Journal of Theoretical Probability, Vol. 19, No. 3, July 2006 (© 2006)
DOI: 10.1007/s10959-006-0047-9

Translated Poisson Approximation for Markov Chains

A. D. Barbour[1,3] and Torgny Lindvall[2]

Received October 12, 2004; revised May 30, 2005

The paper is concerned with approximating the distribution of a sum W of integer valued random variables Y_i, $1 \leqslant i \leqslant n$, whose distributions depend on the state of an underlying Markov chain X. The approximation is in terms of a translated Poisson distribution, with mean and variance chosen to be close to those of W, and the error is measured with respect to the total variation norm. Error bounds comparable to those found for normal approximation with respect to the weaker Kolmogorov distance are established, provided that the distribution of the sum of the Y_i's between the successive visits of X to a reference state is aperiodic. Without this assumption, approximation in total variation cannot be expected to be good.

1. INTRODUCTION

The Stein–Chen method is now well established in the study of approximation by a Poisson or compound Poisson distribution (Arratia et al.[(1)], Barbour et al.[(3)]). It has turned out to be very efficient for treating sums of the form $W := W_n := \sum_{i=1}^{n} Y_i$, where the variables $Y_1, Y_2, \ldots$ are non-negative, integer-valued, rarely different from 0, and have a short range of dependence. A basic example is the following: let $Y_1, Y_2, \ldots$ be independent and taking values 0 or 1 only, with $p_i := \mathbb{P}(Y_i = 1)$ generally small, to make a Poisson approximation plausible. Then the method offers

[1] Angewandte Mathematik, Winterthurerstrasse 190, 8057 Zürich, Switzerland. E-mail: a.d.barbour@math.unizh.ch
[2] School of Mathematical Sciences, Chalmers and GU, 41296 Göteborg, Sweden. E-mail: lindvall@math.chalmers.se
[3] To whom correspondence should be addressed.

a proof of the celebrated Le Cam theorem, which is transparent and relatively simple (cf. Ref. 3, I.(1.23)), and gives the optimal constant:

$$\|\mathcal{L}(W) - \mathrm{Po}\,(\lambda)\| \leqslant 2\lambda^{-1} \sum_{i=1}^{n} p_i^2 \leqslant 2 \max_{1\leqslant i\leqslant n} p_i, \tag{1.1}$$

where $\lambda := \mathbb{E}W = \sum_{i=1}^{n} p_i$. Here, $\mathcal{L}(X)$ denotes the distribution of a random element X, $\mathrm{Po}\,(\lambda)$ the Poisson distribution with mean λ, and $\|\nu\|$ the total variation norm of a signed bounded measure ν; we need this only for differences of probability measures Q, Q' on the integers $\mathbb{Z}$, when

$$\|Q - Q'\| := \sum_i |Q(i) - Q'(i)| = 2 \sup_{A\subset\mathbb{Z}} |Q(A) - Q'(A)|.$$

Clearly, if the p_i's are not required to be small, there is little content in (1.1). This is to be expected, since then $\mathbb{E}W = \lambda$ and $\mathrm{Var}\,W = \lambda - \sum_{i=1}^{n} p_i^2$ need no longer be close to one another, whereas Poisson distributions have equal mean and variance. This makes it more natural to try to find a family of distributions for the approximation within which both mean and variance can be matched, as is possible using the normal family in the classical central limit theorem. One choice is to approximate with a member of the family of translated Poisson distributions $\{\mathrm{TP}\,(\mu,\sigma^2),\ (\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+\}$, where

$$\begin{aligned}\mathrm{TP}\,(\mu,\sigma^2)\{j\} &:= \mathrm{Po}\,(\sigma^2+\delta)\{j - \lfloor \mu-\sigma^2 \rfloor\}\\ &= \mathrm{Po}\,(\lambda')\{j-\gamma\}, \quad j\in\mathbb{Z},\end{aligned}$$

where

$$\begin{aligned}&\gamma := \gamma(\mu,\sigma^2) := \lfloor \mu-\sigma^2 \rfloor, \qquad \delta := \delta(\mu,\sigma^2) := \mu-\sigma^2-\gamma\\ &\text{and} \quad \lambda' := \lambda'(\mu,\sigma^2) := \sigma^2+\delta.\end{aligned} \tag{1.2}$$

The $\mathrm{TP}\,(\mu,\sigma^2)$ distribution is just that of a Poisson with mean $\lambda' := \lambda'(\mu,\sigma^2) := \sigma^2+\delta$, then shifted along the lattice by an amount $\gamma := \gamma(\mu,\sigma^2) := \lfloor \mu-\sigma^2 \rfloor$. In particular, it has mean $\lambda'+\gamma = \mu$ and variance λ' such that $\sigma^2 \leqslant \lambda' < \sigma^2+1$; note that $\lambda' = \sigma^2$ only if $\mu-\sigma^2\in\mathbb{Z}$. For sums of independent, integer-valued random variables Y_i, this idea has been exploited by Vaitkus and Čekanavičius,[12] (1998), and also in Refs. 2, 4, and 7, using Stein's method, leading to error rates of the same order as in the classical central limit theorem, but now with respect to the much stronger total variation norm, as long as some 'smoothness' of the distribution of W can be established.

As in the Poisson case, the introduction of Stein's method raises the possibility of making similar approximations for sums of dependent random variables as well. However, the 'smoothness' needed is a bound of order $O(1/\sqrt{n})$ for $\|\mathcal{L}(W+1)-\mathcal{L}(W)\|$, entailing much more delicate arguments than are required for Poisson approximation. The elementary example of 2-runs in independent Bernoulli trials was treated in Ref. 4, but the argument used there was long and involved. More recently, Röllin[10] (2005) has proposed an approach which is effective in a wider range of circumstances, including many local and combinatorial dependence structures, in which one can find an imbedded sum of independent Bernoulli random variables. In this paper, we consider a different kind of dependence, in which the distributions of the random variables Y_i depend on an underlying Markovian environment.

We suppose that $X=(X_i)_{i=0}^{\infty}$ is an aperiodic, irreducible and stationary Markov chain with finite state space $E=\{0,1,\ldots,K\}$. Let $Y_0, Y_1, \ldots$ be integer-valued variables which are independent conditional on X, and, as in a hidden Markov model, such that the conditional distribution $\mathcal{L}(Y_i \mid X)$ depends on the value of X_i alone; we assume further that, for each $0 \leqslant k \leqslant K$, the distributions $\mathcal{L}(Y_i \mid X_i = k)$ are the same for all i. Under these assumptions, and with $W=\sum_{i=1}^{n} Y_i$, we show that $\|\mathcal{L}(W)\mathrm{TP}(\mathbb{E}W, \mathrm{Var}\,W)\|$ is asymptotically small, under reasonable conditions on the conditional distributions $\mathcal{L}(Y_1 \mid X_1 = k)$, $0 \leqslant k \leqslant K$. The detailed results are given in Theorems 4.2–4.4. Roughly speaking, we show that if these conditional distributions are stochastically dominated by a distribution with finite third moment, and if, as smoothness condition, the distribution $Q:=\mathcal{L}\left(\sum_{i=1}^{S_1} Y_i \mid X_0=0\right)$ is aperiodic ($Q\{d\mathbb{Z}\}<1$ for all $d \geqslant 2$), where S_1 is the step at which X first returns to 0, then

$$\|\mathcal{L}(W)-\mathrm{TP}(\mathbb{E}W, \mathrm{Var}\,W)\| = O\left(n^{-1/2}\right). \tag{1.3}$$

An ingredient of our argument, reflecting Röllin's[10] (2005) approach, is again to find an appropriate imbedded sum of independent random variables.

In the next section, we give an introduction to proving translated Poisson approximation by way of the Stein–Chen method. Lemma 2.2 provides a generally applicable formula for bounding the resulting error. In Section 3, we establish bounds on the total variation distance between $\mathcal{L}(W)$ and $\mathcal{L}(W+1)$ using coupling arguments. The results of these two sections are combined in Section 4 to prove the main theorems. Theorem 4.4 gives rather general conditions for (1.3) to hold, whereas Theorem 4.2, in a somewhat more restrictive setting, provides a relatively explicit formula for the approximation error. We then discuss the relationship of our results to those of Čekanavičius and Vaitkus,[6] who studied the degenerate case in which $Y_1 = h(k)$ a.s. on $\{X_1 = k\}$, $0 \leqslant k \leqslant K$. We conclude by showing that, if Q is in fact periodic, $\mathcal{L}(W)$ is usually not well approximated by a translated Poisson distribution.

2. TRANSLATED POISSON APPROXIMATION

Since the TP(μ, σ^2) distributions are just translates of Poisson distributions, the Stein–Chen method can be used to establish total variation approximation. In particular, $W \sim \mathrm{TP}(\mu, \sigma^2)$ if and only if

$$\mathbb{E}\{\lambda' f(W+1) - (W-\gamma) f(W)\} = 0 \tag{2.1}$$

for all bounded functions $f\colon \mathbb{Z} \to \mathbb{R}$, where $\lambda' = \lambda'(\mu, \sigma^2)$ and $\gamma = \gamma(\mu, \sigma^2)$ are as defined in (1.2). Define f_C^* for $C \subset \mathbb{Z}_+$ by

$$\begin{aligned} &f_C^*(k) = 0, \quad k \leqslant 0, \\ &\lambda' f_C^*(k+1) - k f_C^*(k) = \mathbf{1}_C(k) - \mathrm{Po}\,(\lambda')\{C\}, \quad k \geqslant 0, \end{aligned}$$

as in the Stein–Chen method. It then follows that

$$\|f_C^*\| \leqslant (\lambda')^{-1/2} \quad \text{and} \quad \|\Delta f_C^*\| \leqslant (\lambda')^{-1}$$

(see Ref. 3, Lemma I.1.1), where $\Delta f(j) := f(j+1) - f(j)$ and, for bounded functions $g\colon \mathbb{Z} \to \mathbb{R}$, we let $\|g\|$ denote the supremum norm. Correspondingly, for $B \subset \mathbb{Z}$ such that $B^* := B - \gamma \subset \mathbb{Z}_+$, the function f_B defined by

$$f_B(j) := f_{B^*}^*(j - \gamma), \quad j \in \mathbb{Z}, \tag{2.2}$$

satisfies

$$\begin{aligned} &\lambda' f_B(w+1) - (w-\gamma) f_B(w) \\ &\quad = \lambda' f_{B^*}^*(w - \gamma + 1) - (w-\gamma) f_{B^*}^*(w-\gamma) \\ &\quad = \mathbf{1}_{B^*}(w-\gamma) - \mathrm{Po}\,(\lambda')\{B^*\} \\ &\quad = \mathbf{1}_B(w) - \mathrm{TP}\,(\mu, \sigma^2)\{B\} \end{aligned} \tag{2.3}$$

if $w \geqslant \gamma$, and

$$\lambda' f_B(w+1) - (w-\gamma) f_B(w) = 0 \tag{2.4}$$

if $w < \gamma$, and clearly

$$\|f_B\| \leqslant (\lambda')^{-1/2} \quad \text{and} \quad \|\Delta f_B\| \leqslant (\lambda')^{-1}. \tag{2.5}$$

This can be exploited to prove the closeness in total variation of $\mathcal{L}(W)$ to $\mathrm{TP}(\mu, \sigma^2)$ for an arbitrary integer-valued random variable W. The next two results make use of this.

Lemma 2.1. Let $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1^2, \sigma_2^2 \in \mathbb{R}_+ \setminus \{0\}$ be such that $\gamma_1 = \lfloor \mu_1 - \sigma_1^2 \rfloor \leqslant \gamma_2 = \lfloor \mu_2 - \sigma_2^2 \rfloor$. Then

$$\|\mathrm{TP}(\mu_1, \sigma_1^2) - \mathrm{TP}(\mu_2, \sigma_2^2)\| \leqslant 2\{\sigma_1^{-1}|\mu_1 - \mu_2| + \sigma_1^{-2}(|\sigma_1^2 - \sigma_2^2| + 1)\}.$$

Proof. Both distributions assign probability 1 to $\mathbb{Z} \cap [\gamma_1, \infty)$, so it suffices to consider B such that $B - \gamma_1 \subset \mathbb{Z}_+$. Then, if $W \sim \mathrm{TP}(\mu_2, \sigma_2^2)$, we have

$$\begin{aligned}
&\mathbb{P}(W \in B) - \mathrm{TP}(\mu_1, \sigma_1^2)\{B\} \\
&\quad = \mathbb{E}\{\mathbf{1}_B(W) - \mathrm{TP}(\mu_1, \sigma_1^2)\{B\}\} \\
&\quad = \mathbb{E}\{\lambda_1 f_B(W+1) - (W - \gamma_1) f_B(W)\}
\end{aligned}$$

from (2.3), where $\lambda_l := \lambda'(\mu_l, \sigma_l^2)$, $l = 1, 2$. Applying (2.1), it thus follows that

$$\begin{aligned}
&\mathbb{P}(W \in B) - \mathrm{TP}(\mu_1, \sigma_1^2)\{B\} \\
&\quad = \mathbb{E}\{(\lambda_1 - \lambda_2) f_B(W+1) - (\gamma_2 - \gamma_1) f_B(W)\} \\
&\quad = \mathbb{E}\{(\lambda_1 - \lambda_2) \Delta f_B(W) - (\mu_2 - \mu_1) f_B(W)\}
\end{aligned}$$

and hence, from (2.5), that

$$\begin{aligned}
&|\mathbb{P}(W \in B) - \mathrm{TP}(\mu_1, \sigma_1^2)\{B\}| \\
&\quad \leqslant (\lambda_1)^{-1}(|\sigma_1^2 - \sigma_2^2| + |\delta_1 - \delta_2|) + (\lambda_1)^{-1/2}|\mu_1 - \mu_2|,
\end{aligned}$$

proving the lemma. □

The next lemma provides a very general means to establish total variation bounds; it is our principal tool in Section 4. Note that we make no assumptions about the dependence structure among the random variables $Y_1, \ldots, Y_n$.

Lemma 2.2. Let $Y_1, Y_2, \ldots, Y_n$ be integer valued random variables with finite means, and define $W := \sum_{i=1}^n Y_i$. Let $(a_i)_{i=1}^n$ and $(b_i)_{i=1}^n$ be real numbers such that, for all bounded $f: \mathbb{Z} \to \mathbb{R}$,

$$|\mathbb{E}[Y_i f(W)] - \mathbb{E}[Y_i]\mathbb{E} f(W) - a_i \mathbb{E}[\Delta f(W)]| \leqslant b_i \|\Delta f\|, \quad 1 \leqslant i \leqslant n. \tag{2.6}$$

Then

$$\|\mathcal{L}(W) - \mathrm{TP}\,(\mathbb{E}W, \sigma^2)\| \leqslant 2(\lambda')^{-1}\left(\delta + \sum_{i=1}^{n} b_i\right) + 2\mathbb{P}[W < \mathbb{E}W - \sigma^2],$$

where $\sigma^2 := \sum_{i=1}^{n} a_i$, $\delta = \delta(\mathbb{E}W, \sigma^2)$ and $\lambda' = \sigma^2 + \delta$.

Proof. Adding (2.6) over i, and then adding and subtracting $c\mathbb{E}f(W)$ for $c \in \mathbb{R}$ to be chosen at will, we get

$$|\mathbb{E}[(W-c)f(W)] - (\mathbb{E}W - c - \sigma^2)\mathbb{E}f(W) - \sigma^2\mathbb{E}[f(W+1)]| \leqslant \left(\sum_{i=1}^{n} b_i\right)\|\Delta f\|,$$

where $\sigma^2 = \sum_{i=1}^{n} a_i$ as above. Taking $c = \gamma = \lfloor \mathbb{E}W - \sigma^2 \rfloor$, so that the middle term (almost) disappears, the expression can be rewritten as

$$|\mathbb{E}[(W-\gamma)f(W)] - \lambda'\mathbb{E}[f(W+1)]| \leqslant \left(\delta + \sum_{i=1}^{n} b_i\right)\|\Delta f\|, \tag{2.7}$$

where δ and λ' are as above.

Fixing any set $B \subset \mathbb{Z}_+ + \gamma$, take $f = f_B$ as in (2.2). It then follows from (2.3) that

$$\begin{aligned}
&|\mathbb{P}(W \in B) - \mathrm{TP}\,(\mathbb{E}W, \sigma^2)\{B\}| \\
&\quad = |\mathbb{E}\{(\mathbf{1}_B(W) - \mathrm{TP}\,(\mathbb{E}W, \sigma^2)\{B\})(I[W \geqslant \gamma] + I[W < \gamma])\}| \\
&\quad \leqslant |\mathbb{E}\{(\lambda' f_B(W+1) - (W-\gamma)f_B(W))\, I[W \geqslant \gamma]\}| + \mathbb{P}(W < \gamma) \\
&\quad = |\mathbb{E}\{\lambda' f_B(W+1) - (W-\gamma)f_B(W)\}| + \mathbb{P}(W < \gamma),
\end{aligned} \tag{2.8}$$

this last from (2.4). Hence (2.7) and (2.8) show that, for any $B \subset \mathbb{Z}_+ + \gamma$,

$$\begin{aligned}
&|\mathbb{P}(W \in B) - \mathrm{TP}\,(\mathbb{E}W, \sigma^2)\{B\}| \\
&\quad \leqslant \left(\delta + \sum_{i=1}^{n} b_i\right)\|\Delta f_B\| + \mathbb{P}(W < \gamma) \\
&\quad \leqslant (\lambda')^{-1}\left(\delta + \sum_{i=1}^{n} b_i\right) + \mathbb{P}(W < \gamma).
\end{aligned} \tag{2.9}$$

Now the largest value D of the differences $\{\mathrm{TP}\,(\mathbb{E}W, \sigma^2)\{C\} - \mathbb{P}(W \in C)\}$, $C \subset \mathbb{Z}$, is attained at a set $C_0 \subset \mathbb{Z}_+ + \gamma$, and is thus bounded as in (2.9);

the minimum is attained at $\mathbb{Z}\setminus C_0$ with the value $-D$. Hence

$$|\mathbb{P}(W\in C)-\mathrm{TP}\,(\mathbb{E}W,\sigma^2)\{C\}|\leqslant(\lambda')^{-1}\left(\delta+\sum_{i=1}^{n}b_i\right)+\mathbb{P}(W<\gamma)$$

for all $C\subset\mathbb{Z}$, and the lemma follows. □

If the random variables Y_i have finite variances, both λ' and $\mathrm{Var}\,W$ are typically of order $O(n)$, so that letting $\bar{b}:=n^{-1}\sum_{i=1}^{n}b_i$ and applying Chebyshev's inequality to bound the final probability, we find that then $\|\mathcal{L}(W)-\mathrm{TP}\,(\mathbb{E}W,\sigma^2)\|$ is of order $O(n^{-1}+\bar{b})$. Hence we are interested in choosing $a_1,a_2,\dots$ so that $b_1,b_2,\dots$ are small. For independent $Y_1,Y_2,\dots$, it is easy to convince oneself that the choice

$$a_i=\mathbb{E}[Y_i\,W]-\mathbb{E}[Y_i]\mathbb{E}[W] \tag{2.10}$$

is a good one, and this also emerges in our Markovian context. Notice that (2.10) implies that $\sigma^2=\mathrm{Var}\,W$.

Establishing (2.6) in the Markovian setting, for a_i chosen as in (2.10), is the core of the paper; it is accomplished in Section 4. For the estimates made in that analysis, it is useful to introduce a coupling of X with an independent copy $X'=(X'_i)_{i=0}^{\infty}$. The relevant properties of the coupling are given in the next section. From now on, we assume that the conditional distributions $\mathcal{L}(Y_1\,|\,X_1=k)$, $0\leqslant k\leqslant K$, each have finite variance.

3. THE MARKOV CHAIN COUPLING

Let $X=(X_i)_{i=0}^{\infty}$ and $X'=(X'_i)_{i=0}^{\infty}$ be independent copies of an aperiodic, irreducible and stationary Markov chain with state space $E=\{0,1,\dots,K\}$. To understand their crucial role, recall (2.6), and note that

$$\begin{aligned}\mathbb{E}[Y_if(W)]-\mathbb{E}[Y_i]\mathbb{E}[f(W)]&=\mathbb{E}[Y_if(W)]-\mathbb{E}[Y_if(W')]\\&=\mathbb{E}[Y_i(f(W)-f(W'))].\end{aligned} \tag{3.1}$$

Here $W'=\sum_{i=1}^{n}Y'_i$, and $Y'_1,\dots,Y'_n$ are chosen from the conditional distributions $(\mathcal{L}(Y_i\,|\,X'_i),\ 1\leqslant i\leqslant n)$, independently of each other and of X and $Y:=(Y_1,\dots,Y_n)$. Also, recall (2.10), and note that then

$$a_i=\mathbb{E}[Y_i(W-W')]. \tag{3.2}$$

Of course (3.1) and (3.2) follow from the independence of (X,Y) and (X',Y').

We refer to Lindvall[8] (Part II.1) for proofs of the statements to be made now; we shall be brief.

Let 0 be our reference state, and let $S=(S_m)_{m=0}^{\infty}$ and $S'=(S'_m)_{m=0}^{\infty}$ be the points in increasing order of the sets

$$\{k\in\mathbb{Z}_+;\ X_k=0\} \quad \text{and} \quad \{k\in\mathbb{Z}_+;\ X'_k=0\},$$

respectively. Then S and S' are stationary renewal processes. Define $Z_0, Z_1, \ldots, Z'_0, Z'_1, \ldots$ by

$$S_m=\sum_{j=0}^{m} Z_j, \qquad S'_m=\sum_{j=0}^{m} Z'_j.$$

Then all the Z variables are independent, and the recurrence times $Z_1, Z'_1, Z_2, Z'_2, \ldots$ are identically distributed, while the delays Z_0, Z'_0 have the well-known distribution that renders S and S' stationary.

Now define $\tilde{S}=(\tilde{S}_m)_{m=0}^{\infty}$ to be the time points at which both S and S' have a renewal, i.e.,

$$\{k\in\mathbb{Z}_+;\ X_k=X'_k=0\}.$$

Then $\tilde{S}$ is again a stationary renewal process, and we set $\tilde{S}_m=\sum_{j=0}^{m}\tilde{Z}_j$.

Let $X^*=(X_i^*)_{i=0}^{\infty}$ be an irreducible, finite state space Markov chain with reference state 0, and let the associated $(S_m^*)_{m=0}^{\infty}, (Z_j^*)_{j=0}^{\infty}$ have the obvious meanings. For $j\geqslant 0$, write

$$D_j=\min\{S_m^*-j;\ S_m^*\geqslant j\}.$$

Due to the finiteness of the state space, it is easily proved that there exists a $\rho>1$ such that, as $m\to\infty$,

$$\max_k \mathbb{P}(D_j\geqslant m \mid X_j^*=k)=O(\rho^{-m}), \tag{3.3}$$

$$\mathbb{P}(Z_0^*\geqslant m)=O(\rho^{-m}), \qquad \mathbb{P}(Z_1^*\geqslant m)=O(\rho^{-m}), \tag{3.4}$$

(cf. Ref. 8, II.4, p. 30 ff.). Of course, the maximum in (3.3) does not depend on j. When applied to $((X_i, X'_i))_{i=0}^{\infty}$, the state space is $E\times E$; notice that the aperiodicity of X is needed to make $((X_i, X'_i))_{i=0}^{\infty}$ irreducible.

For the rest of this section, drop the assumption that X and X' are stationary, but rather let $X_0=X'_0=0$, denoting the associated probability by $\mathbb{P}^0$. We shall have much use for an estimate of

$$\beta(n):=\left\|\mathbb{P}^0\left[\left(X_n, 1+\sum_{i=1}^{n} Y_i\right)\in\cdot\right]-\mathbb{P}^0\left[\left(X_n, \sum_{i=1}^{n} Y_i\right)\in\cdot\right]\right\|. \tag{3.5}$$

It is natural to conjecture that $\beta(n)=O(1/\sqrt{n})$, since that would be true if the sums $\sum_{i=1}^{n} Y_i$ formed a random walk independent of X, under an aperiodicity assumption (cf. Ref. 8, II.12 and II.14).

Let us say that the distribution of an integer-valued variable V is *strongly aperiodic* if

$$\text{g.c.d.}\{k+i;\ \mathbb{P}(V=i)>0\}=1 \quad \text{for all } k. \tag{3.6}$$

It is crucial to our argument to assume as smoothness condition that

$$\text{the distribution of } \sum_{i=1}^{S_1} Y_i \text{ is strongly aperiodic} \tag{3.7}$$

a condition that we are actually able to weaken later (see Theorem 4.4). It then follows from (3.7) that also

$$\sum_{i=1}^{\tilde{S}_1}(Y_i - Y_i') \text{ is strongly aperiodic.} \tag{3.8}$$

For the estimate of (3.5), notice that

$$\begin{aligned} &\left\| \mathbb{P}^0\left[\left(X_n, 1+\sum_{i=1}^{n} Y_i\right)\in\cdot\right] - \mathbb{P}^0\left[\left(X_n, \sum_{i=1}^{n} Y_i\right)\in\cdot\right]\right\| \\ &\quad = \left\| \mathbb{P}^0\left[\left(X_n, 1+\sum_{i=1}^{n} Y_i\right)\in\cdot\right] - \mathbb{P}^0\left[\left(X_n', \sum_{i=1}^{n} Y_i'\right)\in\cdot\right]\right\|. \end{aligned} \tag{3.9}$$

Now let

$$\tau = \min\left\{k;\ 1+\sum_{i=1}^{\tilde{S}_k} Y_i = \sum_{i=1}^{\tilde{S}_k} Y_i'\right\}.$$

We note that $\sum_{i=1}^{\tilde{S}_k}(Y_i - Y_i')$, $k \geqslant 0$, is a random walk, with step size distribution given by (3.8): it has expectation 0, finite second moment, and is strongly aperiodic. For such a random walk, Karamata's Tauberian theorem may be used to prove that the probability that at least m steps are needed to hit the state -1 is of magnitude $O(1/\sqrt{m})^{(4)}$ (Ref 5, Theorem 10.25), and hence

$$\mathbb{P}^0(\tau \geqslant m) = O(1/\sqrt{m}). \tag{3.10}$$

Now make a coupling as follows:

$$X_i'' = \begin{cases} X_i' & \text{for } i < \tilde{S}_\tau, \\ X_i & \text{for } i \geqslant \tilde{S}_\tau \end{cases}$$

and define Y_i'', $i \geqslant 0$, accordingly. Recall (3.9). Standard coupling arguments yield

$$\begin{aligned} &\left\| \mathbb{P}^0\left[\left(X_n, 1+\sum_{i=1}^n Y_i\right) \in \cdot\right] - \mathbb{P}^0\left[\left(X_n, \sum_{i=1}^n Y_i\right) \in \cdot\right]\right\| \\ &\quad = \left\| \mathbb{P}^0\left[\left(X_n, 1+\sum_{i=1}^n Y_i\right) \in \cdot\right] - \mathbb{P}^0\left[\left(X_n'', \sum_{i=1}^n Y_i''\right) \in \cdot\right]\right\| \\ &\quad \leqslant 2\mathbb{P}^0(\tilde{S}_\tau > n). \end{aligned} \tag{3.11}$$

Let $\tilde{\mu} = \mathbb{E}[\tilde{S}_1]$ and $\alpha = 1/(2\tilde{\mu})$. We get

$$\begin{aligned} \mathbb{P}^0(\tilde{S}_\tau > n) &= \mathbb{P}^0(\tilde{S}_\tau > n,\ \tau \geqslant \alpha n) + \mathbb{P}^0(\tilde{S}_\tau > n,\ \tau < \alpha n) \\ &\leqslant \mathbb{P}^0(\tau \geqslant \alpha n) + \mathbb{P}^0(\tilde{S}_{\lfloor \alpha n \rfloor + 1} > n). \end{aligned}$$

But the latter probability is of order $O(1/n)$, due to Chebyshev's inequality, and the former of order $O(1/\sqrt{n})$, by (3.10). Hence (3.5), (3.9), and (3.11) imply that

$$\beta(n) = O(1/\sqrt{n}) \quad \text{as } n \to \infty. \tag{3.12}$$

4. MAIN THEOREM

We now turn to the approximation of $\mathcal{L}(W)$, with W as defined in the Markovian setting introduced in Section 1; the notation is as in the previous section, and the assumption that X and X' are stationary is back in force.

In order to state the main lemma, we need some further terminology. For each $1 \leqslant i \leqslant n$, we define

$$T_i^+ := \min\{n, \min\{\tilde{S}_k;\ \tilde{S}_k \geqslant i\}\}$$

and

$$T_i^- := \begin{cases} \max\{\tilde{S}_k;\ \tilde{S}_k \leqslant i\}, & \text{if } \tilde{S}_0 \leqslant i, \\ 1, & \text{if } \tilde{S}_0 > i. \end{cases}$$

We then set

$$A_i = \sum_{j=T_i^-}^{T_i^+} Y_j, \qquad W_i^- = \sum_{j=1}^{T_i^- -1} Y_i, \quad \text{and} \quad W_i^+ = \sum_{j=T_i^+ +1}^{n} Y_j \tag{4.1}$$

with the understanding that $W_i^- = 0$ if $T_i^- = 1$ and $W_i^+ = 0$ if $T_i^+ = n$. We also define A_i', $W_i'^-$ and $W_i'^+$ by replacing Y_j by Y_j'. For use in the argument to come, we introduce independent copies $X^{(l)}$ of the X-chain, $0 \leqslant l \leqslant K$, with $\mathcal{L}(X^{(l)}) = \mathcal{L}(X \mid X_0 = l)$. By sampling the corresponding Y-variables conditional on the realizations $X^{(l)}$, we then construct the associated partial sum processes $U^{(l)}$ by setting $U_m^{(l)} := \sum_{s=1}^m Y_s^{(l)}$. Similarly, we define pairs of processes $(\bar{X}^{(l)}, \bar{U}^{(l)})$ in the same way, but based on the *time-reversed* chain $\bar{X}$ starting with $\bar{X}_0 = l$ (Ref. 9, Theorem 1.9.1). For any $m \geqslant 1$ and $0 \leqslant l \leqslant K$, we then write

$$h_r(l,m) := \mathbb{P}(U_m^{(l)} \geqslant r+1), \;\; r \geqslant 0, \qquad h_r(l,m) := -\mathbb{P}(U_m^{(l)} \leqslant r), \quad r < 0,$$

and specify $\bar{h}_r(l,m)$ analogously, using the time-reversed processes $\bar{U}^{(l)}$; we then set $H(m) := \max\left\{\sum_{r\in\mathbb{Z}} \|h_r(\cdot,m)\|, \sum_{r\in\mathbb{Z}} \|\bar{h}_r(\cdot,m)\|\right\}$.

Lemma 4.1. With the a_i chosen as in (2.10), the inequality (2.6) is satisfied with

$$\begin{aligned} b_i := \varphi(n) \Big\{ & \tfrac{1}{2}\mathbb{E}[|Y_i(A_i - A_i')|(|A_i| + |A_i'|)] \\ & + \mathbb{E}\{|Y_i(A_i - A_i')|(H(T_i^+ - i) + H(i - T_i^-))\} \\ & + \mathbb{E}\{|Y_i(A_i - A_i')|\}\{\mathbb{E}|A_i| + \mathbb{E}(H(T_i^+ - i) + H(i - T_i^-))\}\Big\}, \end{aligned}$$

where $\varphi(n) = O(n^{-1/2})$ under assumption (3.7).

Proof. The analysis of (2.6) in our Markovian setting is rather technical, and we divide it into three steps. For the first step, we recall (3.1), giving

$$\begin{aligned} & \mathbb{E}[Y_i f(W)] - \mathbb{E}[Y_i]\mathbb{E}[f(W)] \\ & \quad = \mathbb{E}[Y_i(f(W) - f(W'))] \\ & \quad = \mathbb{E}[Y_i(f(W_i^- + A_i + W_i^+) - f(W'^-_i + A_i' + W'^+_i))] \\ & \quad = \mathbb{E}[Y_i(f(W_i^- + A_i + W_i^+) - f(W_i^- + A_i' + W_i^+))], \end{aligned} \tag{4.2}$$

a careful proof of the last equality making use of a conditioning on

$$\sigma\{T_i^-, T_i^+, \text{ and } X_j, X_j', Y_j, Y_j' \text{ for } T_i^- \leqslant j \leqslant T_i^+\}$$

and of the symmetry of X and X'. Let

$$\mathcal{F}_i = \sigma\{T_i^-, T_i^+, \text{ and } X_j, X_j', Y_j, Y_j' \text{ for } 0 \leqslant j \leqslant T_i^+\}.$$

Then the last expectation in (4.2) is equal to

$$\mathbb{E}\{\mathbb{E}[Y_i(f(W_i^- + A_i + W_i^+) - f(W_i^- + A_i' + W_i^+))|\mathcal{F}_i]\}. \tag{4.3}$$

Now, for $r, m \in \mathbb{Z}$, define

$$V(r,m) := I[0 \leqslant r \leqslant m-1] - I[-1 \geqslant r \geqslant m]$$

and observe that

$$\begin{aligned}
&f(W_i^- + A_i + W_i^+) - f(W_i^- + A_i' + W_i^+) \\
&\quad = \sum_{r\in\mathbb{Z}} \Delta f(W_i^- + A_i' + W_i^+ + r) V(r, A_i - A_i') \\
&\quad = (A_i - A_i')\Delta f(W_i^- + W_i^+) \\
&\qquad + \sum_{r\in\mathbb{Z}} [\Delta f(W_i^- + A_i' + W_i^+ + r) - \Delta f(W_i^- + W_i^+)] V(r, A_i - A_i') \\
&\quad = (A_i - A_i')\Delta f(W_i^- + W_i^+) \\
&\qquad + \sum_{r\in\mathbb{Z}} V(r, A_i - A_i') \sum_{s\in\mathbb{Z}} \Delta^2 f(W_i^- + W_i^+ + s) V(s, A_i' + r).
\end{aligned} \tag{4.4}$$

So, from (4.2) to (4.4), in estimating $\mathbb{E}[Y_i f(W)] - \mathbb{E}[Y_i]\mathbb{E}[f(W)]$, we have isolated the term

$$\mathbb{E}[Y_i(A_i - A_i')\Delta f(W_i^- + W_i^+)], \tag{4.5}$$

together with an error involving the second differences in (4.4). The remainder of the first step consists of bounding the magnitude of this error.

To do so, assume first that $1 \leqslant i \leqslant n/2$. Then the second differences in (4.4) are all of the form $\Delta^2 f(\cdot + W_i^+)$, where the "$\cdot$"-part is measurable with respect to $\mathcal{F}_i$, and W_i^+ is the contribution from the Markov chain starting from 0 at time T_i^+. Write

$$\begin{aligned}
\Delta^2 f(\cdot + W_i^+) = \Delta^2 f(\cdot + W_i^+) I(T_i^+ - i > n/4) \\
+ \Delta^2 f(\cdot + W_i^+) I(T_i^+ - i \leqslant n/4).
\end{aligned} \tag{4.6}$$

Due to (3.3) and (3.12), and observing also that $\|\Delta^2 f\| \leqslant 2\|\Delta f\|$, we obtain

$$|\mathbb{E}[\Delta^2 f(\cdot + W_i^+) \mid \mathcal{F}_i]| \leqslant 2\|\Delta f\|(C\rho^{-n/4} + \beta(n/4)) =: \varphi(n)\|\Delta f\| \tag{4.7}$$

for some constant $C>0$, and thus $\varphi(n)$ is of magnitude $O\left(1/\sqrt{n}\right)$ under assumption (3.7), in view of (3.12).

Now observe that all the variables in (4.4) except W_i^+ are $\mathcal{F}_i$-measurable. What remains in order to use (4.3) is a careful count of the second difference terms in (4.4), of which there are at most $\frac{1}{2}|A_i-A_i'|(|A_i|+|A_i'|)$. Using (4.2)–(4.5) and (4.7), we have found that

$$
\begin{aligned}
&|\mathbb{E}[Y_i(f(W)-f(W'))]-\mathbb{E}[Y_i(A_i-A_i')\Delta f(W_i^-+W_i^+)]| \\
&\quad \leqslant \tfrac{1}{2}\varphi(n)\|\Delta f\|\mathbb{E}[|Y_i(A_i-A_i')|(|A_i|+|A_i'|)],
\end{aligned}
\tag{4.8}
$$

where $\varphi(n)=O(1/\sqrt{n})$. This is responsible for the first term in the expression for b_i in the statement of the lemma, and completes the proof of the first step.

The next step is to work on $\mathbb{E}[Y_i(A_i-A_i')\Delta f(W_i^-+W_i^+)]$. Although the random variables $Y_i(A_i-A_i')$, W_i^-, and W_i^+ are dependent, they are conditionally independent given T_i^- and T_i^+, and then $\mathcal{L}(W_i^+\mid T_i^+=s)=\mathcal{L}(U_{n-s}^{(0)})$ for $i\leqslant s\leqslant n$, and $\mathcal{L}(W_i^-\mid T_i^-=s)=\mathcal{L}(\bar{U}_{s-1}^{(0)})$ for $1\leqslant s\leqslant i$. This suggests writing

$$
\begin{aligned}
&\mathbb{E}[Y_i(A_i-A_i')\Delta f(W_i^-+W_i^+)] \\
&\quad =\mathbb{E}[Y_i(A_i-A_i')\Delta f(\bar{U}_{i-1}^{(0)}+U_{n-i}^{(0)})]+\eta_i \\
&\quad =\mathbb{E}[Y_i(A_i-A_i')]\mathbb{E}\{\Delta f(\bar{U}_{i-1}^{(0)}+U_{n-i}^{(0)})\}+\eta_i
\end{aligned}
\tag{4.9}
$$

with η_i to be bounded.

We start by writing

$$
\mathbb{E}[Y_i(A_i-A_i')\Delta f(W_i^-+W_i^+)]=\mathbb{E}\{\mathbb{E}[Y_i(A_i-A_i')\Delta f(W_i^-+W_i^+)\mid \mathcal{G}_i]\}
$$

with $\mathcal{G}_i:=\sigma(W_i^-,Y_i(A_i-A_i'),T_i^+)$. Now $Y_i(A_i-A_i')$ is $\mathcal{G}_i$-measurable, and

$$
\begin{aligned}
&\mathbb{E}\{\Delta f(W_i^-+U_{n-i}^{(0)})-\Delta f(W_i^-+W_i^+)\mid \mathcal{G}_i\} \\
&\quad =\mathbb{E}\left\{\sum_{r\in\mathbb{Z}}\Delta^2 f(W_i^-+U_{n-T_i^+}^{(0)}+r)\,V(r,U_{n-i}^{(0)}-U_{n-T_i^+}^{(0)})\,\middle|\,\mathcal{G}_i\right\} \\
&\quad =\sum_{r\in\mathbb{Z}}\mathbb{E}\left\{\Delta^2 f(W_i^-+U_{n-T_i^+}^{(0)}+r)\,h_r(X_{n-T_i^+}^{(0)},T_i^+-i)\,\middle|\,\mathcal{G}_i\right\},
\end{aligned}
$$

where the last line follows because, conditional on $X^{(0)}_{n-T_i^+}$, $U^{(0)}_{n-i}-U^{(0)}_{n-T_i^+}$ is independent of W_i^- and $U^{(0)}_{n-T_i^+}$. This in turn implies that

$$\begin{aligned}
&|\mathbb{E}\{\Delta f(W_i^+ + U^{(0)}_{n-i}) - \Delta f(W_i^- + W_i^+)\,|\,\mathcal{G}_i\}| \\
&\quad\leqslant \|\Delta f\| \sum\nolimits_{r\in\mathbb{Z}} \|h_r(\cdot, T_i^+ - i)\| \\
&\quad\times\ \mathbb{E}\left\{\|\mathcal{L}((X^{(0)}_{n-T_i^+}, U^{(0)}_{n-T_i^+}+1)) - \mathcal{L}((X^{(0)}_{n-T_i^+}, U^{(0)}_{n-T_i^+}))\|\,|\,\mathcal{G}_i\right\} \\
&\quad\leqslant \|\Delta f\| H(T_i^+ - i)\varphi(n),
\end{aligned}$$

where the last line follows exactly as for (4.7). Thus it follows that

$$\begin{aligned}
&|\mathbb{E}\{Y_i(A_i - A_i')\Delta f(W_i^- + W_i^+)\} - \mathbb{E}\{Y_i(A_i - A_i')\Delta f(W_i^- + U^{(0)}_{n-i})\}| \\
&\quad\leqslant \|\Delta f\|\varphi(n)\mathbb{E}\{Y_i|A_i - A_i'|H(T_i^+ - i)\} =: \eta_{i1}.
\end{aligned}\tag{4.10}$$

An analogous argument, replacing W_i^- by $\bar{U}^{(0)}_{i-1}$, uses the expression

$$\begin{aligned}
&\mathbb{E}\{\Delta f(\bar{U}^{(0)}_{i-1} + U^{(0)}_{n-i}) - \Delta f(W_i^- + U^{(0)}_{n-i})\,|\,\mathcal{G}_i'\} \\
&\quad= \sum_{r\in\mathbb{Z}} \mathbb{E}\left\{\Delta^2 f(\bar{U}^{(0)}_{T_i^- - 1} + U^{(0)}_{n-i} + r)\,\bar{h}_r(\bar{X}^{(0)}_{T_i^- - 1}, i - T_i^-)\,\middle|\,\mathcal{G}_i'\right\},
\end{aligned}$$

where $\mathcal{G}_i' := \sigma(Y_i(A_i - A_i'), T_i^-)$, which we bound using $\varphi(n)$ as a bound for $\|\mathcal{L}(U^{(0)}_{n-i}+1) - \mathcal{L}(U^{(0)}_{n-i})\|$. This yields

$$\begin{aligned}
&|\mathbb{E}\{Y_i(A_i - A_i')\Delta f(W_i^- + U^{(0)}_{n-i})\} - \mathbb{E}\{Y_i(A_i - A_i')\Delta f(\bar{U}^{(0)}_{i-1} + U^{(0)}_{n-i})\}| \\
&\quad\leqslant \|\Delta f\|\varphi(n)\mathbb{E}\{Y_i|A_i - A_i'|H(i - T_i^-)\} =: \eta_{i2},
\end{aligned}\tag{4.11}$$

so that (4.9) holds with $\eta_i = \eta_{i1} + \eta_{i2}$, accounting for the second term in b_i, and completing the second step.

It now remains only to bound the difference between $\mathbb{E}\{\Delta f(\bar{U}^{(0)}_{i-1} + U^{(0)}_{n-i})\}$ and $\mathbb{E}\{\Delta f(W)\}$. This is accomplished much as before, by writing $W = W_i^- + A_i + W_i^+$. This gives

$$\begin{aligned}
&|\mathbb{E}\Delta f(W) - \mathbb{E}\Delta f(W_i^- + W_i^+)| \\
&\quad= \left|\mathbb{E}\left\{\sum_{r\in\mathbb{Z}} \mathbb{E}[\Delta^2 f(W_i^- + W_i^+ + r)V(r, A_i)\,|\,T_i^-, T_i^+, A_i]\right\}\right| \\
&\quad\leqslant \|\Delta f\|\varphi(n)\mathbb{E}|A_i|,
\end{aligned}\tag{4.12}$$

where the last line again follows as for (4.7), and then

$$\begin{aligned}
&|\mathbb{E}\Delta f(W_i^- + W_i^+) - \mathbb{E}\Delta f(\bar{U}^{(0)}_{i-1} + U^{(0)}_{n-i})| \\
&\quad\leqslant \|\Delta f\|\varphi(n)\mathbb{E}\{H(i - T_i^-) + H(T_i^+ - i)\}.
\end{aligned}\tag{4.13}$$

Multiplying the bounds in (4.12) and (4.13) by $\mathbb{E}\{Y_i|A_i - A_i'|\}$ gives the remaining elements of b_i, and the lemma is proved for $1 \leqslant i \leqslant n/2$ by combining (4.8)–(4.13).

For $n/2 < i \leqslant n$, recall that X and X' are stationary. It is well known that then $(X_{n-j})_{j=0}^n$ and $(X'_{n-j})_{j=0}^n$ are also stationary; these reversed processes inherit all the relevant properties of X and X'. In carrying out the analysis above for the reversed processes, we meet no obstacle, and hence the formula for the b_i holds also for $i > n/2$, with $\varphi(n)$ defined in (4.7) now replaced by its reversed process counterpart. This proves the lemma. □

The bound in Lemma 4.1 can be combined with Lemma 2.2 to prove the total variation approximation that we are aiming for, under appropriate conditions. The expression for b_i simplifies substantially, if we assume that

$$\max\{\mathbb{P}(Y_1 \geqslant r \mid X_1 = l), \mathbb{P}(Y_1 \leqslant -r \mid X_1 = l)\} \leq \mathbb{P}(Z \geqslant r) \tag{4.14}$$

for all $r \geqslant 0$ and $0 \leqslant l \leqslant K$, for a nonnegative random variable Z with $\mathbb{E}Z^3 < \infty$. If this is the case, then

$$\begin{aligned} H(m) \leqslant 2m\mathbb{E}Z, \qquad & \mathbb{E}(|A_i| \mid X, X') \leqslant 2(T_i^+ - T_i^- + 1)\mathbb{E}Z, \\ \mathbb{E}(A_i^2 \mid X, X') \ &\leqslant \ 2(T_i^+ - T_i^- + 1)^2\mathbb{E}Z^2, \\ \mathbb{E}(|Y_i A_i| \mid X, X') \ &\leqslant \ 2(T_i^+ - T_i^- + 1)\mathbb{E}Z^2 \end{aligned}$$

and

$$\mathbb{E}(|Y_i| A_i^2 \mid X, X') \leqslant 2(T_i^+ - T_i^- + 1)^2\mathbb{E}Z^3.$$

From these bounds, together with the fact that A_i and A_i' are independent conditional on X, X', it follows that

$$\begin{aligned} b_i &\leqslant \varphi(n)\{4\mathbb{E}Z^3\mathbb{E}\tau_i^2 + 8\mathbb{E}Z\mathbb{E}Z^2\mathbb{E}\tau_i^2 + 4\mathbb{E}Z^2\mathbb{E}\tau_i(2\mathbb{E}Z\mathbb{E}\tau_i + 2\mathbb{E}Z\mathbb{E}\tau_i)\} \\ &\leqslant 28\varphi(n)\mathbb{E}Z^3\mathbb{E}\tau_i^2, \end{aligned} \tag{4.15}$$

where $\tau_i := T_i^+ - T_i^- + 1$. Note that, since X and X' are in equilibrium, both chains can be taken to run for all positive and negative times, so that then $\mathbb{E}\tau_i^2 \leqslant \mathbb{E}\tau^2$, where τ is the length of that interval between successive times at which both X and X' are in the state 0 which contains the time point 0. $\mathbb{E}\tau_i^2$ is in general smaller than $\mathbb{E}\tau^2$, because T_i^- and T_i^+ are restricted to lie between 1 and n. Then the bound (4.15), combined with Lemma 2.2, leads to the following theorem.

Theorem 4.2. Under assumptions (3.7) and (4.14), and with stationary X, it follows that

$$\|\mathcal{L}(W) - \mathrm{TP}\,(\mathbb{E}W, \mathrm{Var}\,W)\| \leqslant 4(1 + 14n\varphi(n)\mathbb{E}\tau^2\mathbb{E}Z^3)/\mathrm{Var}\,W.$$

Note that

$$\mathrm{Var}\,W = \sum_{i=1}^{n} \mathbb{E}[\mathrm{Var}\,(Y_i \mid X_i)] + \mathrm{Var}\left(\sum_{i=1}^{n} \mathbb{E}(Y_i \mid X_i)\right),$$

so that the bound in Theorem 4.2 is of order $O\left(n^{-1} + \varphi(n)\right) = O(n^{-1/2})$ under these assumptions, unless $\mathcal{L}(Y_1)$ is degenerate, in which case W is a.s. constant. Note also that replacing each Y_i by $Y_i - c$, for any $c \in \mathbb{Z}$, results only in a translation, and does not change $\|\mathcal{L}(W) - \mathrm{TP}\,(\mathbb{E}W, \mathrm{Var}\,W)\|$, and this can be exploited if necessary when choosing the random variable Z in (4.14).

The assumption that X be stationary is not critical.

Theorem 4.3. Suppose that the assumptions of Theorem 4.2 hold, except that the initial distribution $\mathcal{L}(X_0)$ is not the stationary distribution. Then it is still the case that $\|\mathcal{L}(W) - \mathrm{TP}\,(\mathbb{E}W, \mathrm{Var}\,W)\| = O(n^{-1/2})$.

Proof. Let X' be in equilibrium and independent of X, and use it as in Section 3 to construct an equilibrium process X'' which is identical with X after the time T_1^+ at which X and X' first coincide in the state 0. Then Theorem 4.2 can be applied to W'', constructed from X'', and also

$$W = A_1 + W_1^+ \quad \text{and} \quad W'' = A_1'' + W_1^+$$

with A_1 and A_1'' defined as before. Let $g: \mathbb{Z} \to \mathbb{R}$ be any bounded function, and observe that

$$\begin{aligned}
|\mathbb{E}g(W) - \mathbb{E}g(W'')| &= |\mathbb{E}\{g(A_1 + W_1^+) - g(A_1'' + W_1^+)\}| \\
&\leqslant \left| \mathbb{E}\left\{ \mathbb{E}\left(I[A_1 > A_1''] \sum_{j=1}^{A_1 - A_1''} \Delta g(W_1^+ + A_1'' + j - 1) \mid T_1^+, A_1, A_1'' \right) \right.\right. \\
&\quad \left.\left. - \mathbb{E}\left(I[A_1 < A_1''] \sum_{j=1}^{A_1'' - A_1} \Delta g(W_1^+ + A_1 + j - 1) \mid T_1^+, A_1, A_1'' \right) \right\} \right|. \qquad (4.16)
\end{aligned}$$

Now, arguing as for (4.7) in the second inequality, we have

$$|\mathbb{E}\{\Delta g(W_1^+ + A_1'' + j) \mid T_1^+, A_1, A_1''\}| \leqslant \|g\| \, \|\mathcal{L}(W_1^+ + 1) - \mathcal{L}(W_1^+)\| \leqslant \|g\| \varphi(n)$$

with $\varphi(n) = O(1/\sqrt{n})$, implying from (4.16) that

$$|\mathbb{E}g(W) - \mathbb{E}g(W'')| \leqslant \mathbb{E}|A_1 - A_1''| \varphi(n) \|g\| \leqslant 2\mathbb{E}T_1^+ \mathbb{E}Z\varphi(n)\|g\|. \tag{4.17}$$

Although the distribution of T_1^+ is not the same as if both X and X' were at equilibrium, it has moments which are uniformly bounded for all initial distributions ν, in view of (3.3) and (3.4), and hence, from (4.17) and because $\varphi(n) = O(1/\sqrt{n})$, it follows that $\|\mathcal{L}(W) - \mathcal{L}(W'')\| = O(n^{-1/2})$.

On the other hand,

$$|\mathbb{E}W - \mathbb{E}W''| \;\leq\; \mathbb{E}|A_1 - A_1''| \;\leq\; 2\mathbb{E}T_1^+ \mathbb{E}Z$$

and also

$$|\mathrm{Var}\, W - \mathrm{Var}\, W''| \;\leq\; \mathrm{Var}\,(W - W'') + 2\sqrt{\mathrm{Var}\, W \,\mathrm{Var}\,(W - W'')}$$

with

$$\mathrm{Var}\,(W - W'') \;\leq\; \mathbb{E}\{|A_1 - A_1''|^2\} \;\leq\; 4\mathbb{E}\{(T_1^+)^2\}\mathbb{E}\{Z^2\} \;=:\; 4D^2,$$

giving

$$|\mathrm{Var}\, W - \mathrm{Var}\, W''| \;\leq\; 8D \max\{\sqrt{\mathrm{Var}\, W}, D\}.$$

Hence, from Lemma 2.1, it follows that

$$\|\mathrm{TP}\,(\mathbb{E}W, \mathrm{Var}\, W) - \mathrm{TP}\,(\mathbb{E}W'', \mathrm{Var}\, W'')\| = O(n^{-1/2})$$

also, completing the proof. □

Assumption (3.7), that the distribution $Q := \mathcal{L}\left(\sum_{i=1}^{S_1} Y_i \,\middle|\, X_0 = 0\right)$ be strongly aperiodic, can actually be relaxed; it is enough to assume that Q is aperiodic.

Theorem 4.4. Suppose that the assumptions of Theorem 4.2 hold, except that assumption (3.7) is weakened to assuming that Q is aperiodic. Then it is still the case that $\|\mathcal{L}(W) - \mathrm{TP}\,(\mathbb{E}W, \mathrm{Var}\, W)\| = O(n^{-1/2})$.

Proof. Define a new Markov chain $\widehat{X}$ by splitting the state 0 in X into two states, 0 and -1. For each j, set

$$\widehat{X}_j = \begin{cases} X_j, & \text{if } X_j \geqslant 1. \\ -R_j, & \text{if } X_j = 0, \end{cases}$$

where $(R_j,\ j \geqslant 0)$ are independent Bernoulli $\text{Be}(1/2)$ random variables; then set $\widehat{Y}_j = Y_j,\ j \geqslant 0$, and define $\widehat{W} = \sum_{j=1}^{n} \widehat{Y}_j$. Clearly, $W = \widehat{W}$ a.s., so that we can use the construction based on the chain $\widehat{X}$ to investigate $\mathcal{L}(W)$. However, choosing 0 as reference state also for $\widehat{X}$, we have

$$\widehat{Q} := \mathcal{L}\left(\sum_{j=1}^{\widehat{S}_1} \widehat{Y}_j \,\middle|\, \widehat{X}_0 = 0 \right) = \mathcal{L}\left(\sum_{m=1}^{M} V_m \right),$$

where $V_1, V_2, \ldots$ are independent and identically distributed with distribution Q, and M is independent of the V_j's, and has the geometric distribution $\text{Ge}(1/2)$. Since Q is aperiodic, it follows that $\widehat{Q}$ assigns positive probability to all large enough integer values, and is thus strongly aperiodic. Hence Theorems 4.2 and 4.3 can be applied to W, because of its construction as $\widehat{W}$ by way of $\widehat{X}$ and $\widehat{Y}$. □

Čekanavičius and Mikalauskas[6] have also studied total variation approximation in this context, in the degenerate case in which $Y_1 = h(k)$ a.s. on $\{X_1 = k\}$, $0 \leqslant k \leqslant K$. They use characteristic function arguments, based on earlier work of Ref. 11, and their approximations are in terms of signed measures, rather than translated Poisson distributions. In their Theorem 2.2, they give one approximation with error of order $O(n^{-1/2})$, and another, more complicated approximation with error of order $o(n^{-1/2})$. However, their formulation is probabilistically opaque, and their proofs give no indication as to the magnitude of the implied constants in the error bounds, or as to their dependence on the parameters of the problem. In fact, their 'smoothness' condition (2.8) requires that the Markov chain X has a certain structure, irrespective of the values of h, which is unnatural. For example, the X-chain with $K = 2$ which has transition matrix

$$\begin{pmatrix} \frac{9}{10} & \frac{1}{10} & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{4.18}$$

fails to satisfy their condition, although, for many score functions h, (1.3) is still true; for instance, our Theorem 4.4 applies to prove (1.3) if $h(0) = 3$ and $h(1) = h(2) = 1$. However, Q is not aperiodic when $h(0) = 3$, $h(1) = 1$

and $h(2) = 2$, and, without this smoothness condition being satisfied, Theorem 4.4 cannot be applied. This is in fact just as well, since the equilibrium distribution of W then assigns probability much greater than 2/3 to the set $3\mathbb{Z} \cup \{3\mathbb{Z}+1\}$, whereas the probability assigned to this set by the translated Poisson distribution with the corresponding mean and variance approaches 2/3 as $n \to \infty$.

In fact, if Q is periodic, it is rather the exception than the rule that $\mathcal{L}(W)$ and $\mathrm{TP}(\mathbb{E}W, \mathrm{Var}\, W)$ should be close in total variation. To see this, let Q have period d. Fix any $k \in E$, and take any $i \in \mathbb{Z}_+$ and any realization of the process such that $X_0(\omega) = 0$ and $X_i(\omega) = k$; let $R_{ki}(\omega) := \sum_{l=1}^{i} Y_l(\omega)$ modulo d. Then it is immediate that $R_{ki}(\omega) = r_k$ is a constant depending only on k, since, continuing two such realizations along the same X-path and with the same Y values until the process next hits 0, the two Y-sums then have to have the same remainder 0 modulo d. The same considerations show that $\mathcal{L}(Y_i \mid X_i = k)$ is concentrated on a set $d\mathbb{Z} + \rho_k$ for some $\rho_k \in \{0, 1, \ldots, d-1\}$, and that the transition matrix $P = (p_{kj})$ of the X-chain satisfies the condition

$$r_k + \rho_j \equiv r_j \mod d, \quad \text{whenever} \quad p_{kj} > 0. \tag{4.19}$$

Moreover, for the same r- and ρ-values, *any* choice of P consistent with (4.19) yields a distribution Q with period d.

Now the distribution $\mathrm{TP}(\mu, \sigma)$ assigns probability approaching $1/d$ as $\sigma \to \infty$ to any set of the form $d\mathbb{Z} + r$, $r \in \{0, 1, \ldots, d-1\}$. On the other hand, using $\mathbb{P}^\lambda$ to denote probabilities computed with λ as the distribution of X_0, we have

$$\begin{aligned}\mathbb{P}^\lambda[W \equiv r \mod d] &= \sum_{i \in E} \lambda_i \mathbb{P}[W \equiv r \mod d \mid X_0 = i] \\ &= \sum_{i \in E} \lambda_i \mathbb{P}[X_n \in E_{r-r_i} \mid X_0 = i],\end{aligned}$$

where $E_r := \{k \in E : r_k = r\}$ and differences in the indices are evaluated modulo d. This, as $n \to \infty$, approaches the value

$$\sum_{i \in E} \lambda_i \pi(E_{r-r_i}) = \sum_{s=0}^{d-1} \lambda(E_s) \pi(E_{r-s}),$$

where π is the stationary distribution of the X-chain. Hence $\mathcal{L}^\lambda(W)$ becomes far from any translated Poisson distribution as $n \to \infty$ unless

$$\sum_{s=0}^{d-1} \lambda(E_s) \pi(E_{r-s}) = 1/d \quad \text{for all } 0 \leqslant r \leqslant d-1. \tag{4.20}$$

It is immediate that (4.20) cannot hold for *all* choices of λ unless

$$\pi(E_r) = 1/d \quad \text{for each } r \in \{0, 1, \ldots, d-1\}. \tag{4.21}$$

What is more, it cannot hold in the stationary case, when $\lambda = \pi$, unless (4.21) holds. This follows from multiplying both sides of (4.20) (with $\lambda = \pi$) by t_j^r and adding over r, where t_j, $0 \leqslant j \leqslant d-1$, are the complex dth roots of unity, with $t_0 := 1$. Writing $\pi(t) := \sum_{s=0}^{d-1} \pi(E_s) t^s$, this implies that $\{\pi(t_j)\}^2 = 0$ for $1 \leqslant j \leqslant d-1$, and hence that the polynomial $\pi(t)$ is proportional to the polynomial $\sum_{s=0}^{d-1} t^s$, which implies (4.21). Indeed, the (circulant) matrix Π with elements $\Pi_{rs} = \pi(E_{r-s})$ has d distinct eigenvectors corresponding to the eigenvalues $\pi(t_j)$, so that if $\pi(t_j) \neq 0$ for all j, then (4.20) has $\lambda(E_s) = 1/d$ for all s as its only solution.

But condition (4.19) depends only on the communication structure of P, and not on the exact values of its positive elements, whereas for (4.21) to be true needs careful choice of the values of these elements. Hence, for most choices of P leading to a periodic Q, meaning those in which $\pi(E_r) = 1/d$ for all r is *not* true, $\mathcal{L}^\lambda(W)$ and $\mathrm{TP}(\mathbb{E}^\lambda W, \mathrm{Var}^\lambda W)$ are not asymptotically close for $\lambda = \pi$, or if λ is concentrated on a single point, or indeed, if $\pi(t_j) \neq 0$ for all j, for any λ not satisfying $\lambda(E_s) = 1/d$ for all s. In consequence, for most choices of P leading to a periodic Q, the conclusions of Theorems 4.2 and 4.3 are very far from true.

In example (4.18), Q has period 3 when $h(0) = 3$, $h(1) = 1$ and $h(2) = 2$. Clearly, we have $\rho_0 = 0$, $\rho_1 = 1$ and $\rho_2 = 2$; we then also have $r_0 = r_2 = 0$ and $r_1 = 1$, so that $E_0 = \{0, 2\}$, $E_1 = \{1\}$ and $E_2 = \emptyset$. It is easy to check that (4.19) is satisfied, and that it would still be satisfied if p_{21} were also positive. The matrices P consistent with condition (4.19) for these values of the ρ_k and r_k thus take the form

$$\begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & 0 & 1 \\ \beta & 1-\beta & 0 \end{pmatrix}$$

for $0 \leqslant \alpha, \beta \leqslant 1$, so that $\pi = (\beta, \alpha, \alpha)/\{\beta + 2\alpha\}$; in (4.18), $\alpha = 1/10$ and $\beta = 1$. However, since $\pi(E_2)$ is necessarily zero, condition (4.21) is never satisfied. Furthermore, $\lambda(E_2)$ must also be zero, and $\pi(t_j) = 0$ can only occur for t_j a complex cube root of unity if $\beta = \alpha$. Thus, in this example, the conclusions of Theorems 4.2 and 4.3 are never true; furthermore, if $\alpha \neq \beta$, translated Poisson approximation cannot be good for *any* initial distribution λ.

As a second example, take $K = 3$ and P of the form

$$\begin{pmatrix} 1-\alpha & \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1-\beta & \beta \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

for $0 \leqslant \alpha, \beta \leqslant 1$, so that $\pi = (\beta, \alpha\beta, \alpha, \alpha\beta)/\{\alpha + \beta + 2\alpha\beta\}$. This matrix satisfies (4.19) for Y-distributions satisfying $\rho_0 = \rho_2 = 0$ and $\rho_1 = \rho_3 = 1$ with d=2, and then $r_0 = r_3 = 0$ and $r_1 = r_2 = 1$, so that $E_0 = \{0, 3\}$ and $E_1 = \{1, 2\}$. Hence $\pi(E_0) = \pi(E_1) = 1/2$ only if $\alpha = \beta$, and, if $\alpha \neq \beta$, $\pi(-1) \neq 0$. Thus, if $\alpha \neq \beta$, the conclusions of Theorems 4.2 and 4.3 are far from true, and indeed translated Poisson approximation cannot possibly be good for any initial distribution λ which does not give equal weight to E_0 and E_1.

The assumption that X has finite state space E greatly simplifies our arguments, because uniform bounds on hitting and coupling times, such as those given in (3.3) and (3.4), are immediate. Results similar to ours can be expected to hold also for countably infinite E, provided that the chain X is such that uniform bounds analogous to (3.3) and (3.4) are valid, and if the distributions of the Y_i are such that, for instance, (4.14) also holds. However, a full analysis of the case in which E is countably infinite would be a substantial undertaking.

ACKNOWLEDGMENTS

The authors wish to thank the referees for many helpful suggestions. ADB is grateful for support from the Institute for Mathematical Sciences of the National University of Singapore, from Monash University School of Mathematics and Statistics, and from Schweizer Nationalfonds Projekt Nr 20–67909.02.

REFERENCES

1. Arratia, R., Goldstein, L., and Gordon, L. (1990). Poisson approximation and the Chen–Stein method. *Stat. Sci.* 5, 403–434.
2. Barbour, A. D., and Čekanavičius, V. (2002). Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30**, 509–545.
3. Barbour, A. D., Holst, L., and Janson, S. (1992). *Poisson Approximation,* Oxford University Press.
4. Barbour, A. D., and Xia, A. (1999). Poisson perturbations. *ESAIM P&S* **3**, 131–150.
5. Breiman, L. (1968). *Probability,* Addison–Wesley, Reading, MA.
6. Čekanavičius, V., and Mikalauskas, M. (1999). Signed Poisson approximations for Markov chains. *Stoch. Process. Appl.* **82**, 205–227.
7. Čekanavičius, V., and Vaitkus, P. (2001). Centered Poisson approximation by the Stein's method. *Lithuanian Math. J.* **41**, 319–329.

8. Lindvall, T. (2002). *Lectures on the Coupling Method,* Dover Publications, New York.
9. Norris, J. R. (1997). *Markov Chains,* Cambridge University Press, Cambridge.
10. Röllin, A. (2005). Approximation by the translated Poisson distribution. *Bernoulli* **12**, 1115–1128.
11. Siraždinov, S. H., and Formanov, Š. K. (1979). *Limit Theorems for Sums of Random Vectors Forming a Markov Chain,* Fan, Tashkent (in Russian).
12. Vaitkus, P., and Čekanavičius, V. (1998). On a centered Poisson approximation. *Lithuanian Math. J.* **38**, 391–404.