The Annals of Probability
1999, Vol. 27, No. 1, 565–596

COMPOUND POISSON APPROXIMATION FOR MARKOV CHAINS USING STEIN'S METHOD

BY TORKEL ERHARDSSON

Royal Institute of Technology, Stockholm

Let η be a stationary Harris recurrent Markov chain on a Polish state space $(S,\mathscr{S})$, with stationary distribution μ. Let $\Psi_n := \sum_{i=1}^{n} I\{\eta_i \in S_1\}$ be the number of visits to $S_1 \in \mathscr{S}$ by η, where S_1 is "rare" in the sense that $\mu(S_1)$ is "small." We want to find an approximating compound Poisson distribution for $\mathscr{L}(\Psi_n)$, such that the approximation error, measured using the total variation distance, can be *explicitly* bounded with a bound of order not much larger than $\mu(S_1)$. This is motivated by the observation that approximating Poisson distributions often give larger approximation errors when the visits to S_1 by η tend to occur in clumps and also by the compound Poisson limit theorems of classical extreme value theory.

We here propose an approximating compound Poisson distribution which in a natural way takes into account the regenerative properties of Harris recurrent Markov chains. A total variation distance error bound for this approximation is derived, using the compound Poisson Stein equation of Barbour, Chen and Loh and certain couplings. When the chain has an atom S_0 (e.g., a singleton) such that $\mu(S_0) > 0$, the bound depends only on much studied quantities like hitting probabilities and expected hitting times, which satisfy Poisson's equation. As "by-products" we also get upper and lower bounds for the error in the approximation with Poisson or normal distributions. The above results are illustrated by numerical evaluations of the error bound for some Markov chains on finite state spaces.

1. Introduction. In this paper we are concerned with the following problem: let η be a stationary Markov chain (by a "chain" we mean a Markov process in *discrete time*) on a Polish state space $(S,\mathscr{S})$. Assume that η is Harris recurrent with a unique stationary distribution μ (this includes irreducible positive recurrent Markov chains on countable state spaces; see below). Let Ψ_n be the number of visits made by η, during n consecutive time points, to a subset S_1 of the state space S, which is "rare" in the sense that $\mu(S_1)$ is "small." What can be said about the distribution of Ψ_n? In particular, can we go beyond purely asymptotic results and find approximating simpler distributions to $\mathscr{L}(\Psi_n)$, for which the error in the approximation can be explicitly bounded?

To clarify the last point: there are two different kinds of results that we could try to establish about $\mathscr{L}(\Psi_n)$. On one hand, we might consider a suitably chosen sequence of Markov chains $\{\eta^{(n)};\ n \in Z^+\}$, with a correspond-

Received March 1998; revised September 1998.

AMS 1991 *subject classifications*. Primary 60E15, 60J05; secondary 60G70.

Key words and phrases. Compound Poisson, approximation, error bound, stationary Markov chain, Harris recurrence, regenerative, "rare" set, Stein equation, coupling, hitting probabilities, expected hitting times.

ing sequence of "rare" sets $\{S_1^{(n)};\ n \in Z^+\}$; this could be referred to as a *scaling* of η. If the sets $\{S_1^{(n)};\ n \in Z^+\}$ get increasingly "rarer" as $n \to \infty$, then we could try to prove that, as $n \to \infty$, $\mathscr{L}(\Psi_n^{(n)})$ converges weakly to some nondegenerate limiting distribution. This would give us a *limit theorem*. On the other hand, we might try to find an error bound for an approximation of $\mathscr{L}(\Psi_n)$ with some suitable simpler distribution, by which we mean a bound for the distance in some appropriate metric between $\mathscr{L}(\Psi_n)$ and this approximating distribution, for each fixed n.

An often used metric on the space of probability measures is the *total variation distance* $d_{\mathrm{TV}}(\cdot, \cdot)$, defined for any two probability measures ν_1 and ν_2 by

$$d_{\mathrm{TV}}(\nu_1, \nu_2) := \sup_{A \in N} |\nu_1(A) - \nu_2(A)|.$$

A limit theorem is implied by the convergence to 0 of the corresponding total variation distance error bound. In this paper we will be concerned with finding suitable approximations of $\mathscr{L}(\Psi_n)$, and total variation distance error bounds for such approximations.

A motivation for the interest in such bounds is that many problems can be formulated in terms of visits made by a stationary Markov chain to a "rare" set. An important area where such problems arise is the *extreme value theory* for stationary Markov chains, where the state space is typically R and the "rare" set S_1 is (z, ∞), for some $z \in R$. Here a number of important results are known, although almost exclusively limit theorems. Some other applications can be found in Erhardsson (1998a, b). In Erhardsson (1998a) the number of overlapping occurrences of fixed sequences in a finite-state Markov chain is considered, and also the number of visits to "rare" sets by birth–death chains. In Erhardsson (1998b) the number of components of the uncovered set in the one-dimensional Johnson–Mehl model is studied. As was shown in Erhardsson (1996), this quantity can be interpreted as the number of visits to a "rare" set by a Markov chain.

A general observation concerning approximations for $\mathscr{L}(f(\eta))$, where $f(\cdot)$ is any functional, is that it is not a priori obvious which approximating distributions should be preferred. A natural first-order approximation is the limiting distribution of $\mathscr{L}(f(\eta))$ under some scaling of η. Convergence in distribution can be very slow, however, and this is not revealed by the limit theorem. In such cases it is of interest to find second-order approximations; at the price of a few additional degrees of freedom in the approximating distribution, one might achieve a better fit and greatly reduce the size of the error bound.

Returning to $\mathscr{L}(\Psi_n)$, if visits to S_1 occur at times which are not too close, and if the dependence in the chain is not too strong, then it is reasonable to expect Ψ_n to be approximately Poisson distributed. This should hold not only for Markov chains, but for stationary random sequences in general. Poisson limit theorems which formalize this idea can be found, in the context of extreme value theory for stationary random sequences, in Leadbetter, Lindgren and Rootzén (1983).

With the appearance of the Stein–Chen method for Poisson approximation, a powerful tool became available for proving total variation distance error bounds in the same direction; see Stein (1972), Chen (1975), Arratia, Goldstein and Gordon (1989) and Barbour, Holst and Janson (1992). Theorem 3.2.1 in Erhardsson (1996) is an example of a Poisson approximation error bound, proved using the Stein–Chen method. Such error bounds often converge to 0 under some natural scaling of η. However, it also turns out that this convergence can be quite slow if the visits by the Markov chain η to S_1 tend to occur in clumps. Sometimes this clumping can be so pronounced that not even a Poisson limit theorem can be proved.

If visits to S_1 tend to occur in clumps, then the compound Poisson distribution should be a better choice for an approximating distribution. The book by Aldous (1989) is devoted to this topic in a more general context. He there argues, in an intentionally heuristic manner, that the random set $\{t; \eta_t \in S_1\}$ could in many cases be well approximated by a so-called Boolean model, which can be thought of as generated by first placing out points (*germs*) according to a stationary Poisson point process on R, and then, in the vicinity of each germ, a random set (a *grain*); the grains are i.i.d. random sets, and they are independent of the point process. This approximation, which Aldous calls the "Poisson clumping heuristic," leads in the setting described above to a compound Poisson approximation for $\mathscr{L}(\Psi_n)$. Aldous designates as an important research topic the problem of finding error bounds for such approximations.

As for rigorous results, compound Poisson limit theorems for $\mathscr{L}(\Psi_n)$ have been proved in the context of extreme value theory for stationary random sequences on the state space $(R, \mathscr{B}(R))$. According to these theorems, if η is a stationary random sequence with extremal index γ (see Remark 6.2 below), which satisfies a certain mixing condition and some regularity conditions, then the point process of exceedances of η above a level $z > 0$ converges weakly, after a suitable scaling, to a compound Poisson point process. The regularity conditions involve the parameters of the limit process. In general it may be difficult to show that these conditions are met and determine the parameters explicitly. For details, see Leadbetter and Rootzén (1988) and Rootzén (1988).

In order to establish also total variation distance error bounds for compound Poisson approximations, it was asked whether an efficient Stein method for compound Poisson approximation could be found. One idea is to use the Stein–Chen method for (discrete) Poisson process approximation; see Section 10.4 in Barbour, Holst and Janson (1992). This method often works well but does not give the best possible bounds if the quantity $n\mu(S_1)$ is large. The compound Poisson Stein equation derived in Barbour, Chen and Loh (1992) constituted an important step forward and gave rise to hopes that it could yield approximation error bounds superior to those previously suggested. Compound Poisson approximations based on this result are suggested in Roos (1993), and these are, in Roos and Stark (1996), applied to the problem of compound Poisson approximation of $\mathscr{L}(\Psi_n)$, in a special case.

However, in their approach the regenerative structure of such Markov chains is not taken into account. Therefore, we believe that for the problem of compound Poisson approximation of $\mathscr{L}(\Psi_n)$, there is room for an alternative approach.

The main results of the present paper, which is a shortened version of Chapter 3 in Erhardsson (1997), are the following. We propose a new approximating compound Poisson distribution $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ for $\mathscr{L}(\Psi_n)$. This approximating distribution reflects in a natural way the fact that a Harris recurrent Markov chain is either regenerative or can be embedded in another Harris recurrent Markov chain with this property. In principle, $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ is the $\mathrm{Po}(\lambda^*)$ distribution, where λ^* is the expected number of cycles with at least one visit to S_1, compounded with the conditional distribution of the number of visits to S_1 by η during a cycle, given that the number of visits is at least 1. When S_1 is an atom (i.e., $P(\eta_1 \in B|\eta_0)$ is constant on $\{\eta_0 \in S_1\}$ for each $B \in \mathscr{F}$), the compounding distribution is geometric, so $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ is a Pólya–Aeppli distribution.

Moreover, we derive an explicit total variation distance error bound for the approximation of $\mathscr{L}(\Psi_n)$ with $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$. The fundamental tool used for this is the compound Poisson Stein equation in Barbour, Chen and Loh (1992), together with certain couplings, for which different choices are possible. When the Markov chain η has an atom S_0 such that $\mu(S_0) > 0$, the error bound takes on the following appearance (Theorem 4.3):

$$
\begin{aligned}
&d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) \\
&\quad \le 2H_1(\lambda_1^*, \lambda_2^*, \dots)\left(E_{S_1}(\tau_{S_0}) + E_{S_1}\big(\tau_{S_0}(\eta^R)\big) + \frac{E(\bar{\tau}_{S_0})}{\mu(S_0)}\right) \\
&\qquad \times n\mu(S_1)^2 + 2P(\tau_{S_1} < \tau_{S_0}).
\end{aligned}
$$

Here, $E_{S_1}(\tau_{S_0})$ and $E_{S_1}(\tau_{S_0}(\eta^R))$ are the expected first hitting times of the set S_0 for the chain η and the reversed chain η^R, respectively, with initial distributions $\mu(\cdot|S_1)$; $E(\bar{\tau}_{S_0})$ is the expected first hitting time of S_0 for η, with initial distribution μ, and $P(\tau_{S_1} < \tau_{S_0})$ is the probability that η will hit S_1 before it hits S_0, again with initial distribution μ. The first factor in the bound is the so-called magic factor, which is under a certain condition bounded by 1, and, if $n\mu(S_1)$ is large, much smaller than that; it gives an improvement over the previously mentioned Poisson process approximation approach. The quantities appearing in the error bound, expected first hitting times and hitting probabilities, have been subject to much study and can be found as solutions to Poisson's equation; in particular, when the state space is finite, they are the unique solutions to certain linear equation systems.

Some related questions are also treated. If $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots))$ can be explicitly bounded, we give upper and lower error bounds for the approximation of $\mathscr{L}(\Psi_n)$ with an arbitrary compound Poisson distribution (in particular, a Poisson distribution). We also consider the approximation of $\mathscr{L}(\Psi_n)$ with a normal distribution, which is of interest if $n\mu(S_1)$ is large; we combine

the main results in the paper with a Berry–Esseen theorem for normal approximation of compound Poisson distributions, to get a Berry–Esseen theorem for normal approximation of $\mathscr{L}(\Psi_n)$. Furthermore, there is a connection to the hitting time τ_{S_1} of a "rare" subset S_1 for the Markov chain η, since $P(\tau_{S_1} > n) = P(\Psi_n = 0)$. Many results about the convergence of $\mathscr{L}(\tau_{S_1})$ to an exponential distribution under a suitable scaling have been given in recent years; here, we give some error bounds for such exponential approximations.

The rest of the paper is organized as follows. In Section 2 we give some preliminary definitions and notation, concerning Markov chains in general and Harris recurrent Markov chains in particular. In Section 3 we define the approximating compound Poisson distribution $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$, and give some of its relevant properties. In Section 4 we state and prove the main theorems on total variation distance error bounds for compound Poisson approximation of $\mathscr{L}(\Psi_n)$. In Section 5 we explain how to calculate the hitting probabilities of subsets of the state space and expected first hitting times as solutions to Poisson's equation. Section 6 contains the results on, in particular, upper and lower error bounds for arbitrary compound Poisson and normal approximations for $\mathscr{L}(\Psi_n)$. Finally, in Section 7 we calculate numerically the parameters of the approximating compound Poisson distribution $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ and the bound for $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots))$ in some examples, where η is a Markov chain on a finite state space.

Further applications of the results in this paper to particular Markov models can be found in Erhardsson (1998a, b).

2. Preliminaries. We will consider the following situation. Let $(S, \mathscr{F})$ be a Polish state space equipped with the Borel σ-algebra. Let $p\colon S \times \mathscr{F} \to [0,1]$ be a stochastic transition probability on S with an invariant probability measure μ. Let $(\Omega, \mathscr{G}, P)$ be a probability space on which there is defined a random sequence $\eta\colon \Omega \to S^Z$ (i.e., a random element $\{\eta_t;\ t \in Z\}$ of the sequence space $(S^Z, \mathscr{F}^Z)$ equipped with the product σ-algebra), which is a stationary Markov chain on S, with transition probability p and stationary distribution μ. Define the shift operator $\theta\colon S^Z \to S^Z$ by $\theta(\{\dots, x_{-1}, x_0, x_1, \dots\}) := \{\dots, x_0, x_1, x_2, \dots\}$, and denote the kth iterate of θ by θ_k.

We denote the P-distribution of any random variable X defined on $(\Omega, \mathscr{G})$ by $\mathscr{L}(X)$. For each integrable random variable X defined on $(\Omega, \mathscr{G})$, each $A \in \mathscr{F}$ such that $\mu(A) > 0$, and each $B \in \mathscr{G}$, we define

$$E_A(X) := E(X|\eta_0 \in A), \qquad P_A(B) := P(B|\eta_0 \in A).$$

Furthermore, the transition probability p induces, for each $x \in S$, a probability measure on the sequence space $(S^N, \mathscr{F}^N)$, under which the coordinate process $\{\eta_t;\ t \in N\}$ is a p-Markov chain with initial distribution δ_x; see Section 2 in Chapter 1 of Revuz (1984). We denote this probability measure by $P_x(\cdot)$, and expectation with respect to $P_x(\cdot)$ by $E_x(\cdot)$. When used together with $P_x(\cdot)$, η will denote the coordinate process of $(S^N, \mathscr{F}^N)$.

We define, for each $A \in \mathscr{F}$, $t \in Z$ and $k \in N$, the functional $\tau_A^{t,k}: S^Z \to N$ as the time until the kth visit to A for a sequence $s := \{\dots, x_{-1}, x_0, x_1, \dots\} \in S^Z$, starting from but not including time t, by

$$\tau_A^{t,k}(s) := \inf\{j > \tau_A^{t,k-1}(s);\ x_{t+j} \in A\} \qquad \forall\, k \geq 1,$$
$$\tau_A^{t,0}(s) := 0.$$

It obviously holds that $\tau_A^{t,k}(\cdot) = \tau_A^{0,k} \circ \theta_t(\cdot)$. For brevity we will use the notation $\tau_A(\cdot) := \tau_A^{0,1}(\cdot)$, and the random variable $\tau_A^{t,k}(\eta)$ will be denoted by $\tau_A^{t,k}$. In an analogous way we define the functional $\bar{\tau}_A^{t,k}: S^Z \to N$ as the time until the kth visit to a subset A for a sequence $s \in S^Z$, starting from *and including* time t, by

$$\bar{\tau}_A^{t,k}(s) := \inf\{j > \tau_A^{t,k-1}(s);\ x_{t+j} \in A\} \qquad \forall\, k \geq 2,$$
$$\bar{\tau}_A^{t,1}(s) := \inf\{j \geq 0;\ x_{t+j} \in A\},$$
$$\bar{\tau}_A^{t,0}(s) := 0.$$

We will use the notation $\bar{\tau}_A(\cdot) := \bar{\tau}_A^{0,1}(\cdot)$, and $\bar{\tau}_A^{t,k} := \bar{\tau}_A^{t,k}(\eta)$.

We define, for each $A \in \mathscr{F}$, $B \in \mathscr{F}$ and $t \in Z$, the functional $N_{A,B}^t: S^Z \to N$ as the number of visits by $s \in S^Z$ to A before the first visit to B, starting from and including time t, by

$$N_{A,B}^t(s) := \sum_{i=0}^{\bar{\tau}_B^{t,1}(s)-1} I\{x_{t+i} \in A\}.$$

We will use the notation $N_{A,B}(\cdot) := N_{A,B}^0(\cdot)$ and $N_{A,B}^t := N_{A,B}^t(\eta)$.

We denote by η^R the reverse Markov chain of η, that is, the random element of $(S^Z, \mathscr{F}^Z)$ defined by $\eta_t^R := \eta_{-t}$ for each $t \in Z$. It is well known that since η is a stationary Markov chain with stationary distribution μ, the same holds for η^R. We denote by p^R the transition probability of η^R; this is clearly a μ-modification of the P-regular conditional distribution of η_0 given η_1, which exists since $(S, \mathscr{F})$ is Polish. Moreover, if η is Harris recurrent (see below), then so is η^R; see Theorem 4.8 in Chapter 4 of Revuz (1984).

For any two probability measures ν and ν' on any probability space $(\Omega, \mathscr{G})$, we define the *total variation distance* $d_{\mathrm{TV}}(\nu, \nu')$ in the usual way as

$$d_{\mathrm{TV}}(\nu, \nu') := \sup_{A \in \mathscr{G}} |\nu(A) - \nu'(A)|.$$

We say that a set $A \in \mathscr{F}$ is an *atom* for η if $p(\cdot, B)$ is constant on A for each $B \in \mathscr{F}$.

We will assume that the Markov chain η is Harris recurrent. Detailed references on Harris recurrence and numerous other related and important topics in the field of general state space Markov chains, are in Meyn and Tweedie (1993), Nummelin (1984), Revuz (1984) and Section VI.3 of Asmussen (1987). We give here some definitions and fundamental properties of Harris recurrent Markov chains, some of which will be needed in the following sections.

ψ-irreducibility and Harris recurrence. A Markov chain η is said to be *φ-irreducible* if there exists a measure φ on $(S,\mathscr{F})$ satisfying $\varphi(S) > 0$, called an irreducibility measure, such that

$$P_x(\tau_B < \infty) > 0 \qquad \forall\, x \in S,\, B \in \{A \in \mathscr{F};\, \varphi(A) > 0\}.$$

If η is φ-irreducible for some irreducibility measure φ, then there exists a (not necessarily unique) maximal irreducibility measure ψ on $(S,\mathscr{F})$, such that a measure φ on $(S,\mathscr{F})$ is an irreducibility measure for η if and only if $\varphi \ll \psi$; for short, we will say that η is ψ-irreducible.

By a Harris recurrent Markov chain η we mean a ψ-irreducible chain which also satisfies the condition that

$$P_x(\tau_B < \infty) = 1 \qquad \forall\, x \in B,\, B \in \{A \in \mathscr{F};\, \psi(A) > 0\}. \tag{2.1}$$

It follows from Proposition 9.1.1 and Theorem 9.1.4 in Meyn and Tweedie (1993) that under ψ-irreducibility (2.1) is in fact equivalent to the following seemingly stronger condition:

$$P_x\big(\tau_B^{0,k} < \infty\big) = 1 \qquad \forall\, k \in Z^+,\, x \in S,\, B \in \{A \in \mathscr{F};\, \psi(A) > 0\}. \tag{2.2}$$

Also, Corollary VI.3.12 in Asmussen (1987) tells us that a Markov chain η is Harris recurrent if and only if there exists a set $S_R \in \mathscr{F}$, a probability measure λ on $(S,\mathscr{F})$ and constants $\beta > 0$ and $m \in Z^+$, such that the following two conditions are satisfied:

$$P_x(\tau_{S_R} < \infty) = 1 \qquad \forall\, x \in S; \tag{2.3}$$

$$P_x(\eta_m \in B) \geq \beta\lambda(B) \qquad \forall x \in S_R,\, B \in \mathscr{F}. \tag{2.4}$$

S_R is called a regeneration set. Condition (2.4) is often called a minorization condition, and a set $S_R \in \mathscr{F}$ which satisfies (2.4) for some probability measure λ and constants β and m is called a small set; see Section 5.2 in Meyn and Tweedie (1993). From Theorem 5.2.2 in Meyn and Tweedie (1993) it follows that if η is ψ-irreducible, then each $A \in \mathscr{F}$ such that $\psi(A) > 0$ contains a small set A' such that $\psi(A') > 0$. Therefore, if η is Harris recurrent, then each $A \in \mathscr{F}$ such that $\psi(A) > 0$ contains a regeneration set A' such that $\psi(A') > 0$; in particular, each singleton A such that $\psi(A) > 0$ is a regeneration set.

For a Markov chain on a countable state space S, Harris recurrence means that S contains exactly one nonempty closed irreducible recurrent subset and that this subset will be hit in a.s. finite time, starting from any transient state.

Invariant measures. According to Theorem 10.4.9 in Meyn and Tweedie (1993), for a Harris recurrent Markov chain η there exists a unique (up to

constant multiples) invariant measure ν, which is equivalent to the maximal irreducibility measure ψ, and satisfies

$$\nu(\cdot) = \int_B E_x\left(\sum_{i=0}^{\tau_B - 1} I\{\eta_i \in \cdot\} \right) d\nu(x) \qquad \forall\, B \in \{A \in \mathscr{F};\ \nu(A) > 0\}. \tag{2.5}$$

A Harris recurrent Markov chain η is called positive if the invariant measure ν is finite, so that there exists a stationary distribution. This will be the case throughout the paper and holds if and only if

$$\int_B E_x(\tau_B)\, d\nu(x) < \infty \qquad \forall\, B \in \{A \in \mathscr{F};\ \nu(A) > 0\}.$$

In the case considered in this paper, when η is a stationary Harris recurrent Markov chain with stationary distribution μ, it also holds that, for each measurable function $g\colon S^Z \to [0, \infty)$,

$$\begin{aligned} &E\left(I\{\eta_0 \in B\} \sum_{i=0}^{\tau_B - 1} g(\theta_i \circ \eta) \right) \\ &\qquad = E(g(\eta)) \qquad \forall\, B \in \{A \in \mathscr{F};\ \mu(A) > 0\}. \end{aligned} \tag{2.6}$$

This follows from the Palm inversion formula; see Section 2.2 in Rolski (1981).

Regenerative properties. By a (classic sense) regenerative process in discrete time, we mean a random sequence $\{\eta_t;\ t \in N\}$ for which there exists a sequence of nonnegative finite random times $\{T_k;\ k \in N\}$ such that:

1. $\mathscr{L}(\{\eta_{T_k + r};\ r \in N\}, \{T_{k+l} - T_k;\ l \in N\})$ does not depend on k, for each $k \in N$.
2. $\sigma(\{\eta_{T_k + r};\ r \in N\}, \{T_{k+l} - T_k;\ l \in N\})$ is independent of $\sigma(T_i;\ i = 0, 1, \ldots, k)$ and $\sigma(\eta_t;\ t = 0, 1, \ldots, \eta_{T_k - 1})$, for each $k \in N$.

In particular, $\{T_k;\ k \in N\}$ is a renewal process, and the segments $\{\eta_{T_{k-1}}, \ldots, \eta_{T_k} - 1;\ k \in Z^+\}$ are called cycles. The process η is wide sense regenerative if, instead of property 2, it holds that $\sigma(\{\eta_{T_k + r};\ r \in N\}, \{T_{k+l} - T_k;\ l \in N\})$ is independent of $\sigma(T_i;\ i = 0, 1, \ldots, k)$ for each $k \in Z$; see Kalashnikov (1994).

A Harris recurrent Markov chain η is either classic sense regenerative or it can be embedded in another Markov chain which is classic sense, or at least wide sense, regenerative. The first case occurs when η has an atom A (see above) such that $\mu(A) > 0$; then clearly (2.3) and (2.4) hold with $S_R = A$ and $m = 1$ and the strong Markov property implies that we can choose $T_k := \bar{\tau}_A^{0, k+1} + 1$ for each $k \in N$.

The second case occurs when (2.3) and (2.4) hold with a regeneration set S_R which is not an atom, but still holds that $m = 1$. Then, using a technique called "splitting," η can be embedded in another Harris recurrent Markov chain on the state space $S \times \{0, 1\}$, for which $S_R \times \{1\}$ is an atom; see Section 4.4 in Nummelin (1984) or Example 1.4.1 in Kalashnikov (1994).

Finally, the third case occurs when (2.3) and (2.4) hold with a regeneration set S_R which is not an atom, and $m > 1$. Using more complicated "splitting" construction, η can be embedded in another Markov chain which is wide sense regenerative with 1-dependent cycles, meaning that nonadjacent cycles are independent. For details, see Example 1.4.1 in Kalashnikov (1994).

3. The approximating compound Poisson distribution.

DEFINITION 3.1. By CP($\lambda_1, \lambda_2, \dots$) or the compound Poisson distribution with parameters $\{\lambda_k;\ k \in Z^+\}$, where $\lambda_k \geq 0$ for each $k \in Z^+$ and $0 < \lambda := \sum_{k=1}^{\infty} \lambda_k < \infty$, we mean the distribution with the following two equivalent definitions:

(i) $\mathscr{L}(\sum_{i=1}^{M} T_i)$, where the variables $\{T_i;\ i \in Z^+\}$ and M are independent, $P(T_i = k) = \lambda_k/\lambda$ for each $k \in Z^+$ and $i \in Z^+$, and $M \sim \text{Po}(\lambda)$. Here $\mathscr{L}(T_1)$ is called the compounding distribution.

(ii) $\mathscr{L}(\sum_{k=1}^{\infty} kU_k)$, where the variables $\{U_k;\ k \in \mathbb{Z}^+\}$ are independent, and $U_k \sim \text{Po}(\lambda_k)$ for each $k \in Z^+$. (If $\lambda_k = 0$, then $U_k \equiv 0$.)

In the case when the compounding distribution of CP($\lambda_1, \lambda_2, \dots$) is geometric with parameter θ [i.e., $\lambda_k/\lambda = (1-\theta)^{k-1}\theta$ for each $k \in Z^+$], we refer to CP($\lambda_1, \lambda_2, \dots$) as the Pólya–Aeppli($\lambda, \theta$) distribution; see Section 8.2 in Johnson and Kotz (1969).

A more general definition of a compound Poisson distribution is the following [see Section A.19 in Aldous (1989)]. A random variable W is said to have a compound Poisson distribution POIS(ν), where ν is a measure on $(0, \infty)$ such that $\int_0^{\infty} (1 \wedge x)\, d\nu < \infty$, if the Laplace transform of W is

$$E(\exp(-sW)) = \exp\left(-\int_0^{\infty} (1 - e^{-sx})\, d\nu\right) \qquad \forall\ s \in (0, \infty).$$

If ν is finite, POIS(ν) $= \mathscr{L}(\sum_{i=1}^{M} T_i)$, where the variables $\{T_i;\ i \in Z^+\}$ and M are independent, $\mathscr{L}(T_i) = \nu/\nu(0, \infty)$ for each $i \in Z^+$ and $M \sim \text{Po}(\nu(0, \infty))$. In this paper we consider only distributions POIS(ν) for which ν is finite and has support on Z^+ (i.e., $\nu = \sum_{k=1}^{\infty} \lambda_k \delta_k$).

Our purpose is, for a suitable choice of parameters $\{\lambda_k;\ k \in Z^+\}$, to find an upper bound for $d_{\text{TV}}(\mathscr{L}(\Psi_n), \text{CP}(\lambda_1, \lambda_2, \dots))$, where $\Psi_n := \sum_{i=1}^{n} I\{\eta_i \in S\}$. To this end, we will use the following fundamental results from Barbour, Chen and Loh (1992), which we state as propositions. Proposition 3.1 is part of their Theorem 3; Proposition 3.2 follows from their Theorems 4 and 5.

PROPOSITION 3.1. *Let $g\colon N \to R$ be bounded, and let $\{\lambda_k;\ k \in Z^+\}$ satisfy the conditions of Definition* 3.1. *Then there exists a bounded solution $f\colon N \to R$ of the equation*

$$wf(w) - \sum_{k=1}^{\infty} k\lambda_k f(w+k) = g(w) \qquad \forall\ w \in N, \tag{3.1}$$

if and only if g is such that $Eg(W) = 0$, where $W \sim \mathrm{CP}(\lambda_1, \lambda_2, \dots)$. The solution is unique except at $w = 0$.

PROPOSITION 3.2. *Let $f_A\colon N \to R$ be the unique bounded solution of* (3.1) *when $g(\cdot) = I_A(\cdot) - P(W \in A)$, where $A \subset N$ and $W \sim \mathrm{CP}(\lambda_1, \lambda_2, \dots)$. Then,*

$$H_1(\lambda_1, \lambda_2, \dots) := \sup_{A \subset N} \sup_{w \geq 1} |f_A(w+1) - f_A(w)| \leq \left(\frac{1}{\lambda_1} \wedge 1 \right) e^{\lambda}. \tag{3.2}$$

In the case when $\{k\lambda_k;\ k \in Z^+\}$ is monotone decreasing towards 0, then,

$$H_1(\lambda_1, \lambda_2, \dots) \leq \frac{1}{\lambda_1 - 2\lambda_2} \left(\frac{1}{4(\lambda_1 - 2\lambda_2)} + \log^+ 2(\lambda_1 - 2\lambda_2) \right) \wedge 1. \tag{3.3}$$

In particular, the bound (3.3) *is bounded above by* 1 *and therefore smaller than the bound* (3.2), *which is bounded below by* 1.

Equation (3.1) with $g(\cdot) = I_A(\cdot) - P(W \in A)$, where $A \subset N$ and $W \sim \mathrm{CP}(\lambda_1, \lambda_2, \dots)$, is called a (compound Poisson) Stein equation. In the case when $\{k\lambda_k;\ k \in Z^+\}$ is monotone decreasing towards 0, if we define the function $h\colon N \to R$ by

$$h(w) := \sum_{k=0}^{w} f(k) \qquad \forall\ w \in N,$$

then the left-hand side of (3.1) can be expressed as $-\mathscr{A}h$, where $\mathscr{A}$ is the generator of an "immigration (in groups)-death" process Z, which is a pure jump process in continuous time with stationary distribution $\mathrm{CP}(\lambda_1, \lambda_2, \dots)$. It is well known [see Lemma 10 in Barbour, Chen and Loh (1992) and Lemma 1 in Barbour (1988)] that a solution to the equation $-\mathscr{A}h = g$, where $g(\cdot) = I_A(\cdot) - P(W \in A)$ for $A \subset N$ and $W \sim \mathrm{CP}(\lambda_1, \lambda_2, \dots)$, is given by $h_A\colon N \to R$, defined by

$$h_A(w) := \int_0^\infty \big(P(Z_t \in A | Z_0 = w) - P(W \in A) \big)\, dt \qquad \forall\ w \in N.$$

Thus, it holds that $f_A(w) = h_A(w) - h_A(w-1)$ for each $w \in Z^+$. Using couplings of "immigration (in groups)-death" processes with randomly perturbed initial values, the bound (3.3) for the quantity $H_1(\lambda_1, \lambda_2, \dots)$ can be derived. If $\{k\lambda_k;\ k \in Z^+\}$ is not monotone decreasing towards 0, then this Markovian generator interpretation is not valid. Using a different approach, one can still prove the bound (3.2).

We will use the Stein equation in the following way ("Stein's method"). Assume that Y is a random variable and that $W \sim \mathrm{CP}(\lambda_1, \lambda_2, \dots)$. Then, from (3.1),

$$\begin{aligned} d_{\mathrm{TV}}(\mathscr{L}(Y), \mathrm{CP}(\lambda_1, \lambda_2, \dots)) &= \sup_{A \subset N} \left| E\big(I\{Y \in A\} - P(W \in A) \big) \right| \\ &= \sup_{A \subset N} \left| E\left(Y f_A(Y) - \sum_{k=1}^{\infty} k\lambda_k f_A(Y+k) \right) \right|. \end{aligned} \tag{3.4}$$

We will get bounds for $d_{\mathrm{TV}}(\mathscr{L}(Y), \mathrm{CP}(\lambda_1, \lambda_2, \ldots))$ by constructing bounds for the right-hand side in (3.4), making use of (3.2) or (preferably) (3.3).

We first address the question of the exact choice for the approximating compound Poisson distribution. To our knowledge, the only previous suggestion that has been made in a Stein's method context is the one in Roos (1993); see Remark 3.3 below. The alternative that we propose here is in our view very natural for the study of Markov chains. The idea, in the spirit of Aldous's "Poisson clumping heuristic," is to partition the times of visits to S_1 by η into weakly dependent clumps, in the following way.

Choose a set $S_0 \in \mathscr{F}$ such that $\mu(S_0) > 0$ and $S_0 \cap S_1 = \varnothing$; preferably, $\mu(S_0)$ should be rather large. We know from Section 2 that any $S_0 \in \mathscr{F}$ such that $\mu(S_0) > 0$ contains a *regeneration set* S_R such that $\mu(S_R) > 0$. When η has an *atom* it is often convenient to choose an atom as S_0, but this is not necessary. We say that a clump of visits to S_1 by the Markov chain η begins at time t if η has an excursion away from S_0 which begins at time t, that is, if $\eta_t \in S_0$ and $\eta_{t+1} \notin S_0$. The size of such a clump is the number of visits to S_1 made by the chain between times t and $\tau_{S_0}^{t,1}$ (thus, the size may be 0). In particular, if S_0 is an atom, then a clump begins at t if and only if a cycle begins at t, and the size of the clump is the number of visits to S_1 during that cycle.

To indicate whether a clump begins at time t, and in such a case, the size of the clump, we define random variables $\{Z_t;\ t \in Z\}$ by

$$Z_t := I\{\eta_t \in S_0\} N_{S_1, S_0}^{t+1} \qquad \forall\ t \in Z.$$

The P-distribution of Z_t is the following:

$$P(Z_t = k) = P\left(\eta_0 \in S_0,\ \tau_{S_1}^{0,k} < \tau_{S_0} < \tau_{S_1}^{0,k+1}\right) \qquad \forall\ k \in Z^+.$$

We next define the quantity $W_n := \sum_{i=1}^n Z_i$, for which the basic coupling inequality gives

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathscr{L}(W_n)\big) &\le P(\Psi_n \neq W_n) \\ &\le P\left(\{\tau_{S_1} < \tau_{S_0}\} \cup \left\{\tau_{S_1}^{n,1} < \tau_{S_0}^{n,1}\right\}\right) \le 2P(\tau_{S_1} < \tau_{S_0}). \end{aligned}$$

Since the right-hand side should be small, we switch our interest from Ψ_n to W_n. The natural approximating compound Poisson distribution for $\mathscr{L}(W_n)$ is $\mathscr{L}(\sum_{i=1}^M T_i)$, where the variables $\{T_i;\ i \in Z^+\}$ and M are independent, the $\{T_i;\ i \in Z^+\}$ all have the distribution given by $P(T_i = k) = P(Z_i = k | Z_i > 0)$ for each $k \in Z^+$ and $M \sim \mathrm{Po}(nP(Z_0 > 0))$. In words, we consider the total number of nonempty clumps as Poisson distributed (with the same mean) and the sizes of the nonempty clumps as independent random variables, which are also independent of their total number. Summing up, we are led to the following definition.

DEFINITION 3.2. Let η be a stationary Harris recurrent Markov chain, and let $S_0 \in \mathscr{F}$ and $S_1 \in \mathscr{F}$ be such that $\mu(S_0) > 0$, $\mu(S_1) > 0$, and $S_0 \cap S_1 = \varnothing$. Let also $Z_0 := I\{\eta_0 \in S_0\} N_{S_1, S_0}^1$. By the approximating compound Poisson

distribution $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ for $\mathscr{L}(\Psi_n)$ (with respect to S_0), we mean the compound Poisson distribution with parameters $\lambda_k = \lambda_k^* := nP(Z_0 = k)$ for each $k \in Z^+$, and $\lambda = \lambda^* := \sum_{k=1}^{\infty} \lambda_k^* = nP(Z_0 > 0)$.

REMARK 3.1. In the special case when $P_{(\cdot)}(\tau_{S_1} < \tau_{S_0})$ is constant on S_1, which will happen, for example, if S_1 is an atom, the strong Markov property implies that

$$P(Z_t = k) = P\big(\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}\big) P_{S_1}(\tau_{S_1} < \tau_{S_0})^{k-1} P_{S_1}(\tau_{S_0} < \tau_{S_1}) \qquad \forall\, k \in Z^+.$$

Hence, $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ is in this case the Pólya–Aeppli(λ^*, θ) distribution, with parameters $\theta := P_{S_1}(\tau_{S_0} < \tau_{S_1})$ and $\lambda^* = nP(\eta_0 \in S_0,\ \tau_{S_1} < \tau_{S_0}) = n\mu(S_1) P_{S_1}(\tau_{S_0} < \tau_{S_1})$, where the last equality follows from Lemma 3.1 below with $k = 1$.

REMARK 3.2. A result from extreme value theory [Theorem 2.4.3 in Leadbetter and Rootzén (1988)] gives, for any stationary random sequence η on the state space $(R, \mathscr{B}(R))$, sufficient conditions for the point process of exceedances of η above a level z to converge weakly to a compound Poisson point process, under a certain scaling. If η is a stationary aperiodic Harris recurrent Markov chain with stationary distribution μ and an atom S_0 such that $\mu(S_0) > 0$, and if $S_1 := \{x \in R;\ x > u_n\}$ for some sequence $\{u_n;\ n \in Z^+\}$, then these sufficient conditions can be written as

$$\lim_{n \to \infty} \lambda^* = \lambda; \qquad \lim_{n \to \infty} \lambda_k^* = \lambda_k \qquad \forall\, k \in Z^+, \tag{3.5}$$

where $\{\lambda_k^*;\ k \in Z^+\}$ are the same as in Definition 3.2, and $\{\lambda_k;\ k \in Z^+\}$ are constants satisfying the conditions of Definition 3.1; see Theorem 2.6.1 in Leadbetter and Rootzén (1988). If (3.5) holds, then the limiting compound Poisson point process has constant jump intensity λ and a cluster size distribution which is the compounding distribution of $\mathrm{CP}(\lambda_1, \lambda_2, \dots)$.

REMARK 3.3. In Roos (1993), a compound Poisson distribution $\mathrm{CP}(\lambda_1^R, \lambda_2^R, \dots)$ is suggested for the approximation of $\mathscr{L}(\sum_{i=1}^n I_i)$, where $\{I_i;\ i = 1, \dots, n\}$ is any collection of indicator variables (in particular, one could take $I_i = I\{\eta_i \in S_1\}$ for each $i = 1, \dots, n$). The index set $\Gamma := \{1, \dots, n\}$ is there partitioned, for each $j \in \Gamma$, into four disjoint subsets: $\Gamma = \{j\} \cup \Gamma_j^{vs} \cup \Gamma_j^{vw} \cup \Gamma_j^b$. The partitioning should be chosen so that, informally,

$$\Gamma_j^{vs} = \big\{k \in \Gamma \setminus \{j\};\ I_k \text{ depends "very strongly" on } I_j\big\};$$

$$\Gamma_j^{vw} = \Big\{k \in \Gamma \setminus \{j\};\ I_k \text{ depends "very weakly" on } \big\{I_l;\ l \in \{j\} \cup \Gamma_j^{vs}\big\}\Big\}.$$

The distribution $\mathrm{CP}(\lambda_1^R, \lambda_2^R, \dots)$ is then defined by

$$\lambda_k^R := \begin{cases} \dfrac{1}{k} \displaystyle\sum_{i=1}^{n} E\left(I_i I\left\{ I_i + \sum_{j \in \Gamma_i^{vs}} I_j = k \right\} \right), & k = 1, \dots, \max_{l \in \Gamma} |\Gamma_l^{vs}| + 1; \\ 0, & k > \max_{l \in \Gamma} |\Gamma_l^{vs}| + 1. \end{cases}$$

See Roos (1993, 1994) for further details.

Last in this section we give a few theorems relating the parameters $\{\lambda_k^*; k \in Z^+\}$ to the upper bounds (3.2) and (3.3). We will need the following elementary result.

LEMMA 3.1. *Let η be a stationary random sequence with one-dimensional distribution μ, and let $A \in \mathscr{F}$ and $B \in \mathscr{F}$ be such that $\mu(A) > 0$, $\mu(B) > 0$ and $A \cap B = \varnothing$. Then,*

$$P\left(\eta_0 \in A, \tau_B^{0,k} < \tau_A\right) = P\left(\eta_0 \in B, \tau_B^{0,k-1} < \tau_A < \tau_B^{0,k}\right) \qquad \forall\, k \in Z^+. \tag{3.6}$$

PROOF. Define

$$I_{i,k} := I\left\{\eta_i \in A, \tau_B^{i,k} < \tau_A^{i,1}\right\} \qquad \forall\, i \in N, k \in Z^+;$$
$$J_{i,k} := I\left\{\eta_i \in B, \tau_B^{i,k-1} < \tau_A^{i,1} < \tau_B^{i,k}\right\} \qquad \forall\, i \in N, k \in Z^+.$$

We want to show that $E(I_{i,k}) = E(J_{i,k})$. It holds that $E(I_{i,k}) = E((1/n)\sum_{i=1}^n I_{i,k})$, that $E(J_{i,k}) = E((1/n)\sum_{i=1}^n J_{i,k})$ and also that $|\sum_{i=1}^n I_{i,k} - \sum_{i=1}^n J_{i,k}| \le 1$. We therefore get

$$\left| E(I_{i,k}) - E(J_{i,k}) \right| \le \frac{1}{n} E\left(\left| \sum_{i=1}^{n} I_{i,k} - \sum_{i=1}^{n} J_{i,k} \right| \right) \le \frac{1}{n} \qquad \forall\, n \in Z^+,$$

from which (3.6) follows, letting $n \to \infty$. □

THEOREM 3.1. *Let $\{\lambda_k^*; k \in Z^+\}$ be as in Definition 3.2 and define for brevity of notation $\hat{P}_{S_1}(\tau_{S_1} < \tau_{S_0}) := \mu\text{-ess sup}_{x \in S_1} P_x(\tau_{S_1} < \tau_{S_0})$ and $\check{P}_{S_1}(\tau_{S_1} < \tau_{S_0}) := \mu\text{-ess inf}_{x \in S_1} P_x(\tau_{S_1} < \tau_{S_0})$. A sufficient condition for $\{k\lambda_k^*; k \in Z^+\}$ to be monotone decreasing towards 0 is*

$$\hat{P}_{S_1}(\tau_{S_1} < \tau_{S_0}) \le \frac{1}{3 - 2\check{P}_{S_1}(\tau_{S_1} < \tau_{S_0})}. \tag{3.7}$$

Furthermore, a necessary condition for (3.7) to hold is that $\check{P}_{S_1}(\tau_{S_1} < \tau_{S_0}) \le \frac{1}{2}$. Also, if $P_{(\cdot)}(\tau_{S_1} < \tau_{S_0})$ is constant on S_1, a necessary and sufficient condition for $\{k\lambda_k^*; k \in Z^+\}$ to be monotone decreasing towards 0 is that $P_{S_1}(\tau_{S_1} < \tau_{S_0}) \le \frac{1}{2}$.

PROOF. According to the definition, we have

$$\begin{aligned}
&\frac{k\lambda_k^*}{(k+1)\lambda_{k+1}^*} \\
&\quad = \frac{kP(Z_i = k)}{(k+1)P(Z_i = k+1)} \\
&\quad \geq \frac{kP\big(\eta_0 \in S_0,\, \tau_{S_1}^{0,k} < \tau_{S_0}\big)\big(1 - \hat{P}_{S_1}(\tau_{S_1} < \tau_{S_0})\big)}{(k+1)P\big(\eta_0 \in S_0,\, \tau_{S_1}^{0,k} < \tau_{S_0}\big)\hat{P}_{S_1}(\tau_{S_1} < \tau_{S_0})\big(1 - \check{P}_{S_1}(\tau_{S_1} < \tau_{S_0})\big)} \\
&\quad \geq \frac{1 - \hat{P}_{S_1}(\tau_{S_1} < \tau_{S_0})}{2\hat{P}_{S_1}(\tau_{S_1} < \tau_{S_0})\big(1 - \check{P}_{S_1}(\tau_{S_1} < \tau_{S_0})\big)} \geq 1 \qquad \forall\, k \in Z^+,
\end{aligned}$$

where the last inequality is equivalent to condition (3.7). Furthermore, it can be shown by elementary calculations that $(3 - 2x)^{-1} \geq x$ for $x \in [0, \frac{1}{2}]$ and that $(3 - 2x)^{-1} < x$ for $x \in (\frac{1}{2}, 1)$, so the second part of the theorem follows. Finally, if $P_{(\cdot)}(\tau_{S_1} < \tau_{S_0})$ is constant on S_1, then

$$\begin{aligned}
\frac{k\lambda_k^*}{(k+1)\lambda_{k+1}^*} &= \frac{kP(Z_i = k)}{(k+1)P(Z_i = k+1)} \\
&= \frac{k}{(k+1)P_{S_1}(\tau_{S_1} < \tau_{S_0})} \geq 1 \qquad \forall\, k \in Z^+,
\end{aligned}$$

if and only if $P_{S_1}(\tau_{S_1} < \tau_{S_0}) \leq \frac{1}{2}$. □

THEOREM 3.2. *Let* $\{\lambda_k^*;\, k \in Z^+\}$ *be as in Definition* 3.2. *Then*,

$$\lambda_1^* - 2\lambda_2^* \geq n\mu(S_1)\big(1 - 4P_{S_1}(\tau_{S_1} < \tau_{S_0})\big).$$

Also, if $P_{(\cdot)}(\tau_{S_1} < \tau_{S_0})$ *is constant on* S_1, *then*,

$$\lambda_1^* - 2\lambda_2^* = n\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})^2\big(1 - 2P_{S_1}(\tau_{S_1} < \tau_{S_0})\big).$$

PROOF. For the first result,

$$\begin{aligned}
\lambda_1^* - 2\lambda_2^* &= n\big(P(Z_0 = 1) - 2P(Z_0 = 2)\big) \\
&= n\big(P(Z_0 > 0) - P(Z_0 > 1) - 2P(Z_0 = 2)\big) \\
&\geq n\big(P(Z_0 > 0) - 3P(Z_0 > 1)\big) \\
&\geq n\big(\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1}) - 3\mu(S_1)P_{S_1}(\tau_{S_1} < \tau_{S_0})\big) \\
&= n\mu(S_1)\big(1 - 4P_{S_1}(\tau_{S_1} < \tau_{S_0})\big),
\end{aligned}$$

where we used that, according to Lemma 3.1,

$$P(Z_0 > 1) = P\left(\eta_0 \in S_0,\, \tau_{S_1}^{0,2} < \tau_{S_0}\right) = P\left(\eta_0 \in S_1,\, \tau_{S_1} < \tau_{S_0} < \tau_{S_1}^{0,2}\right)$$
$$\leq \mu(S_1) P_{S_1}(\tau_{S_1} < \tau_{S_0}).$$

The second result is immediate from Remark 3.1. □

4. Total variation distance error bounds. In the first theorem of this section we give a total variation distance bound for $d_{\mathrm{TV}}(\mathscr{L}(\sum_{i=1}^n Z_i), \mathrm{CP}(\lambda_1, \lambda_2, \dots))$, where $\{Z_i;\ i = 1, \dots, n\}$ are *arbitrarily* but *identically* distributed nonnegative integer valued random variables, and $\lambda_k := nP(Z_1 = k)$ for each $k \in Z^+$. Remember that a pair of random variables (X, Y) defined on the same probability space is called a coupling of two probability distributions ν_1 and ν_2, if $\mathscr{L}(X) = \nu_1$ and $\mathscr{L}(Y) = \nu_2$.

THEOREM 4.1. *Let $\{Z_i;\ i = 1, \dots, n\}$ be identically distributed nonnegative integer valued random variables, such that $E(Z_1) < \infty$. Let $\lambda_k := nP(Z_1 = k)$ for each $k \in Z^+$. Let, for each $j \in \{1, \dots, n\}$ and $k \in Z^+$ such that $P(Z_1 = k) > 0$, the random variables $\{(Z_i^{j,k}, \tilde{Z}_i^{j,k});\ i = 1, \dots, n\}$ be a coupling of $\mathscr{L}(Z_i;\ i = 1, \dots, n)$ and $\mathscr{L}(Z_i;\ i = 1, \dots, n | Z_j = k)$, such that*

$$\mathscr{L}\left(Z_i^{j,k};\ i = 1, \dots, n\right) = \mathscr{L}(Z_i;\ i = 1, \dots, n),$$
$$\mathscr{L}\left(\tilde{Z}_i^{j,k};\ i = 1, \dots, n\right) = \mathscr{L}\left(Z_i;\ i = 1, \dots, n | Z_j = k\right).$$

Then,

$$d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{i=1}^n Z_i\right), \mathrm{CP}(\lambda_1, \lambda_2, \dots)\right)$$
$$\leq \frac{H_1(\lambda_1, \lambda_2, \dots)}{n} \sum_{j=1}^n \sum_{k=1}^\infty k\lambda_k E\left(\left|\sum_{i=1}^n Z_i^{j,k} - \sum_{\substack{i=1 \\ i \neq j}}^n \tilde{Z}_i^{j,k}\right|\right).$$

PROOF. Let $W \sim \mathrm{CP}(\lambda_1, \lambda_2, \dots)$. From (3.4) and the fact that $\sum_{k=1}^n k\lambda_k = nE(Z_1) < \infty$, we get

$$d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{i=1}^n Z_i\right), \mathrm{CP}(\lambda_1, \lambda_2, \dots)\right)$$
$$= \sup_{A \subset N} \left|P\left(\sum_{i=1}^n Z_i \in A\right) - P(W \in A)\right|$$
$$= \sup_{A \subset N} \left|E\left(\sum_{j=1}^n Z_j f_A\left(\sum_{i=1}^n Z_i\right) - \sum_{k=1}^\infty k\lambda_k f_A\left(\sum_{i=1}^n Z_i + k\right)\right)\right|$$
$$= \sup_{A \subset N} \left|\sum_{j=1}^n E\left(Z_j f_A\left(\sum_{i=1}^n Z_i\right)\right) - \sum_{k=1}^\infty k\lambda_k E\left(f_A\left(\sum_{i=1}^n Z_i + k\right)\right)\right|.$$

The definition of $\{\lambda_k;\ k \in Z^+\}$ implies that

$$E\left(Z_j f_A\left(\sum_{i=1}^n Z_i\right)\right) = \sum_{k=1}^\infty E\left(Z_j f_A\left(\sum_{i=1}^n Z_i\right)\middle| Z_j = k\right) P(Z_j = k)$$
$$= \frac{1}{n}\sum_{k=1}^\infty k\lambda_k E\left(f_A\left(\sum_{i=1}^n Z_i\right)\middle| Z_j = k\right) \qquad \forall\, j \in \{1,\dots,n\},$$

and this result together with the couplings gives

$$\sup_{A\subset N}\left|\sum_{j=1}^n E\left(Z_j f_A\left(\sum_{i=1}^n Z_i\right)\right) - \sum_{k=1}^\infty k\lambda_k E\left(f_A\left(\sum_{i=1}^n Z_i + k\right)\right)\right|$$
$$= \sup_{A\subset N}\left|\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^\infty k\lambda_k E\left(f_A\left(\sum_{\substack{i=1\\ i\neq j}}^n \tilde{Z}_i^{j,k} + k\right)\right)\right.$$
$$\left. - \frac{1}{n}\sum_{j=1}^n\sum_{k=1}^\infty k\lambda_k E\left(f_A\left(\sum_{i=1}^n Z_i^{j,k} + k\right)\right)\right|$$
$$\le \sup_{A\subset N}\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^\infty k\lambda_k E\left(\left|f_A\left(\sum_{i=1}^n Z_i^{j,k} + k\right) - f_A\left(\sum_{\substack{i=1\\ i\neq j}}^n \tilde{Z}_i^{j,k} + k\right)\right|\right).$$

Finally, from Proposition 3.2 we get that

$$\sup_{A\subset N}\left|f_A\left(\sum_{i=1}^n Z_i^{j,k} + k\right) - f_A\left(\sum_{\substack{i=1\\ i\neq j}}^n \tilde{Z}_i^{j,k} + k\right)\right|$$
$$\le H_1(\lambda_1, \lambda_2, \dots)\left|\sum_{i=1}^n Z_i^{j,k} - \sum_{\substack{i=1\\ i\neq j}}^n \tilde{Z}_i^{j,k}\right|,$$

which completes the proof. □

We now return to the situation considered in Section 3, where η is a stationary Harris recurrent Markov chain, $S_0 \in \mathscr{F}$ and $S_1 \in \mathscr{F}$ are such that $\mu(S_0) > 0$, $\mu(S_1) > 0$ and $S_0 \cap S_1 = \varnothing$, and the variables $\{Z_t;\ t \in Z\}$ are defined as

$$Z_t := I\{\eta_t \in S_0\} N_{S_1,S_0}^{t+1} \qquad \forall\, t \in Z.$$

Let, from now on, for each $j \in \{1,\dots,n\}$ and $k \in Z^+$ such that $P(Z_0 = k) > 0$, the random sequence $\{(\eta_t^{j,k}, \tilde{\eta}_t^{j,k});\ t \in Z\}$ be a coupling of $\mathscr{L}(\eta_t;\ t \in Z)$ and $\mathscr{L}(\eta_t;\ t \in Z | Z_j = k)$, such that

$$\mathscr{L}\left(\eta_t^{j,k};\ t \in Z\right) = \mathscr{L}(\eta_t;\ t \in Z), \qquad \mathscr{L}\left(\tilde{\eta}_t^{j,k};\ t \in Z\right) = \mathscr{L}\left(\eta_t;\ t \in Z | Z_j = k\right).$$

Also, define the random variables $\{Z_t^{j,k};\ t \in Z\}$ and $\{\tilde{Z}_t^{j,k};\ t \in Z\}$ by

$$Z_t^{j,k} := I\{\eta_t^{j,k} \in S_0\} N_{S_1,S_0}^{t+1}(\eta^{j,k}) \qquad \forall\ t \in Z;$$

$$\tilde{Z}_t^{j,k} := I\{\tilde{\eta}_t^{j,k} \in S_0\} N_{S_1,S_0}^{t+1}(\tilde{\eta}^{j,k}) \qquad \forall\ t \in Z.$$

THEOREM 4.2. *Let η be a stationary Harris recurrent Markov chain. Let $S_0 \in \mathscr{F}$ and $S_1 \in \mathscr{F}$ be such that $\mu(S_0) > 0$, $\mu(S_1) > 0$ and $S_0 \cap S_1 = \varnothing$ and let $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ be as in Definition 3.2. Let, for each $j \in \{1, \dots, n\}$ and $k \in Z^+$ such that $P(Z_0 = k) > 0$, the random sequence $(\eta^{j,k}, \tilde{\eta}^{j,k})$ be such a coupling of $\mathscr{L}(\eta_i;\ i \in Z)$ and $\mathscr{L}(\eta_i;\ i \in Z | Z_j = k)$ as described in the preceding paragraph. Then,*

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) \\ &\quad \le \frac{H_1(\lambda_1^*, \lambda_2^*, \dots)}{n} \sum_{j=1}^{n} \sum_{k=1}^{\infty} k \lambda_k^* E\left(\left| \sum_{i=1}^{n} Z_i^{j,k} - \sum_{\substack{i=1 \\ i \ne j}}^{n} \tilde{Z}_i^{j,k} \right| \right) \\ &\qquad + 2P(\tau_{S_1} < \tau_{S_0}). \end{aligned} \tag{4.1}$$

PROOF. The triangle inequality implies

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) \\ &\qquad \le d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathscr{L}(W_n)\big) + d_{\mathrm{TV}}\big(\mathscr{L}(W_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big), \end{aligned}$$

where $W_n := \sum_{i=1}^{n} Z_i$. As noted in the previous section, the first term can be bounded using the basic coupling inequality

$$d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathscr{L}(W_n)\big) \le P(\Psi_n \ne W_n) \le 2P(\tau_{S_1} < \tau_{S_0}).$$

For the second term, Theorem 4.1 can be applied, since it follows from (2.5) that $E(Z_1) = \mu(S_1) < \infty$. □

How should the couplings and the set S_0 be chosen in order to make the bound (4.1) small and/or easily computed? The answer depends on the special structure of the particular Markov chain considered. In Erhardsson (1998b) an example of a "tailor-made" set of couplings is given for a Markov chain η which is a function of an underlying i.i.d. sequence. Below, we derive a rather simple theorem, valid if S_0 is an atom for the Markov chain η [by which we mean, as in Section 2, that $p(\cdot, B)$ is constant on S_0 for each $B \in \mathscr{F}$].

THEOREM 4.3. *Let η be a stationary Harris recurrent Markov chain. Let $S_0 \in \mathscr{F}$ and $S_1 \in \mathscr{F}$, where S_0 is an* atom, *be such that $\mu(S_0) > 0$, $\mu(S_1) > 0$ and $S_0 \cap S_1 = \varnothing$, and let* $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ *be as in Definition* 3.2. *Then,*

$$
\begin{aligned}
&d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) \\
&\quad \le 2H_1(\lambda_1^*, \lambda_2^*, \dots)\left(E\big(\tau_{S_0} Z_0\big) + \mu(S_1)\frac{E\big(\bar{\tau}_{S_0}\big)}{\mu(S_0)}\right) \\
&\qquad \times n\mu(S_1) + 2P(\tau_{S_1} < \tau_{S_0}) \qquad (4.2) \\
&\quad = 2H_1(\lambda_1^*, \lambda_2^*, \dots)\left(E_{S_1}(\tau_{S_0}) + E_{S_1}\big(\tau_{S_0}(\eta^R)\big) + \frac{E\big(\bar{\tau}_{S_0}\big)}{\mu(S_0)}\right) \\
&\qquad \times n\mu(S_1)^2 + 2P(\tau_{S_1} < \tau_{S_0}).
\end{aligned}
$$

PROOF. We use Theorem 4.2 together with certain couplings $\{(\eta_i^{j,k}, \tilde{\eta}_i^{j,k}); i \in Z\}$, which are constructed as follows. First, if S_0 is not a singleton, then replace it by a singleton (without loss of generality, since $S_0 \cap S_1 = \varnothing$). Extend the probability space $(\Omega, \mathscr{G}, P)$ so that it contains, apart from the Markov chain η, a collection of random sequences $\{\tilde{\eta}^k; k \in Z^+, P(Z_0 = k) > 0\}$, which are independent of each other and independent of η, and satisfy $\mathscr{L}(\tilde{\eta}^k) = \mathscr{L}(\eta | Z_0 = k)$. Next, define, for each $k \in Z^+$ such that $P(Z_0 = k) > 0$, the random sequence η^k by

$$
\eta_t^k := \begin{cases} \eta_t, & -\bar{\tau}_{S_0}(\eta^R) \le t \le \bar{\tau}_{S_0}; \\ \tilde{\eta}^k_{t+\bar{\tau}_{S_0}(\eta^R)}, & t < -\bar{\tau}_{S_0}(\eta^R); \\ \tilde{\eta}^k_{t-\bar{\tau}_{S_0}+\tau_{S_0}(\tilde{\eta}^k)}, & t > \bar{\tau}_{S_0}. \end{cases}
$$

For each $k \in Z^+$ such that $P(Z_0 = k) > 0$, it holds that $\mathscr{L}(\eta^k) = \mathscr{L}(\eta)$, since, for each $a \in Z$ and $b \in Z$ such that $a \le 0 \le b$, and each $\{A_t \in \mathscr{F}; t = a, \dots, b\}$,

$$
\begin{aligned}
P\left(\bigcap_{t=a}^{b}\{\eta_t^k \in A_t\}\right) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P\left(\bigcap_{t=a}^{b}\{\eta_t^k \in A_t\}, \bar{\tau}_{S_0}(\eta^R) = i, \bar{\tau}_{S_0} = j\right) \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P\left(\bigcap_{t=-i}^{j}\{\eta_t \in A_t\}, \bar{\tau}_{S_0}(\eta^R) = i, \bar{\tau}_{S_0} = j\right) \\
&\qquad \times P\left(\bigcap_{t=a+i}^{-1}\{\tilde{\eta}_t^k \in A_{t-i}\}, \bigcap_{t=1}^{b-j}\left\{\tilde{\eta}^k_{\tau_{S_0}(\tilde{\eta}^k)+t} \in A_{t+j}\right\}\right) \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P\left(\bigcap_{t=-i}^{j}\{\eta_t \in A_t\}, \bar{\tau}_{S_0}(\eta^R) = i, \bar{\tau}_{S_0} = j\right) \\
&\qquad \times P\left(\bigcap_{t=a+i}^{-1}\{\eta_t \in A_{t-i}\} | \eta_0 \in S_0\right)
\end{aligned}
$$

$$\times P_{S_0}\left(\bigcap_{t=1}^{b-j}\{\eta_t \in A_{t+j}\}\right)$$

$$= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P\left(\bigcap_{t=a}^{b}\{\eta_t \in A_t\}, \bar{\tau}_{S_0}(\eta^R) = i, \bar{\tau}_{S_0} = j\right)$$

$$= P\left(\bigcap_{t=a}^{b}\{\eta_t \in A_t\}\right).$$

Similarly, it can be shown that η^k is independent of $\tau_{S_0}(\tilde{\eta}^k)$. We introduce the obvious notation

$$Z_t^k := I\{\eta_t^k \in S_0\}N_{S_1,S_0}^{t+1}(\eta^k) \qquad \forall\ t \in Z;$$

$$\tilde{Z}_t^k := I\{\tilde{\eta}_t^k \in S_0\}N_{S_1,S_0}^{t+1}(\tilde{\eta}^k) \qquad \forall\ t \in Z$$

and we finally define

$$\eta_t^{j,k} := \eta_{t-j}^k \qquad \forall\ t \in Z;$$

$$\tilde{\eta}_t^{j,k} := \tilde{\eta}_{t-j}^k \qquad \forall\ t \in Z.$$

The definition of the couplings is thus complete. We now evaluate the first term on the right-hand side of (4.1), using these couplings:

$$\frac{H_1(\lambda_1^*, \lambda_2^*, \dots)}{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} k\lambda_k^* E\left(\left|\sum_{i=1}^{n} Z_i^{j,k} - \sum_{\substack{i=1\\ i\neq j}}^{n}\tilde{Z}_i^{j,k}\right|\right)$$

$$= \frac{H_1(\lambda_1^*, \lambda_2^*, \dots)}{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} k\lambda_k^* E\left(\left|\sum_{i=1}^{n} Z_{i-j}^{k} - \sum_{\substack{i=1\\ i\neq j}}^{n}\tilde{Z}_{i-j}^{k}\right|\right)$$

$$\leq \frac{H_1(\lambda_1^*, \lambda_2^*, \dots)}{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} k\lambda_k^* E\left(\sum_{i=-\bar{\tau}_{S_0}(\eta^R)}^{\bar{\tau}_{S_0}-1} Z_i^k + \sum_{i=1-j}^{(1-j+\bar{\tau}_{S_0}(\eta^R)-1)\wedge -1}\tilde{Z}_i^k\right.$$

$$\left. + \sum_{i=(n-j-\bar{\tau}_{S_0}+1)\vee 1}^{n-j}\tilde{Z}_i^k + \sum_{i=n-j-\tau_{S_0}(\bar{\eta}^k)+1}^{n-j} Z_i^k\right)$$

$$= \frac{H_1(\lambda_1^*, \lambda_2^*, \dots)}{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} k\lambda_k^*\left(\mathscr{T}_1^{j,k} + \mathscr{T}_2^{j,k} + \mathscr{T}_3^{j,k} + \mathscr{T}_4^{j,k}\right).$$

We will calculate the terms in this expression one by one. For the first term,

$$\mathscr{T}_1^{j,k} = E\left(\sum_{i=-\bar{\tau}_{S_0}(\eta^R)}^{\bar{\tau}_{S_0}-1} Z_i^k\right) = E\left(\sum_{i=-\bar{\tau}_{S_0}(\eta^R)}^{\bar{\tau}_{S_0}-1} I\{\eta_i \in S_1\}\right)$$

$$= E\left(N_{S_1,S_0}(\eta^R) + N_{S_1,S_0} - I\{\eta_0 \in S_1\}\right)$$

$$= E\left(I\{\eta_0 \in S_0\}(\tau_{S_0} - 1)N_{S_1,S_0}^1\right) \leq E\left(\tau_{S_0}Z_0\right),$$

where we used (2.6) in the last equality. This result and Definition 3.2 imply that

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} k\lambda_k^*\mathscr{T}_1^{j,k} = nE(Z_0)E\big(\tau_{S_0}Z_0\big) = n\mu(S_1)E\big(\tau_{S_0}Z_0\big).$$

For the second term, we use the fact that $\tilde{\eta}^k$ is independent of $\bar{\tau}_{S_0}(\eta^R)$,

$$\begin{aligned}\mathscr{T}_2^{j,k} &= E\left(\sum_{i=1-j}^{(1-j+\bar{\tau}_{S_0}(\eta^R)-1)\wedge -1}\tilde{Z}_i^k\right) = \sum_{r=1}^{\infty}E\left(\sum_{i=1-j}^{(1-j+r-1)\wedge -1}\tilde{Z}_i^k\right)P\big(\bar{\tau}_{S_0}(\eta^R)=r\big)\\ &= \sum_{r=1}^{\infty}\sum_{i=1-j}^{(1-j+r-1)\wedge -1}E\big(Z_i|Z_0=k\big)P\big(\bar{\tau}_{S_0}(\eta^R)=r\big),\end{aligned}$$

which implies that

$$\begin{aligned}\sum_{k=1}^{\infty}k\lambda_k^*\mathscr{T}_2^{j,k} &= \sum_{k=1}^{\infty}k\lambda_k^*\sum_{r=1}^{\infty}\sum_{i=1-j}^{(1-j+r-1)\wedge -1}E\big(Z_i|Z_0=k\big)P\big(\bar{\tau}_{S_0}(\eta^R)=r\big)\\ &= \sum_{r=1}^{\infty}\sum_{i=1-j}^{(1-j+r-1)\wedge -1}\sum_{k=1}^{\infty}k\lambda_k^*E\big(Z_i|Z_0=k\big)P\big(\bar{\tau}_{S_0}(\eta^R)=r\big).\end{aligned}$$

Since also

$$\begin{aligned}&\sum_{i=1-j}^{(1-j+r-1)\wedge -1}\sum_{k=1}^{\infty}k\lambda_k^*E\big(Z_i|Z_0=k\big)\\ &= \sum_{i=1-j}^{(1-j+r-1)\wedge -1}nE(Z_0Z_i)\\ &= \sum_{i=1-j}^{(1-j+r-1)\wedge -1}\sum_{l=1}^{-i}nE\Big(Z_0Z_iI\big\{\tau_{S_0}^{i,1}=l\big\}\Big)\\ &= \sum_{i=1-j}^{(1-j+r-1)\wedge -1}\sum_{l=1}^{-i}nE\Big(Z_iI\big\{\tau_{S_0}^{i,1}=l\big\}\Big)E_{S_0}\big(Z_{-(i+l)}\big)\\ &= \sum_{i=1-j}^{(1-j+r-1)\wedge -1}\sum_{l=1}^{-i}nE\Big(Z_iI\big\{\tau_{S_0}^{i,1}=l\big\}\Big)\frac{E\big(I\{\eta_0\in S_0\}Z_{-(i+l)}\big)}{\mu(S_0)}\\ &\le \sum_{i=1-j}^{(1-j+r-1)\wedge -1}nE(Z_i)\frac{E\big(Z_{-(i+l)}\big)}{\mu(S_0)} \le nr\frac{\mu(S_1)^2}{\mu(S_0)},\end{aligned}$$

we get

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty}k\lambda_k^*\mathscr{T}_2^{j,k} \le n\frac{\mu(S_1)^2}{\mu(S_0)}E\big(\bar{\tau}_{S_0}(\eta^R)\big).$$

For the third term we use the fact that $\tilde{\eta}^k$ is independent of $\bar{\tau}_{S_0}$,

$$\mathscr{T}_3^{j,k} = E\left(\sum_{i=(n-j-\bar{\tau}_{S_0}+1)\vee 1}^{n-j} \tilde{Z}_i^k\right) = \sum_{r=1}^{\infty} E\left(\sum_{i=(n-j-r+1)\vee 1}^{n-j} \tilde{Z}_i^k\right) P\left(\bar{\tau}_{S_0} = r\right)$$

$$= \sum_{r=1}^{\infty} \sum_{i=(n-j-r+1)\vee 1}^{n-j} E\left(Z_i | Z_0 = k\right) P\left(\bar{\tau}_{S_0} = r\right)$$

and calculations similar to those for the second term [see Erhardsson (1997)] give

$$\frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{\infty} k \lambda_k^* \mathscr{T}_3^{j,k} \leq n \frac{\mu(S_1)^2}{\mu(S_0)} E\left(\bar{\tau}_{S_0}\right).$$

Finally, for the fourth term we use the fact that η^k is independent of $\tau_{S_0}(\tilde{\eta}^k)$,

$$\mathscr{T}_4^{j,k} = E\left(\sum_{i=n-j-\tau_{S_0}(\tilde{\eta}^k)+1}^{n-j} Z_i^k\right) = \sum_{r=k+1}^{\infty} E\left(\sum_{i=n-j-r+1}^{n-j} Z_i^k\right) P\left(\tau_{S_0}(\tilde{\eta}^k) = r\right)$$

$$= \sum_{r=k+1}^{\infty} \sum_{i=n-j-r+1}^{n-j} E\left(Z_i\right) P\left(\tau_{S_0} = r | Z_0 = k\right)$$

$$= \mu(S_1) \sum_{r=k+1}^{\infty} r P\left(\tau_{S_0} = r | Z_0 = k\right) = \mu(S_1) E\left(\tau_{S_0} | Z_0 = k\right),$$

implying that

$$\frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{\infty} k \lambda_k^* \mathscr{T}_4^{j,k} = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{\infty} k \lambda_k^* \mu(S_1) E\left(\tau_{S_0} | Z_0 = k\right)$$

$$= \mu(S_1) \sum_{k=1}^{\infty} k \lambda_k^* E\left(\tau_{S_0} | Z_0 = k\right) = n \mu(S_1) E\left(\tau_{S_0} Z_0\right).$$

All that now remains is to show that $E(\bar{\tau}_{S_0}(\eta^R)) = E(\bar{\tau}_{S_0})$, and that

$$E\left(\tau_{S_0} Z_0\right) = \mu(S_1)\left(E_{S_1}(\tau_{S_0}) + E_{S_1}\left(\tau_{S_0}(\eta^R)\right)\right).$$

The first of these assertions holds since

$$E\left(\bar{\tau}_{S_0}(\eta^R)\right) = \sum_{i=0}^{\infty} P\left(\bar{\tau}_{S_0}(\eta^R) > i\right) = \sum_{i=0}^{\infty} P\left(\bar{\tau}_{S_0} > i\right) = E\left(\bar{\tau}_{S_0}\right);$$

the second assertion follows from (2.6), since

$$E\left(\tau_{S_0} Z_0\right) = E\left(I\{\eta_0 \in S_0\} \sum_{i=0}^{\tau_{S_0}-1} \left(\tau_{S_0}^{i,1} + \tau_{S_0}\left((\theta_i \circ \eta)^R\right)\right) I\{\eta_i \in S_1\}\right). \qquad \square$$

We note that when S_1 is an atom, the quantity $E(\tau_{S_0} Z_0)$ in (4.2) can be written in a different form, which is sometimes more tractable.

THEOREM 4.4. *Let η be a stationary Harris recurrent Markov chain, and let $S_0 \in \mathscr{F}$ and $S_1 \in \mathscr{F}$ be such that $\mu(S_0) > 0$, $\mu(S_1) > 0$ and $S_0 \cap S_1 = \emptyset$. Assume also that $P_{(\cdot)}(\tau_{S_1} < \tau_{S_0})$, $E_{(\cdot)}(\tau_{S_1} I\{\tau_{S_1} < \tau_{S_0}\})$ and $E_{(\cdot)}(\tau_{S_0} I\{\tau_{S_0} < \tau_{S_1}\})$ are constant on S_1. Then,*

$$E(\tau_{S_0} Z_0) = \frac{E\big(\tau_{S_1} I\{\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}\}\big) + 2\mu(S_1) E_{S_1}\big(\tau_{S_1} I\{\tau_{S_1} < \tau_{S_0}\}\big)}{P_{S_1}(\tau_{S_0} < \tau_{S_1})} + \frac{\mu(S_1) E_{S_1}\big(\tau_{S_0} I\{\tau_{S_0} < \tau_{S_1}\}\big)}{P_{S_1}(\tau_{S_0} < \tau_{S_1})}.$$

PROOF. The strong Markov property gives

$$\begin{aligned}
&E\big(\tau_{S_0} I\{Z_0 = k\}\big) \\
&\quad = E\Bigg(\Bigg(\tau_{S_1} + \sum_{i=2}^{k} \big(\tau_{S_1}^{0,i} - \tau_{S_1}^{0,i-1}\big) + \big(\tau_{S_0} - \tau_{S_1}^{0,k}\big)\Bigg) \\
&\qquad\qquad \times I\{\eta_0 \in S_0\} \prod_{i=1}^{k} I\big\{\tau_{S_1}^{0,i} < \tau_{S_0}\big\} I\big\{\tau_{S_1}^{0,k+1} > \tau_{S_0}\big\}\Bigg) \\
&\quad = E\big(\tau_{S_1} I\{\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}\}\big) P_{S_1}(\tau_{S_1} < \tau_{S_0})^{k-1} P_{S_1}(\tau_{S_0} < \tau_{S_1}) \\
&\qquad + (k-1) P(\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}) \\
&\qquad\quad \times E_{S_1}\big(\tau_{S_1} I\{\tau_{S_1} < \tau_{S_0}\}\big) P_{S_1}(\tau_{S_1} < \tau_{S_0})^{k-2} P_{S_1}(\tau_{S_0} < \tau_{S_1}) \\
&\qquad + P(\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}) P_{S_1}(\tau_{S_1} < \tau_{S_0})^{k-1} E_{S_1}\big(\tau_{S_0} I\{\tau_{S_0} < \tau_{S_1}\}\big), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k \in Z^+.
\end{aligned}$$

From this we get, using also Lemma 3.1,

$$\begin{aligned}
E(\tau_{S_0} Z_0) &= \sum_{k=1}^{\infty} k E\big(\tau_{S_0} I\{Z_0 = k\}\big) \\
&= \frac{E\big(\tau_{S_1} I\{\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}\}\big) + 2\mu(S_1) E_{S_1}\big(\tau_{S_1} I\{\tau_{S_1} < \tau_{S_0}\}\big)}{P_{S_1}(\tau_{S_0} < \tau_{S_1})} \\
&\quad + \frac{\mu(S_1) E_{S_1}\big(\tau_{S_0} I\{\tau_{S_0} < \tau_{S_1}\}\big)}{P_{S_1}(\tau_{S_0} < \tau_{S_1})}.
\end{aligned}$$

□

REMARK 4.1. As explained in Section 2, any Harris recurrent Markov chain which satisfies the minorization condition (2.4) with $m = 1$ can be embedded in another Harris recurrent Markov chain, which has an atom; therefore, this case is covered by Theorem 4.3. Moreover, a Markov chain η which satisfies (2.4) with $m > 1$ can be embedded in another Markov chain which is wide sense regenerative with 1-dependent cycles (see Section 2). It is not too difficult to see how the couplings in the proof of Theorem 4.3 could be generalized to handle such a process, at the price of added complexity in the

bound. Indeed, it should be possible to generalize Theorem 4.3 to handle any stationary weak sense regenerative (not necessarily Markovian) random sequence with d-dependent cycles ($d \in N$).

REMARK 4.2. Assume that a scaling of the Markov chain η is chosen such that $\liminf_{n\to\infty} \mu(S_0) > 0$, and

$$0 < \liminf_{n\to\infty} n\mu(S_1) \le \limsup_{n\to\infty} n\mu(S_1) < \infty.$$

Using the fact that

$$P(\tau_{S_1} < \tau_{S_0}) = P\big(\bar{\tau}_{S_1} < \bar{\tau}_{S_0}\big) \le \sum_{i=0}^{\infty} P\big(\eta_i \in S_1, \bar{\tau}_{S_0} > i\big)$$

$$= \sum_{i=0}^{\infty} P\big(\eta_0^R \in S_1, \tau_{S_0}(\eta^R) > i\big) = \mu(S_1) E_{S_1}\big(\tau_{S_0}(\eta^R)\big),$$

we get from Theorem 4.3 and Proposition 3.2 that, for some explicit constant $C < \infty$ and n large enough,

$$d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) \le C\Big(E_{S_1}(\tau_{S_0}) + E_{S_1}\big(\tau_{S_0}(\eta^R)\big) + E\big(\bar{\tau}_{S_0}\big)\Big)\mu(S_1),$$

and C can be chosen particularly small if also $\{k\lambda_k^*;\ k \in Z^+\}$ is monotone decreasing towards 0. On the other hand, if it holds that $\liminf_{n\to\infty} \mu(S_0) > 0$, that $\{k\lambda_k^*;\ k \in Z^+\}$ is monotone decreasing towards 0 and that $\limsup_{n\to\infty} P_{S_1}(\tau_{S_1} < \tau_{S_0}) < \frac{1}{4}$ (if S_1 is an atom, then $\frac{1}{4}$ can be replaced by $\frac{1}{2}$ in the last condition), and if also $\lim_{n\to\infty} n\mu(S_1) = \infty$, then we get, using Proposition 3.2 and Theorem 3.2, for some explicit constant $C < \infty$ and n large enough,

$$d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big)$$

$$\le C \log(n\mu(S_1))\Big(E_{S_1}(\tau_{S_0}) + E_{S_1}\big(\tau_{S_0}(\eta^R)\big) + E\big(\bar{\tau}_{S_0}\big)\Big)\mu(S_1).$$

In the latter situation it is natural to look for an approximating normal distribution for $\mathscr{L}(\Psi_n)$ (and an error bound). This will be given in Theorem 6.2.

5. Hitting probabilities and expected hitting times. We now consider the question of how to calculate explicitly the parameters in the approximating compound Poisson distribution of Definition 3.2 and the quantities appearing in the total variation distance upper bound of Theorem 4.3. For this, we may use the following well-known results, which we give here as propositions. Variations of Proposition 5.1 can be found in many places in the literature; see, for example, Proposition 1.8 in Chapter 2 of Revuz (1984). Proposition 5.2 follows from Proposition 5.1 and the Markov property.

PROPOSITION 5.1. *Let p be the transition probability of a Markov chain η, and let $S_\dagger \in \mathscr{F}$ be such that $P_{(\cdot)}(\tau_{S_\dagger} < \infty) \equiv 1$. Then, for any two measurable*

functions $f\colon S_\dagger^c \to R$ *and* $f_\dagger\colon S_\dagger \to R$, *there exists at most one bounded measurable function* $h\colon S \to R$ *satisfying the following conditions*:

$$h(x) = \begin{cases} \int_S h(y)\, p(x, dy) + f(x), & x \in S_\dagger^c; \\ f_\dagger(x), & x \in S_\dagger. \end{cases}$$

(*If* $f \equiv 0$, h is called *p-harmonic outside* $S_\dagger$; otherwise h is said to satisfy *Poisson's equation with charge f outside* $S_\dagger$.)

PROPOSITION 5.2. *Let* η *be a stationary Harris recurrent Markov chain with transition probability p. Let* $A \in \mathscr{F}$ *and* $B \in \mathscr{F}$ *be such that* $\mu(A) > 0$, $\mu(B) > 0$ *and* $A \cap B = \varnothing$. *Then*:

(i) *The function* $P_{(\cdot)}(\bar{\tau}_B < \bar{\tau}_A)\colon S \to [0,1]$ *is the unique bounded measurable function on S which satisfies*

$$P_x(\bar{\tau}_B < \bar{\tau}_A) = \begin{cases} \int_S P_y(\bar{\tau}_B < \bar{\tau}_A) p(x, dy), & x \in (A \cup B)^c, \\ I_B(x), & x \in A \cup B. \end{cases}$$

(ii) *For each fixed* $s \in (0,1)$, *the function* $E_{(\cdot)}(s^{N_{B,A}})\colon S \to [0,1]$ *is the unique bounded measurable function on S which satisfies*

$$E_x(s^{N_{B,A}}) = \begin{cases} \int_S E_y(s^{N_{B,A}})\, p(x, dy), & x \in (A \cup B)^c, \\ s\int_S E_y(s^{N_{B,A}})\, p(x, dy), & x \in B, \\ 1, & x \in A. \end{cases}$$

(iii) *The function* $E_{(\cdot)}(\bar{\tau}_A)\colon S \to [0, \infty)$ *satisfies*

$$E_x(\bar{\tau}_A) = \begin{cases} \int_S E_y(\bar{\tau}_A)\, p(x, dy) + 1, & x \in A^c, \\ 0, & x \in A. \end{cases}$$

Moreover, if $E_{(\cdot)}(\bar{\tau}_A)$ *is bounded, then it is the unique bounded measurable function which satisfies this equation.*

Using Proposition 5.2, provided that the corresponding Poisson's equations can be solved, we can calculate explicitly the quantities appearing in the total variation distance error bound of Theorem 4.3 and the generating function for $\mathscr{L}(Z_0)$. When the state space S is finite, these equations reduce to systems of linear equations of dimensions at most $|S_0^c|$. In Section 7 we give some numerical examples of the latter case.

6. Poisson, normal and exponential approximation. In some situations, it might be of interest to consider other approximating distributions for

$\mathscr{L}(\Psi_n)$ than the compound Poisson distribution $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$. A simple choice is the $\mathrm{Po}(n\mu(S_1))$ distribution, or some compound Poisson distribution with easily calculated parameters. To find an upper bound for $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{Po}(n\mu(S_1)))$, one can use some variation on the Stein–Chen method for Poisson approximation; see the introduction and the references given there. However, the next theorem tells us that if a bound for $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots))$ can be explicitly calculated, then we can also find upper and lower bounds for $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1, \lambda_2, \dots))$, where $\mathrm{CP}(\lambda_1, \lambda_2, \dots)$ is an arbitrary compound Poisson distribution. [For another upper bound, derived in a different way, see Lemma 2.4.2 in Roos (1993).]

THEOREM 6.1. *Let η be a stationary Harris recurrent Markov chain. Let $S_0 \in \mathscr{F}$ and $S_1 \in \mathscr{F}$ be such that $\mu(S_0) > 0$, $\mu(S_1) > 0$, and $S_0 \cap S_1 = \varnothing$, and let $\mathrm{CP}(\lambda_1, \lambda_2, \dots)$ be any compound Poisson distribution. Then,*

$$\begin{aligned}
&\left|\exp\bigl(-n\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})\bigr) - \exp(-\lambda)\right| - d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)) \\
&\quad \le d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1, \lambda_2, \dots)) \\
&\quad \le 1 - \exp\left(-\sum_{k=1}^{\infty} |\lambda_k - \lambda_k^*|\right) + d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)).
\end{aligned}$$

In particular,

$$\begin{aligned}
&n\mu(S_1)P_{S_1}(\tau_{S_1} < \tau_{S_0})\exp(-n\mu(S_1)) - d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)) \\
&\quad \le d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{Po}(n\mu(S_1))) \\
&\quad \le 3n\mu(S_1)P_{S_1}(\tau_{S_1} < \tau_{S_0}) + d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)).
\end{aligned}$$

PROOF. The triangle inequality gives

$$\begin{aligned}
&d_{\mathrm{TV}}(\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots), \mathrm{CP}(\lambda_1, \lambda_2, \dots)) - d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)) \\
&\quad \le d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1, \lambda_2, \dots)) \\
&\quad \le d_{\mathrm{TV}}(\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots), \mathrm{CP}(\lambda_1, \lambda_2, \dots)) + d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)).
\end{aligned}$$

We want to find an upper and a lower bound for $d_{\mathrm{TV}}(\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots), \mathrm{CP}(\lambda_1, \lambda_2, \dots))$. For the upper bound, let independent random variables $\{N_k;\ k \in Z^+\}$ and $\{N_k';\ k \in Z^+\}$ be defined on the same probability space, with distributions $N_k \sim \mathrm{Po}(\lambda_k \wedge \lambda_k^*)$ and $N_k' \sim \mathrm{Po}(|\lambda_k - \lambda_k^*|)$ for each $k \in Z^+$. Define also the index set $\Gamma \subset Z^+$ by $\Gamma := \{k \in Z^+;\ \lambda_k \ge \lambda_k^*\}$. According to Definition 3.1, it holds that

$$\mathscr{L}\left(\sum_{k=1}^{\infty} kN_k + \sum_{k\in\Gamma} kN_k'\right) = \mathrm{CP}(\lambda_1, \lambda_2, \dots),$$

$$\mathscr{L}\left(\sum_{k=1}^{\infty} kN_k + \sum_{k\in\Gamma^c} kN_k'\right) = \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots).$$

Using these identities and the basic coupling inequality we get

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots), \mathrm{CP}(\lambda_1, \lambda_2, \dots)\big) \\ &\quad \le P\bigg(\sum_{k=1}^{\infty} kN_k + \sum_{k\in\Gamma} kN_k' \ne \sum_{k=1}^{\infty} kN_k + \sum_{k\in\Gamma^c} kN_k'\bigg) \\ &\quad \le P\bigg(\sum_{k=1}^{\infty} kN_k' > 0\bigg) = P\bigg(\sum_{k=1}^{\infty} N_k' > 0\bigg) = 1 - \exp\bigg(-\sum_{k=1}^{\infty} |\lambda_k - \lambda_k^*|\bigg). \end{aligned}$$

For the lower bound, assuming that $W_1 \sim \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ and $W_2 \sim \mathrm{CP}(\lambda_1, \lambda_2, \dots)$, we get, from the definition of total variation distance,

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots), \mathrm{CP}(\lambda_1, \lambda_2, \dots)\big) \\ &\quad = \sup_{A\subset N} |P(W_1 \in A) - P(W_2 \in A)| \\ &\quad \ge |P(W_1 = 0) - P(W_2 = 0)| \\ &\quad = |\exp\big(-n\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})\big) - \exp(-\lambda)|. \end{aligned}$$

The second part of the theorem follows from the fact that if $\lambda_1 = \lambda = n\mu(S_1)$ and $\lambda_k = 0$ for $k = 2, 3, 4, \dots,$ then,

$$\begin{aligned} &1 - \exp\bigg(-\sum_{k=1}^{\infty} |\lambda_k - \lambda_k^*|\bigg) \\ &\quad = 1 - \exp\big(-n(\mu(S_1) - P(Z_0 = 1) + P(Z_0 > 1))\big) \\ &\quad = 1 - \exp\big(-n(\mu(S_1) - P(Z_0 > 0) + 2P(Z_0 > 1))\big) \\ &\quad = 1 - \exp\big(-n(\mu(S_1)P_{S_1}(\tau_{S_1} < \tau_{S_0}) + 2P(Z_0 > 1))\big) \\ &\quad \le 1 - \exp\big(-3n\mu(S_1)P_{S_1}(\tau_{S_1} < \tau_{S_0})\big), \end{aligned}$$

where we used that, according to Lemma 3.1,

$$\begin{aligned} P(Z_0 > 1) &= P\big(\eta_0 \in S_0, \tau_{S_1}^{0,2} < \tau_{S_0}\big) = P\big(\eta_0 \in S_1, \tau_{S_1} < \tau_{S_0} < \tau_{S_1}^{0,2}\big) \\ &\le \mu(S_1)P_{S_1}(\tau_{S_1} < \tau_{S_0}), \end{aligned}$$

and from the mean value theorem. □

REMARK 6.1. Assume that a scaling of the Markov chain η is chosen such that

$$0 < a_0 := \liminf_{n\to\infty} n\mu(S_1) \le a_1 := \limsup_{n\to\infty} n\mu(S_1) < \infty.$$

According to Theorem 6.1, if

$$\limsup_{n\to\infty} \frac{d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big)}{P_{S_1}(\tau_{S_1} < \tau_{S_0})} < a_0 \exp(-a_1), \tag{6.1}$$

then, for some explicit constants $0 < C \le C' < \infty$ and n large enough,

$$CP_{S_1}(\tau_{S_1} < \tau_{S_0}) \le d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{Po}(n\mu(S_1))\big) \le C'P_{S_1}(\tau_{S_1} < \tau_{S_0}).$$

This result quantifies the essential drawback of the first-order Poisson approximation for $\mathscr{L}(\Psi_n)$: the approximation error is too large if the visits to S_1 have a strong tendency to occur in clumps. The quantity $P_{S_1}(\tau_{S_1} < \tau_{S_0})$ is a natural measure of this tendency.

REMARK 6.2. Let η be a stationary random sequence on the state space $(R, \mathscr{B}(R))$, with one-dimensional distribution μ. Assume that for each $\tau > 0$ there exists a sequence $\{u_n;\ n \in Z^+\}$ such that $\lim_{n\to\infty} n\mu(S_1) = \tau$, where $S_1 := \{x \in R;\ x > u_n\}$. In classical extreme value theory, such a random sequence is said to have extremal index $\gamma \in [0, 1]$ if

$$\begin{aligned} &\lim_{n\to\infty} P\big(\max\{\eta_i;\ i = 1, \dots, n\} \notin S_1\big) \\ &\qquad = \lim_{n\to\infty} P(\Psi_n = 0) = \exp(-\gamma\tau) \qquad \forall\, \tau > 0; \end{aligned} \tag{6.2}$$

see Section 2.2 in Leadbetter and Rootzén (1988). If η has extremal index $\gamma = 1$ and satisfies a certain mixing condition, then

$$\lim_{n\to\infty} d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{Po}(\tau)\big) = 0 \qquad \forall\, \tau > 0. \tag{6.3}$$

Assume that η is a stationary Harris recurrent Markov chain which has an atom S_0 such that $\mu(S_0) > 0$ and $S_0 \cap S_1 = \varnothing$. In this case (6.2) holds if and only if

$$\lim_{n\to\infty} P_{S_1}(\tau_{S_0} < \tau_{S_1}) = \gamma \qquad \forall\, \tau > 0; \tag{6.4}$$

see Section 2.6 in Leadbetter and Rootzén (1988). If (6.4) holds with $\gamma = 1$ and if

$$\begin{aligned} \limsup_{n\to\infty} &\frac{d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) + |n\mu(S_1) - \tau|}{P_{S_1}(\tau_{S_1} < \tau_{S_0})} \\ &< \tau \exp(-\tau) \qquad \forall\, \tau > 0, \end{aligned}$$

then (6.3) holds with the same rate of convergence as in (6.4). This follows from the same argument as in Remark 6.1.

We next consider the case when the approximating compound Poisson distribution $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ is such that $\lambda^* = n\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})$ is large. It is then reasonable to expect $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ to be close to a normal distribution. To quantify this assertion, we define m_r for each $r \in Z^+$ as the rth moment of the compounding distribution of $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$; that is,

$$m_r := E\big(Z_0^r | Z_0 > 0\big) = \sum_{k=1}^{\infty} k^r \frac{\lambda_k^*}{\lambda^*} \qquad \forall\, r \in Z^+.$$

THEOREM 6.2. *Let η be a stationary Harris recurrent Markov chain such that $m_3 < \infty$. Then, the following Berry–Esseen theorem holds*:

$$\sup_{x \in R} \left| P\left(\frac{\Psi_n - n\mu(S_1)}{\sqrt{(m_2/m_1)n\mu(S_1)}} < x \right) - \Phi(x) \right| \leq d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) + \frac{0.8 m_3 m_1^{1/2}}{m_2^{3/2}\sqrt{n\mu(S_1)}}.$$

PROOF. Definitions 3.1 and 3.2 tell us that $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots) = \mathscr{L}(\sum_{i=1}^M T_i)$, where the variables $\{T_i;\ i \in Z^+\}$ and M are independent, $P(T_i = k) = P(Z_0 = k | Z_0 > 0) = \lambda_k^*/\lambda^*$ for each $k \in Z^+$ and each $i \in Z^+$, and $M \sim \mathrm{Po}(\lambda^*)$. It holds that

$$E\left(\sum_{i=1}^M T_i \right) = m_1 \lambda^* = \sum_{k=1}^\infty k \lambda_k^* = n\mu(S_1),$$

$$\mathrm{Var}\left(\sum_{i=1}^M T_i \right) = m_2 \lambda^* = \frac{m_2}{m_1} n\mu(S_1),$$

and the triangle inequality gives

$$\sup_{x \in R} \left| P\left(\frac{\Psi_n - n\mu(S_1)}{\sqrt{(m_2/m_1)n\mu(S_1)}} < x \right) - \Phi(x) \right| \leq d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) + \sup_{x \in R} \left| P\left(\frac{\sum_{i=1}^M T_i - n\mu(S_1)}{\sqrt{(m_2/m_1)n\mu(S_1)}} < x \right) - \Phi(x) \right|.$$

The second term on the right-hand side in this expression can be bounded using Theorem 1 in Michel (1993), which yields the desired result. □

In the (easily derived) last theorem of this section, we give error bounds for two related kinds of approximations. First, if $\mathscr{L}(\Psi_n)$ can be approximated by $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ with good accuracy, then it is reasonable to expect the quantity τ_{S_1} to be approximately exponentially distributed; see, for example, Section 5.1.5 in Kalashnikov (1994). Secondly, in some situations (e.g., in extreme value theory) it is natural to define a strictly decreasing sequence of "rare" sets $\{S_1^{(m)};\ m \in N\}$ and ask for an approximating distribution for $\mathscr{L}(M_n)$, where $M_n := \max\{m \in N;\ \Psi_n^{(m)} > 0\}$.

THEOREM 6.3. *Let η be a stationary Harris recurrent Markov chain. Let S_0 and S_1 be such that $\mu(S_0) > 0$, $\mu(S_1) > 0$ and $S_0 \cap S_1 = \emptyset$. Then,*

$$\begin{aligned}\Big|P(\tau_{S_1} > x) - \exp\big(-x\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})\big)\Big| \\ \le d_{\mathrm{TV}}\big(\mathscr{L}(\Psi_{[x]}), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) \\ + \mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})\exp\big(-[x]\,\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})\big) \qquad \forall\, x \in [1, \infty).\end{aligned}$$

Also,

$$\begin{aligned}\Big|P(M_n \le x) - \exp\Big(-n\mu\big(S_1^{([x]+1)}\big)P_{S_1^{([x]+1)}}(\tau_{S_0} < \tau_{S_1^{([x]+1)}})\Big)\Big| \\ \le d_{\mathrm{TV}}\big(\mathscr{L}\big(\Psi_n^{([x]+1)}\big), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)\big) \qquad \forall\, x \in [0, n).\end{aligned}$$

PROOF. Use the triangle inequality, the mean value theorem and the following facts:

$$\begin{aligned}\{\tau_{S_1} > x\} &= \big\{\tau_{S_1} > [x]\big\} = \big\{\Psi_{[x]} = 0\big\} \qquad \forall\, x \in [1, \infty), \\ \{M_n \le x\} &= \big\{M_n \le [x]\big\} = \big\{\Psi_n^{([x]+1)} = 0\big\} \qquad \forall\, x \in [0, n).\end{aligned}$$

□

7. Numerical examples. In this section we calculate numerically the parameter values of the approximating compound Poisson distribution $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$, and the value of the bound for $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots))$ given in Theorem 4.3, for some examples where η is an irreducible Markov chain on a finite state space S. From Section 2, we know that η is then Harris recurrent and that any singleton in S is a regeneration set. In the examples that we consider, we take $S = \{1, 2, \dots, 8\}$. We define a transition matrix q on S by

$$q := \begin{pmatrix} 0.334 & 0.215 & 0.173 & 0.119 & 0.065 & 0.086 & 0.003 & 0.005 \\ 0.289 & 0.133 & 0.211 & 0.133 & 0.067 & 0.156 & 0.007 & 0.004 \\ 0.356 & 0.184 & 0.075 & 0.043 & 0.151 & 0.183 & 0.002 & 0.006 \\ 0.41 & 0.162 & 0.108 & 0.075 & 0.14 & 0.097 & 0.005 & 0.003 \\ 0.316 & 0.239 & 0.044 & 0.218 & 0.076 & 0.098 & 0.004 & 0.005 \\ 0.44 & 0.176 & 0.044 & 0.242 & 0.088 & 0 & 0.005 & 0.005 \\ 0.18 & 0.06 & 0.19 & 0.09 & 0.13 & 0.1 & 0.13 & 0.12 \\ 0.2 & 0.16 & 0.07 & 0.1 & 0.14 & 0.1 & 0.09 & 0.14 \end{pmatrix},$$

and using q we define the transition matrix p of the Markov chain η in the following way:

$$p(i,j) := \begin{cases} \dfrac{1 - \alpha q(i, S_1)}{1 - q(i, S_1)} q(i,j), & i \in S_1^c,\ j \in S_1^c, \\ \alpha q(i,j), & i \in S_1^c,\ j \in S_1, \\ q(i,j), & i \in S_1,\ j \in S, \end{cases}$$

TABLE 1
Parameter values for the first example

α	1	0.5	0.25	0.1	0.01	0.001
$\mu(S_1)$	0.006140	0.003063	0.001529	6.113×10^{-4}	6.110×10^{-5}	6.109×10^{-6}
θ	0.8386	0.8492	0.8546	0.8578	0.8598	0.8600

where $0 \leq \alpha \leq 1$, and $S_1 \subset S$ is a "rare" set. In the first example we take $S_1 = \{8\}$ and $S_0 = \{1\}$. In this case, from Remark 3.1, it follows that $\mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots)$ is the Pólya–Aeppli(λ^*, θ) distribution, with parameters $\theta := P_{S_1}(\tau_{S_0} < \tau_{S_1})$ and $\lambda^* = n\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})$; the compounding distribution is geometric with mean θ^{-1}. The values of these parameters can be explicitly calculated simply by solving linear equation systems (see Section 5). Using Maple to perform these calculations for some different values of α, we get the values in Table 1.

As a bound for the total variation distance $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots))$, we may use (4.2) in Theorem 4.3. All quantities appearing in this bound can be calculated by solving linear equation systems (with dimension at most $|S_0^c| = 7$). Also, Theorem 3.1 implies that, since $P_{S_1}(\tau_{S_1} < \tau_{S_0}) \leq \frac{1}{2}$, we may use the good upper bound (3.3) for the "magic" factor $H_1(\lambda_1^*, \lambda_2^*, \dots)$. For the same values of α as previously considered and some typical values of n, we get the values in Table 2.

In the second example, we consider the same Markov chain η as above. We now, however, define $\alpha' := 2\alpha$, and we take $S_1 = \{7, 8\}$. In this case $\mathrm{CP}\{\lambda_1^*, \lambda_2^*, \dots)$ is no longer a Pólya–Aeppli distribution, but it still holds that $\lambda^* = n\mu(S_1)\theta = n\mu(S_1)P_{S_1}(\tau_{S_0} < \tau_{S_1})$ and that the mean of the compounding distribution is θ^{-1}. Furthermore, the generating function g of the compounding distribution is rational and can be computed as the solution of a linear equation system. Calculating the parameters as before, for some different values of $\alpha' = 2\alpha$, we get the values in Table 3.

As a bound for the total variation distance $d_{\mathrm{TV}}(\mathscr{L}(\Psi_n), \mathrm{CP}(\lambda_1^*, \lambda_2^*, \dots))$, we again use (4.2) in Theorem 4.3. Theorem 3.1 again implies that we may use the good upper bound (3.3) for the "magic" factor $H_1(\lambda_1^*, \lambda_2^*, \dots)$ [since condi-

TABLE 2
Values of the bound (4.2) *for the first example*

		α					
		1	0.5	0.25	0.1	0.01	0.001
n	10	0.04106	0.01831	0.008611	0.003315	3.238×10^{-4}	3.231×10^{-5}
	10^2	0.1215	0.03816	0.01354	0.004101	3.317×10^{-4}	3.238×10^{-5}
	10^3	0.5982	0.2004	0.06171	0.01196	4.100×10^{-4}	3.317×10^{-5}
	10^4	1.279	0.5089	0.2013	0.05584	0.001194	4.100×10^{-5}
	10^5	1.980	0.8361	0.3586	0.1164	0.005545	1.194×10^{-4}

TABLE 3
Parameter values for the second example

α'	**1**	**0.5**	**0.25**	**0.1**	**0.01**	**0.001**
$\mu(S_1)$	0.005796	0.002906	0.001455	5.826×10^{-4}	5.829×10^{-5}	5.830×10^{-6}
θ	0.7534	0.7570	0.7588	0.7599	0.7606	0.7606

TABLE 4
Values of the bound (4.2) *for the second example*

		α'					
		1	**0.5**	**0.25**	**0.1**	**0.01**	**0.001**
n	10	0.03616	0.01616	0.007598	0.002924	2.855×10^{-4}	2.848×10^{-5}
	10^2	0.1079	0.03410	0.01208	0.003642	2.927×10^{-4}	2.855×10^{-5}
	10^3	0.6763	0.2075	0.05695	0.01082	3.645×10^{-4}	2.927×10^{-5}
	10^4	1.714	0.6829	0.2621	0.06649	0.001082	3.645×10^{-5}
	10^5	2.810	1.214	0.5210	0.1657	0.006637	1.082×10^{-4}

tion (3.7) is satisfied]. Computing this bound for the second example, for the same values of α' as previously considered, and some typical values of n, we get the values in Table 4.

We finally note that, using Maple, it is not difficult to show that if n is chosen so that $\lim_{\alpha \to 0} n\mu(S_1) = c > 0$, then, in both of the examples above, the total variation distance bound divided by α converges to a constant as $\alpha \to 0$. For c values 0.1, 1 and 10, the values of this limiting constant are, in the first example, 0.04655, 0.1748 and 0.6779, respectively; in the second example, 0.08431, 0.3306 and 1.763, respectively.

Acknowledgment. I thank my advisor, Professor Lars Holst, for originally introducing me to the Stein–Chen method and for valuable discussions.

REFERENCES

ALDOUS, D. (1989). *Probability Approximations Via the Poisson Clumping Heuristic*. Springer, New York.

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Probab.* **17** 9–25.

ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, Chichester.

BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25** 175–184.

BARBOUR, A. D., CHEN, L. H. Y. and LOH, W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20** 1843–1866.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press.

CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

ERHARDSSON, T. (1996). On the number of high excursions of linear growth processes. *Stochastic Process. Appl.* **65** 31–53.

ERHARDSSON, T. (1997). Compound Poisson approximation for Markov chains. Ph.D. thesis, Royal Institute of Technology, Stockholm.

ERHARDSSON, T. (1998a). Compound Poisson approximation for pattern counts and birth-death chains. Unpublished manuscript.

ERHARDSSON, T. (1998b). Compound Poisson approximation for the Johnson–Mehl model. Unpublished manuscript.

JOHNSON, N. L. and KOTZ, S. (1969). *Distributions in Statistics*: *Discrete Distributions*. Wiley, New York.

KALASHNIKOV, V. (1994). *Topics on Regenerative Processes*. CRC Press, Boca Raton, FL.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

LEADBETTER, M. R. and ROOTZÉN, H. (1988). Extremal theory for stochastic processes. *Ann. Probab.* **16** 431–478.

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

MICHEL, R. (1993). On Berry–Esseen results for the compound Poisson distribution. *Insurance Math. Econom.* **13** 35–37.

NUMMELIN, E. (1984). General irreducible Markov chains and non-negative operators. Cambridge Univ. Press.

REVUZ, D. (1984). *Markov Chains*, 2nd ed. North-Holland, Amsterdam.

ROLSKI, T. (1981). Stationary random processes associated with point processes. *Lecture Notes in Statist.* **5** 18–27. Springer, New York.

ROOS, M. (1993). Stein–Chen method for compound Poisson approximation. Ph.D. thesis, Univ. Zürich.

ROOS, M. (1994). Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Probab.* **4** 1177–1187.

ROOS, M. and STARK, D. (1996). Compound Poisson approximation of the number of visits to a small set in a Markov chain. Unpublished manuscript.

ROOTZÉN, H. (1988). Maxima and exceedances of stationary Markov chains. *Adv. in Appl. Probab.* **20** 371–390.

STEIN, C. M. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 583–602. Univ. California Press, Berkeley.

DEPARTMENT OF MATHEMATICS
ROYAL INSTITUTE OF TECHNOLOGY
S-100 44 STOCKHOLM
SWEDEN
E-MAIL: ter@math.kth.se