude that $H = \lim_{k\to\infty} H_{\delta_k}$ exists in $L^2(\pi_2)$. Another application of the same estimate shows that $H_\varepsilon \to H$ as $\varepsilon \to 0$. □

Return to the decomposition

$$S_n(g) = M_n(\varepsilon) + R_n(\varepsilon) + \varepsilon S_n(u_\varepsilon) \tag{5.3}$$

where

$$M_n(\varepsilon) = \sum_{k=0}^{n-1} H_\varepsilon(X_k, X_{k+1}) \quad \text{and} \quad R_n(\varepsilon) = u_\varepsilon(X_0) - u_\varepsilon(X_n).$$

Since

$$\|H_\varepsilon(X_k, X_{k+1}) - H(X_k, X_{k+1})\|_{L^2(P_\pi)} = \|H_\varepsilon - H\|_{L^2(\pi_2)}$$

it follows that for each n we have the $L^2(P_\pi)$-limit

$$M_n \equiv \sum_{k=0}^{n-1} H(X_k, X_{k+1}) = \lim_{\varepsilon\to 0} \sum_{k=0}^{n-1} H_\varepsilon(X_k, X_{k+1}) = \lim_{\varepsilon\to 0} M_n(\varepsilon).$$

By Lemma 5.2 $\|\varepsilon S_n(u_\varepsilon)\| \le n\varepsilon^{1-\alpha}$ for each n, and so $\varepsilon S_n(u_\varepsilon) \to 0$ in $L^2(P_\pi)$ as $\varepsilon \to 0$. The L^2-limit $R_n = \lim_{\varepsilon\to 0} R_n(\varepsilon)$ must then also exist since all other terms in (5.3) converge (for fixed n).

Thus after letting $\varepsilon \to 0$, (5.3) has turned into

$$S_n(g) = M_n + R_n \tag{5.4}$$

where M_n is an L^2 martingale, and the increment processes

$$\{M_n - M_{n-1} = H(X_{n-1}, X_n) : n \ge 1\} \quad \text{and} \quad \{R_n - R_{n-1} : n \ge 1\}$$

are stationary and ergodic. Next we derive a moment bound on the remaining error.

Lemma 5.5. $E_\pi\big[\,|R_n|^2\,\big] \le Cn^{2\alpha}$.

Proof. Write

$$R_n = S_n(g) - M_n = M_n(\varepsilon) - M_n + R_n(\varepsilon) + \varepsilon S_n(u_\varepsilon). \tag{5.5}$$

Choose k so that $2^{k-1} \le n \le 2^k$ and take $\varepsilon = \delta_k = 2^{-k}$ Then by an earlier calculation,

$$\begin{aligned}\|H_\varepsilon - H\|_{L^2(\pi_2)} &\le \sum_{j=k}^{\infty} \big\|H_{\delta_j} - H_{\delta_{j+1}}\big\|_{L^2(\pi_2)} \le C \sum_{j=k}^{\infty} 2^{-(\frac12-\alpha)j} \\ &\le C2^{-(\frac12-\alpha)k} \le Cn^{\alpha-\frac12}.\end{aligned}$$

By orthogonality of martingale increments and stationarity,

$$\|M_n(\varepsilon) - M_n\|^2_{L^2(P_\pi)} = n\,\|H_\varepsilon - H\|^2_{L^2(\pi_2)} \le Cn^{2\alpha}.$$

Next

$$\|R_n(\varepsilon)\|_{L^2(P_\pi)} \le 2\|u_\varepsilon\|_{L^2(\pi)} \le C\varepsilon^{-\alpha} \le Cn^{\alpha}$$

and

$$\|\varepsilon S_n(u_\varepsilon)\|_{L^2(P_\pi)} \le Cn\varepsilon^{1-\alpha} \le Cn^\alpha.$$

These estimates applied to the right-hand side of (5.5) give the conclusion. □

We can record a CLT for the stationary process. Define

$$\sigma^2 = \|M_1\|^2_{L^2(P_\pi)} = \int_{S^2} H^2\, d\pi_2. \tag{5.6}$$

THEOREM 5.6. *Let π be an ergodic invariant distribution for the Markov transition p on the measurable state space $(S, \mathcal{B})$, and let $g \in L^2(\pi)$ satisfy $\int g\, d\pi = 0$. Suppose Assumption 5.1 is in force. Then $n^{-1/2}S_n(g)$ converges weakly to a centered normal distribution with variance σ^2, and σ^2 is the limiting variance:*

$$\sigma^2 = \lim_{n\to\infty} E_\pi\big[\big(n^{-1/2}S_n(g)\big)^2\big]. \tag{5.7}$$

Proof. By Lemma 5.5 $n^{-1/2}R_n \to 0$ in L^2, so by (5.4) it suffices to prove the convergence for $n^{-1/2}M_n$. Utilizing the Markov property and the ergodic theorem,

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^n E_\pi[(M_k - M_{k-1})^2|\mathcal{F}_{k-1}] &= \frac{1}{n}\sum_{k=1}^n E_{X_{k-1}}[H(X_0,X_1)^2]\\ \underset{\text{a.s.}}{\longrightarrow} \int E_x[H(X_0,X_1)^2]\,\pi(dx) &= E_\pi[H(X_0,X_1)^2] = \sigma^2.\end{aligned}$$

Since $\varepsilon\sqrt{n} > A$ eventually for any fixed $0 < A < \infty$, we can bound as follows:

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^n E_\pi\big[(M_k - M_{k-1})^2\mathbf{1}\{|M_k - M_{k-1}| \ge \varepsilon\sqrt{n}\}\,\big|\,\mathcal{F}_{k-1}\big]\\ &\le \limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^n E_{X_{k-1}}\big[H(X_0,X_1)^2\mathbf{1}\{|H(X_0,X_1)| \ge A\}\big]\\ &\overset{\text{a.s.}}{=} E_\pi\big[H(X_0,X_1)^2\mathbf{1}\{|H(X_0,X_1)| \ge A\}\big] = \int_{|H|\ge A} H^2\, d\pi_2.\end{aligned}$$

The last quantity vanishes as $A \nearrow \infty$ since $H \in L^2(\pi_2)$. We have verified the conditions of a martingale central limit theorem.

For the limiting variance, (5.4), Lemma 5.5 and orthognality of martingale increments imply both

$$\sup_n \|n^{-1/2}S_n(g)\| \le 2\sup_n \|n^{-1/2}M_n\| + Cn^{2\alpha-1} < \infty$$

and

$$\|n^{-1/2}S_n(g) - n^{-1/2}M_n\| \to 0.$$

Limit (5.7) now follows from this general observation utilizing the Schwarz inequality:

$$|\,E(\eta^2) - E(\zeta^2)\,| = |\,E(\eta-\zeta)(\eta+\zeta)\,| \le \|\eta-\zeta\|(\|\eta\|+\|\zeta\|). \qquad \square$$

In order to improve Theorem 5.6 into a functional central limit theorem and to get results under a fixed starting state x rather than for the stationary process, next we strengthen the moment assumption on g by a tiny bit.

6. Stronger result from a stronger moment assumption

Recall the definition $Y_n(t) = n^{-1/2}S_{[nt]}(g)$ of the process Y whose paths are in the space $D[0,\infty)$, and also recall the definition 5.6 of σ^2. B denotes standard one-dimensional Brownian motion.

Theorem 6.1. *Let π be an ergodic invariant distribution for the Markov transition p on the measurable state space $(S,\mathcal{B})$. Let $p>2$, and $g \in L^p(\pi)$ satisfy $\int g\,d\pi = 0$. Suppose Assumption 5.1 is in force. Then for π-almost every x this happens: as $n\to\infty$, under the measure P_x the process Y_n converges weakly to σB on the path space $D[0,\infty)$.*

Observe first that the verifications of the hypotheses of the martingale invariance principle in the proof of Theorem 5.6 involved (countably many) limits that hold P_π-almost surely. Since

$$P_\pi(A) = \int_S P_x(A)\,\pi(dx)$$

these limits hold also P_x-almost surely, for π-almost every x. Thus the processes $\bar{M}_n(t) = n^{-1/2}M_{[nt]}$ satisfy the invariance principle under the measure P_x for π-almost every x. For each $0<T<\infty$

$$\sup_{0\le t\le T}|Y_n(t) - \bar{M}_n(t)| \le n^{-1/2}\max_{k\le nT+1}|R_k|.$$

Consequently in order to transfer the invariance principle from $\bar{M}_n$ to Y_n, we need to show that for π-almost every x,

$$n^{-1/2}\max_{k\le n}|R_k| \underset{n\to\infty}{\longrightarrow} 0 \ \text{ in } P_x\text{-probability.} \tag{6.1}$$

Limit (6.1) is done in several steps.

Fix $0<\gamma<1$, and let $m=\lceil n^{1-\gamma}\rceil$ and $\ell = \lceil n^\gamma \rceil$. Since $R_n = S_n(g) - M_n$, we can decompose

$$n^{-1/2}\max_{i\le n}|R_i| \le n^{-1/2}\max_{0\le k\le m}|R_{k\ell}| \tag{6.2}$$

$$+\, n^{-1/2}\max_{0\le k<m}\ \max_{k\ell\le i\le (k+1)\ell}|M_i - M_{k\ell}| \tag{6.3}$$

$$+\, n^{-1/2}\max_{0\le k<m}\ \max_{k\ell\le i\le (k+1)\ell}|S_i(g) - S_{k\ell}(g)|\,. \tag{6.4}$$

We give separate arguments for the three lines above. The terms on (6.2) and (6.4) will be handled by proving almost sure convergence under P_π along a subsequence.

We start with line (6.3) where we can apply the martingale invariance principle. First note that, for any $\delta > 0$ and large enough n,

$$n^{-1/2} \max_{0\le k<m} \max_{k\ell \le i \le (k+1)\ell} |M_i - M_{k\ell}| \le \sup_{s,t\le 2: |t-s|\le n^{\gamma-1}+n^{-1}} |\bar{M}_n(t) - \bar{M}_n(s)|$$
$$\le \sup_{s,t\le 2:|t-s|\le\delta} |\bar{M}_n(t) - \bar{M}_n(s)|.$$

The set

$$\{x \in D[0,\infty) : \sup_{s,t\le 2:|t-s|\le\delta} |x(t) - x(s)| \ge \varepsilon\}$$

is closed in $D[0,\infty)$. Consequently by weak convergence

$$\limsup_{n\to\infty} P_x\Big\{ \max_{0\le k<m} \max_{k\ell\le i\le (k+1)\ell} |M_i - M_{k\ell}| \ge n^{1/2}\varepsilon \Big\}$$
$$\le P\Big\{ \sup_{s,t\le 2:|t-s|\le\delta} |\sigma B(t) - \sigma B(s)| \ge \varepsilon \Big\} \longrightarrow 0 \qquad \text{as } \delta \searrow 0.$$

Thus line (6.3) converges to 0 in P_x-probability.

We turn to line (6.2). We will use the following maximal inequality.

LEMMA 6.2. *Suppose Z_n is a process such that $Z_0 = 0$, the increments $\{Z_n - Z_{n-1} : n \ge 1\}$ are stationary, and $E(Z_n^2) \le An$ for some constant A and all n. Then for all $n \ge 1$, $\lambda > 0$ and integers $k \ge 0$,*

$$P\Big\{ \max_{0\le j\le n} |Z_j| \ge \lambda \Big\} \le \frac{2^{6k} A n^{1+2^{-k}}}{\lambda^2}.$$

Proof. Induction on k. Case $k = 0$:

$$P\Big\{\max_{0\le j\le n} |Z_j| \ge \lambda\Big\} \le \sum_{j=1}^{n} P\{|Z_j| \ge \lambda\} \le \sum_{j=1}^{n} A j \lambda^{-2} \le A n^2 \lambda^{-2}.$$

Suppose the claim holds for a given $k \ge 0$. Let $m = \lceil \sqrt{n}\, \rceil$. Then by the stationarity of increments, the induction hypothesis, and finally by $m \le 2\sqrt{n}$,

$$P\Big\{ \max_{0\le j\le n} |Z_j| \ge 2\lambda \Big\} \le P\Big\{ \max_{0\le k<m} \max_{0\le j\le m} |Z_{km} + Z_{km+j} - Z_{km}| \ge 2\lambda \Big\}$$
$$\le P\Big\{ \max_{0\le k\le m} |Z_{km}| \ge \lambda \Big\} + mP\Big\{ \max_{0\le j\le m} |Z_j| \ge \lambda \Big\}$$
$$\le \frac{2^{6k}(Am)m^{1+2^{-k}}}{\lambda^2} + m \cdot \frac{2^{6k} A m^{1+2^{-k}}}{\lambda^2}$$
$$= 2 \cdot \frac{2^{6k} A m^{2+2^{-k}}}{\lambda^2} \le \frac{2^{1+6k+2+2^{-k}} A n^{1+2^{-k-1}}}{\lambda^2} \le \frac{2^{6(k+1)} A n^{1+2^{-k-1}}}{(2\lambda)^2}.$$

This proves the case $k+1$. □

By Lemma 5.5

$$E_\pi|R_{k\ell}| \le C(k\ell)^{2\alpha} \le C\ell^{2\alpha}k.$$

Let $\beta > 1$. By the maximal inequality, for a fixed $\varepsilon > 0$,

$$P_\pi\{\, n^{-1/2} \max_{0\le k\le m} |R_{k\ell}| \ge \varepsilon \} \le C(\beta)\ell^{2\alpha} m^\beta n^{-1} \le C(\beta) n^{-(1-2\alpha\gamma-\beta(1-\gamma))}. \tag{6.5}$$

Since $g \in L^p(\pi)$, we can bound line (6.4) by

$$\begin{aligned} &P_\pi\Big\{ \max_{0\le k<m} \max_{k\ell\le i\le(k+1)\ell} |S_i(g) - S_{k\ell}(g)| \ge \varepsilon\sqrt{n}\,\Big\} \\ &\le P_\pi\{ \max_{i\le 2n} |g(X_i)| \ge \varepsilon\sqrt{n}/\ell\} \\ &\le 2n\pi\{\,|g| \ge \varepsilon\sqrt{n}/\ell\,\} \le 2n\cdot C\ell^p n^{-p/2} \le Cn^{-(p/2-1-\gamma p)}. \end{aligned} \tag{6.6}$$

Now consider the two exponents

$$1 - 2\alpha\gamma - \beta(1-\gamma) = 2\gamma(\tfrac{1}{2} - \alpha) - (\beta - 1)(1-\gamma)$$

and $p/2 - 1 - \gamma p$ from (6.5) and (6.6). Since $p > 1$ and $\alpha < \frac{1}{2}$, we can choose first $\gamma > 0$ small enough and then $\beta > 1$ close enough to 1 so that both expressions are positive. Take $n_j = j^r$ for a large enough integer r that makes the bounds in (6.5) and (6.6) summable over j. Then by the Borel-Cantelli lemma, we have proved the limit (6.1) along the subsequence n_j. But this suffices for the entire statement, because for $n_{j-1} \le n \le n_j$

$$\max_{k\le n} \frac{|R_k|}{\sqrt{n}} \le \left(\frac{j}{j-1}\right)^{r/2} \max_{k\le n_j} \frac{|R_k|}{\sqrt{n_j}}. \tag{6.7}$$

Theorem 6.1 is thereby proved.

References

[1] M. Maxwell and M. Woodroofe. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.*, 28(2):713–724, 2000.

[2] F. Rassoul-Agha and T. Seppäläinen. An almost sure invariance principle for additive functionals of Markov chains. `http://front.math.ucdavis.edu/math.PR/0411603`.

T. Seppäläinen, Mathematics Department, University of Wisconsin-Madison, Van Vleck Hall, Madison, WI 53706, USA.

E-mail address: `seppalai@math.wisc.edu`