JOURNAL OF MULTIVARIATE ANALYSIS 52, 83–106 (1995)

Some Continuous Edgeworth Expansions for Markov Chains with Applications to Bootstrap

SOMNATH DATTA* AND WILLIAM P. MCCORMICK

University of Georgia

This paper deals with the first order Edgeworth expansions for sums related to an ergodic Markov chain with general state space. In the first part of the paper, we establish certain continuity, w.r.t. the transition probability function and the initial distribution, in these expansions. In the second part, we illustrate the use of our continuous expansions in the area of bootstrap. We consider bootstrapping the distribution of the (sample) mean of a fixed real function of a Markov chain. Under a conditional non-latticeness condition, the bootstrap is shown to be second order accurate. As a second application we obtain Edgeworth expansions for the bootstrap approximation to the sampling distribution of the m.l.e. of a particular transition probability in a finite Markov chain. It is shown that the bootstrap is second order accurate and is therefore superior to the normal approximation, if the transition probability is irrational. In the other case, the exact asymptotic upper bound constant in the $O(n^{-1/2})$ rate of bootstrap approximation is determined.

1. INTRODUCTION

Let $(\mathbf{X}, \mathbf{B})$ be a measurable space and $\{X_t : t = 1, 2, ...\}$ be a Markov chain with state space $\mathbf{X}$. Let $p = p(x, A)$, with $x \in \mathbf{X}$ and $A \in \mathbf{B}$, be its transition probability function. For $k \geqslant 1$, let

$$p^{(k)}(x, A) = \int_{\mathbf{X}^{k-1} \times A} \prod_{i=1}^{k-1} p(x_i, dx_{i+1}), \qquad \text{with} \quad x_1 = x,$$

denote the k-step transition probability function for the chain.

Assume that for some positive integer k_0, and $0 < \delta < 1$,

$$\sup_{x, y, A} |p^{(k_0)}(x, A) - p^{(k_0)}(y, A)| \leqslant \delta, \tag{1.1}$$

Received November 11, 1992; revised June 8, 1994.

AMS 1990 subject classifications: 62G09, 62M05, 60J05.

Key words and phrases: bootstrap, continuous Edgeworth expansions, Markov chain, asymptotic accuracy, second order correctness.

* Research supported in part by a grant from the University of Georgia.

0047-259X/95 $6.00

where the supremum is taken over all states $x, y \in \mathbf{X}$ and sets $A \in \mathbf{B}$. It is well known (see, e.g., Nagaev, 1961) that in this case there exists a (unique) stationary distribution p_0 for the chain satisfying

$$\sup_{x, A} |p^{(k)}(x, A) - p_0(A)| \leqslant \gamma \rho^k, \tag{1.2}$$

for all $k \geqslant 1$ where $\gamma = \delta^{-1}$ and $\rho = \delta^{1/k_0}$.

Let f be a real measurable function on $\mathbf{X}$ satisfying $\int f(x)\, p_0(dx) = 0$ and $\int f^2(x)\, p_0(dx) < \infty$. Then the asymptotic variance of the normalized sum $n^{-1/2} \sum_{t=1}^{n} f(X_t)$ exists and is given by

$$\sigma^2 = \mathbf{E} f^2(X_1) + 2 \sum_{t=1}^{\infty} \mathbf{E} f(X_1) f(X_{t+1}),$$

where $\mathbf{E}$ stands for the expectation under the stationary distribution.

Classical Edgeworth expansions for the distribution of the standarized sum

$$S_n = n^{-1/2} \sigma^{-1} \sum_{t=1}^{n} f(X_t) \tag{1.3}$$

received a beautiful treatment using the operator theory by Nagaev (1961). In this paper, we establish certain continuity in these expansions with respect to the distribution of the chain (as determined by the initial distribution and the transition probability function). Inter alia, these results are nontrivial extensions of those in Datta and McCormick (1993) where Edgeworth expansions for sums related to a given Markov chain are obtained under somewhat weaker conditions than those in Nagaev (1961).

Our study of continuous Edgeworth expansions was motivated by their applications to the theory of bootstrap asymptotics. In Section 3, we consider bootstrapping of a univariate mean related to a Markov chain with general state space. Our bootstrap samples are generated from an estimated Markov chain; thus they are different from the so-called moving block bootstrap. The latter is a general resampling procedure which is usable in other weakly dependent, e.g., non-Markovian, models as well. See Künsch (1989), Lahiri (1991), and Liu and Singh (1992) for further details. On the other hand, our approach is, in some sense, simpler and more natural for the Markovian setup. An added theoretical advantage of the present approach is that the expansions for the original statistics and its bootstrap version can be studied in one shot, via a continuous Edgeworth expansion. In particular, for the case of a mean we show that, under some general conditions, the bootstrap is correct up to $o(n^{-1/2})$.

We also study the asymptotic accuracy of the (parametric) bootstrap approximation to the m.l.e. of a particular transition probability of a finite state Markov chain. We establish that bootstrap is second order accurate, and therefore superior to the classical normal approximation, if the particular transition probability under study is an irrational number. In the case when this transition probability is rational we obtain the asymptotic value of the constant in the $O(n^{-1/2})$ rate of the bootstrap approximation. All the proofs are presented in Section 4.

2. Continuous Edgeworth Expansions

Let us consider more than one possible transition function p for the chain, each satisfying (1.1). Note that the stationary distribution p_0 for the chain is a function of p. Also, we will allow the function f in S_n to depend on p. However, to retain notational simplicity, we will not exhibit this dependence unless otherwise necessary. *Throughout this section we will assume that* $\int f(x)\,p_0(dx)=0$, $\sup_x \int |f(y)|^3\,p(x,dy)<\infty$, *and* $\sigma>0$, *for every p under consideration.* This condition will not be repeated.

Following Nagaev (1961), define the following functionals of p:

$$M=\frac{1}{1-\rho}\,(2k_0 3^{k_0}(1+2\rho)+3), \tag{2.1}$$

where k_0 and ρ are as in (1.1),

$$M_i=\sup_x \int |f(y)|^i\,p(x,dy), \qquad \text{for} \quad i=1,2,3, \tag{2.2}$$

and

$$\mu_3=\mathbf{E}f^3(X_1)+3\sum_{t=1}^{\infty}\mathbf{E}f^2(X_1)f(X_{t+1})+3\sum_{t=1}^{\infty}\mathbf{E}f(X_1)f^2(X_{t+1})$$
$$+6\sum_{t,s=1}^{\infty}\mathbf{E}f(X_1)f(X_{t+1})f(X_{t+s+1}), \tag{2.3}$$

where $\mathbf{E}$ denotes the expectation when initial distribution is p_0. In addition, for an initial distribution ω satisfying $\int |f|\,d\omega<\infty$, define

$$\mu_\omega(=\mu_\omega(p))=\sum_{t=1}^{\infty}\mathbf{E}_\omega f(X_t),$$

where E_ω denotes the expectation when the inital distribution is ω. It turns out that both μ_3 and μ_ω are finite.

Let p' be a $\mathbf{B}\times\mathbf{B}$ measurable function such that $p'(x,\cdot)$ is a density, w.r.t. p_0, of the component of $p(x,\cdot)$ which is absolutely continuous w.r.t. p_0. For an integer k and sets $C, D \in \mathbf{B}$ with $\int_C p^{(k-1)}(x, D)\, p_0(dx) > 0$, let $\mu_{k, C, D}$ denote the conditional distribution of the sum $\sum_{i=1}^{k} f(X_i)$ given the event $\{X_1 \in C,\ X_k \in D\}$, under the stationary initial distribution.

A continuous Edgeworth expansion will be applicable to situations when one has a triangular array of summands related to Markov chains and both the inital distribution and the transition probability function change with the sample size. Consequently, we introduce a sequence (ω_n, p_n) of initial distributions and transition probability functions for the chain, as well as its limit (ω, p). We will write $\omega_n \Rightarrow \omega$ to mean that ω_n converges to ω in total variation. Similarly, the convergence

$$\sup_{x, A} |p_n(x, A) - p(x, A)| \to 0$$

will be denoted by $p_n \Rightarrow p$.

First we state a continuous version of the first order Edgeworth expansion in the non-lattice case. Recall that we allow the function f to depend on p. In the statement of the following theorem f_n and f will denote the functions corresponding to p_n and p, respectively. Similarly, σ and μ_3 correspond to the limiting p and μ_ω corresponds to the limiting p and ω. The object of interest is $F_{n\omega_n}$, the distribution function of S_n (with f_n in place of f) when the initial distribution is ω_n and the transition probability function is p_n.

THEOREM 2.1. *Suppose p is a transition probability function satisfying* (1.1) *such that there exist a positive integer k and sets $C_i \in \mathbf{B}$, $1 \leqslant i \leqslant 4$, satisfying the following conditions*:

(i) μ_{k, C_2, C_3} *is a non-lattice distribution*

(ii) $p_0(C_1) > 0, p_0(C_4) > 0, \int_{C_2} p^{(k-1)}(x, C_3)\, p_0(dx) > 0,$ (2.4)

(iii) $0 < \inf_{x \in C_1, y \in C_2} p'(x, y),$
$0 < \inf_{x \in C_3, y \in C_4} p'(x, y) \leqslant \sup_{x \in \mathbf{X}, y \in C_4} p'(x, y) < \infty.$ (2.5b)

Let ω be a probability on $\mathbf{B}$ *satisfying $\int |f|\, d\omega < \infty$. If $p_n \Rightarrow p$, $\int f_n\, d\omega_n \to \int f\, d\omega$, and $\omega_n \Rightarrow \omega$ then*

$$F_{n, \omega_n}(x) = \Phi(x) + n^{-1/2} \left\{ \frac{\mu_3}{6\sigma^3}(1 - x^2) - \frac{\mu_\omega}{\sigma} \right\} \phi(x) + o(n^{-1/2}), \quad (2.6)$$

uniformly in x, provided in addition that

(iv) $\sup_x \int |f_n(y) - f(y)|\, p(x, dy) \to 0,$

(v) $\limsup_{n \to \infty} \sup_x \int |f_n(y)|^3\, g(|f_n(y)|)\, p_n(x, dy) < \infty$

for some function g satisfying $g(x) \to \infty$ when $x \to \infty$, and

(vi) $\limsup_{n \to \infty} \int f_n^2(x)\, d\omega_n(x) < \infty$.

Note that in the above theorem we do not require that the distribution of f_n be non-lattice. In fact, f_n in our second bootstrap application will be lattice valued for all n. (Only the limiting f needs to be non-lattice) Also note that the expression $n^{-1/2}(\mu_\omega/\sigma)\,\phi(x)$ reflects the influence of the initial distribution on the expansion. In the case of the stationary distribution taken as the initial distribution this term vanishes.

Next, we consider the case when f takes values on a lattice given by $\{a+mh\colon m \in \mathbf{Z}\}$, for some $-\infty < a < \infty$ and $h > 0$. Here $\mathbf{Z}$ is the set of integers. Since f depends on p so do a and h. Suppose, for each n, $f_n = f(p_n)$ takes values on a lattice $\{a_n - mh_n\colon m \in \mathbf{Z}\}$ and $\sigma_n = \sigma(p_n)$. As before, μ_3, σ, μ_ω, h, etc. correspond to the limiting p and ω.

THEOREM 2.2. *Let p be a transition probability function satisfying* (1.1) *such that conditions* (ii) *and* (iii) *of Theorem* 2.1 *are satisfied and μ_{k, C_2, C_3} is a lattice distribution with span h. Let ω be an initial distribution with $\int |f|\, d\omega < \infty$. If $p_n \Rightarrow p$, $\int f_n\, d\omega_n \to \int f\, d\omega$, and if $\omega_n \Rightarrow \omega$ then*

$$F_{n,\omega_n}(x) = \Phi(x) + n^{-1/2} \times \left\{ \frac{\mu_3}{6\sigma^3}(1-x^2) - \frac{\mu_\omega}{\sigma} + \frac{h}{\sigma} Q\left(\frac{n^{1/2}(x\sigma_n - n^{1/2}a_n)}{h_n} \right) \right\} \times \phi(x) + o(n^{-1/2}), \tag{2.7}$$

uniformly in x, provided $h_n \to h$ and conditions (iv)–(vi) *of Theorem* 2.1 *hold. Here $Q(t) = [t] - t + 1/2$, for $t \in \mathbf{R}$.*

Note that the above theorem is applicable even if h_n is not the span of the values of f_n. We only require the limiting conditional distribution μ_{k, C_2, C_3} to have span h. Our second bootstrap application presents such a situation.

Remark 2.1. It can be seen from the proofs that Theorems 2.1 and 2.2 remain true if the conditions $\int f\, d\omega_n \to \int f\, d\omega$ and $\omega_n \Rightarrow \omega$ are replaced by $\mu_{\omega_n} \to \mu_\omega$.

3. APPLICATIONS TO BOOTSTRAP: ASYMPTOTIC ACCURACY

As a statistical application of our continuous edgeworth expansion results, we now consider bootstrapping the sampling distribution of the mean of a real function of a Markov chain. After this, we consider a second application where we bootstrap the sampling distribution of the m.l.e. of a transition probability of a finite state Markov chain.

Following the seminal paper by Efron (1979), bootstrap has received tremendous attention in recent years. Limit theorems for bootstrap showing its (asymptotic) validity and, in some cases, its superiority (in terms of rate of convergence) over the classical asymptotic theory (e.g., normal approximation) were obtained by many authors; see e.g., Bickel and Freedman (1981), Singh (1981), Beran (1982), Babu (1984), Babu and Singh (1984), Ghosh *et al.* (1984), Abramovitch and Singh (1985), Hall (1988), Bhattacharya and Qumsiyeh (1989), and Helmers (1991). The above references are only a few of numerous works done in this area. Although bootstrap was originally described for the i.i.d. setup, similar ideas can be used for some non-i.i.d. situations as well. Bootstrapping for non-i.i.d. models was considered by Freedman (1984), Liu (1988), Künsch (1989), Kulperger and Praskasa Rao (1989), Athreya and Fuh (1989, 1992), Bawasa *et al.* (1990), Lahiri (1991), Liu and Singh (1992), and Datta and McCormick (1992), among others

Let us suppose that $\{X_t\}$ is a Markov chain as in Section 1 (satisfying all the conditions therein) with a general state space $\mathbf{X}$. Let g (need not be centered) be a given real valued function on the state space such that the conditions (i)–(iii) of Theorem 2.1 are satisfied with $f = g - \mathbf{E}g$. Also assume that

$$\sup_x \int |g(y)|^3\, p(x, dy) < \infty.$$

Letting σ^2 denote the (asymptotic) variance of the sum $n^{-1/2} \sum_l^n g(X_t)$, we form the pivot

$$R_n = n^{-1/2}\sigma^{-1}\left(\sum_l^n g(X_t) - \mathbf{E}g\right). \tag{3.1}$$

Suppose we are interested in approximating the sampling distribution (under the stationary distribution) of R_n. The first step for this is to construct a consistent estimator $\hat{p}$ of p, based on $X_1, X_2, ..., X_n$; i.e., one which satisfies $\hat{p} \Rightarrow p$, a.s. (A more exact description of $\hat{p}$ will depend on the particular situation at hand; see Remark 3.2 below.) Let $\hat{p}_0$ be the stationary initial distribution corresponding to $\hat{p}$ (which exists, a.s., for large n). Now generate bootstrap samples in the obvious way: i.e., generate $X_1^* \sim \hat{p}_0$, and having observed $X_1^*, ..., X_i^*$, generate $X_{i+1}^* \sim \hat{p}(X_i^*, \cdot)$ for $i = 1, ..., n-1$. Let P^* denote the bootstrap probability and E^* its (stationary) expectation. Let R_n^* denote the bootstrap pivot

$$R_n^* = n^{-1/2}(\sigma^*)^{-1}\left(\sum_1^n g(X_t^*) - E^*g\right). \tag{3.2}$$

Note that $E^*g = \int g(x)\,\hat{p}_0(dx)$ and

$$\sigma^{*2} = \int (g(x) - E^*g)^2\,\hat{p}_0(dx) + 2\sum_{t=1}^{\infty} \int (g(x) - E^*g) \times (g(y) - E^*g)\,\hat{p}^t(x, dy)\,\hat{p}_0(dx).$$

As an immediate application of Theorem 2.1, we get the following result. We use $\| \ \|_\infty$ for the sup norm of real functions on $\mathbf{R}$.

THEOREM 3.1. *Let p satisfy* (1.1) *and conditions* (i)–(iii) *of Theorem* 2.1 *with $f = g - \mathbf{E}g$. Suppose $\hat{p} \Rightarrow p$ a.s. and* $\limsup_{n\to\infty} \sup_x \int |g(x)|^{3+\varepsilon}\,\hat{p}(x, dy) < \infty$, *a.s., for some $\varepsilon > 0$. Then, as $n \to \infty$,*

$$\|P\{R_n \leqslant x\} - P^*\{R_n^* \leqslant x\}\|_\infty = o(n^{1/2}), \qquad a.s.$$

Remark 3.1. One may take, instead of the above mentioned $\hat{p}_0$, any other estimator $\tilde{p}_0$ (say) of the initial distribution provided $\tilde{p}_0 \Rightarrow p_0$ and $\int g\,d\tilde{p}_0 \to \int g\,dp_0$, a.s. Also, it is possible to replace σ^* in the definition of R_n^* by

$$\tilde{\sigma}^{*2} = \int (g(x) - E^*g)^2\,\hat{p}_0(dx) + 2\sum_{t=1}^{k_n} \int (g(x) - E^*g)(g(y) - E^*g)\,\hat{p}_0(dx)\,\hat{p}^t(x, dy)$$

provided $k_n/\log n \to \infty$.

Remark 3.2. Consider a parametric setup where the transition probability function is indexed by a finite dimensional parameter θ. Suppose, in addition, that the parametrization $\theta \rightsquigarrow p_\theta$ is continuous in $\Rightarrow$ convergence topology on the range. Then to get a consistent estimator of p, it suffices to get one for θ.

On the other hand, in a nonparametric context, it is often possible to construct a nonparametric estimator of p through density estimation. See, e.g., Roussas (1969), Rosenblatt (1970), Rüschendorf (1977), Yakowitz (1979), and Roussas (1990).

Remark 3.3. It may not always be possible (or easy) to construct estimator $\hat{p}$ to satisfy $\hat{p} \Rightarrow p$, a.s. This condition can be dispensed with. A careful investigation of the proofs of Theorems 2.1 and 3.1 will show that all we really need is that the following conditions hold almost surely (besides the other conditions of Theorem 3.1):

(i) $\{p: n \geqslant n_0\}$ satisfies condition (U1) of Section 4, for some n_0.

(ii) $E^*g^i(X_1)g^j(X_{t+1})g^k(X_{t+s+1}) \to Eg^i(X_1)g^j(X_{t+1})g^k(X_{t+s+1})$, for all $0 \leqslant i, j, k$, with $i+j+k \leqslant 3$, and $s \geqslant 1$.

(iii) $\limsup_{n\to\infty} \sup_{a\leqslant|\theta|\leqslant b} \sup_x \|\hat{p}^{(m)}(\theta, x, \cdot)\| < 1$, for some $m \geqslant 1$, and all $0 < a < b < \infty$.

(The complex measure $\hat{p}^{(m)}(\theta, x, \cdot)$ is defined before Lemma 4.10 with p replaced by $\hat{p}$ throughout.)

In particular, these conditions are easily verified for the i.i.d. case yielding Singh's (1981) result for non-lattice summands. For verifying (iii) in the i.i.d. case one needs to use the facts that $\hat{p}(x, \cdot) \equiv \hat{p}_0$ and in this case $\|\hat{p}^{(m)}(\theta, x, \cdot)\| \equiv |\int e^{i\theta f(y)} \hat{p}_0(dy)|$.

Remark 3.4. In a recent paper, Rajarshi (1990) considered a bootstrap procedure for the mean of a Markov chain using the density estimation approach described in Remark 3.2. He asserted the first order correctness (i.e., asymptotic validity) of the procedure and conjectured that it is also second order correct. Our Theorem 3.1 gives a set of sufficient conditions for this.

Remark 3.5. In this paper, we answer the bootstrap accuracy question for a standardized mean. A similar question can be addressed for a studentized mean and perhaps, more generally, for a studentized statistic which is (essentially) a function of several means related to a Markov chain. See, e.g., Babu and Singh (1984) and Abramovitch and Singh (1985) for the corresponding developments in the i.i.d. case.

We will now consider a situation where the original sample $X_1, X_2, \ldots, X_{n+1}$ is a realization, up to some time point $n+1$, of an irreducible and aperiodic Markov chain $\{X_t : t \geqslant 1\}$ taking values in a finite state space, taken to be $\mathbf{X} = \{1, 2, \ldots, N\}$, for some integer $N \geqslant 3$. It is well known (see Doob, 1953) that such a chain is ergodic and satisfies (1.1).

Let $p = (p_{ij} : 1 \leqslant i, j \leqslant N)$ denote the matrix of transition probabilities for the chain. A natural estimator of p is $\hat{p} = (\hat{p}_{ij} : 1 \leqslant i, j \leqslant N)$, obtained by the method of maximum likelihood, where

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} 1_{\{n_i > 0\}} + \delta_{ij} 1_{\{n_i = 0\}}, \qquad 1 \leqslant i, j \leqslant N, \tag{3.3}$$

with

$$n_i = \sum_{t=1}^{n} 1_{\{X_t = i\}} \qquad \text{and} \qquad n_{ij} = \sum_{t=1}^{n} 1_{\{X_t = i,\, X_{t+1} = j\}}. \tag{3.4}$$

Consider the following bootstrap scheme for approximating the sampling distribution of $\hat{p}_{ij}$: Generate bootstrap samples $X_1^*, X_2^*, ..., X_{n+1}^*$ following a Markov chain with transition probability matrix $\hat{p}$ and some initial distribution $\hat{\omega}$. Then form n_{ij}^*, n_i^*, and $\hat{p}_{ij}^*$ by (3.3) and (3.4) with X^*'s in place of X's. Then $\hat{p}_{ij}^*$ will be a bootstrap estimator of p_{ij}, and the sampling distribution of $(\hat{p}_{ij}^* - \hat{p}_{ij})$ over all possible size $(n+1)$ bootstrap samples (given $X_1, X_2, ..., X_{n+1}$) will be used as a bootstrap approximation to the distribution of $(\hat{p}_{ij} - p_{ij})$.

The above scheme was proposed by Basawa *et al.* (1990) for a finite Markov chain and independently by Athreya and Fuh (1989) for the more general case of countable state space. (Actually, Basawa *et al.* made a slight modification to make $\hat{p}$ ergodic. However, it was not really necessary for the asymptotic results.) Both of these papers proved that the above procedure is asymptotically valid, in the respective setups, but did not provide any asymptotic accuracy result of the procedure. In this paper, we obtain a number of results regarding the accuracy of the above bootstrap scheme. The proofs of the last two results make heavy use of the continuous Edgeworth expansions of the previous section. When p_{ij} is irrational, the bootstrap (for a standardized statistic) is accurate up to $o(n^{-1/2})$, and is therefore better than the normal approximation.

In general, a proper choice of $\hat{\omega}$ for generating X_1^* will be important to obtain the second order accuracy for the bootstrap. If we want to estimate the sampling distribution under the stationary distribution then a natural and correct choice is $\hat{\omega}(i) = n_i/n$. However, as our results will show, for the statistics of interest here the choice of the initial distribution plays no role, even up to the second order. This is so because, for the related summands, $\mu_\omega = 0$, for all of the initial distribution ω.

Once again, let P and P^* denote the probabilities governing the original Markov chain $\{X_t\}$ and the bootstrapped chain $\{X_t^*\}$ (given $X_1, X_2, ..., X_{n+1}$), respectively. Throughout the rest of this section, i and j will denote two arbitrarily fixed states such that $0 < p_{ij} < 1$. All the conclusions of the following theorems are valid (P) almost surely. Note that under P, X_1 may have any distribution.

THEOREM 3.2. *For any x and $y \in \mathbf{X}$, as $n \to \infty$,*

(a) $$\begin{aligned}&\|P\{n^{1/2}(\hat{p}_{ij} - p_{ij}) \leqslant \cdot \mid X_1 = x\} \\ &- P^*\{n^{1/2}(\hat{p}_{ij}^* - \hat{p}_{ij}) \leqslant \cdot \mid X_1^* = y\}\|_\infty \\ &= O(n^{-1/2}(\log n)^{1/2}),\end{aligned}$$

(a′) $$\begin{aligned}&\|P\{n_i^{1/2}(\hat{p}_{ij} - p_{ij}) \leqslant \cdot \mid X_1 = x\} \\ &- P^*\{n_i^{*1/2}(\hat{p}_{ij}^* - \hat{p}_{ij}) \leqslant \cdot \mid X_1^* = y\}\|_\infty \\ &= O(n^{-1/2}(\log n)^{1/2}),\end{aligned}$$

(b) $$\|P\{n^{-1/2}n_i(\hat{p}_{ij}-p_{ij})\leqslant \cdot \mid X_1=x\}$$

$$-P^*\{n^{-1/2}n_i^*(\hat{p}_{ij}^*-\hat{p}_{ij})\leqslant \cdot \mid X_1^*=y\}\|_\infty$$

$$=O(n^{-1/2}(\log\log n)^{1/2}),$$

(c) $$\left\|P\left\{\frac{n^{-1/2}n_i(\hat{p}_{ij}-p_{ij})}{(p_i p_{ij}(1-p_{ij}))^{1/2}}\leqslant \cdot \mid X_1=x\right\}\right.$$

$$\left.-P^*\left\{\frac{n^{-1/2}n_i^*(\hat{p}_{ij}^*-\hat{p}_{ij})}{(\hat{p}_i\hat{p}_{ij}(1-\hat{p}_{ij}))^{1/2}}\leqslant \cdot \mid X_1^*=y\right\}\right\|_\infty$$

$$=O(n^{-1/2}).$$

Next we analyze the $O(n^{-1/2})$ rate for the standarized statistics in more detail. We will need some technical conditions for doing the Edgeworth expansions. These conditions were introduced in Datta and McCormick (1993). Some special terminology will be necessary for this purpose. Let $S\subset\bigcup_{n=2}^{\infty}\mathbf{X}^n$ be the path space, where $\underline{s}=(s_1, s_2, ..., s_n)\in S$ iff $p_{s_k, s_{k+1}}>0$, $1\leqslant k\leqslant n-1$. We refer to the elements of S as paths, and for the path $\underline{s}=(s_1, s_2, ..., s_n)$, we define its length, denoted by $|\underline{s}|$, as $n-1$. For the fixed states i and j under discussion, we will say two paths $\underline{s}_1$ and $\underline{s}_2$, starting at i, are of the same type if $\underline{s}_1=(i, s_2, ..., s_{n_1})$ and $\underline{s}_2=(i, t_2, ..., t_{n_2})$ and either both s_2 and t_2 equal j or neither of them equals j (i.e., either $s_2=t_2=j$ or $s_2\neq j$ and $t_2\neq j$). We say that a path is a loop at $x\in\mathbf{X}$ if it starts and ends with x, but otherwise does not pass through x. Two loops of the same type at i are said to be essentially different if they have different lengths.

THEOREM 3.3. *Let p_{ij} be irrational and let there be two essentially different loops at i. Then as $n\to\infty$,*

$$\left\|P\left\{\frac{n^{1/2}n_i(\hat{p}_{ij}-p_{ij})}{(p_i p_{ij}(1-p_{ij}))^{1/2}}\leqslant \cdot \,\middle|\, X_1=x\right\}-P^*\left\{\frac{n^{-1/2}n_i^*(\hat{p}_{ij}^*-\hat{p}_{ij})}{(\hat{p}_i\hat{p}_{ij}(1-\hat{p}_{ij}))^{1/2}}\right.\right.$$

$$\left.\left.\leqslant \cdot \,\middle|\, X_1^*=y\right\}\right\|_\infty=o(n^{-1/2}),$$

for any $x, y\in\mathbf{X}$.

THEOREM 3.4. *Let $p_{ij} = k/m$, where k and m are two relatively prime positive integers. Suppose, for some state $z \neq i$, there exist two loops at z, $\underline{s}_1$ and $\underline{s}_2$, are not passing through i, with g.c.d. $(|\underline{s}_1|, |\underline{s}_2|) = 1$. Then*

$$\limsup_{n \to \infty} n^{1/2} \left\| P\left\{ \frac{n^{-1/2} n_i (\hat{p}_{ij} - p_{ij})}{(p_i p_{ij}(1 - p_{ij}))^{1/2}} \leqslant \cdot \;\middle|\; X_1 = x \right\} - P^* \left\{ \frac{n^{-1/2} n_i^* (\hat{p}_{ij}^* - \hat{p}_{ij})}{(\hat{p}_i \hat{p}_{ij}(1 - \hat{p}_{ij}))^{1/2}} \leqslant \cdot \;\middle|\; X_1^* = y \right\} \right\|_\infty$$
$$= \frac{1}{m(2\pi p_i p_{ij}(1 - p_{ij}))^{1/2}},$$

for any $x, y \in \mathbf{X}$.

4. PROOFS.

Let $\mathbf{G}$ be the space of all bounded $\mathbf{B}$ measurable functions on $\mathbf{X}$ equipped with the sup norm. Consider the following linear operators from $\mathbf{G}$ to $\mathbf{G}$ defined for every p satisfying (1.1):

$$Pg(\cdot) = \int g(y)\, p(\cdot, dy), \quad P_1 g(\cdot) = \int g(y)\, p_0(dy), \quad \text{and}$$

$$P(\theta)\, g(\cdot) = \int g(y)\, e^{i\theta f(y)} p(\cdot, dy), \qquad \text{for all} \quad g \in \mathbf{G}, \tag{4.1}$$

where θ is a real number and f was introduced in Section 1. We denote the resolvents of P and $P(\theta)$ by R and $R(\cdot, \theta)$, respectively. Let $\mathbf{C}$ denote the complex plane and

$$\mathbf{I}_1 = \{z \in \mathbf{C}: |z - 1| = \rho_1\} \qquad \text{and} \qquad \mathbf{I}_2 = \{z \in \mathbf{C}: |z| = \rho_2\}, \tag{4.2}$$

where $\rho_1 = (1 - \rho)/3$ and $\rho_2 = (1 + 2\rho)/3$. It is shown in Nagaev (1961) that if $|\theta| < (2M^2 M_1)^{-1}$ then $\mathbf{I}_1$ and $\mathbf{I}_2$ both lie in the resolvent set of $P(\theta)$. Consequently one can define the projection operators

$$P_j(\theta) = \frac{1}{2\pi i} \int_{I_j} R(z, \theta)\, dz, \qquad j = 1, 2, \tag{4.3}$$

for such θ's. It is easy to see that $P(0) = P$. Furthermore, it follows from (1.10) in Nagaev (1957) and Cauchy's integral formula that $P_1(0) = P_1$.

Let us denote $\int g\, d\mu$ by $\langle g, \mu\rangle$ for any $g \in \mathbf{G}$ and a totally finite complex measure μ on $\mathbf{B}$. For $|\theta| < (2M^2 M_1)^{-1}$, define

$$\lambda(\theta) = \frac{\langle P(\theta)\, P_1(\theta)\, \psi, p_0\rangle}{\langle P_1(\theta)\, \psi, p_0\rangle}, \tag{4.4}$$

$$\Psi(\theta) = \log \lambda(\theta) + \tfrac{1}{2}\sigma^2\theta^2, \tag{4.5}$$

and for any probability measure ω on $\mathbf{B}$,

$$\mu_\omega(\theta) = \langle P_1(\theta)\, \psi, \omega(\theta)\rangle, \tag{4.6}$$

where $\psi \in \mathbf{G}$ is the constant function taking value 1 and $d\omega(\theta) = e^{i\theta f}\, d\omega$.

For every $\varepsilon > 0$, define $\delta(\varepsilon, \omega, p) = \sup S(\varepsilon, \omega, p)$, where

$$\begin{aligned} S(\varepsilon, \omega, p) = \{0 < \delta < (2M^2 M_1)^{-1} &: |\Psi(\theta) + \tfrac{1}{6} i\mu_3\theta^3| < \varepsilon\sigma^3|\sigma|^3, \\ &|\Psi(\theta)| < \tfrac{1}{4}\sigma^2\theta^2,\ |\tfrac{1}{6} i\mu_3\theta^3| < \tfrac{1}{4}\sigma^2\theta^2,\ |\mu_\omega(\theta) - 1 - i\mu_\omega\theta| < \varepsilon|\theta|, \\ &\text{for all } 0 < |\theta| < \delta\}. \end{aligned}$$

(Interpret δ to be zero in the case S is empty.)

Let $\phi_{n\omega}$ be the characteristic function of the sum $\sum_{t=1}^n f(X_t)$, when X_1 has distribution ω. First we establish a uniform two term expansion of $\phi_{n\omega}$ in a $n^{1/2}$-neighborhood of the origin. Before that, consider the following conditions on the class of possible transition probability functions for the chain.

DEFINITION A class $\mathbf{P}$ of transition probabilities will be called a uniformly class if

(U1) all $p \in \mathbf{P}$ satisfy (1.1) with a common choice of k_0 and δ,

(U2) for some function g: $[0, \infty) \to [0, \infty)$ (not dependent on p), with $g(x) \to \infty$ as $x \to \infty$,

$$\sup_{p \in \mathbf{P}} \sup_{x \in \mathbf{X}} \int |f(y)|^3\, g(|f(y)|)\, p(x, dy) < \infty,$$

and

(U3) $\inf_{p \in \mathbf{P}} \sigma \overset{\text{definition}}{=} \sigma_*$ is positive.

LEMMA 4.1. *Let* $\mathbf{P}$ *be a uniformity class of transition probabilities and* Ω *be a class of initial distributions such that* $\sup_{\omega \in \Omega, p \in \mathbf{P}} \int |f|^2\, d\omega < \infty$. *Then for any* $0 < \varepsilon < 1$,

$$\delta_* \overset{\text{definition}}{=} \inf_{\omega \in \Omega,\, p \in \mathbf{P}} \delta(\varepsilon, \omega, p)$$

is positive, and for $|\theta| < n^{1/2}\sigma_*\delta_*$,

$$\sup_{\omega \in \Omega, p \in \mathbf{P}} \left| \phi_{n\omega}\left(\frac{n^{-1/2}\theta}{\sigma}\right) - e^{-\theta^2/2} - n^{-1/2}\left\{\frac{\mu_3}{6\sigma^3}(i\theta)^3 + \frac{\mu_\omega}{\sigma}(i\theta)\right\} e^{-\theta^2/2}\right|$$
$$\leqslant Cn^{-1/2}|\theta|\{(\varepsilon + n^{-1/2})\, e^{-\theta^2/4} + \rho_2^n\}, \tag{4.7}$$

where C *is some finite constant which depends on* Ω *and* $\mathbf{P}$ *only through* M, $\sup M_1$, $\sup \int |f|\, d\omega$, $\inf \sigma$, $\sup |\mu_3|$, *and* $\sup |\mu_\omega|$.

To prove Lemma 4.1 we need to establish some preliminary results. Throughout this section $\| \ \|$ is used to denote either the operator norm or the total variation norm. The context will make the use clear. The notation $\| \ \|_\infty$ is used to denote the sup norm of real functions on the reals.

For every p, define functions

$$a(\theta) = \langle P(\theta)\, P_1(\theta)\, \psi, p_0 \rangle \qquad \text{and} \qquad b(\theta) = \langle P_1(\theta)\, \psi, p_0 \rangle,$$

where $\psi \equiv 1$. Let $g^{(j)}$ stand for the jth derivative of a (j-times differentiable) function g, $j \geqslant 1$, with $g^{(0)} = g$. We note the facts (see Lemma 1.2 and the proof of Lemma 1.4 in Nagaev (1957) for the first three relations and Lemma 1.2 of Nagaev (1961) for the last one) that

$$\Psi^{(i)}(0) = 0, \quad i = 0, 1, 2, \qquad \text{and} \qquad \Psi^{(3)}(0) = -i\mu_3. \tag{4.8}$$

LEMMA 4.2. *Under assumption* (U2) *the operator-valued map* $P(\theta)$ *is differentiable three-times and* $P^{(3)}(\theta)$ *is continuous at* 0 *uniformly in* $p \in \mathbf{P}$; *that is,*

$$\sup_{p \in \mathbf{P}} \|P^{(3)}(\theta) - P^{(3)}(0)\| \to 0, \qquad as \quad \theta \to 0.$$

Proof. It is easy to check that $P(\theta)$ is differentiable three times with

$$P^{(3)}(\theta)\, h(x) = -i \int e^{i\theta f(y)} h(y) f^3(y)\, p(x, dy).$$

Take $K = K(\varepsilon)$ large enough so that $\inf_{x > K} g(x) > 2/\varepsilon$. Then

$$\|P^{(3)}(\theta) - P^{(3)}(0)\| \leqslant \int_{[|f(y)| \leqslant K]} |e^{i\theta f(y)} - 1|\, |f(y)|^3\, p(x, dy)$$
$$+ 2 \int_{[|f(y)| > K]} |f(y)|^3\, p(x, dy),$$

$$\leqslant e^{|\theta K|}|\theta K| \sup_{p \in \mathbf{P}, x \in \mathbf{X}} \int |f(y)|^3 p(x, dy)$$

$$+ \varepsilon \sup_{p \in \mathbf{P}, x \in \mathbf{X}} \int |f(y)|^3 g(|f(y)|) p(x, dy), \quad (4.9)$$

since for any complex z, $|e^z - 1| \leqslant |z| e^{|z|}$. The lemma clearly follows from (4.9). ∎

The next four lemmas are in the same spirit as Lemma 4.2. Lemma 4.4 follows by direct differentiation and Lemma 4.3; Lemmas 4.5 and 4.6 can be proved using Lemma 4.3 and arguments similar to those in the proofs of Lemmas 4.2 and 4.3. These results are proved in detail in Datta and McCormick (1991). In this paper we only present an abridged proof of Lemma 4.3.

LEMMA 4.3. *Under assumptions* (U1) *and* (U2), *the operator valued map* $R(z, \theta)$, *for* $z \in I_1$, *is differentiable three times in* θ *and* $R^{(3)}(z, \theta)$ *is continuous at* 0 *uniformly in* $p \in \mathbf{P}$ *and* $z \in I_1$, *with* I_1 *given in* (4.2).

Proof. Using the second Neumann series for the resolvent (see Kato, 1984) one can represent $R(z, \theta)$, for all sufficiently small (uniformly in $p \in \mathbf{P}$) θ, as

$$R(z, \theta) = \sum_0^{\infty} (R(z) A)^n R(z) \qquad \text{with} \quad A = P(\theta) - P(0). \quad (4.10)$$

Then, by direct differentiation, the uniform continuity of $R^{(3)}(z, \theta)$ reduces to that of $R(z, \theta)$ in view of Lemma 4.2. This in turn can be established from (4.10) using similar arguments to those in the proof of Lemma 4.2. ∎

LEMMA 4.4. *Under the same assumptions as in Lemma* 4.3, $b(\theta)$ *is differentiable three times and* $b^{(3)}(\theta)$ *is continuous at* 0 *uniformly in* $p \in \mathbf{P}$.

LEMMA 4.5. *Under the same assumptions as in Lemma* 4.3, $a(\theta)$ *is differentiable three times and* $a^{(3)}(\theta)$ *is continuous at* 0 *uniformly in* $p \in \mathbf{P}$.

In view of Lemmas 4.4 and 4.5 and the fact that $b(0) = 1$, we have the following important result.

COROLLARY 4.1. *Suppose* (U1) *and* (U2) *hold. Then* $\Psi(\theta)$ *is differentiable three times and* $\Psi^{(3)}(\theta)$ *is continuous at* 0 *uniformly in* $p \in \mathbf{P}$.

LEMMA 4.6. *Suppose* (U1), (U2), *and* (2.15) *hold. Then*

$$\mu_\omega^{(1)}(\theta) \to \mu_\omega^{(1)}(0), \qquad as \quad \theta \to 0,$$

uniformly in $p \in \mathbf{P}$ *and* $\omega \in \Omega$.

LEMMA 4.7. *Under assumptions* (U1) *and* (U2), $\sup_{p \in \mathbf{P}} |\mu_3| < \infty$.

Proof. For $s < t$,

$$\begin{aligned}|\mathbf{E}f(X_1)f(X_{s+1})f(X_{t+s+1})| &= \Big|\iiint f(x)f(y)f(z)[p^{(t)}(y, dz) \\ &\qquad - p_0(dz)]\, p^{(s)}(x, dy)\, p_0(dx)\Big| \\ &\leqslant \left\{\iint |f(x)f(y)|\, p^{(s)}(x, dy)\, p_0(dx)\right\} \\ &\quad \times \left\{\sup_y \left(\int f^2(z)[p^{(t)}(y, dz) + p_0(dz)]^{1/2}\right)\right\} \\ &\quad \times \left\{\sup_y (\|p^{(t)}(y, \cdot) - p_0(\cdot)\|)^{1/2}\right\}, \\ &\leqslant C\rho^{t/2},\end{aligned}$$

for some constant C uniformly in $p \in \mathbf{P}$, and $\rho < 1$ as in (U1). It therefore follows that

$$\sup_{p \in \mathbf{P}} \sum_{s,\, t=1}^{\infty} |\mathbf{E}f(X_1)f(X_{s+1})f(X_{s+t+1})| < \infty.$$

The other terms in the definition of μ_3 an be handled similarly. ∎

LEMMA 4.8. *Under assumptions* (U1), (U2), *and that* $\sup_{\omega \in \Omega,\, p \in \mathbf{P}} \int |f|^2\, d\omega < \infty$, *we have*

$$\sup_{\omega \in \Omega,\, p \in \mathbf{P}} |\mu_\omega| < \infty.$$

Proof. The proof is similar to that of Lemma 4.7. ∎

We are now in a position to prove Lemma 4.1.

Proof of Lemma 4.1. We first consider the proposition that $\delta_* = \delta_*(\varepsilon) = \inf \delta(\varepsilon, \omega, p) > 0$. In view of (U2), Lemma 4.7, and the Taylor expansion (see (4.8)), the above proposition follows from uniform continuity (at zero) of $\Psi^{(3)}$ and $\mu_{\omega}^{(1)}$. But these are provided in Corollary 4.1 and Lemma 4.6, respectively.

Next, to prove the uniform expansion write, for $|\theta| < n^{1/2}\sigma_*\delta_*$,

$$\begin{aligned}
&\left|\lambda^n\left(n^{-1/2}\frac{\theta}{\sigma}\right) - e^{-\theta^2/2}\left\{1 + n^{-1/2}\frac{\mu_3}{6\sigma^3}(i\theta)^3\right\}\right| \\
&\quad\leqslant e^{-\theta^2/2}\left\{\left|n\Psi\left(n^{-1/2}\frac{\theta}{\sigma}\right) - n^{-1/2}\frac{\mu_3}{6\sigma^3}(i\theta)^3\right|\right. \\
&\quad\quad \left. + n^{-1}\frac{\mu_3^2}{72\sigma^6}\theta^6\right\}e^{\theta^2/4},
\end{aligned} \tag{4.11}$$

where we have used the inequality

$$|e^\alpha - 1 - \beta| \leqslant (|\alpha - \beta| + \tfrac{1}{2}|\beta|^2)\, e^\gamma$$

for all complex α, β and non-negative $\gamma \geqslant \max(|\alpha|, |\beta|)$ (cf. Feller, 1970, p. 534), with

$$\alpha = n\Psi\left(n^{-1/2}\frac{\theta}{\sigma}\right), \quad \beta = n^{-1/2}\frac{\mu_3}{6\sigma^3}(i\theta)^3, \quad \text{and} \quad \gamma = \theta^2/4.$$

Note that for all $|\theta| < n^{1/2}\sigma_*\delta_*$, the restriction $\gamma \geqslant \max(|\alpha|, |\beta|)$ is satisfied, by definition of δ_*.

Using the definition of δ_* we see that for $|\theta| < n^{1/2}\sigma_*\delta_*$, the RHS (4.11) is no more than

$$e^{-\theta^2/4}\left\{\varepsilon|\theta|^3 + n^{-1/2}\frac{\mu_3^2}{72\sigma^6}\theta^6\right\} n^{-1/2}.$$

Next we have, for $|\theta| < (2M^2M_1)^{-1}$, by (1.12) in Nagaev (1961), that

$$\phi_{n\omega}(\theta) = \lambda^n(\theta)\langle P_1(\theta)\,\psi, \omega(\theta)\rangle + \langle P^n(\theta)\, P_2(\theta)\,\psi, \omega(\theta)\rangle \tag{4.12}$$

and by (1.41) in Nagaev (1961), for $|n^{-1/2}\theta/\sigma| < (2M^2M_1)^{-1}$,

$$\begin{aligned}
&\left|\left\langle P^n\left(n^{-1/2}\frac{\theta}{\sigma}\right) P_2\left(n^{-1/2}\frac{\theta}{\sigma}\right)\psi, \omega\left(n^{-1/2}\frac{\theta}{\sigma}\right)\right\rangle\right| \\
&\quad\leqslant n^{-1/2}(2M^2M_1^* + m_{11}^*)\frac{|\theta|}{\sigma}\rho_2^n,
\end{aligned} \tag{4.13}$$

where

$$M_1^* = \sup_{p \in \mathbf{P}} M_1 \qquad \text{and} \qquad m_{11}^* = \sup_{\omega \in \Omega,\, p \in \mathbf{P}} \int |f| \, d\omega.$$

Thus we obtain after some algebra that, for $|\theta| < n^{1/2} \sigma_* \delta_*$,

$$\begin{aligned}
&\left| \phi_{n\omega}\left(n^{-1/2} \frac{\theta}{\sigma}\right) - e^{-\theta^2/2} - n^{-1/2} \left\{ \frac{\mu_3}{6\sigma^3} (i\theta)^3 + \frac{\mu_\omega}{\sigma} (i\theta) \right\} e^{-\theta^2/2} \right| \\
&\leqslant n^{-1/2} \left\{ e^{-\theta^2/2} \left(1 + n^{-1/2} \frac{|\mu_3|}{6\sigma^3} |\theta|^3 \right) \right. \\
&\quad \left. + e^{-\theta^2/4} \left(\varepsilon |\theta|^3 + n^{-1/2} \frac{\mu_3^2}{72\sigma^6} |\theta|^6 \right) n^{-1/2} \right\} \frac{|\theta|}{\sigma} \varepsilon \\
&\quad + n^{-1/2} e^{-\theta^2/4} \left(\varepsilon |\theta|^3 + n^{-1/2} \frac{\mu_3^2}{72\sigma^6} \theta^6 \right) \left(1 + n^{-1/2} \frac{|\mu_\omega|}{\sigma} |\theta| \right) \\
&\quad + n^{-1/2} (2M^2 M_1^* + m_{11}^*) \frac{|\theta|}{\sigma} \rho_2^n + n^{-1} \frac{|\mu_3 \mu_\omega|}{6\sigma^4} \theta^4 e^{-\theta^2/2},
\end{aligned}$$

where we also used the definition of δ_*. Clearly, by (U1)–(U3) and Lemmas (4.7) and (4.8), the last bound is no more than

$$C n^{-1/2} |\theta| \{ (\varepsilon + n^{-1/2}) e^{-\theta^2/4} + \rho_2^n \}$$

for some constant C as in the statement of the lemma. This proves the lemma. ▮

LEMMA 4.9. *Under the conditions of Theorem* 2.1, *there exists* n_0 *such that* $\mathbf{P} = \{p_n : n \geqslant n_0\}$ *forms a uniformity class. Moreover,* $\sigma_n \to \sigma$, $\mu_{3,n} \to \mu_3$, *and* $\mu_{\omega_n} \to \mu_\omega$.

Proof. Clearly, condition (v) of Theorem 2.1 shows that for some n_0, $\mathbf{P}$ satisfies (U2). It is easy to conclude from $p_n \Rightarrow p$ that $\mathbf{P}$ satisfies (U1) and $p_{0,n} \Rightarrow p_0$.

Next, by arguments similar to those in the proof of Lemma 4.7, we get constants C and $r < 1$ not depending on n such that

$$|\mathbf{E}_n f_n(X_1) f_n(X_{t+1})| \leqslant C r^t, \qquad \text{for all} \quad t \geqslant 1, n \geqslant 1. \tag{4.14}$$

Now we will show that

$$\mathbf{E}_n f_n(X_1) f_n(X_{t+1}) \to \mathbf{E} f(X_1) f(X_{t+1}), \qquad \text{for} \quad t \geqslant 0.$$

First, consider the case with $t=0$. For any $K>0$ which is a continuity point of the distribution function of $f^2(X_1)$ under p_0 we get

$$\mathbf{E}f_n^2(X_1)\, 1_{\{f_n^2(X_1) \leqslant K\}} \to \mathbf{E}f^2(X_1)\, 1_{\{f^2(X_1) \leqslant K\}} \tag{4.15}$$

in view of condition (iv) of Theorem 2.1. Also,

$$\begin{aligned} |\mathbf{E}_n f_n^2(X_1)\, 1_{\{f_n^2(X_1)| \leqslant K\}} - \mathbf{E}f_n^2(X_1)\, 1_{\{f_n^2(X_1) \leqslant K\}}| \\ \leqslant K\|p_{0,n} - p_0\| \to 0, \quad \text{as } n \to \infty. \end{aligned} \tag{4.16}$$

Since $\mathbf{E}_n f_n^2(X_1)\, 1_{\{f_n^2(X_1) > K\}} \leqslant K^{-1/2} \sup_x \int |f_n(x)|^3\, p_n(x, dy)$, the conclusion for $t=0$ follows from (4.15) and (4.16) and condition (v) of Theorem 2.1. The proof for other values of t are similar.

Finally, by (4.14), (4.15), and DCT we get that $\sigma_n \to \sigma$, which in turn implies (U3) for **P**. Thus we have proved that **P** is a uniformity class.

The convergences of $\mu_{3,n}$ and μ_{ω_n} can be proved along the same line as that of σ_n. ∎

Before stating the next lemma we need to define the following sequence of complex measures on (**X**, **B**):

$$p^{(1)}(\theta, x, A) = \int_A e^{i\theta f(y)} p(x, dy),$$

$$p^{(m)}(\theta, x, A) = \int_{\mathbf{X}} p^{(m-1)}(\theta, y, A)\, p^{(1)}(\theta, x, dy), \qquad \text{for} \quad m > 1,$$

where $A \in \mathbf{B}$, $\theta \in (-\infty, \infty)$, and $x \in \mathbf{X}$. Define $p_n^{(m)}$ similarly.

LEMMA 4.10. *Let $p_n \Rightarrow p$ and let condition* (iv) *of Theorem* 2.1 *hold. Then, for any $m \geqslant 1$ and $0 < b < \infty$,*

$$\sup_{|\theta| \leqslant b,\, x \in \mathbf{X}} \|p_n^{(m)}(\theta, x, \cdot) - p^{(m)}(\theta, x, \cdot)\| \to 0, \qquad \text{as} \quad n \to \infty.$$

Proof. For $m=1$, we get

$$\begin{aligned} &\|p_n^{(1)}(\theta, x, \cdot) - p^{(1)}(\theta, x, \cdot)\| \\ &\qquad \leqslant \|p_n(x, \cdot) - p(x, \cdot)\| + |\theta| \int |f_n(y) - f(y)|\, p(x, dy) \to 0, \end{aligned}$$

uniformly in $\theta \in [-b, b]$, and $x \in \mathbf{X}$ by the assumptions of the lemma.

Next, we proceed by induction and assume that the assertion is true for $m = m_0$. Then by a similar triangulation as before,

$$\begin{aligned}
&\| p_n^{(m_0+1)}(\theta, x, \cdot) - p^{(m_0+1)}(\theta, x, \cdot) \| \\
&\quad \leqslant \| p_n^{(1)}(\theta, x, \cdot) - p^{(1)}(\theta, x, \cdot) \| \\
&\qquad + \sup_x \| p_n^{(m_0)}(\theta, x, \cdot) - p^{(m_0)}(\theta, x, \cdot) \| \to 0,
\end{aligned}$$

uniformly in $\theta \in [-b, b]$ and x, by the assertion for $m = 1$ and m_0. ∎

Proof of Theorem 2.1. Let

$$G_n(x) = \Phi(x) + n^{-1/2}\phi(x)\left\{\frac{\mu_{3,n}}{6\sigma_n^3}(1 - x^2) - \frac{\mu_{\omega_n}}{\sigma_n}\right\}$$

and

$$\gamma_n(\theta) = \left\{1 + n^{-1/2}\frac{\mu_{3,n}}{6\sigma_n^3}(i\theta)^3 + n^{-1/2}\frac{\mu_{\omega_n}}{\sigma_n}(i\theta)\right\}.$$

Then by the Esseen lemma (cf. Feller, 1970, p. 533), we get, for any $\varepsilon > 0$ and $a > 0$,

$$\begin{aligned}
n^{1/2}\|F_{n\omega_n} - G_n(x)\|_\infty &\leqslant \frac{n^{1/2}}{\pi}\int_{|\theta| < n^{1/2}\sigma_n\delta_*} \frac{1}{|\theta|}\left|\phi_{n\omega_n}\left(n^{-1/2}\frac{\theta}{\sigma_n}\right) - \gamma_n(\theta)\right| d\theta \\
&\quad + \frac{n^{1/2}}{\pi}\int_{n^{1/2}\sigma_n\delta_* \leqslant |\theta| \leqslant n^{1/2}\sigma_n a} \frac{1}{|\theta|} \\
&\quad \times \left|\phi_{n\omega_n}\left(n^{-1/2}\frac{\theta}{\sigma_n}\right) - \gamma_n(\theta)\right| d\theta + O(a^{-1}). \qquad (4.17)
\end{aligned}$$

The first term is handled by Lemma 4.1, since Lemma 4.9 $\mathbf{P} = \{p_n : n \geqslant n_0\}$ is a uniformity class. To handle the middle term use the inequality

$$|\phi_{n\omega_n}(\theta)| \leqslant (\sup_x \| p_n^{(r)}(\theta, x, \cdot)\|)^{[n/r]},$$

with $r = (m+2)k$, where we choose m so large that $\inf_{x \in \mathbf{X}} p^{(m-1)k+1}(x, C_1) \geqslant \frac{1}{2}p_0(C_1)$. Here k and C_1 are as in the conditional non-latticeness assumption for the limiting p. Next use Lemma 4.10 to conclude

$$\sup_x \| p_n^{(r)}(\theta, x, \cdot)\| \to \sup_x \| p^{(r)}(\theta, x, \cdot)\|. \qquad (4.18)$$

Going through the proof of Lemma 3.2 in Datta and McCormick (1993) one can see that

$$\sup_x \| p^{(r)}(\theta, x, \cdot)\| \leqslant (1 - L\{1 - |\phi_{k, C_2, C_3}(\theta)|^2\})$$

for some $L > 0$. Here ϕ_{k, c_2, c_3} is the characteristic function of the nonlattice distribution μ_{k, c_2, c_3}. Consequently,

$$\sup_{\theta \in K} \text{RHS}(4.18) < 1,$$

for any compact K not containing 0. Therefore the middle term of RHS (4.17) converges to zero. As before, now let $a \uparrow \infty$ and $\varepsilon \downarrow 0$ to conclude that LHS (4.17) $\to 0$, as $n \to \infty$.

Finally, (2.6) obtains from above by using that $\sigma_n \to \sigma$, $\mu_{3,n} \to \mu_3$, and $\mu_{\omega_n} \to \mu_\omega$. ∎

Proof of Theorem 2.2. The latticeness of the summands is handled in the same way as in the i.i.d. case. First, we compare $F^{\#}_{n\omega_n}$ with $G^{\#}_n$, where $\#$ denotes convolution with $U(-n^{-1/2}(2\sigma_n)^{-1}, n^{-1/2}(2\sigma_n)^{-1})$. By Esseen's lemma we get an inequality similar to (4.17) for $n^{1/2} \| F^{\#}_{n\omega_n} - G^{\#}_n \|$. The first term of the bound in this case is dominated by that in (4.17) and is therefore handled in the same way by Lemma 4.1. We tackle the middle term of the Esseen bound using the following facts:

(i) we can find constants ε, A, and B, such that, for all n,

$$|\phi_{n\omega_n}(\theta)| \leqslant A e^{-n\theta^2 \sigma_n^2/4} + \theta B \rho_2^n,$$

because by Lemma 4.9 $\{p_n : n \geqslant n_0\}$ forms a uniformity class, for some n_0,

(ii) $\sup_{\varepsilon \leqslant |\theta| \leqslant \pi/h} |\phi_{k, c_2, c_3}(\theta)| < 1$,

since μ_{k, c_2, c_3} is a lattice distribution with span h; consequently,

$$\sup_{\varepsilon \leqslant |\theta| \leqslant \pi/h_n} \sup_x \| p_n^{(r)}(\theta, x, \cdot) \| \to \sup_{\varepsilon \leqslant |\theta| \leqslant \pi/h} \sup_x \| p^{(r)}(\theta, x, \cdot) \| < 1,$$

$$\text{with} \quad r = (m+2)\, k.$$

Finally, the comparison of $F^{\#}_{n\omega_n}$ and $F_{n\omega_n}$, as well as that of $G^{\#}_n$ and G_n, is identical to the i.i.d. case (cf. Feller, 1970, and Datta, 1992). ∎

Proof of Theorem 3.1. The first order Edgeworth expansion for $P\{R_n \leqslant x\}$ is that of the RHS of (2.6), with $f = g - \mathbf{E}g$; this is given as Theorem 3.1 in Datta and McCormick (1993).

Next, fix a sample path (in the probability-one set) for which all the a.s. conditions of Theorem 3.1 hold. Then, for that path, the Edgeworth expansion for $P^*\{R_n^* \leqslant x\}$ is given by the RHS of (2.6) also, because Theorem 2.1 applies with $p_n = \hat{p}$ and $\omega_n = \hat{p}_0$. Therefore, the conclusion now follows, as usual, by the triangle inequality. Since in this case $f = g - \mathbf{E}g$, it is particularly easy to check the conditions of Theorem 2.1. Also, take $g(x)$ (in the notation of Theorem 2.1) to be $|x|^\varepsilon$. ∎

For the proofs of Theorems 3.2–3.4, we will work with the Markov chain $Y_t = (X_t, X_{t+1})$, $t \geqslant 1$, on $S_2 = \{(u, v): 1 \leqslant u, v \leqslant N, p_{uv} > 0\}$, where X_t is the original Markov chain. To keep the notation simple, we will refer to (i.e., index) the transition probabilities for Y_t also by p. We will select our f to be $f(Y_t) = 1_{\{X_t = i, X_{t+1} = j\}} - p_{ij} 1_{\{X_t = i\}}$. Then

$$\sum_1^n f(Y_t) = n_i(\hat{p}_{ij} - p_{ij}) \qquad \text{and} \qquad \sigma^2 = p_i p_{ij}(1 - p_{ij}).$$

Given a sample path, we take p_n (in the notation of Section 2) to be $\hat{p}$. Then

$$\sum_1^n f_n(Y_t^*) = n_i^*(\hat{p}_{ij}^* - \hat{p}_{ij}) \qquad \text{and} \qquad \sigma_n^2 = \hat{p}_{0i}\hat{p}_{ij}(1 - \hat{p}_{ij}),$$

where $\hat{p}_{0i}$ is the stationary probability of state i under $\hat{p}$. (Note that, with probability one, eventually $\hat{p}$ will be ergodic and hence $\hat{p}_0$ will be defined.) It is not hard to argue that, almost surely,

$$|\hat{p}_i - \hat{p}_{0i}| = O(n^{-1}(\log n)),$$

where $\hat{p}_i = n_i/n$. Therefore it is possible to replace $\hat{p}_{0i}$ by $\hat{p}_i$ in the formula for σ_n^2 without affecting the expansion up to $o(n^{-1/2})$; i.e.,

$$\|P^*\{Z_n^* \leqslant \cdot \mid X_1^* = y\} - P^*\{\tilde{Z}_n^* \leqslant \cdot \mid X_1^* = y\}\|_\infty = o(n^{-1/2}), \tag{4.19}$$

almost surely, where

$$Z_n^* = \frac{n^{-1/2} n_i^*(\hat{p}_{ij}^* - \hat{p}_{ij})}{(\hat{p}_i \hat{p}_{ij}(1 - \hat{p}_{ij}))^{1/2}} \qquad \text{and} \qquad \tilde{Z}_n^* = \frac{n^{-1/2} n_i^*(\hat{p}_{ij}^* - \hat{p}_{ij})}{(\hat{p}_{0i}\hat{p}_{ij}(1 - \hat{p}_{ij}))^{1/2}}.$$

Proof of Theorem 3.2. We first prove part (c). Use Berry–Esseen for Markov chain (e.g., Theorem 1 in Nagaev, 1961) twice—once for $Z_n = n^{-1/2} n_i(\hat{p}_{ij} - p_{ij})/(p_i p_{ij}(1 - p_{ij}))^{1/2}$ and the other time for $\tilde{Z}_n^*$—then use the triangle inequality and (4.19).

Next, part (b) follows from part (c) by the law of iterated logarithm (see Chung, 1967 for $\hat{p}_i$ and $\hat{p}_{ij}$ and the triangle inequality.

Parts (a) and (a′) follow from part (b) by Berry–Esseen for $\hat{p}_i$ and $\hat{p}_i^* = n_i^*/n$. See the proof of Theorem 2.1(a) in Datta and McCormick (1992) for details of similar arguments. ∎

Proof of Theorem 3.3. Edgeworth expansions for $Z_n = n^{-1/2} n_i(\hat{p}_{ij} - p_{ij})/(p_i p_{ij}(1 - p_{ij}))^{1/2}$ were obtained in Theorem 4.1 of Datta and McCormick (1993):

$$P\{Z_n \leqslant u \mid X_1 = x\}$$
$$= \Phi(u) + n^{-1/2} \frac{(1 - 2p_{ij} + 3p_{ij}S_{ij})}{6(p_i p_{ij}(1 - p_{ij}))^{1/2}} \phi(u)(1 - u^2) + o(n^{-1/2}), \quad (4.20)$$

uniformly in $u \in \mathbf{R}$, where

$$S_{ij} = \sum_{k=1}^{\infty} (p_{ji}^{(k-1)} - p_{ii}^{(k)}).$$

In the course of the proof of (4.20), it was shown in Datta and McCormick (1993) that, under the conditions of Theorem 3.3, one can construct sets $C_1 - C_4$ and integer k such that $f(Y_t)$ under p satisfies the conditions (i)–(iii) of Theorem 2.1. Clearly, with $p_n = \hat{p}$, any ω_n, and the choice of f described earlier, the conditions (iv)–(vi) of Theorem 2.1 are satisfied whenever $\hat{p} \Rightarrow p$. However, by the strong law for ergodic Markov chains, the latter convergence takes place almost surely. Also, as stated earlier, $\mu_\omega = 0$ for any ω. Therefore, by Theorem 2.1 and Remark 2.1, $P^*\{\tilde{Z}_n \leqslant \cdot \mid X_1^* = y\}$ has the same expansion as $P\{Z_n \leqslant \cdot \mid X_1 = x\}$, i.e., (4.20), up to $o(n^{-1/2})$, almost surely (P). Now use the fact (4.19) to end the proof. ∎

Proof of Theorem 3.4. Note that in this case the limiting $f(Y_t)$ is a lattice with span $h = 1/m$. It can be checked that under the conditions of Theorem 3.4 one can construct sets $C_1 - C_4$ and integer k satisfying the conditions of Theorem 2.2 with f (see Data and McCormick, 1990). Therefore by two applications of Theorem 2.2 (once with $p_n = p$ for all n and once with $p_n = \hat{p}$ along almost all sample paths) we get

$$\begin{aligned} P\{Z_n \leqslant u \mid X_1 = x\} &= \Phi(u) + n^{-1/2} \frac{(1 - 2p_{ij} + 3p_{ij}S_{ij})}{6(p_i p_{ij}(1 - p_{ij}))^{1/2}} \phi(u)(1 - u^2) \\ &\quad + \frac{n^{-1/2}}{m(p_i p_{ij}(1 - p_{ij}))^{1/2}} Q(m(np_i p_{ij}(1 - p_{ij}))^{1/2} u) \\ &\quad \times \phi(u) + o(n^{-1/2}), \end{aligned} \quad (4.21)$$

uniformly in u, and

$$\begin{aligned} P^*\{\tilde{Z}_n^* \leqslant u \mid X_1^* = y\} &= \Phi(u) + n^{-1/2} \frac{(1 - 2p_{ij} + 3p_{ij}S_{ij})}{6(p_i p_{ij}(1 - p_{ij}))^{1/2}} \phi(u)(1 - u^2) \\ &\quad + \frac{n^{-1/2}}{m(p_i p_{ij}(1 - p_{ij}))^{1/2}} Q(\hat{m}(n\hat{p}_{0i}\hat{p}_{ij}(1 - \hat{p}_{ij}))^{1/2} u) \\ &\quad \times \phi(u) + o(n^{-1/2}), \end{aligned} \quad (4.22)$$

uniformly in u, almost surely, where $\hat{m} = \hat{k}/\hat{p}_{ij}$, with $\hat{k}$ = nearest integer to $m\hat{p}_{ij}$.

Since

$$\limsup_{n \to \infty} n^{1/2} |\hat{m} p_i p_{ij}(1 - p_{ij}) - m\hat{p}_{0i}\hat{p}_{ij}(1 - \hat{p}_{ij})| = \infty,$$

almost surely (by LIL), we have by arguments similar to those in the proof of Theorem 3.2 of Datta (1992) that

$$\limsup_{n \to \infty} \| \{ Q(\hat{m}(n\hat{p}_{0i}\hat{p}_{ij}(1 - \hat{p}_{ij}))^{1/2} u) - Q(m(np_i p_{ij}(1 - p_{ij}))^{1/2} u) \} \phi(u) \|_\infty = (2\pi)^{-1/2},$$

almost surely. The proof now ends by the triangle inequality together with (4.19). ∎

References

Abramovitch, L., and Singh K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.

Athreya, K. B., and Fuh, C. D. (1992). Bootstrapping Markov chains. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billiard, Eds.), pp. 49–64. Wiley, New York.

Athreya, K. B., and Fuh, C. D. (1989). Bootstrapping Markov chains: Countable case, Tech. Rep. B-89-7. Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC.

Babu, G. J. (1984). Bootstrapping statistics with linear combination of chi-squares as weak limits. *Sankhyā Ser. A* **46** 85–93.

Babu, G. J., and Singh, K. (1984). On one-term Edgeworth correction by Efron's bootstrap. *Sankhyā Ser. A* **46** 219–232.

Basawa, I. V., Green, T. A., McCormick, W. P., and Taylor, R. L. (1990). Asymptotic bootstrap validity for finite Markov chains. *Comm. Statist. Theory Methods* **19** 1493–1510.

Beran, R. (1982). Estimated sampling distributions: The bootstrap and competitors. *Ann. Statist.* **10** 212–225.

Bhattacharya, R. N., and Qumsiyeh, M. (1989). Second order and L^p-comparisons between the bootstrap and empirical Edgeworth expansion methodologies. *Ann. Statist.* **17** 160–169.

Bickel, P. J., and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

Chung, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*, 2nd ed. Springer-Verlag, New York/Berlin.

Datta, S. (1992). A note on continuous Edgeworth expansions and the bootstrap. *Sankhyā Ser. A* **54** 171–182.

Datta, S., and McCormick, W. P. (1993). On the first order Edgeworth expansion for a Markov chain. *J. Multivariate Anal.* **44** 345–359.

Datta, S., and McCormick, W. P. (1992). Bootstrap for a finite state Markov chain based on i.i.d. resampling. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billiard, Eds.), pp. 77–97. Wiley, New York.

DATTA, S., AND MCCORMICK, W. P. (1991). Asymptotic expansions for sums related to families of Markov chains with applications to bootstrap, Tech. Rep. 91-30. Dept. of Statistics, Univ. of Georgia, Athens, GA.

DATTA, S., AND MCCORMICK, W. P. (1990). On first order Edgeworth expansions for a Markov chain, Tech. Rep. 164. Dept of Statistics, Univ. of Georgia, Athens.

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

FELLER, W. (1970). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed. Wiley, New York.

FREEDMAN, D. (1984). On bootstrapping two-stage least squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842.

GOSH, M. PARR, W.C., SINGH, K., AND BABU, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12** 827–842.

HALL, P. (1988). Rate of convergence in bootstrap approximations. *Ann. Statist.* **16** 1665–1684.

HELMERS, R. (1991). On the Edgeworth expansion and the bootstrap approximation for a studentized U-statistics. *Ann Statist.* **19** 470–484.

KATO, T. (1984). *Perturbation Theory for Linear Operators*. Springer-Verlag, New York.

KULPERGER, R. J., AND PRASKASA RAO, B. L. S. (1989). Bootstrapping a finite state Markov chain. *Sankhyā Ser. A* **51** 178–191.

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

LAHIRI, S. N. (1991). Second order optimality of stationary bootstrap. *Statist. Probab. Lett.* **11** 335–341.

LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708.

LIU, R. Y. AND SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billiard, Eds.), pp. 225–248. Wiley, New York.

NAGAEV, S. V. (1961). More exact statements of limit theorems for homogeneous Markov chains. *Theory Probab. Appl.* **6** 62–81.

NAGAEV, S. V. (1957). Some limit theorems for stationary Markov chains. *Theory Probab. Appl.* **4** 378–406.

RAJARSHI, M. B. (1990). Bootstrap in Markov-sequences based on estimates of transition density. *Ann. Inst. Statist. Math.* **42** 253–268.

ROSENBLATT, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference* (M. Puri, Ed.), Cambridge Univ. Press, Oxford.

ROUSSAS, G. G. (1990). Estimation of transition distribution function and its quantiles in Markov processes: Strong consistency and asymptotic normality, Tech. Rep. 197. Division of Statistics, Univ. of Calif., Davis.

ROUSSAS, G. G. (1969). Nonparametric estimation of the transition distribution function of a Markov Process. *Ann. Math. Statist.* **40** 1386–1400.

RÜSCHENDORF, L. (1977). Consistency of estimators for multivariate density functions and for the mode. *Sankhyā Ser. A* **39** 243–250.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

YAKOWITZ, S. (1979). Nonparametric estimation of Markov transition functions. *Ann. Statist.* **7** 671–679.