

## A SELF-NORMALIZED CENTRAL LIMIT THEOREM FOR MARKOV RANDOM WALKS

CHENG-DER FUH,\* *National Central University*

TIAN-XIAO PANG,\*\* *Zhejiang University*

### Abstract

Motivated by the study of the asymptotic normality of the least-squares estimator in the (autoregressive) AR(1) model under possibly infinite variance, in this paper we investigate a self-normalized central limit theorem for Markov random walks. That is, let  $\{X_n, n \geq 0\}$  be a Markov chain on a general state space  $\mathcal{X}$  with transition probability  $P$  and invariant measure  $\pi$ . Suppose that an additive component  $S_n$  takes values on the real line  $\mathbb{R}$ , and is adjoined to the chain such that  $\{S_n, n \geq 1\}$  is a Markov random walk. Assume that  $S_n = \sum_{k=1}^n \xi_k$ , and that  $\{\xi_n, n \geq 1\}$  is a nondegenerate and stationary sequence under  $\pi$  that belongs to the domain of attraction of the normal law with zero mean and possibly infinite variance. By making use of an asymptotic variance formula of  $S_n/\sqrt{n}$ , we prove a self-normalized central limit theorem for  $S_n$  under some regularity conditions. An essential idea in our proof is to bound the covariance of the Markov random walk via a sequence of weight functions, which plays a crucial role in determining the moment condition and dependence structure of the Markov random walk. As illustrations, we apply our results to the finite-state Markov chain, the AR(1) model, and the linear state space model.

*Keywords:* Self-normalized; central limit theorem; Markov random walk; Poisson equation; domain of attraction of the normal law

2010 Mathematics Subject Classification: Primary 60F05  
Secondary 60J25

### 1. Introduction

Let  $\{X_n, n \geq 0\}$  be a Markov chain on a general state space  $\mathcal{X}$  with  $\sigma$ -algebra  $\mathcal{A}$ . Suppose that an additive component  $S_n = \sum_{k=1}^n \xi_k$  takes values on the real line  $\mathbb{R}$  and is adjoined to the chain such that  $\{(X_n, S_n), n \geq 0\}$  is a Markov chain on  $\mathcal{X} \times \mathbb{R}$  with

$$\begin{aligned} P((X_n, S_n) \in A \times (B + s) \mid (X_{n-1}, S_{n-1}) = (x, s)) \\ &= P((X_1, S_1) \in A \times B \mid (X_0, S_0) = (x, 0)) \\ &= P(x, A \times B) \end{aligned} \tag{1}$$

for all  $x \in \mathcal{X}, s \in \mathbb{R}, A \in \mathcal{A}$ , and  $B \in \mathcal{B}$  (the Borel  $\sigma$ -algebra on  $\mathbb{R}$ ). We call  $\{S_n, n \geq 1\}$  a Markov random walk. For an initial distribution  $\nu$  on  $X_0$ , let  $P_\nu$  denote the probability measure under the initial distribution  $\nu$  on  $X_0$ , and let  $E_\nu$  denote the corresponding expectation. In the

Received 6 January 2011; revision received 17 October 2011.

\* Postal address: Graduate Institute of Statistics, National Central University, Zhongli, Taiwan.

Email address: stcheng@stat.sinica.edu.tw

\*\* Postal address: Department of Mathematics, Zhejiang University, Hangzhou 310027, P. R. China.

Email address: txpang@zju.edu.cn

case where  $\nu$  is degenerate at  $x$ , we write  $P_x$  instead of  $P_\nu$  and  $E_x$  instead of  $E_\nu$ . Moreover, let  $\{X_n, n \geq 0\}$  be a positive recurrent Markov chain with stationary probability measure  $\pi$ . A simple example of Markov random walks is given as follows.

**Example 1.** (The (autoregressive) AR(1) model.) Let

$$X_{n+1} = \rho X_n + \varepsilon_{n+1}, \quad n = 0, 1, 2, \dots, \tag{2}$$

where  $|\rho| < 1$  and  $\varepsilon_n$  is a sequence of independent and identically distributed (i.i.d.) random variables with distribution  $N(0, \sigma^2)$ . Moreover, we assume that  $\varepsilon_{n+1}$  and  $X_n$  are independent. Then  $\{X_n, n \geq 0\}$  forms a Markov chain with transition probability kernel

$$p(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y - \rho x)^2}{2\sigma^2}\right\},$$

and stationary probability  $\pi$  which has probability density  $N(0, \sigma^2/(1-\rho^2))$ . The least-squares estimator (maximum likelihood estimator) of  $\rho$  is

$$\hat{\rho} = \frac{\sum_{k=1}^n X_k X_{k-1}}{\sum_{k=1}^n X_{k-1}^2}.$$

It is known (cf. Anderson (1959)) that, as  $n \rightarrow \infty$ ,

$$\frac{\hat{\rho} - \rho}{\sqrt{(1 - \rho^2)/n}} \rightarrow N(0, 1) \quad \text{in distribution.}$$

Note that

$$\hat{\rho} - \rho = \frac{\sum_{k=1}^n \varepsilon_k X_{k-1}}{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}} \frac{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}}{\sum_{k=1}^n X_{k-1}^2}.$$

By letting  $\xi_n = \varepsilon_n X_{n-1}$ , then  $S_n = \sum_{k=1}^n \xi_k$  is a Markov random walk. The other cases can be defined in a similar way.

The problem of interest here is whether the asymptotic normality of  $(\hat{\rho} - \rho)/\sqrt{(1 - \hat{\rho}^2)/n}$  still holds when the  $\varepsilon_n$  have heavy tails and only the sample is available.

Example 1 motivates our study of the self-normalized central limit theorem for Markov random walks. In the literature, central limit theorems for partial sums of Markov chains have been studied under various assumptions; see Nagaev (1957), Lifshits (1978), and Maxwell and Woodroffe (2000), among others. Note that these results all hold under the assumption of a finite stationary second moment.

On the other hand, it is known that the celebrated self-normalized limit theorems put a totally new countenance on classical limit theorems. Similar results may still hold under a less strong, or even no, moment condition if the normalizing constants in the classical limit theorems are replaced by an appropriate sequence of random variables. In the i.i.d. case, we refer the reader to Griffin and Kuelbs (1989) for the law of the iterated logarithm, Csörgő *et al.* (1994) for studentized increments, Bentkus and Götze (1996) for Berry–Esseen inequalities, Lin (1996) for the Chung-type law of the iterated logarithm, Shao (1997) for large deviations, Giné *et al.* (1997) for a necessary and sufficient condition of the asymptotic normality, Csörgő *et al.* (2003a), (2003b) for the Darling–Erdős and Donsker theorems, respectively, and Jing *et al.* (2003) for

Cramér-type large deviations. For a survey on recent developments in this area, the reader is referred to Lai and Shao (2007) or de la Peña *et al.* (2009) for details.

Regarding self-normalized limit theorems for dependent random variables with a finite second moment assumption, Peligrad and Shao (1994) established the self-normalized central limit theorem for  $\alpha$ -mixing and an associated sequence of random variables, Chen (1999) proved the self-normalized law of the iterated logarithm for functionals of the Harris recurrent Markov chain, and Faure (2002) obtained self-normalized large deviations for Markov chains. For the self-normalized limit theorems on dependent random variables without a finite second moment assumption, McElroy and Politis (2007) established a self-normalized weak convergence result for a class of new sequences constructed by taking the product of a long-memory sequence and a stable sequence such that the variance is infinite and the covariance is finite.

It is worth noting that, for dependent random variables, the dependence structure usually relies on moment conditions and/or mixing rates. For instance, under the assumption that the sequence is  $\rho$ -mixing with some mixing rate, and in the domain of attraction of the normal law, Bradley (1988) proved that the covariance part (or the variance part) of truncated partial sums is proportional to the product of the sample size and the truncated second moment. Moreover, Diebold and Inoue (2001) gave a definition of long memory which involves the rate of growth for the variances of partial sums  $\text{var}(S_n) = O(n^{2d+1})$ ,  $0 < d < 1$ . They also made a connection between the moment condition and the dependence assumption.

There are three aspects to this study. First, when the second moment exists, we provide an asymptotic variance formula, whose consistent estimator can be used as the self-normalized term. Second, when the second moment does not exist, we prove a self-normalized central limit theorem for Markov random walks. Third, the conditions posed in our theorems are weak enough to cover several practical examples, to which the variance can be calculated explicitly. Note that a difficulty in studying self-normalized central limit theorems for dependent random variables is that one needs to investigate the variance–covariance structure under the condition that the second moment does not exist. By bounding the variance–covariance of the Markov random walk by a sequence of weight functions defined on the truncated state space (see Assumption 2 for details), we obtain the result. This idea plays an essential role in determining the moments and dependence of Markov random walks, and may also be applicable in other cases.

The remainder of this paper is organized as follows. In Section 2 we state our main results on self-normalized central limit theorems for the Markov random walk  $S_n = \sum_{k=1}^n \xi_k$  under suitable conditions, where  $\{\xi_n, n \geq 1\}$  is a stationary sequence under  $\pi$  that belongs to the domain of attraction of the normal law with zero mean. In Section 3 we apply our results to the finite-state Markov chain, the AR(1) model, and the linear state space model. The proofs of the theorems are given in Section 4.

## 2. Main results

Let  $\{X_n, n \geq 0\}$  be the Markov chain defined in (1) on a state space  $\mathcal{X}$ . For simplicity, we define  $P(x, \cdot) = P(x, \cdot \times \mathbb{R})$  for all  $x \in \mathcal{X}$ . We say that  $\{(X_n, S_n), n \geq 0\}$  satisfies the minorization condition if there exist a measure  $\Psi$  on  $\mathcal{X} \times \mathbb{R}$  and a measurable function  $h$  on  $\mathcal{X}$  such that  $\int \pi(dx)h(x) > 0$ ,  $\Psi(\mathcal{X} \times \mathbb{R}) = 1$ ,  $\int \Psi(dx \times \mathbb{R})h(x) > 0$ , and the kernel  $T(x, A \times B) = P(x, A \times B) - h(x)\Psi(A \times B)$  is nonnegative for all  $A \in \mathcal{A}$  and  $B \in \mathcal{B}$ . It is known (see Ney and Nummelin (1987)) that, under this assumption, we can extend Nummelin's splitting technique to Markov random walks. To this end, we introduce the following notation. A set  $\alpha \in \mathcal{A}$  is called an *atom* if there exists a measure  $\nu$  on  $\mathcal{A}$  such that  $P(x, \cdot) = \nu(\cdot)$

for  $x \in \alpha$ . Let  $\Delta$  be an atom of the Markov chain, define

$$\tau_\Delta = \inf\{n \geq 1, X_n \in \Delta\}$$

to be the regeneration time, where

$$\tau_\Delta(0) = 0, \quad \tau_\Delta(1) = \tau_\Delta, \quad \text{and} \quad \tau_\Delta(k + 1) = \inf\{n > \tau_\Delta(k), X_n \in \Delta\}, \quad k \geq 1.$$

Let  $\{X_n, n \geq 0\}$  be an irreducible (with respect to a maximal irreducible measure  $\varphi$  on  $(\mathcal{X}, \mathcal{A})$ ), aperiodic Markov chain such that  $\{(X_n, S_n), n \geq 0\}$  satisfies the minorization condition. It is known (see Chapter 17 of Meyn and Tweedie (1993)) that, under the assumptions that

$$E_\Delta \left( \sum_{k=1}^{\tau_\Delta-1} \xi_k \right)^2 < \infty \quad \text{and} \quad E_\Delta \tau_\Delta < \infty, \tag{3}$$

the central limit theorem holds for  $\sum_{k=1}^n \xi_k / \sqrt{n}$ . Moreover, Fuh and Hu (2007) provided a representation of the asymptotic variance  $\Sigma$  of  $\sum_{k=1}^n \xi_k / \sqrt{n}$  as

$$\Sigma = \int_{\mathcal{X}} [G(x) - \Gamma^2(x)]\pi(dx) + \int_{\mathcal{X} \times \mathcal{X}} [\Gamma(y) - \delta_x + \delta_y]^2 P(x, dy)\pi(dx), \tag{4}$$

where  $\Gamma(x) = E_x \xi_1$ ,  $G(x) = E_x \xi_1^2$ ,  $P(x, dy)$  stands for the transition probability, and  $\delta_x$  is a measurable function from  $\mathcal{X}$  to  $\mathbb{R}$  satisfying the Poisson equation

$$(I - P)\delta_x = P\Gamma(x), \tag{5}$$

where  $I$  denotes the identity kernel.

When  $S_n = \sum_{k=0}^n f(X_k)$ , where  $f$  is a real-valued function defined on  $\mathcal{X}$ , the variance formula (4) becomes

$$\Sigma = \int_{\mathcal{X} \times \mathcal{X}} [f(y) - \delta_x + \delta_y]^2 P(x, dy)\pi(dx). \tag{6}$$

To state our main results, we need to introduce truncated random variables. That is, let

$$l(t) = E_\pi(\xi_1^2 \mathbf{1}\{|\xi_1| \leq t\}), \quad b = \inf\{t \geq 1: l(t) > 0\}, \tag{7}$$

where  $\mathbf{1}\{\cdot\}$  denotes the indicator function, and

$$\eta_j = \inf \left\{ s: s \geq b + 1, \frac{l(s)}{s^2} \leq \frac{1}{j} \right\} \quad \text{for } j = 1, 2, \dots \tag{8}$$

Let

$$\tilde{\xi}_{i,n} = \xi_i \mathbf{1}\{|\xi_i| \leq \eta_n\} - E_\pi(\xi_i \mathbf{1}\{|\xi_i| \leq \eta_n\}), \quad i = 1, 2, \dots, \tag{9}$$

and define  $\Gamma_n(x) = E_x \tilde{\xi}_{1,n}$ . For each given  $n$ , let  $\delta_{x,n}$  be a measurable function from  $\mathcal{X}$  to  $\mathbb{R}$  satisfying the corresponding Poisson equation

$$(I - P)\delta_{x,n} = P\Gamma_n(x). \tag{10}$$

In this paper, let  $C, C_1, C_2, \dots$  stand for positive constants whose values can differ from line to line, and write  $a_n \sim b_n$  if  $\lim_{n \rightarrow \infty} a_n/b_n = 1$ . The following assumptions will be used throughout the rest of this paper.

**Assumption 1.** Let  $\{X_n, n \geq 0\}$  be a positive recurrent, irreducible (with respect to some measure on  $\mathcal{A}$ ), aperiodic Markov chain on a state space  $\mathcal{X}$  such that  $\{(X_n, S_n), n \geq 1\}$  satisfies the minorization condition. Assume that, under the stationary probability  $\pi$ ,  $\{\xi_n, n \geq 1\}$  is a nondegenerate and stationary sequence that belongs to the domain of attraction of the normal law with zero mean. Furthermore, assume that  $E_\Delta \tau_\Delta^2 < \infty$ .

**Assumption 2.** Define

$$V_n(x) = 1 + E(\xi_1^2 \mathbf{1}\{|\xi_1| \leq \eta_n\} \mid X_1 = x).$$

Assume that, for any given family of measurable functions  $g_n: \mathcal{X} \rightarrow \mathbb{R}$  such that  $g_n \leq V_n$ , there exist  $\mathcal{X}_n$ , a subset of  $\mathcal{X}$ , which is asymptotically equal to  $\mathcal{X}$ ,  $\beta > 1$ , and a large enough  $M > 0$  such that, for all  $m > 1$  and  $n \geq M$ ,

$$\sup_{x \in \mathcal{X}_n} \frac{|E_x g_n(X_m) - \int g_n(y)\pi(dy)|}{V_n(x)} \leq Cm^{-\beta}. \tag{11}$$

Since  $\delta_{x,n}$ , defined in (10), plays an essential role in the truncated version of (4), but is unknown, we need to find an estimator of  $\delta_{x,n}$ . A natural estimator is the kernel density estimator for stationary Markov processes (see Basawa and Prakasa Rao (1980)). To define the kernel estimator, we let the state space  $\mathcal{X} = \mathbb{R}$ . Let  $p(x)$  be the probability density function of the invariant probability  $\pi$  with respect to the Lebesgue measure, and let  $q(x, y)$  be the two-dimensional joint probability density of the Markov chain  $\{X_n, n \geq 0\}$  with respect to the Lebesgue measure. Let  $K$  be a probability density function defined on  $\mathbb{R}$ , and let  $h_n$  be a sequence of positive constants. For given observations  $\{X_0, X_1, \xi_1, \dots, X_n, \xi_n\}$ , define

$$\begin{aligned} \hat{p}_n(x) &= \frac{1}{nh_n} \sum_{j=0}^n K\left(\frac{x - X_j}{h_n}\right), & \hat{q}_n(x, y) &= \frac{1}{nh_n} \sum_{j=0}^{n-1} K\left(\frac{x - X_j}{\sqrt{h_n}}\right) K\left(\frac{y - X_{j+1}}{\sqrt{h_n}}\right), \\ \hat{p}_n(x, y) &= \frac{\hat{q}_n(x, y)}{\hat{p}_n(x)}, & \hat{\Gamma}_n(x) &= \frac{1}{nh_n \hat{p}_n(x)} \sum_{j=1}^n \xi_j K\left(\frac{x - X_j}{h_n}\right). \end{aligned}$$

Let  $\hat{P}_n$  be the transition kernel induced by the transition probability density function  $\hat{p}_n(x, y)$ .

**Assumption 3.** Let the state space  $\mathcal{X} = \mathbb{R}$ . Assume that the probability kernel function  $K$  satisfies  $K(x) \leq C < \infty$  for all  $x \in \mathbb{R}$ , and that  $|x|K(x) \rightarrow 0$  as  $|x| \rightarrow \infty$ . Furthermore, assume that  $h = h_n \rightarrow 0$ , and  $nh \rightarrow \infty$  nondecreasingly as  $n \rightarrow \infty$ .

**Assumption 4.** Assume that  $p(x)$  and  $E(\xi_1 \mid X_1 = x)$  are continuous with respect to  $x$ , and that  $q(x, y)$  is continuous with respect to  $x$  and  $y$ . Furthermore, assume that the solution  $\delta$  of

$$(I - P)\delta = P\Gamma \tag{12}$$

is a continuous function of  $p$  (the density function associated with  $P$ ) and  $\Gamma$ ,  $P$ -almost surely.

**Remarks 1.** (a) The minorization condition was first introduced in Nummelin (1978). If a Markov chain is Harris recurrent then the minorization condition holds for the  $n$ -step transition probability. It is known that, under the irreducible assumption, Assumption 1 implies that  $\{(X_n, \xi_n), n \geq 0\}$  is Harris recurrent; see Theorem 4.1(iv) of Ney and Nummelin (1987).

(b) Assumption 2 is a  $V_n$  ergodic condition, where the sup is taken over a truncated state space. Note that the weight function  $V_n$ , generated by the truncated second moment of  $\xi_1$ , depends on the sample size  $n$ . This condition will be used to study the case in which the stationary second moment of  $\xi_1$  does not exist. It is weak enough to cover several practical examples in Section 3. Assumption 3 is a standard condition for kernel estimation. Assumption 4 is a smooth condition of the Poisson equation.

(c) Since  $E_\pi V_n(x)$  is a slowly varying function at  $\infty$ , the condition  $E_\Delta \tau_\Delta^2 < \infty$  can be removed if (11) is replaced with

$$\sup_{x \in \mathcal{X}} \frac{|E_x g_n(X_m) - \int g_n(y)\pi(dy)|}{V_n(x)} \leq Cm^{-\beta},$$

by first taking  $g_n(x) = x$ , and then applying Equation (3.1) of Bolthausen (1982) and Remark 1 of Bertail and Cléménçon (2006) to the Harris recurrent Markov chains in Assumption 1. Moreover, noting that the AR(1) model (2) is geometric ergodic with weight function  $|x| + 1$  when the innovations in the model are in the domain of attraction of the normal law, i.e.

$$\sup_{x \in \mathcal{X}} \frac{|E_x X_m - \int y\pi(dy)|}{|x| + 1} \leq Cm^{-\beta_0}$$

for some  $\beta_0 > 1$ , we have the strong mixing coefficients  $\alpha(n) \leq Cn^{-\beta_0}$  by Equation (3.1) of Bolthausen (1982). Hence, the condition  $E_\Delta \tau_\Delta^2 < \infty$  is fulfilled in the AR(1) model by Remark 1 of Bertail and Cléménçon (2006).

**Theorem 1.** *Let  $\{S_n, n \geq 1\}$  be a Markov random walk satisfying Assumptions 1 and 2, where  $S_n = \sum_{k=1}^n \xi_k$ .*

(a) *Suppose that Assumptions 3 and 4 hold. If  $\text{var}_\pi(\xi_1) < \infty$  then, as  $M \rightarrow \infty$  and  $n \rightarrow \infty$ ,*

$$\frac{S_n}{\sqrt{|\Delta_{n,M}|}} \rightarrow N(0, 1) \text{ in distribution,} \tag{13}$$

where, for any  $M > 0$ ,

$$\Delta_{n,M} = \sum_{k=1}^n \xi_k^2 + n \int_{-M}^M \int_{-M}^M [(\hat{\delta}_{x,n} - \hat{\delta}_{y,n})^2 - 2(\hat{\delta}_{x,n} - \hat{\delta}_{y,n})\hat{\Gamma}_n(y)]\hat{q}_n(x, y) \, dy \, dx.$$

(b) *If  $\text{var}_\pi(\xi_1) = \infty$  then, as  $n \rightarrow \infty$ ,*

$$\frac{S_n}{\sqrt{\sum_{k=1}^n \xi_k^2}} \rightarrow N(0, 1) \text{ in distribution.} \tag{14}$$

**Remarks 2.** (a) Note that  $V_n(x)$  reduces to  $V(x) = 1 + E(\xi_1^2 \mid X_1 = x)$  when  $\text{var}_\pi(\xi_1) < \infty$ . This means that the weight function does not depend on  $n$  under the finite second moment assumption. The details can be found in the proof of Theorem 1 in Section 4.

(b) Equation (14) shows that  $\sum_{k=1}^n \xi_k^2$  can be used as the self-normalized term when  $\text{var}_\pi(\xi_1)$  is infinite, while (13) reveals that the covariance part after truncation plays an essential role when  $\text{var}_\pi(\xi_1)$  is finite.

(c) Theorem 1 gives the result for the nonparametric case with or without the finite second moment assumption. In the parametric case, when the second moment is finite, the transition probability and stationary distribution can be obtained from the corresponding estimator. The self-normalized factors can then be determined by direct computation via (4), in which the unknown parameters can be replaced by their consistent estimators (see Example 3 below). The result is summarized in Proposition 1 below.

**Proposition 1.** *Let  $\{S_n, n \geq 1\}$  be a Markov random walk satisfying Assumption 1 and (3). Let the asymptotic variance of  $S_n/\sqrt{n}$  be  $\Sigma = f(\theta_1, \dots, \theta_m)$ , where  $m$  is a positive integer and  $f(\cdot)$  is a continuous function  $\mathbb{P}$ -almost surely with unknown parameters  $\theta_1, \dots, \theta_m$ . Let  $\hat{\theta}_i$  be a consistent estimator of  $\theta_i$  for  $i = 1, \dots, m$ . Then, as  $n \rightarrow \infty$ ,*

$$\frac{S_n}{\sqrt{nf(\hat{\theta}_1, \dots, \hat{\theta}_m)}} \rightarrow N(0, 1) \text{ in distribution.}$$

The following result gives a central limit theorem for a Markov random walk with or without the finite second moment assumption. This result will be used in the proof of Theorem 1 and is of independent interest.

**Theorem 2.** *Let  $\{S_n, n \geq 1\}$  be a Markov random walk satisfying Assumptions 1 and 2. Then, as  $n \rightarrow \infty$ ,*

$$\frac{S_n}{\sqrt{\text{var}_\pi(\sum_{k=1}^n \xi_k \mathbf{1}\{|\xi_k| \leq \eta_n\})}} \rightarrow N(0, 1) \text{ in distribution,}$$

where  $\eta_n$  is defined in (8).

### 3. Some examples

To illustrate our results, in this section we study three examples: finite-state Markov chains, AR(1) models, and linear state space models. In these examples, we first consider the self-normalized central limit theorem for  $S_n = \sum_{k=1}^n \xi_k$ , and then investigate the asymptotic normality of the estimator  $\hat{\rho}$  of  $\rho$  in (2), with or without the finite second moment assumption. Note that the solution of the Poisson equation, and, hence, the asymptotic variance formula, can be described explicitly in each case. Before presenting these examples, we state the following lemma (see Csörgő *et al.* (2003b)), which will be used in Sections 3 and 4.

**Lemma 1.** *Let  $X$  be a random variable, and define  $l(x) = E(X^2 \mathbf{1}\{|X| \leq x\})$ . The following statements are equivalent:*

- (a)  $X$  is in the domain of attraction of the normal law,
- (b)  $x^2 P(|X| > x) = o(l(x))$ ,
- (c)  $x E(|X| \mathbf{1}\{|X| > x\}) = o(l(x))$ ,
- (d)  $E(|X|^n \mathbf{1}\{|X| \leq x\}) = o(x^{n-2}l(x))$  for  $n > 2$ .

**Example 2.** Let  $\{X_n, n \geq 0\}$  be an ergodic Markov chain on a finite state space  $\mathcal{X} = \{1, 2, 3\}$ , with transition probability  $P = \{p(i, j)\}$  and stationary distribution  $\pi$ . Assume that  $\xi_n$  takes values on the real line  $\mathbb{R}$ , with  $E_\pi \xi_1 = 0$  and  $E_\pi \xi_1^2 < \infty$ . Fuh and Hu (2007) established the

variance formula

$$\Sigma = \sum_{i=1}^3 [G(i) - \Gamma^2(i)]\pi_i + \sum_{i,j=1}^3 [\Gamma(j) - \delta_i + \delta_j]^2 p(i, j)\pi_i,$$

where  $\Gamma(i) = E(\xi_1 \mid X_0 = i)$ ,  $G(i) = E(\xi_1^2 \mid X_0 = i)$ , and  $\delta_i - \delta_j$  is the difference of two solutions to the Poisson equation (5), which is uniquely determined, and can be explicitly expressed as

$$\begin{aligned} \delta_1 - \delta_2 &= a_{11}\Gamma(1) + a_{12}\Gamma(2) + a_{13}\Gamma(3) =: \sum_{j=1}^3 g_{1j}(p(1, 1), \dots, p(3, 3))\Gamma(j), \\ \delta_2 - \delta_3 &= a_{21}\Gamma(1) + a_{22}\Gamma(2) + a_{23}\Gamma(3) =: \sum_{j=1}^3 g_{2j}(p(1, 1), \dots, p(3, 3))\Gamma(j), \\ \delta_3 - \delta_1 &= a_{31}\Gamma(1) + a_{32}\Gamma(2) + a_{33}\Gamma(3) =: \sum_{j=1}^3 g_{3j}(p(1, 1), \dots, p(3, 3))\Gamma(j), \end{aligned}$$

where

$$\begin{aligned} a_{11} &= -1, \\ a_{12} &= -\frac{p(2, 2)(p(3, 1) + p(3, 2)) + p(2, 3)p(3, 2)}{p(2, 1)(p(3, 1) + p(3, 2)) + p(3, 1)p(2, 3)}, \\ a_{13} &= -\frac{p(2, 3)(p(3, 1) + p(3, 2)) + p(2, 3)p(3, 3)}{p(2, 1)(p(3, 1) + p(3, 2)) + p(3, 1)p(2, 3)}, \\ a_{21} &= 0, \\ a_{22} &= \frac{p(2, 2)p(3, 1) - p(2, 1)p(3, 2)}{p(2, 1)(p(3, 1) + p(3, 2)) + p(3, 1)p(2, 3)}, \\ a_{23} &= \frac{p(2, 3)p(3, 1) - p(2, 1)p(3, 3)}{p(2, 1)(p(3, 1) + p(3, 2)) + p(3, 1)p(2, 3)}, \\ a_{31} &= 1, \\ a_{32} &= \frac{p(3, 2)(p(2, 1) + p(2, 2)) + p(2, 3)p(3, 2)}{p(2, 1)(p(3, 1) + p(3, 2)) + p(3, 1)p(2, 3)}, \\ a_{33} &= \frac{p(3, 3)(p(2, 1) + p(2, 3)) + p(2, 3)p(3, 2)}{p(2, 1)(p(3, 1) + p(3, 2)) + p(3, 1)p(2, 3)}. \end{aligned}$$

Let

$$\begin{aligned} n_i &= \sum_{k=1}^n \mathbf{1}\{X_k = i\}, & \hat{p}_i &= \frac{n_i}{n} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}\{X_k = i\}, \\ \hat{q}(i, j) &= \frac{1}{n} \sum_{k=1}^{n-1} \mathbf{1}\{X_k = i, X_{k+1} = j\}, & \hat{p}(i, j) &= \frac{\hat{q}(i, j)}{\hat{p}_i}, \\ \hat{\Gamma}(j) &= \frac{\sum_{k=1}^n \xi_k \mathbf{1}\{X_k = j\}}{\sum_{k=1}^n \mathbf{1}\{X_k = j\}}, & \hat{G}(j) &= \frac{\sum_{k=1}^n \xi_k^2 \mathbf{1}\{X_k = j\}}{\sum_{k=1}^n \mathbf{1}\{X_k = j\}}. \end{aligned}$$



It is easy to see that  $\hat{p}_i \rightarrow p_i := P_\pi(X_1 = i)$  in probability and  $\hat{q}(i, j) \rightarrow q(i, j)$  in probability as  $n \rightarrow \infty$ . This yields  $\hat{p}(i, j) \rightarrow p(i, j)$  in probability as  $n \rightarrow \infty$ . In addition, by (1),  $\{\xi_n \mathbf{1}\{X_n = j\}, n \geq 0\}$  forms a Markov chain for any  $j \in \mathcal{X}$ . Then, in view of the ratio limit theorem for Markov chains (see Meyn and Tweedie (1993, p. 424)), we have

$$\hat{\Gamma}(j) \rightarrow \frac{E_\pi(\xi_1 \mathbf{1}\{X_1 = j\})}{p_j} \quad \text{in probability} \tag{15}$$

as  $n \rightarrow \infty$ , and the right-hand side of (15) equals

$$\frac{\sum_{i=1}^3 \int_{-\infty}^{\infty} x \, dF_i(x) p(j, i) p_j}{p_j} = \sum_{i=1}^3 \int_{-\infty}^{\infty} x \, dF_i(x) p(j, i) = \Gamma(j), \tag{16}$$

where  $F_i(x)$  denotes the distribution of  $\xi_1$  under  $X_1 = i$ . By making use of the same argument used in (15) and (16), we obtain  $\hat{G}(i) \rightarrow G(i)$  in probability. Since  $g_{ik}$  is a smooth function  $P$ -almost surely, for  $i, j = 1, 2, 3$ ,

$$\widehat{\delta_i - \delta_j} = \sum_{k=1}^3 g_{ik}(\hat{p}(1, 1), \dots, \hat{p}(3, 3)) \hat{\Gamma}(k)$$

is a consistent estimator of  $\delta_i - \delta_j$ . It is easy to see that Assumption 1 and (3) are fulfilled in the finite-state Markov chain. Therefore, by Proposition 1,  $\sum_{k=1}^n \xi_k$  obeys the self-normalized central limit theorem, i.e. as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \xi_k}{\sqrt{n(\sum_{i=1}^3 [\hat{G}(i) - \hat{\Gamma}^2(i)] \hat{p}_i + \sum_{i,j=1}^3 [\hat{\Gamma}(j) - \hat{\delta}_i + \hat{\delta}_j]^2 \hat{p}(i, j) \hat{p}_i)}} \rightarrow N(0, 1) \quad \text{in distribution.}$$

If  $\xi_n$  is in the domain of attraction of the normal law with infinite variance, it is easy to see that both Assumptions 1 and 2 are fulfilled in the finite-state Markov chain. Hence, by Theorem 1(b) we have, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \xi_k}{\sqrt{\sum_{k=1}^n \xi_k^2}} \rightarrow N(0, 1) \quad \text{in distribution.}$$

**Example 3.** Consider the AR(1) model. Let

$$X_{n+1} = \rho X_n + \varepsilon_{n+1}, \tag{17}$$

where  $|\rho| < 1$  and  $\varepsilon_n$  is a sequence of i.i.d. real-valued random variables with  $E|\varepsilon_1| < \infty$ .

(a) We first consider the case in which  $\xi_n = X_n$  and  $S_n = \sum_{k=1}^n X_k$ . Under the assumption that the  $\varepsilon_n$  are i.i.d. random variables with distribution  $N(0, \sigma^2)$ , it is easy to see that the conditions in Theorem 2 of Fuh and Hu (2007) hold; therefore, the Poisson equation (5) becomes

$$\delta_x - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y - \rho x)^2}{2\sigma^2}\right\} \delta_y \, dy = \rho x. \tag{18}$$

A solution to (18) is obtained by differentiating both sides with respect to  $x$ , i.e.

$$\delta'_x - \rho \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y - \rho x)^2}{2\sigma^2}\right\} \frac{y - \rho x}{\sigma^2} \delta_y \, dy = \rho, \tag{19}$$

where a prime denotes the first derivative. Then the solutions of (19) are  $\delta_x = \rho x / (1 - \rho) + C$ . Note that  $\delta_x - \delta_y$  is uniquely determined; therefore, the asymptotic variance (6) of  $S_n / \sqrt{n}$  becomes

$$\begin{aligned} \Sigma &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \delta_x + \delta_y)^2 P(x, dy) \pi(dx) \\ &= \frac{1}{(1 - \rho)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \rho x)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y - \rho x)^2}{2\sigma^2}\right\} \\ &\quad \times \frac{1}{\sqrt{2\pi}\sigma/\sqrt{1 - \rho^2}} \exp\left\{-\frac{x^2}{2\sigma^2/(1 - \rho^2)}\right\} dx \, dy \\ &= \frac{\sigma^2}{(1 - \rho)^2}, \end{aligned}$$

which is the same as that in Theorem 7.1.2 of Brockwell and Davis (1991). Define

$$\hat{\rho} = \frac{\sum_{k=1}^n X_k X_{k-1}}{\sum_{k=1}^n X_{k-1}^2} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \hat{\rho} X_{k-1})^2. \tag{20}$$

Note that  $\hat{\rho}$  is a consistent estimator of  $\rho$  (see Anderson (1959)). In view of

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{k=1}^n (X_k - \rho X_{k-1} + (\rho - \hat{\rho}) X_{k-1})^2 \\ &= \frac{1}{n} \sum_{k=1}^n \varepsilon_k^2 + \frac{2(\rho - \hat{\rho})}{n} \sum_{k=1}^n \varepsilon_k X_{k-1} + (\rho - \hat{\rho})^2 \frac{1}{n} \sum_{k=1}^n X_{k-1}^2. \end{aligned}$$

Theorems 4.1 and 4.2 of Anderson (1959), and the Cauchy–Schwarz inequality, we conclude that  $\hat{\sigma}^2$  is a consistent estimator of  $\sigma^2$ . Hence, as  $n \rightarrow \infty$ ,

$$\frac{\hat{\sigma}^2 / (1 - \hat{\rho})^2}{\Sigma} \rightarrow 1 \quad \text{in probability.}$$

It is easy to see that Assumption 1 and (3) are fulfilled in the AR(1) model (17). Then by Proposition 1 we have, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n X_k}{\sqrt{n \hat{\sigma}^2 / (1 - \hat{\rho}^2)}} \rightarrow N(0, 1) \quad \text{in distribution.} \tag{21}$$

Note that (21) is also obtained, via a different method, in Theorem 7.1.2 of Brockwell and Davis (1991).

Under the assumption that  $\varepsilon_n$  is in the domain of attraction of the normal law with infinite variance, Kulik (2006) proved a self-normalized central limit theorem of  $\sum_{k=1}^n X_k$  for linear processes.

(b) Consider the AR(1) model in (17). Assume that  $|\rho| < 1$  and that the  $\varepsilon_n$  are i.i.d. real-valued random variables with zero mean and finite second moment. Under the condition that  $X_0$  is a constant, Anderson (1959) established the consistency of the estimator  $\hat{\rho}$ , defined in (20), of  $\rho$ . Next, we will show that if the  $\varepsilon_n$  have continuous probability density functions then  $\hat{\rho} - \rho$  obeys a self-normalized central limit theorem, i.e. as  $n \rightarrow \infty$ ,

$$\frac{\hat{\rho} - \rho}{\sqrt{(1 - \hat{\rho}^2)/n}} \rightarrow N(0, 1) \quad \text{in distribution.} \tag{22}$$

Note that the induced Markov chain from model (17) has a continuous probability density function, which implies that the solution of the Poisson equation (12) is a continuous function of  $p(x, y)$  and  $\Gamma(x)$ , P-almost surely, that is, Assumption 4 is fulfilled. Define  $\xi_n = \varepsilon_{n+1}X_n$ . Then  $\xi_n$  has finite variance. To apply Theorem 1(a) for  $\sum_{k=1}^n \xi_k = \sum_{k=1}^n \varepsilon_k X_{k-1}$ , we first note that Assumptions 1 and 3 are obviously satisfied. For Assumption 2, in view of Remarks 2(a), it is easy to see that, for  $x \in \mathcal{X}_n$ ,

$$V(x) = 1 + E(\xi_1^2 \mid X_1 = x) = 1 + E((\varepsilon_2 X_1)^2 \mid X_1 = x) = 1 + x^2 \sigma^2 \geq Cx^2,$$

where  $\sigma^2$  denotes the variance of  $\varepsilon_n$ . This implies that Assumption 2 holds (see Meyn and Tweedie (1993, p. 380)). Since  $\Gamma(x) = E_x \xi_1 = E_x \varepsilon_2 X_1 = E(\varepsilon_2(\rho x + \varepsilon_1) \mid X_0 = x) = 0$ , the solution  $\delta_x$  of the Poisson equation (5) is 0, and, hence,  $\delta_x - \delta_y = 0$ . Consequently, it follows from Theorem 1(a) that, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \varepsilon_k X_{k-1}}{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2 + n o(1)}} \rightarrow N(0, 1) \quad \text{in distribution.}$$

Moreover, in view of the strong law of large numbers for Markov random walks (see Theorem 17.1.7 of Meyn and Tweedie (1993)),

$$\frac{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}{n} \rightarrow E_\pi \varepsilon_1^2 X_0^2 = \frac{\sigma^4}{1 - \rho^2} \quad \text{P-almost surely,} \tag{23}$$

which implies that, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \varepsilon_k X_{k-1}}{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}} \rightarrow N(0, 1) \quad \text{in distribution.} \tag{24}$$

Combining (23), (24), and the fact that, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n X_{k-1}^2}{n} \rightarrow \frac{\sigma^2}{1 - \rho^2} \quad \text{in probability,}$$

given in Anderson (1959), we have, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \sqrt{n}(\hat{\rho} - \rho) &= \sqrt{n} \frac{\sum_{k=1}^n \varepsilon_k X_{k-1}}{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}} \frac{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}}{\sum_{i=1}^n X_{k-1}^2} \\ &\rightarrow N(0, 1 - \rho^2) \quad \text{in distribution,} \end{aligned}$$

which means that  $\hat{\rho}$  is a consistent estimator of  $\rho$ , and, hence, (22) holds.

(c) Consider the AR(1) model (17) with  $|\rho| < 1$ . Let  $\varepsilon_n$  be a sequence of i.i.d. random variables that belongs to the domain of attraction of the normal law with infinite variance. Under the assumptions that  $X_0$  is a random variable with finite variance or is in the same domain of attraction as the normal law of  $\varepsilon_n$ , and that the truncated second moment for  $\{\varepsilon_n, n \geq 1\}$  satisfies the condition that there exists  $1 < \alpha < 2$  such that, for  $i \neq j$  and large  $x$ ,

$$E(\varepsilon_i \varepsilon_j)^2 \mathbf{1}\{|\varepsilon_i \varepsilon_j| \leq x\} \leq C(E \varepsilon_i^2 \mathbf{1}\{|\varepsilon_i| \leq x\})^\alpha, \tag{25}$$

we will show that the self-normalized central limit theorem for  $\hat{\rho}$  still holds. That is,

$$\frac{\hat{\rho} - \rho}{\sqrt{(1 - \hat{\rho}^2)/n}} \rightarrow N(0, 1) \quad \text{in distribution.} \tag{26}$$

Note that condition (25) covers the following case (cf. Example 2 of Hall and Seneta (1988)):

$$E \varepsilon_i^2 \mathbf{1}\{|\varepsilon_i| \leq x\} \sim C \exp\{(\log x)^{\alpha'}\} \quad \text{for some } 0 < \alpha' < 1.$$

Note that condition (25) is not a necessary condition for (26) to hold.

*Proof of (26).* Let

$$A_n = \sum_{k=1}^n X_k X_{k-1} - \rho \sum_{k=1}^n X_{k-1}^2 = \sum_{k=1}^n \varepsilon_k X_{k-1} \quad \text{and} \quad B_n = \sum_{k=1}^n X_{k-1}^2.$$

Then we have

$$\begin{aligned} \sqrt{n}(\hat{\rho} - \rho) &= \frac{\sqrt{n}A_n}{B_n} \\ &= \sqrt{n} \frac{\sum_{k=1}^n \varepsilon_k X_{k-1}}{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}} \frac{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}}{B_n} \\ &= \sqrt{n} \frac{\sum_{k=1}^n \varepsilon_k X_{k-1}}{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}} \frac{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}}{nl^*(\eta_n^*)} \frac{nl^*(\eta_n^*)}{B_n}, \end{aligned}$$

where  $\eta_n^*$  and  $l^*(\eta_n^*)$  are defined in (7) and (8) with  $\xi_1$  replaced by  $\varepsilon_1$ .

To prove (26), we need to show that, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \varepsilon_k X_{k-1}}{\sqrt{\sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}} \rightarrow N(0, 1) \quad \text{in distribution,} \tag{27}$$

$$\frac{(1 - \rho^2) \sum_{k=1}^n \varepsilon_k^2 X_{k-1}^2}{nl^{*2}(\eta_n^*)} \rightarrow 1 \quad \text{in probability,} \tag{28}$$

and

$$\frac{B_n}{nl^*(\eta_n^*)} \rightarrow \frac{1}{1 - \rho^2} \quad \text{in probability.} \tag{29}$$

To prove (27), define  $\xi_n = \varepsilon_{n+1} X_n$ . Then  $\xi_n$  has infinite variance. To apply Theorem 1(b) for  $\sum_{k=1}^n \xi_k = \sum_{k=1}^n \varepsilon_k X_{k-1}$ , we need to check that Assumptions 1 and 2 hold. For Assumption 1, we first note that  $E_\Delta \tau_\Delta^2 < \infty$  is fulfilled by Remarks 1(c). The remaining parts of

Assumption 1 are satisfied by Theorem 1.1 of Hall and Seneta (1988). For Assumption 2, we take  $\mathcal{X}_n = [-n^a, n^a]$ , where  $a \in (0, \frac{1}{2})$  is a given constant. It is easy to see that, for  $x \in \mathcal{X}_n$ ,

$$\begin{aligned} V_n(x) &= 1 + E(\xi_1^2 \mathbf{1}\{|\xi_1| \leq \eta_n^*\} \mid X_1 = x) \\ &= 1 + E((\varepsilon_2 X_1)^2 \mathbf{1}\{|\varepsilon_2 X_1| \leq \eta_n^*\} \mid X_1 = x) \\ &:= 1 + x^2 l^* \left( \frac{\eta_n^*}{x} \right). \end{aligned}$$

Since  $\eta_n^* = \sqrt{nl^*(\eta_n^*)}$  and  $|x| \leq n^a$  ( $0 < a < \frac{1}{2}$ ), we have  $\eta_n^*/|x| \geq n^{1/2-a} \sqrt{l^*(\eta_n^*)} \geq Cn^{1/2-a}$ . Furthermore,  $l^*(\eta_n^*/|x|) \geq C_1$  for large  $n$ . Therefore, either  $X_0$  has finite variance or  $X_0$  and  $\{\varepsilon_i, i \geq 1\}$  are in the same domain of attraction as the normal law with infinite variance. We have  $V_n(x) \geq 1 + C_1 x^2 \geq C_2(1 + |x|)$  whenever  $|x| \geq 1$  or  $|x| < 1$ , which implies that Assumption 2 holds (see Meyn and Tweedie (1993, p. 380)). Hence, (27) is proved.

To prove (28), we first construct the following new AR(1) model:

$$\begin{aligned} X'_1 &= \rho[X_0 \mathbf{1}\{|X_0| \leq \eta_n^*\} - E_\pi X_0 \mathbf{1}\{|X_0| \leq \eta_n^*\}] \\ &\quad + [\varepsilon_1 \mathbf{1}\{|\varepsilon_1| \leq \eta_n^*\} - E \varepsilon_1 \mathbf{1}\{|\varepsilon_1| \leq \eta_n^*\}] \end{aligned} \tag{30}$$

and, for  $k = 1, \dots, n - 1$ ,

$$X'_{k+1} = \rho X'_k + [\varepsilon_{k+1} \mathbf{1}\{|\varepsilon_{k+1}| \leq \eta_n^*\} - E \varepsilon_{k+1} \mathbf{1}\{|\varepsilon_{k+1}| \leq \eta_n^*\}]. \tag{31}$$

Then it is easy to see that

$$E_\pi X'_1 = 0 \quad \text{and} \quad E_\pi X'^2_1 \sim \frac{l^*(\eta_n^*)}{1 - \rho^2}. \tag{32}$$

Note that, for  $k = 2, \dots, n$ ,

$$\begin{aligned} X_{k-1} &= \sum_{i=1}^{k-1} \varepsilon_i \rho^{k-1-i} + \rho^{k-1} X_0 \\ &= \sum_{i=1}^{k-1} [\varepsilon_i \mathbf{1}\{|\varepsilon_i| \leq \eta_n^*\} - E \varepsilon_i \mathbf{1}\{|\varepsilon_i| \leq \eta_n^*\}] \rho^{k-1-i} \\ &\quad + \rho^{k-1} [X_0 \mathbf{1}\{|X_0| \leq \eta_n^*\} - E_\pi X_0 \mathbf{1}\{|X_0| \leq \eta_n^*\}] \\ &\quad + \sum_{i=1}^{k-1} [\varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_n^*\} - E \varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_n^*\}] \rho^{k-1-i} \\ &\quad + \rho^{k-1} [X_0 \mathbf{1}\{|X_0| > \eta_n^*\} - E_\pi X_0 \mathbf{1}\{|X_0| > \eta_n^*\}] \\ &=: X'_{k-1} + Y_{k-1}, \end{aligned}$$

where

$$\begin{aligned} Y_{k-1} &= \sum_{i=1}^{k-1} [\varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_n^*\} - E \varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_n^*\}] \rho^{k-1-i} \\ &\quad + \rho^{k-1} [X_0 \mathbf{1}\{|X_0| > \eta_n^*\} - E_\pi X_0 \mathbf{1}\{|X_0| > \eta_n^*\}]. \end{aligned} \tag{33}$$

Since

$$\frac{\sum_{k=1}^n \varepsilon_k^2 \mathbf{1}\{|\varepsilon_k| > \eta_n^*\} X_{k-1}^2}{nl^{*2}(\eta_n^*)} \leq \frac{(\sum_{k=1}^n |\varepsilon_k| \mathbf{1}\{|\varepsilon_k| > \eta_n^*\} |X_{k-1}|)^2}{nl^{*2}(\eta_n^*)},$$

by Lemma 1 and  $\eta_n^{*2} = nl^*(\eta_n^*)$ , we have

$$\frac{E(\sum_{k=1}^n |\varepsilon_k| \mathbf{1}\{|\varepsilon_k| > \eta_n^*\} |X_{k-1}|)}{\sqrt{nl^*(\eta_n^*)}} \leq \frac{Co(nl^*(\eta_n^*))}{\eta_n^* \sqrt{nl^*(\eta_n^*)}} = o(1).$$

It is clear that to prove (28), we need only show that, as  $n \rightarrow \infty$ ,

$$\begin{aligned} & \frac{(1 - \rho^2) \sum_{k=1}^n \varepsilon_k^2 \mathbf{1}\{|\varepsilon_k| \leq \eta_n^*\} (X_{k-1}'^2 + 2X_{k-1}' Y_{k-1} + Y_{k-1}^2)}{nl^{*2}(\eta_n^*)} \\ & \rightarrow 1 \quad \text{in probability.} \end{aligned} \tag{34}$$

First, we will show that, as  $n \rightarrow \infty$ ,

$$\frac{(1 - \rho^2) \sum_{k=1}^n \varepsilon_k^2 \mathbf{1}\{|\varepsilon_k| \leq \eta_n^*\} X_{k-1}'^2}{nl^{*2}(\eta_n^*)} \rightarrow 1 \quad \text{in probability.} \tag{35}$$

To this end, define  $\varepsilon'_k = \varepsilon_k \mathbf{1}\{|\varepsilon_k| \leq \eta_n^*\}$  and  $\xi'_k = \varepsilon'_k X'_{k-1}$  for  $k = 1, 2, \dots, n$ . Clearly,  $\xi'_k$  is a random variable in the domain of attraction of the normal law. In view of (32), to prove (35), we need only show that, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n (\xi_k'^2 - E_\pi \xi_k'^2)}{nl^{*2}(\eta_n^*)} \rightarrow 0 \quad \text{in probability.} \tag{36}$$

To prove (36), we write

$$\begin{aligned} & \frac{\sum_{k=1}^n (\xi_k'^2 - E_\pi \xi_k'^2)}{nl^{*2}(\eta_n^*)} \\ & = \frac{\sum_{k=1}^n \xi_k'^2 \mathbf{1}\{|\xi'_k| \leq \eta_n^* \sqrt{l^*(\eta_n^*)}\} - E_\pi (\sum_{k=1}^n \xi_k'^2 \mathbf{1}\{|\xi'_k| \leq \eta_n^* \sqrt{l^*(\eta_n^*)}\})}{nl^{*2}(\eta_n^*)} \\ & \quad + \frac{\sum_{k=1}^n \xi_k'^2 \mathbf{1}\{|\xi'_k| > \eta_n^* \sqrt{l^*(\eta_n^*)}\} - E_\pi (\sum_{k=1}^n \xi_k'^2 \mathbf{1}\{|\xi'_k| > \eta_n^* \sqrt{l^*(\eta_n^*)}\})}{nl^{*2}(\eta_n^*)} \\ & := I + II. \end{aligned}$$

To show that  $II \rightarrow 0$  in probability as  $n \rightarrow \infty$ , let  $l'(\cdot)$  denote the truncated second moment under  $\pi$  for the Markov random walk  $\xi'_i$ . Then it easily follows from assumption (25) that

$$l'(\eta_n^* \sqrt{l^*(\eta_n^*)}) \leq Cl^{*\alpha}(\eta_n^* \sqrt{l^*(\eta_n^*)}), \quad 1 < \alpha < 2. \tag{37}$$

Next, by Potter’s theorem for slowly varying functions (see Bingham *et al.* (1989, p. 25)), for any  $\beta > 0$ , there exist constants  $n_0 > 0$  and  $C = C(\beta, n_0)$  such that, for  $n > n_0$ ,

$$l^{*\alpha}(\eta_n^* \sqrt{l^*(\eta_n^*)}) \leq Cl^{*\alpha}(\eta_n^*) l^{*\beta}(\eta_n^*). \tag{38}$$

Recall that  $\eta_n^{*2} = nl^*(\eta_n^*)$ . By using Lemma 1, (37), and (38), we have

$$\begin{aligned} & \frac{E(\sum_{k=1}^n |\xi'_k| \mathbf{1}\{|\xi'_k| > \eta_n^* \sqrt{l^*(\eta_n^*)}\}) + E_\pi(\sum_{k=1}^n |\xi'_k| \mathbf{1}\{|\xi'_k| > \eta_n^* \sqrt{l^*(\eta_n^*)}\})}{\sqrt{nl^*(\eta_n^*)}} \\ &= no \left( \frac{l'(\eta_n^* \sqrt{l^*(\eta_n^*)})}{\eta_n^* \sqrt{l^*(\eta_n^*)}} \right) \frac{1}{\sqrt{nl^*(\eta_n^*)}} \\ &= o \left( \frac{l^{*(\alpha+\beta)}(\eta_n^*)}{l^{*2}(\eta_n^*)} \right) \\ &= o(1) \end{aligned}$$

by taking  $\beta = 2 - \alpha$ . This, together with the fact that

$$|II| \leq \frac{(\sum_{k=1}^n |\xi'_k| \mathbf{1}\{|\xi'_k| > \eta_n^* \sqrt{l^*(\eta_n^*)}\}) + E_\pi \sum_{k=1}^n |\xi'_k| \mathbf{1}\{|\xi'_k| > \eta_n^* \sqrt{l^*(\eta_n^*)}\})^2}{nl^{*2}(\eta_n^*)}$$

with probability 1, yields, as  $n \rightarrow \infty$ ,

$$II \rightarrow 0 \quad \text{in probability.}$$

To prove that  $I \rightarrow 0$  in probability as  $n \rightarrow \infty$ , we first show that the Markov chain induced by (30) and (31) is geometric ergodic. Let  $\eta'_n$  be as defined in (8) with  $\xi_1$  replaced by  $\xi'_1$ . Then, for  $x \in \mathcal{X}_n$  and large  $n$ ,

$$\begin{aligned} V'_n(x) &= 1 + E(\varepsilon_2'^2 X_1'^2 \mathbf{1}\{|\varepsilon_2' X_1'| \leq \eta'_n\} \mid X_1' = x) \\ &= 1 + x^2 E \varepsilon_2'^2 \mathbf{1}\left\{|\varepsilon_2'| \leq \frac{\eta'_n}{x}\right\} \\ &= 1 + x^2 E \varepsilon_2'^2 \mathbf{1}\left\{|\varepsilon_2| \leq \eta_n^*, |\varepsilon_2| \leq \frac{\eta'_n}{x}\right\} \\ &\geq 1 + Cx^2 \\ &\geq C(1 + |x|) \end{aligned}$$

(see Meyn and Tweedie (1993, p. 380)). Hence, by applying the same argument as used in (64) in Section 4, we have the following covariance estimation:

$$\begin{aligned} & |\text{cov}_\pi(\xi_i'^2 \mathbf{1}\{|\xi'_i| \leq \eta_n^* \sqrt{l^*(\eta_n^*)}\}, \xi_{i+j}'^2 \mathbf{1}\{|\xi'_{i+j}| \leq \eta_n^* \sqrt{l^*(\eta_n^*)}\})| \\ &= Cj^{-\beta} o(1 + \eta_n^{*2} l^*(\eta_n^*) l'(\eta_n^* \sqrt{l^*(\eta_n^*)})). \end{aligned}$$

By (37) and (38), we have, for any  $\varepsilon > 0$ ,

$$\begin{aligned} & P\left(\left| \frac{\sum_{k=1}^n \xi_k'^2 \mathbf{1}\{|\xi'_k| \leq \eta_n^* \sqrt{l^*(\eta_n^*)}\} - E_\pi \sum_{k=1}^n \xi_k'^2 \mathbf{1}\{|\xi'_k| \leq \eta_n^* \sqrt{l^*(\eta_n^*)}\}}{nl^{*2}(\eta_n^*)} \right| > \varepsilon \right) \\ &\leq \frac{no(\eta_n^{*2} l^*(\eta_n^*) l'(\eta_n^* \sqrt{l^*(\eta_n^*)})) + \sum_{i=1}^n (n-i) i^{-\beta} o(1 + \eta_n^{*2} l^*(\eta_n^*) l'(\eta_n^* \sqrt{l^*(\eta_n^*)}))}{\varepsilon^2 n^2 l^{*4}(\eta_n^*)} \\ &= o\left(\frac{l^{*(2+\alpha+\beta)}(\eta_n^*)}{l^{*4}(\eta_n^*)}\right) \\ &= o(1) \end{aligned}$$

by taking  $\beta = 2 - \alpha$  again, which means that, as  $n \rightarrow \infty$ ,

$$I \rightarrow 0 \quad \text{in probability.}$$

We now prove that (34) holds. Using (35) and the Cauchy–Schwarz inequality, we need only show that, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \varepsilon_k^2 \mathbf{1}\{|\varepsilon_k| \leq \eta_n^*\} Y_{k-1}^2}{nl^*(\eta_n^*)} \rightarrow 0 \quad \text{in probability.} \tag{39}$$

To prove (39), we first note that, with probability 1,

$$\left| \sum_{k=1}^n \varepsilon_k^2 \mathbf{1}\{|\varepsilon_k| \leq \eta_n^*\} Y_{k-1}^2 \right| \leq \left( \sum_{k=1}^n |\varepsilon_k| \mathbf{1}\{|\varepsilon_k| \leq \eta_n^*\} |Y_{k-1}| \right)^2. \tag{40}$$

By Lemma 1, we have

$$\begin{aligned} & \mathbb{E} \left( \sum_{k=1}^n |\varepsilon_k| \mathbf{1}\{|\varepsilon_k| \leq \eta_n^*\} |Y_{k-1}| \right) \\ & \leq C \sum_{k=1}^n \left( \sum_{i=1}^{k-1} \mathbb{E} |\varepsilon_i| \mathbf{1}\{|\varepsilon_i| > \eta_n^*\} \rho^{k-1-i} + \rho^{k-1} \mathbb{E} |X_0| \mathbf{1}\{|X_0| > \eta_n^*\} \right) \\ & = o \left( \frac{l^*(\eta_n^*)}{\eta_n^*} \right) \sum_{k=1}^n \left( \sum_{i=1}^{k-1} \rho^{k-1-i} + \rho^{k-1} \right) \\ & = o(\eta_n^*) \end{aligned} \tag{41}$$

whenever  $X_0$  has finite variance or  $X_0$  and  $\{\varepsilon_n, n \geq 1\}$  are in the same domain of attraction as the normal law with infinite variance. Therefore, by (40) and (41), we obtain (39).

To prove (29), we write

$$\frac{B_n}{nl^*(\eta_n^*)} = \frac{\sum_{k=1}^n (X_{k-1}'^2 + 2X_{k-1}' Y_{k-1} + Y_{k-1}^2)}{nl^*(\eta_n^*)},$$

where  $X_{k-1}'$  and  $Y_{k-1}$  are defined in (31) and (33), respectively. Note that  $\{X_n', n \geq 0\}$  is a Markov chain, and, for  $x \in \mathcal{X}_n$  and large  $n$ , we have

$$\begin{aligned} V_n''(x) &= 1 + \mathbb{E}(X_1'^2 \mathbf{1}\{|X_1'| \leq \tilde{\eta}_n\} \mid X_1' = x) \\ &= 1 + x^2 \mathbb{P}(|X_1'| \leq \tilde{\eta}_n) \\ &\geq 1 + Cx^2 \\ &\geq C(1 + |x|), \end{aligned}$$

where  $\tilde{\eta}_n$  is as defined in (8) with  $\xi_1$  replaced by  $X_1'$  (see Meyn and Tweedie (1993, p. 380)). This implies that Assumption 2 holds. Therefore, by applying the same argument as used in (64) below, we obtain the covariance estimation

$$|\text{cov}_\pi(X_i'^2 \mathbf{1}\{|X_i'| \leq \eta_n^*\}, X_{i+j}'^2 \mathbf{1}\{|X_{i+j}'| \leq \eta_n^*\})| = Cj^{-\beta} o(1 + \eta_n^{*2} \tilde{l}(\eta_n^*)),$$



where  $\tilde{l}(\cdot)$  denotes the truncated second moment under  $\pi$  for  $X'_1$ . Then we have, for any  $\varepsilon > 0$ ,

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{\sum_{k=1}^n X_{k-1}'^2 \mathbf{1}\{|X'_{k-1}| \leq \eta_n^*\} - \mathbb{E}_\pi \sum_{k=1}^n X_{k-1}'^2 \mathbf{1}\{|X'_{k-1}| \leq \eta_n^*\}}{nl^*(\eta_n^*)}\right| > \varepsilon\right) \\ & \leq \frac{no(\eta_n^{*2}\tilde{l}(\eta_n^*)) + \sum_{i=1}^n (n-i)i^{-\beta}o(1 + \eta_n^{*2}\tilde{l}(\eta_n^*))}{\varepsilon^2 n^2 l^{*2}(\eta_n^*)} \\ & = o\left(\frac{\tilde{l}(\eta_n^*)}{l^*(\eta_n^*)}\right) \\ & = o(1), \end{aligned}$$

which implies that

$$\frac{\sum_{k=1}^n X_{k-1}'^2 \mathbf{1}\{|X'_{k-1}| \leq \eta_n^*\}}{nl^*(\eta_n^*)} \rightarrow \frac{1}{1 - \rho^2} \quad \text{in probability.} \tag{42}$$

On the other hand, we have, with probability 1,

$$\sum_{k=1}^n Y_{k-1}^2 \leq \left(\sum_{k=1}^n |Y_{k-1}|\right)^2,$$

and

$$\frac{\mathbb{E} \sum_{k=1}^n |Y_{k-1}|}{\sqrt{nl^*(\eta_n^*)}} = o\left(\frac{l^*(\eta_n^*)}{\eta_n^*}\right) \sum_{k=1}^n \sum_{i=1}^{k-1} \rho^{k-1-i} \frac{1}{\sqrt{nl^*(\eta_n^*)}} = o(1),$$

which means that

$$\frac{\sum_{k=1}^n Y_{k-1}^2}{nl^*(\eta_n^*)} \rightarrow 0 \quad \text{in probability.} \tag{43}$$

Similarly, we have

$$\frac{\sum_{k=1}^n X_{k-1}'^2 \mathbf{1}\{|X'_{k-1}| > \eta_n^*\}}{nl^*(\eta_n^*)} \rightarrow 0 \quad \text{in probability.} \tag{44}$$

Combining (42) and (43) with (44), and using the Cauchy–Schwarz inequality, completes the proof of (29) and, thus, (26).

**Example 4.** Consider the linear state space model

$$X_n = \rho X_{n-1} + \varepsilon_n, \quad \xi_n = \theta X_n + \zeta_n,$$

where  $|\rho| < 1$  and  $\theta \in \mathbb{R}$ . Furthermore, assume that  $\{\varepsilon_n, n \geq 1\}$  and  $\{\zeta_n, n \geq 1\}$  are independent. Note that  $S_n = \sum_{k=1}^n \xi_k$  is a Markov random walk.

Let  $\{\varepsilon_n, n \geq 1\}$  be a sequence of i.i.d. random variables with distribution  $N(0, \sigma_1^2)$ , and let  $\{\zeta_n, n \geq 1\}$  be a sequence of i.i.d. random variables with distribution  $N(0, \sigma_2^2)$ . It is easy to see that  $\Gamma(x) = \mathbb{E}_x \xi_1 = \theta \rho x$ ,  $G(x) = \mathbb{E}_x \xi_1^2 = \theta^2(\rho^2 x^2 + \sigma_1^2) + \sigma_2^2$ , and  $P\Gamma(x) = \theta \rho^2 x$ . Then the solutions of the Poisson equation (5) are

$$\delta_x = \frac{\theta \rho^2 x}{1 - \rho} + C \quad \text{for some constant } C.$$

Therefore, the asymptotic variance of  $\sum_{k=1}^n \xi_k / \sqrt{n}$  is

$$\begin{aligned} \Sigma &= \int_{-\infty}^{+\infty} (G(x) - \Gamma^2(x))^2 \pi(dx) \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\Gamma(y) - \delta_x + \delta_y)^2 P(x, dy) \pi(dx) \\ &= \theta^2 \sigma_1^2 + \sigma_2^2 + \frac{\theta^2 \rho^2}{(1 - \rho)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \rho x)^2 P(x, dy) \pi(dx) \\ &= \left(1 + \frac{\rho^2}{(1 - \rho)^2}\right) \theta^2 \sigma_1^2 + \sigma_2^2. \end{aligned} \tag{45}$$

Hence, we have, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \xi_k}{\sqrt{n((1 + \rho^2/(1 - \rho)^2)\theta^2 \sigma_1^2 + \sigma_2^2)}} \rightarrow N(0, 1) \text{ in distribution.} \tag{46}$$

For the self-normalized central limit theorem of  $\sum_{k=1}^n \xi_k$ , we simply replace the unknown parameters in (46) by their consistent estimators (the maximum likelihood estimator for instance—see Fuh (2006)) and apply Proposition 1.

If  $\text{var}(\varepsilon_n) < \infty$  and  $\zeta_n$  is in the domain of attraction of the normal law with  $\text{var}(\zeta_n) = \infty$ , then the above argument is not applicable. We would like to apply Theorem 1(b) to this case. To this end, we need to check that Assumptions 1 and 2 hold.

It is easy to see that Assumption 1 is satisfied. To check that Assumption 2 holds, we consider

$$\begin{aligned} V_n(x) &= 1 + E(\xi_1^2 \mathbf{1}\{|\xi_1| \leq \eta_n\} \mid X_1 = x) \\ &= 1 + E((\theta X_1 + \zeta_1)^2 \mathbf{1}\{|\theta X_1 + \zeta_1| \leq \eta_n\} \mid X_1 = x) \\ &= 1 + \theta^2 x^2 P(|\theta x + \zeta_1| \leq \eta_n) + 2\theta x E(\zeta_1 \mathbf{1}\{|\theta x + \zeta_1| \leq \eta_n\}) \\ &\quad + E(\zeta_1^2 \mathbf{1}\{|\theta x + \zeta_1| \leq \eta_n\}) \\ &\geq 1 + \theta^2 x^2 P(|\theta x + \zeta_1| \leq \eta_n) + 2\theta x E(\zeta_1 \mathbf{1}\{|\theta x + \zeta_1| \leq \eta_n\}). \end{aligned}$$

Since, for large  $n$ ,  $x \in \mathcal{X}_n$ , and  $0 < C < 1$ ,  $\{|\zeta_1| \leq C\eta_n\} \subseteq \{|\zeta_1| \leq \eta_n - |\theta x|\} \subseteq \{|\theta x + \zeta_1| \leq \eta_n\}$  and  $P(|\zeta_1| \leq C\eta_n) \rightarrow 1$  as  $n \rightarrow \infty$ , we have

$$V_n(x) \geq 1 + Cx^2 + o(x) \geq C_1(1 + |x|),$$

which implies that the polynomial ergodicity in Assumption 2 holds. Therefore, by Theorem 1(b) we have, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^n \xi_k}{\sqrt{\sum_{k=1}^n \xi_k^2}} \rightarrow N(0, 1) \text{ in distribution.} \tag{47}$$

Note that in this example, when the variances of  $X_n$  and  $\zeta_n$  are both finite, and the transition probability and stationary distribution are given, then we have the asymptotic variance as shown in (45), and obtain the classical central limit theorem for Markov random walks. Furthermore, the self-normalized central limit theorem can be obtained by replacing the parameters by their consistent estimators; while if the transition probability or the stationary distribution are not

given, we can apply Theorem 1(a) to have the self-normalized central limit theorem. When  $\text{var}_\pi(X_n)$  is finite and  $\text{var}(\zeta_n)$  is infinite, we can also apply Theorem 1(b) to obtain the self-normalized central limit theorem as shown in (47).

In this example we only consider the case of partial sums  $\sum_{k=1}^n \xi_k$ . For the case of the parameter estimation of  $\theta$ , the asymptotic normality of the maximum likelihood estimator  $\hat{\theta}$  of  $\theta$  can be found in Fuh (2006) under the finite second moment assumption. To obtain a self-normalized central limit theorem for  $\hat{\theta} - \theta$ , when  $\text{var}_\pi(X_n) < \infty$  and  $\text{var}(\zeta_n) = \infty$ , is an interesting task.

### 4. Proofs of Theorems 1 and 2

Since the proof of Theorem 1 involves a result in the proof of Theorem 2, we prove Theorem 2 first.

*Proof of Theorem 2.* We will use a splitting chain argument to prove this theorem. Under the minorization condition in Assumption 1, without loss of generality, we may assume that there exists recurrent state  $\Delta$  in  $\mathcal{X}$  such that the Markov chain  $X_n$  visits the state  $\Delta$  infinitely often P-almost surely, and  $\pi(\Delta) > 0$ . Define the nonnegative integer-valued random sequence  $\{i_\Delta(n)\}_{n \geq 1}$  as

$$i_\Delta(n) = \sum_{k=1}^n I_\Delta(X_k) = \max\{k \geq 0; \tau_\Delta(k) \leq n\} \quad \text{for } n = 1, 2, \dots,$$

and let

$$l_\Delta(n) = \tau_\Delta(i_\Delta(n)) = \max\{0 \leq k \leq n; X_k \in \Delta\}.$$

Recall  $\tilde{\xi}_{i,n}$  defined in (9), and let

$$\sum_{i=1}^n \tilde{\xi}_{i,n} = \sum_{k=0}^{i_\Delta(n)-1} \mu_k + \sum_{i=l_\Delta(n)+1}^n \tilde{\xi}_{i,n},$$

where

$$\mu_k = \sum_{i=\tau_\Delta(k)+1}^{\tau_\Delta(k+1)} \tilde{\xi}_{i,n}, \quad k = 1, 2, \dots,$$

forms an i.i.d. sequence of random variables.

Since  $E_\Delta \tilde{\xi}_{i,n} < \infty$  and  $E_\Delta \tau_\Delta < \infty$ , by Wald’s equation for the second moment (see Chow and Teicher (1998, p. 137)) we have

$$E_\Delta \left( \sum_{k=0}^{i_\Delta(n)-1} \mu_k \right)^2 = E_\Delta i_\Delta(n) E \mu_1^2. \tag{48}$$

In addition, taking  $f(\xi_i) = \tilde{\xi}_{i,n} / \sqrt{l(\eta_n)}$  in Lemma 2.3 of Chen (1999), we have

$$E \mu_1^2 = E_\Delta \tau_\Delta \left( \text{var}_\pi(\tilde{\xi}_{1,n}) + 2 \sum_{i=1}^\infty E_\pi(\tilde{\xi}_{1,n} \tilde{\xi}_{i+1,n}) \right), \tag{49}$$

which implies that

$$C_1 E_\Delta \tau_\Delta l(\eta_n) \leq E \mu_1^2 \leq C_2 E_\Delta \tau_\Delta l(\eta_n)$$

by means of (63) below (Assumption 2 is used). Hence, we have

$$C_1 l(\eta_n) E_\Delta i_\Delta(n) E_\Delta \tau_\Delta \leq E_\Delta \left( \sum_{k=0}^{i_\Delta(n)-1} \mu_k \right)^2 \leq C_2 l(\eta_n) E_\Delta i_\Delta(n) E_\Delta \tau_\Delta. \tag{50}$$

Since

$$\frac{E_\Delta \tau_\Delta(i_\Delta(n))}{n} \leq 1 \leq \frac{E_\Delta \tau_\Delta(i_\Delta(n)) + E_\Delta \tau_\Delta}{n} = \frac{E_\Delta \tau_\Delta(i_\Delta(n))}{n} + \frac{1}{n\pi(\Delta)},$$

we have

$$\frac{E_\Delta \tau_\Delta(i_\Delta(n))}{n} \rightarrow 1 \tag{51}$$

as  $n \rightarrow \infty$ . By using Wald’s equation again, we obtain

$$E_\Delta \tau_\Delta(i_\Delta(n)) = E_\Delta \sum_{k=1}^{i_\Delta(n)} \sum_{i=\tau_\Delta(k-1)+1}^{\tau_\Delta(k)} 1 = E_\Delta i_\Delta(n) E_\Delta \tau_\Delta. \tag{52}$$

It follows from (50)–(52) that, when  $n$  is large enough,

$$C_1 n l(\eta_n) \leq E_\Delta \left( \sum_{k=0}^{i_\Delta(n)-1} \mu_k \right)^2 \leq C_2 n l(\eta_n). \tag{53}$$

Moreover, it is easy to see that

$$E_\Delta \left| \sum_{i=l_\Delta(n)+1}^n \tilde{\xi}_{i,n} \right| \leq E_\Delta \left( \sum_{i=1}^{\tau_\Delta} |\tilde{\xi}_{i,n}| \right) < \infty. \tag{54}$$

Using the facts that  $i_\Delta(n)/n \rightarrow \pi(\Delta) > 0$  in probability and  $i_\Delta(n)/E_\Delta i_\Delta(n) \rightarrow 1$  in probability, (48), (53), (54), and the central limit theorem for random sums (see Billingsley (1999, p. 153)), we have

$$\frac{\sum_{i=1}^n \tilde{\xi}_{i,n}}{\sqrt{E_\Delta (\sum_{k=0}^{i_\Delta(n)-1} \mu_k)^2}} \rightarrow N(0, 1) \text{ in distribution.} \tag{55}$$

Now we rewrite  $S_n$  as

$$S_n = \sum_{i=1}^n \tilde{\xi}_{i,n} + \sum_{i=1}^n (\xi_i \mathbf{1}\{|\xi_i| > \eta_n\} - E_\pi \xi_i \mathbf{1}\{|\xi_i| > \eta_n\}) := S_{n1} + S_{n2}.$$

In view of Lemma 1,

$$\frac{E_\pi |S_{n2}|}{\sqrt{n l(\eta_n)}} \leq \frac{2no(l(\eta_n)/\eta_n)}{\sqrt{n l(\eta_n)}} = o(1).$$

Moreover, it follows, from (48), (49), (51), (52), and (57) below, that

$$\frac{E_\Delta (\sum_{k=0}^{i_\Delta(n)-1} \mu_k)^2}{\text{var}_\pi (\sum_{i=1}^n \xi_i \mathbf{1}\{|\xi_i| \leq \eta_n\})} \rightarrow 1. \tag{56}$$

Hence, combining (55) and (56) completes the proof.

*Proof of Theorem 1.* By Theorem 2, we need only show that, under finite and infinite second moment conditions of  $\xi_1$ , the two random normalizing factors in Theorem 1 are consistent estimators of the normalizing factor in Theorem 2. To this end, we first rewrite the normalizing factor in Theorem 2 as

$$\text{var}_\pi \left( \sum_{i=1}^n \tilde{\xi}_{i,n} \right) = n(\Sigma_n + o(l(\eta_n))), \tag{57}$$

where

$$\begin{aligned} \Sigma_n &= \int_{-\infty}^{\infty} (G_n(x) - \Gamma_n^2(x))\pi(dx) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\Gamma_n(y) - \delta_{x,n} + \delta_{y,n})^2 P(x, dy)\pi(dx) \\ &= \int_{-\infty}^{\infty} G_n(x)\pi(dx) \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [-2(\delta_{x,n} - \delta_{y,n})\Gamma_n(y) + (\delta_{x,n} - \delta_{y,n})^2]P(x, dy)\pi(dx), \end{aligned} \tag{58}$$

with  $G_n(x) = E_x(\xi_1^2)$ ,  $\Gamma_n(x) = E_x(\xi_{1,n})$ , and  $\delta_{x,n}$  and  $\delta_{y,n}$  solutions to the Poisson equation (10). Then we prove that  $|\Delta_{n,M}|$  and  $\sum_{i=1}^n \xi_i^2$  are consistent estimators of  $n\Sigma_n$ .

Now let

$$\begin{aligned} g_n(x) &= E(\xi_1 \mathbf{1}\{|\xi_1| \leq \eta_n\} \mid X_1 = x), \\ \tilde{g}_n(x) &= E(\xi_1 \mathbf{1}\{|\xi_1| \leq \eta_n\} \mid X_1 = x) - E_\pi(\xi_1 \mathbf{1}\{|\xi_1| \leq \eta_n\}). \end{aligned}$$

Recall that  $\tilde{\xi}_{i,n} = \xi_i \mathbf{1}\{|\xi_i| \leq \eta_n\} - E_\pi(\xi_i \mathbf{1}\{|\xi_i| \leq \eta_n\})$ ,  $i = 1, 2, \dots$ . Since  $\tilde{\xi}_{i,n}$  and  $\tilde{\xi}_{i+j,n}$  are independent for given  $X_i, X_{i+j}$ , we have

$$\begin{aligned} E_x(\tilde{\xi}_{i,n}\tilde{\xi}_{i+j,n}) &= E(E(\tilde{\xi}_{i,n}\tilde{\xi}_{i+j,n} \mid X_i, X_{i+j}) \mid X_0 = x) \\ &= E(E(\tilde{\xi}_{i,n} \mid X_i, X_{i+j}) E(\tilde{\xi}_{i+j,n} \mid X_i, X_{i+j}) \mid X_0 = x) \\ &= E(E(\tilde{\xi}_{i,n} \mid X_i) E(\tilde{\xi}_{i+j,n} \mid X_{i+j}) \mid X_0 = x) \\ &= E(\tilde{g}_n(X_i)\tilde{g}_n(X_{i+j}) \mid X_0 = x). \end{aligned} \tag{59}$$

Note that  $g_n^2(x) \leq V_n(x)$  via Jensen’s inequality. By making use of the Markovian property of  $\{g_n(X_i), i \geq 1\}$ , and following the same argument as that used in Theorem 16.1.5 and Equation (16.16) of Meyn and Tweedie (1993, p. 388), with  $|\tilde{g}_n| \leq (1 + \int \sqrt{V_n} d\pi)\sqrt{V_n}$  replaced by  $|\tilde{g}_n| \leq \sqrt{V_n} + E_\pi \sqrt{V_n}$ , there exists  $C_1 < \infty$  such that

$$|E_x \tilde{g}_n(X_i)\tilde{g}_n(X_{i+j})| \leq C_1 j^{-\beta} (1 + i^{-\beta} V_n(\ddot{x}) + i^{-\beta} \sqrt{V_n(\ddot{x})}), \tag{60}$$

where  $\beta$  is defined in (11) and  $\ddot{x}$  denotes that  $x$  is restricted to the set  $\mathcal{X}_n$ . Define

$$\ddot{X}_i = \begin{cases} X_i & \text{if } X_i \in \mathcal{X}_n, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$E_\pi \sqrt{V_n(\ddot{X}_i)} \leq \sqrt{E_\pi V_n(\ddot{X}_i)} = \sqrt{1 + E_\pi(\xi_1^2 \mathbf{1}\{|\xi_1| \leq \eta_n\} \mid \ddot{X}_1)} \leq C\sqrt{l(\eta_n)}. \tag{61}$$

Integrating both sides of (60) and applying (61), we obtain

$$|E_\pi \tilde{g}_n(X_i)\tilde{g}_n(X_{i+j})| \leq C_2 j^{-\beta} l(\eta_n) \quad \text{for some } C_2 < \infty. \tag{62}$$

Combining (62) with (59) yields

$$|E_{\pi}(\tilde{\xi}_{i,n}\tilde{\xi}_{i+j,n})| \leq C_2j^{-\beta}l(\eta_n). \tag{63}$$

Let

$$V'_n(x) = 1 + E(\xi_1^4 \mathbf{1}\{|\xi_1| \leq \eta_n\} \mid X_1 = x), \quad g'_n(x) = E(\xi_1^2 \mathbf{1}\{|\xi_1| \leq \eta_n\} \mid X_1 = x).$$

Then we have

$$[g'_n(x)]^2 \leq V'_n(x).$$

Moreover, define

$$\bar{\xi}_{i,n} = \xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\} - E_{\pi}(\xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\}), \quad i = 1, 2, \dots$$

By applying the same argument as used in (63) we have

$$\begin{aligned} |E_{\pi}(\bar{\xi}_{i,n}\bar{\xi}_{i+j,n})| &\leq C_3j^{-\beta} \left( E_{\pi} V'_n(\bar{X}_i) + E_{\pi} \sqrt{V'_n(\bar{X}_i)} \right) \\ &\leq C_4j^{-\beta}(1 + E_{\pi}(\xi_1^4 \mathbf{1}\{|\xi_1| \leq \eta_n\})) \\ &= C_4j^{-\beta}(1 + o(\eta_n^2l(\eta_n))) \end{aligned} \tag{64}$$

for some  $C_3, C_4 < \infty$ .

It is easy to see that

$$\begin{aligned} \text{var}_{\pi} \left( \sum_{i=1}^n \tilde{\xi}_{i,n} \right) &= n \text{var}_{\pi}(\tilde{\xi}_{1,n}) + 2 \sum_{i=1}^{n-1} (n-i) E_{\pi}(\tilde{\xi}_{1,n}\tilde{\xi}_{i+1,n}) \\ &= n \left( \text{var}_{\pi}(\tilde{\xi}_{1,n}) + 2 \sum_{i=1}^{\infty} E_{\pi}(\tilde{\xi}_{1,n}\tilde{\xi}_{i+1,n}) - 2 \sum_{i=n}^{\infty} E_{\pi}(\tilde{\xi}_{1,n}\tilde{\xi}_{i+1,n}) \right. \\ &\quad \left. - \frac{2}{n} \sum_{i=1}^{n-1} i E_{\pi}(\tilde{\xi}_{1,n}\tilde{\xi}_{i+1,n}) \right). \end{aligned}$$

By (63),  $\sum_{i=1}^{\infty} i^{-\beta} < \infty$ , and the Kronecker lemma, we have

$$\begin{aligned} \left| \sum_{i=n}^{\infty} E_{\pi}(\tilde{\xi}_{1,n}\tilde{\xi}_{i+1,n}) + \frac{1}{n} \sum_{i=1}^{n-1} i E_{\pi}(\tilde{\xi}_{1,n}\tilde{\xi}_{i+1,n}) \right| &\leq C \sum_{i=n}^{\infty} i^{-\beta}l(\eta_n) + \frac{C}{n} \sum_{i=1}^{n-1} ii^{-\beta}l(\eta_n) \\ &= o(l(\eta_n)). \end{aligned} \tag{65}$$

By using (65),  $C_1l(\eta_n) \leq \Sigma_n \leq C_2l(\eta_n)$  from (53), and following the same argument as that used to obtain the asymptotic variance in Fuh and Hu (2007), we have

$$\text{var}_{\pi} \left( \sum_{i=1}^n \tilde{\xi}_{i,n} \right) = n(\Sigma_n + o(l(\eta_n))),$$

where  $\Sigma_n$  is defined in (58).

(a) If  $\text{var}_\pi(\xi_1) < \infty$ , we show that the second term in (58) is not negligible, and we find a consistent estimator. To this end, we first show that

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [-2(\delta_{x,n} - \delta_{y,n})\Gamma_n(y) + (\delta_{x,n} - \delta_{y,n})^2]P(x, dy)\pi(dx) \\ & \sim \int_{-M}^M \int_{-M}^M [-2(\delta_{x,n} - \delta_{y,n})\Gamma_n(y) + (\delta_{x,n} - \delta_{y,n})^2]P(x, dy)\pi(dx) \end{aligned} \tag{66}$$

for some fixed, large  $M$ , and then we only need to show that the random normalizing factor in Theorem 1(a) is the consistent estimator of the product of  $n$  and the term appearing on the right-hand side of (66).

It is easy to see that  $\sum_{i=1}^n \xi_i^2 / (n \int_{-\infty}^{\infty} G_n(x)\pi(dx))$  converges to 1 in probability. Next, we will show that  $\hat{\Gamma}_n(x)$ ,  $\hat{q}_n(x, y)$ , and  $\hat{\delta}_{x,n} - \hat{\delta}_{y,n}$  are consistent estimators of  $\Gamma_n(x)$ ,  $q(x, y)$ , and  $\delta_{x,n} - \delta_{y,n}$ , respectively.

To prove that  $\hat{\Gamma}_n(x) \rightarrow \Gamma(x)$  in probability, let  $W_j = (1/h)\xi_j K((x - X_j)/h)$ , and, for given  $\delta > 0$ , define

$$V_j = V_j(n) = \begin{cases} W_j & \text{if } |W_j| \leq n\delta, \\ 0 & \text{otherwise.} \end{cases}$$

Note that  $\{X_n, n \geq 0\}$  is a Markov chain, the partial sum of the sequence  $\{V_n, n \geq 0\}$  is a Markov random walk, and that the function  $K(\cdot)$  is bounded. By applying an argument similar to that used in (63), we obtain

$$|\text{cov}_\pi(V_i, V_{i+j})| \leq Cj^{-\beta}l(n). \tag{67}$$

Under the assumptions in Theorem 1(a), and replacing the covariance inequality in Lemma 4.3 of Roussas (1969b) by (67), the rest of the proof follows the same argument as that in Roussas (1969b, pp. 1393–1394) and is thus omitted.

It follows from (10) and Assumption 4 that  $\delta_{x,n}$  is a continuous function of  $p(x, y)$  and  $\Gamma_n(x)$ , P-almost surely. Using the fact that the set of  $|\delta_x| < \infty$  is full (see Meyn and Tweedie (1993, p. 431)) under Assumptions 1 and 4, we have, for any  $x, y \in \mathcal{X}$ ,

$$(\delta_{x,n} - \delta_{y,n}) - (\delta_x - \delta_y) \rightarrow 0 \quad \text{in probability.}$$

From Roussas (1969a) we know that  $\hat{q}_n(x, y) \rightarrow q(x, y)$  in probability, which together with the smooth condition in Assumption 4 yields, for  $x, y \in \mathcal{X}$ ,

$$(\hat{\delta}_{x,n} - \hat{\delta}_{y,n}) - (\delta_x - \delta_y) \rightarrow 0 \quad \text{in probability.}$$

In addition, we have  $\sup_{[a,b]} \Gamma(x) < \infty$ , P-almost surely for any compact set  $[a, b]$ , since  $E_\pi \Gamma(x) < \infty$ . Therefore, we conclude that, for any given small  $\varepsilon > 0$  and large  $M > 0$ ,

$$\sup_{x \in I_M} |\hat{\Gamma}_n(x) - \Gamma(x)| \leq \varepsilon \quad \text{in probability,} \tag{68a}$$

$$\sup_{(x,y) \in \ddot{D}_M} \left| \frac{\hat{q}_n(x, y)}{q(x, y)} - 1 \right| \leq \varepsilon \quad \text{in probability,} \tag{68b}$$

$$\sup_{(x,y) \in D_M} |(\hat{\delta}_{x,n} - \hat{\delta}_{y,n}) - (\delta_{x,n} - \delta_{y,n})| \leq \varepsilon \quad \text{in probability} \tag{68c}$$

for large enough  $n$ , where  $I_M = [-M, M]$ ,  $D_M = [-M, M] \times [-M, M]$ , and  $\ddot{D}_M = \{(x, y) : (x, y) \in [-M, M] \times [-M, M], q(x, y) \neq 0\}$ .

Under the condition that  $E_{\Delta} \tau_{\Delta}^2 < \infty$  given in Assumption 1, we have

$$\frac{|\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\delta_{x,n} - \delta_{y,n})^2 P(x, dy) \pi(dx)|}{l(\eta_n)} \leq \frac{2E_{\pi} \delta_{x,n}^2}{l(\eta_n)} \leq \frac{E_{\pi}(E_{X_0} \tau_{\Delta})^2 + C}{l(\eta_n)} < \infty,$$

which implies that

$$\frac{|\int_{|x|>M} \int_{|y|>M} (\delta_{x,n} - \delta_{y,n})^2 P(x, dy) \pi(dx)|}{l(\eta_n)} \rightarrow 0, \tag{69a}$$

$$\frac{|\int_{|x|<M} \int_{|y|>M} (\delta_{x,n} - \delta_{y,n})^2 P(x, dy) \pi(dx)|}{l(\eta_n)} \rightarrow 0, \tag{69b}$$

$$\frac{|\int_{|x|>M} \int_{|y|<M} (\delta_{x,n} - \delta_{y,n})^2 P(x, dy) \pi(dx)|}{l(\eta_n)} \rightarrow 0, \tag{69c}$$

as  $M \rightarrow \infty$ . These together with the fact that  $E_{\pi} \Gamma_n^2(y)/l(\eta_n) < \infty$  yield

$$\frac{|\int_{|x|>M} \int_{|y|>M} (\delta_{x,n} - \delta_{y,n}) \Gamma_n(y) P(x, dy) \pi(dx)|}{l(\eta_n)} \rightarrow 0, \tag{70a}$$

$$\frac{|\int_{|x|<M} \int_{|y|>M} (\delta_{x,n} - \delta_{y,n}) \Gamma_n(y) P(x, dy) \pi(dx)|}{l(\eta_n)} \rightarrow 0, \tag{70b}$$

$$\frac{|\int_{|x|>M} \int_{|y|<M} (\delta_{x,n} - \delta_{y,n}) \Gamma_n(y) P(x, dy) \pi(dx)|}{l(\eta_n)} \rightarrow 0, \tag{70c}$$

as  $M \rightarrow \infty$ . Hence, in order to complete the proof, we need to show that

$$\frac{|\iint \ddot{D}_M [(\delta_{x,n} - \delta_{y,n})^2 q(x, y) - (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})^2 \hat{q}_n(x, y)] dy dx|}{l(\eta_n)} \rightarrow 0 \text{ in probability} \tag{71}$$

as  $n \rightarrow \infty$  and

$$\frac{|\iint \ddot{D}_M [\Gamma_n(y) (\delta_{x,n} - \delta_{y,n}) q(x, y) - \hat{\Gamma}_n(y) (\hat{\delta}_{x,n} - \hat{\delta}_{y,n}) \hat{q}_n(x, y)] dy dx|}{l(\eta_n)} \rightarrow 0 \text{ in probability} \tag{72}$$

as  $n \rightarrow \infty$  for given large  $M > 0$ .

For the proof of (71), write, for  $(x, y) \in \ddot{D}_M$  and large  $n$ ,

$$\begin{aligned} & |(\delta_{x,n} - \delta_{y,n})^2 q(x, y) - (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})^2 \hat{q}_n(x, y)| \\ &= |(\delta_{x,n} - \delta_{y,n})^2 [q(x, y) - \hat{q}_n(x, y)] + [(\delta_{x,n} - \delta_{y,n})^2 - (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})^2] \hat{q}_n(x, y)| \\ &\leq q(x, y) (\delta_{x,n} - \delta_{y,n})^2 \left| 1 - \frac{\hat{q}_n(x, y)}{q(x, y)} \right| \\ &\quad + |(\delta_{x,n} - \delta_{y,n}) - (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})| |(\delta_{x,n} - \delta_{y,n}) + (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})| \hat{q}_n(x, y). \end{aligned}$$



Then, by (68), for large enough  $n$ , we have

$$\begin{aligned} & \frac{|\iint_{\ddot{D}_M} [(\delta_{x,n} - \delta_{y,n})^2 q(x, y) - (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})^2 \hat{q}_n(x, y)] dy dx|}{l(\eta_n)} \\ & \leq \varepsilon \iint_{\ddot{D}_M} (\delta_{x,n} - \delta_{y,n})^2 q(x, y) dy dx / l(\eta_n) \\ & \quad + \frac{\varepsilon(1 + \varepsilon)}{l(\eta_n)} \iint_{\ddot{D}_M} |(\delta_{x,n} - \delta_{y,n}) + (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})| q(x, y) dy dx. \end{aligned} \tag{73}$$

Note that (73) is  $o(1)$  in probability, because of the boundness of  $|(\delta_{x,n} - \delta_{y,n}) + (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})|$ , and  $\varepsilon$  can be arbitrary small.

Using the same argument, we have, for  $(x, y) \in \ddot{D}_M$  and large  $n$ ,

$$\begin{aligned} & |\Gamma_n(y)(\delta_{x,n} - \delta_{y,n})q(x, y) - \hat{\Gamma}_n(y)(\hat{\delta}_{x,n} - \hat{\delta}_{y,n})\hat{q}_n(x, y)| \\ & = |\Gamma_n(y)(\delta_{x,n} - \delta_{y,n})[q(x, y) - \hat{q}_n(x, y)] \\ & \quad + [\Gamma_n(y)(\delta_{x,n} - \delta_{y,n}) - \hat{\Gamma}_n(y)(\hat{\delta}_{x,n} - \hat{\delta}_{y,n})]\hat{q}_n(x, y)| \\ & = |\Gamma_n(y)(\delta_{x,n} - \delta_{y,n})[q(x, y) - \hat{q}_n(x, y)] + [\Gamma_n(y) - \hat{\Gamma}_n(y)](\delta_{x,n} - \delta_{y,n})\hat{q}_n(x, y) \\ & \quad + [(\delta_{x,n} - \delta_{y,n}) - (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})]\hat{\Gamma}_n(y)\hat{q}_n(x, y)| \\ & \leq |\Gamma_n(y)||\delta_{x,n} - \delta_{y,n}| \left| 1 - \frac{\hat{q}_n(x, y)}{q(x, y)} \right| q(x, y) + |[\Gamma_n(y) - \hat{\Gamma}_n(y)](\delta_{x,n} - \delta_{y,n})\hat{q}_n(x, y)| \\ & \quad + |(\delta_{x,n} - \delta_{y,n}) - (\hat{\delta}_{x,n} - \hat{\delta}_{y,n})|\hat{\Gamma}_n(y)\hat{q}_n(x, y). \end{aligned} \tag{74}$$

Using (68) in (74) yields (72). Combining (69), (70), (71), and (72), completes the proof of Theorem 1(a).

(b) If  $\text{var}_\pi(\xi_1) = \infty$ , we first show that the second term in (58) can be ignored, and then we show that  $\sum_{i=1}^n \xi_i^2$  is the consistent estimator of  $n \int_{-\infty}^{\infty} G_n(x)\pi(dx)$ .

We analyze the second term of (58) first. By Wald's equation in Fuh and Zhang (2000) we have

$$\frac{|\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\delta_{x,n} - \delta_{y,n})^2 P(x, dy)\pi(dx)|}{l(\eta_n)} \leq \frac{2E_\pi \delta_{x,n}^2}{l(\eta_n)} \leq \frac{E_\pi(E_{X_0} \tau_\Delta)^2 + C}{l(\eta_n)} = o(1). \tag{75}$$

Note that the last term in (75) follows from the conditions that  $E_\Delta \tau_\Delta^2 < \infty$  given in Assumption 1 and  $l(\eta_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Along with

$$\frac{E_\pi \Gamma_n^2(y)}{l(\eta_n)} \leq \frac{E_\pi E_y \xi_{1,n}^2}{l(\eta_n)} \leq 1$$

we obtain, as  $n \rightarrow \infty$ ,

$$\begin{aligned} & \frac{|\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\delta_{x,n} - \delta_{y,n})\Gamma_n(y)P(x, dy)\pi(dx)|}{l(\eta_n)} \\ & \leq \sqrt{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(\delta_{x,n} - \delta_{y,n})^2}{l(\eta_n)} P(x, dy)\pi(dx) \frac{E_\pi \Gamma_n^2(y)}{l(\eta_n)}} \\ & = o(1). \end{aligned} \tag{76}$$

By (75) and (76), we need to prove that

$$\frac{n \int_{-\infty}^{\infty} G_n(x)\pi(dx) - \sum_{i=1}^n \xi_i^2}{nl(\eta_n)} \rightarrow 0 \quad \text{in probability.} \tag{77}$$

We now prove that (77) holds. Since  $n P_{\pi}(|\xi_1| > \eta_n) = o(1)$  by Lemma 1(b), we need only show that

$$\begin{aligned} & \frac{n \int_{-\infty}^{\infty} G_n(x)\pi(dx) - \sum_{i=1}^n \xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\}}{nl(\eta_n)} \\ &= \frac{\sum_{i=1}^n (\mathbb{E}_{\pi}(\xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\}) - \xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\})}{nl(\eta_n)} - \frac{\sum_{i=1}^n (\mathbb{E}_{\pi} \xi_i \mathbf{1}\{|\xi_i| \leq \eta_n\})^2}{nl(\eta_n)} \\ &\rightarrow 0 \quad \text{in probability.} \end{aligned} \tag{78}$$

Obviously, the second term in (78) converges to 0 as  $n \rightarrow \infty$ . Next, we will show that the first term in (78) also converges to 0 in probability as  $n \rightarrow \infty$ . In view of Chebyshev’s inequality,

$$\begin{aligned} & P_{\pi} \left( \frac{|\sum_{i=1}^n [\mathbb{E}_{\pi}(\xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\}) - \xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\}]|}{nl(\eta_n)} > \varepsilon \right) \\ &\leq \frac{\text{var}_{\pi}[\sum_{i=1}^n (\xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\}) - \mathbb{E}_{\pi}(\xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\})]}{\varepsilon^2 n^2 l^2(\eta_n)} \\ &= \frac{n \mathbb{E}_{\pi}[\xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\}) - \mathbb{E}_{\pi}(\xi_i^2 \mathbf{1}\{|\xi_i| \leq \eta_n\})]^2}{\varepsilon^2 n^2 l^2(\eta_n)} \\ &\quad + \frac{\sum_{i=1}^n (n-i) \text{cov}_{\pi}(\bar{\xi}_{1,n}, \bar{\xi}_{i+1,n})}{\varepsilon^2 n^2 l^2(\eta_n)} \\ &=: III + IV. \end{aligned}$$

By Lemma 1 we have

$$III \leq \frac{n \mathbb{E}_{\pi}(\xi_i^4 \mathbf{1}\{|\xi_i| \leq \eta_n\})}{\varepsilon^2 n^2 l^2(\eta_n)} = \frac{no(\eta_n^2 l(\eta_n))}{\varepsilon^2 n^2 l^2(\eta_n)} = o(1).$$

By inequality (64) we have

$$IV \leq \frac{C_4 \sum_{i=1}^n (n-i) i^{-\beta} (1 + o(\eta_n^2 l(\eta_n)))}{\varepsilon^2 n^2 l^2(\eta_n)} = o(1).$$

This completes the proof of Theorem 1(b).

### Acknowledgements

Cheng-Der Fuh’s research was partly supported by NSC 100-2118-M-008-002-MY3, and Tian-Xiao Pang’s research was partly supported by the NSFC (grant number 11001236) and the Zhejiang Provincial Natural Science Foundation of China (grant number Y6100091). We are grateful to an anonymous referee for his/her valuable comments of the first version of this paper.

### References

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30**, 676–687.  
 BASAWA, I. V. AND PRAKASA RAO, B. L. S. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, London.

- BENTKUS, V. AND GÖTZE, F. (1996). The Berry-Esseen bound for Student's statistic. *Ann. Prob.* **24**, 491–503.
- BERTAIL, P. AND CLÉMENTON, S. (2006). Regeneration-based statistics for Harris recurrent Markov chains. In *Dependence in Probability and Statistics* (Lecture Notes Statist. **187**), Springer, New York, pp. 3–54.
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
- BINGHAM, N. H., GOLDIE, C. M. AND TEUGELS, J. L. (1989). *Regular Variation*. Cambridge University Press.
- BOLTHAUSEN, E. (1982). The Berry-Esseen theorem for strongly mixing Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheor.* **60**, 283–289.
- BRADLEY, R. C. (1988). A central limit theorem for stationary  $\rho$ -mixing sequences with infinite variance. *Ann. Prob.* **16**, 313–332.
- BROCKWELL, P. J. AND DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd edn. Springer, New York.
- CHEN, X. (1999). The law of the iterated logarithm for functionals of Harris recurrent Markov chains: self-normalization. *J. Theoret. Prob.* **12**, 421–445.
- CHOW, Y. S. AND TEICHER, H. (1998). *Probability Theory, Independence, Interchangeability, Martingales*. Springer, New York.
- CŠÖRGŐ, M., LIN, Z. Y. AND SHAO, Q. M. (1994). Studentized increments of partial sums. *Sci. China A* **37**, 265–276.
- CŠÖRGŐ, M., SZYSZKOWICZ, B. AND WANG, Q. (2003a). Darling-Erdős theorem for self-normalized sums. *Ann. Prob.* **31**, 676–692.
- CŠÖRGŐ, M., SZYSZKOWICZ, B. AND WANG, Q. (2003b). Donsker's theorem for self-normalized partial sums processes. *Ann. Prob.* **31**, 1228–1240.
- DE LA PEÑA, V. H., LAI, T. L. AND SHAO, Q.-M. (2009). *Self-Normalized Processes*. Springer, Berlin.
- DIEBOLD, F. X. AND INOUE, A. (2001). Long memory and regime switching. *J. Econometrics* **105**, 131–159.
- FAURE, M. (2002). Self-normalized large deviations for Markov chains. *Electron. J. Prob.* **7**, 31pp.
- FUH, C.-D. (2006). Efficient likelihood estimation in state space models. *Ann. Statist.* **34**, 2026–2068.
- FUH, C.-D. AND HU, I. (2007). Estimation in hidden Markov models via efficient importance sampling. *Bernoulli* **13**, 492–513.
- FUH, C.-D. AND ZHANG, C.-H. (2000). Poisson equation, moment inequalities and quick convergence for Markov random walks. *Stoch. Process Appl.* **87**, 53–67.
- GINÉ, E., GÖTZE, F. AND MASON, D. M. (1997). When is the Student  $t$ -statistic asymptotically standard normal? *Ann. Prob.* **25**, 1514–1531.
- GRIFFIN, P. S. AND KUELBS, J. D. (1989). Self-normalized laws of the iterated logarithm. *Ann. Prob.* **17**, 1571–1601.
- HALL, P. AND SENETA, E. (1988). Products of independent, normally attracted random variables. *Prob. Theory Relat. Fields* **78**, 135–142.
- JING, B.-Y., SHAO, Q.-M. AND WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Prob.* **31**, 2167–2215.
- KULIK, R. (2006). Limit theorems for self-normalized linear processes. *Statist. Prob. Lett.* **76**, 1947–1953.
- LAI, T. L. AND SHAO, Q.-M. (2007). Self-normalized limit theorems in probability and statistics. In *Asymptotic Theory in Probability and Statistics with Applications*, eds T. L. Lai, L. Qian and Q.-M. Shao. International Press, Somerville, MA, pp. 3–43.
- LIFSHITS, B. A. (1978). On the central limit theorem for Markov chains. *Theory Prob. Appl.* **23**, 279–296.
- LIN, Z. (1996). A self-normalized Chung-type law of the iterated logarithm. *Theory Prob. Appl.* **41**, 791–798.
- MAXWELL, M. AND WOODROOFE, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Prob.* **28**, 713–724.
- MCLEROY, T. AND POLITIS, D. (2007). Self-normalization for heavy-tailed time series with long memory. *Statistica Sinica* **17**, 199–220.
- MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- NAGAEV, S. V. (1957). Some limit theorems for stationary Markov chains. *Theory Prob. Appl.* **2**, 378–406.
- NEY, P. AND NUMMELIN, E. (1987). Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Prob.* **15**, 561–592.
- NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheor.* **43**, 309–318.
- PELIGRAD, M. AND SHAO, Q. M. (1994). Self-normalized central limit theorem for sums of weakly dependent random variables. *J. Theoret. Prob.* **7**, 309–338.
- ROUSSAS, G. G. (1969a). Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* **21**, 73–87.
- ROUSSAS, G. G. (1969b). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* **40**, 1386–1400.
- SHAO, Q.-M. (1997). Self-normalized large deviations. *Ann. Prob.* **25**, 285–328.