The Annals of Applied Probability
2000, Vol. 10, No. 2, 573–591

COMPOUND POISSON APPROXIMATION FOR COUNTS OF RARE PATTERNS IN MARKOV CHAINS AND EXTREME SOJOURNS IN BIRTH–DEATH CHAINS

BY TORKEL ERHARDSSON

Royal Institute of Technology, Stockholm

We consider the number of overlapping occurrences up to a fixed time of one or several "rare" patterns in a stationary finite-state Markov chain. We derive a bound for the total variation distance between the distribution of this quantity and a compound Poisson distribution, using general results on compound Poisson approximation for Markov chains by Erhardsson. If the state space is $\{0,1\}$ and the pattern is a *head run* $(111\cdots111)$, the bound is completely explicit and improves on an earlier bound given by Geske, Godbole, Schaffner, Skolnick and Wallstrom. In general, the bound can be computed by solving five linear equation systems of dimension at most the number of states plus the sum of the lengths of the patterns. We also give approximations with error bounds for the distributions of the first occurrence time of a head run of fixed length and the longest head run occurring up to a fixed time. Finally, we consider the sojourn time in an "extreme" subset of the state space by a stationary birth–death chain and derive a bound for the total variation distance between the distribution of this quantity and a compound Poisson distribution.

1. Introduction. In this paper we consider two random quantities which both count the number of certain "rare" events occurring in Markov chains. In both cases, our objective is to construct approximating compound Poisson distributions for the distributions of these quantities and explicit error bounds for the approximations. The first quantity is the number of overlapping occurrences up to a fixed time of one or several "rare" patterns in a stationary finite-state Markov chain. The second quantity is the sojourn time in an "extreme" part of the state space by a stationary birth–death chain. A short background is first given.

Occurrences of finite sequences ("patterns," or "words") in a Markov chain on a finite state space ("alphabet") have been studied from various aspects for a long time and by many people. A complete list of references would be very long; for just a few examples, see Chapter B of Aldous (1989), Gordon, Schilling and Waterman (1986), Godbole and Papastavridis (1994), Wang and Ji (1995), Stefanov and Pakes (1997) and the references in these. A number of results have been obtained, including moments, distributional transforms,

Received September 1998; revised August 1999.

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

Key words and phrases. Compound Poisson approximation, error bound, stationary Markov chain, counts of rare patterns, head run, extreme sojourns, birth–death chain.

or limit theorems, for quantities like the following:

1. The first occurrence time of a pattern.
2. The length of the longest pattern of a specified kind occurring up to a fixed time.
3. The number of overlapping or nonoverlapping occurrences of one or several patterns up to a fixed time.
4. The number of occurrences of one or several patterns up to the first occurrence of a particular pattern.

A typical area of application for these results is the analysis of nucleotide sequences in DNA molecules. Sometimes attention is restricted to the cases when the Markov chain is an i.i.d. sequence, or the state space of the chain is $S = \{0, 1\}$ (or both). In the latter case we speak about *coin-tossing* problems, and the particular pattern $111 \cdots 111$ is called a *head run*.

In the present paper, which is based on Chapter 4 of Erhardsson (1997), we consider the number of overlapping occurrences up to a fixed time in the chain of one or several "rare" patterns. We derive an explicit bound for the total variation distance between the distribution of this quantity and a compound Poisson distribution. Why is this a natural thing to do? In view of the rarity of the pattern(s) which we are counting, the first idea would be to find an approximating Poisson distribution, or a Poisson limit theorem. However, patterns with self-overlap (i.e., the end part of the pattern is identical to the initial part) will tend to occur in clumps, which could make an approximation with a Poisson distribution quite poor. A properly chosen compound Poisson distribution is in such cases often a much better alternative.

Among earlier related results, Wang and Ji (1995) derive compound Poisson limit theorems for the number of overlapping occurrences of patterns, while approximating compound Poisson distributions and total variation distance error bounds are given in Geske, Godbole, Schaffner, Skolnick and Wallstrom (hereafter GGSSW) (1995), Schbath (1995), and Roos and Stark (1996). The three latter papers all use some variation on *Stein's method* for compound Poisson approximation to derive the error bounds; for an introduction to Stein's method, see Barbour (1997). GGSSW and Schbath both use Stein's method for discrete Poisson process approximation, an approach which is described in Section 10.4 in Barbour, Holst and Janson (1992). Roos and Stark use the method in Roos (1994), which is based on the compound Poisson Stein equation derived in Barbour, Chen and Loh (1992).

Here, we use the results in Erhardsson (1999), which are also based on the compound Poisson Stein equation, but along different lines than in Roos (1994). It will be shown that an approximating compound Poisson distribution and a total variation distance error bound can be found for the number of head runs, with very little computational effort, by applying Theorem 4.3 in Erhardsson (1999) to a certain embedded Markov chain. The error bound is a significant improvement on the bound in GGSSW (1995) in the case when the expected number of head runs is large; hence, we confirm a conjecture by GGSSW that such an improvement should be possible. Using the related but more general Theorem 5.1 in Erhardsson (2000), we also give approximations with error bounds for the distributions of the first occurrence time of a head run of fixed length and the longest head run occurring up to a fixed time. We then show how the argument used for head runs can be extended to counts of general patterns, giving error bounds which can be computed by solving five linear equation systems of dimension at most the number of states plus the sum of the lengths of the patterns.

The second quantity considered in this paper is the sojourn time in an "extreme" part of the state space up to a fixed time by a stationary birth–death chain. Birth–death chains are reversible and have well-known explicit expressions for probabilities of hitting a subset of the state space before hitting another subset, and for expected first hitting times of subsets. Hence, applying the results in Erhardsson (1999) is rather easy. This enables us to extend some limit theorems given by Serfozo (1980) and Berman (1986) by giving total variation distance error bounds for the corresponding approximations.

The rest of the paper is organized as follows. In Section 2 we give some notation and definitions. In Section 3, we give an approximating compound Poisson distribution and a total variation distance error bound for the number of head runs. In Section 4 we derive approximations and error bounds for the first occurrence time of a head run and the longest head run. In Section 5 we show how the results for head runs can be generalized to the number of occurrences of one or several arbitrary patterns in a finite-state Markov chain. Finally, in Section 6 we give an approximating compound Poisson distribution and an error bound for the sojourn time in an "extreme" subset of the state space by a stationary birth–death chain.

2. Preliminaries. We use the following notation for sets of numbers: $R =$ the real numbers, $Z =$ the integers, $R_+ = [0, \infty)$, $Z_+ = \{0, 1, 2, \ldots\}$ and $Z'_+ = \{1, 2, \ldots\}$.

For any random element X in some measurable space $(S, \mathscr{F})$, we denote the distribution of X by $\mathscr{L}(X)$.

For any measurable space $(S, \mathscr{F})$, we denote by $(S^Z, \mathscr{F}^Z)$ the space of all functions $f\colon Z \to S$, equipped with the σ-algebra generated by the cylinder sets. A random element in $(S^Z, \mathscr{F}^Z)$ is called a *random sequence*. We define, for each $A \in \mathscr{F}$, $t \in Z$, $k \in Z_+$ and $s = \{\ldots, x_{-1}, x_0, x_1, \ldots\} \in S^Z$, the functional $\tau_A^{t,k}\colon S^Z \to Z_+$ by

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

and we define the functional $\bar{\tau}_A^{t,k}\colon S^Z \to Z_+$ by

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

For brevity we will use the notation $\tau_A^t(\cdot) := \tau_A^{t,1}(\cdot)$, $\tau_A(\cdot) := \tau_A^{0,1}(\cdot)$, $\bar{\tau}_A^t(\cdot) := \bar{\tau}_A^{t,1}(\cdot)$ and $\bar{\tau}_A(\cdot) := \bar{\tau}_A^{0,1}(\cdot)$.

Throughout the paper, η denotes a two-sided stationary Markov chain (in discrete time) on a finite state space $(S, \mathscr{F})$ (where $\mathscr{F}$ is the power set of S). Hence, η is a random element in $(S^Z, \mathscr{F}^Z)$. For brevity, we define $P_A(\eta \in \cdot) := P(\eta \in \cdot \mid \eta_0 \in A)$ and $E_A(f(\eta)) := E(f(\eta)|\eta_0 \in A)$ for each $A \in \mathscr{F}$ and each measurable $f\colon S^Z \to R_+$; if $A = \{x\}$ is a singleton, we write $P_x(\eta \in \cdot) := P_{\{x\}}(\eta \in \cdot)$ and $E_x(f(\eta)) := E_{\{x\}}(f(\eta))$. For each $A \in \mathscr{F}$, $t \in Z$ and $k \in Z_+$, we use for brevity the notation $\tau_A^{t,k} := \tau_A^{t,k}(\eta)$, $\tau_A^t := \tau_A^t(\eta)$, $\tau_A := \tau_A(\eta)$, $\bar{\tau}_A^{t,k} := \bar{\tau}_A^{t,k}(\eta)$, $\bar{\tau}_A^t := \bar{\tau}_A^t(\eta)$ and $\bar{\tau}_A := \bar{\tau}_A(\eta)$. We denote by η^R the *reverse* Markov chain of η, that is, the random element in $(S^Z, \mathscr{F}^Z)$ defined by $\eta_t^R := \eta_{-t}$ for each $t \in Z$. It is well known that η^R is also a stationary Markov chain. If η has transition matrix p, then the transition matrix of η^R will be denoted by p^R.

For any two probability measures ν_1 and ν_2 on any measurable space $(S, \mathscr{F})$ we define the *total variation distance* $d_{\mathrm{TV}}(\nu_1, \nu_2)$ in the usual way as

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

We denote by $\mathrm{CP}(\lambda_1, \lambda_2, \dots)$ the *compound Poisson* distribution with parameters $\{\lambda_k; k \in Z'_+\}$, where $\lambda_k \geq 0$ for each $k \in Z'_+$ and $0 < \lambda := \sum_{k=1}^{\infty} \lambda_k < \infty$. By this we mean the distribution $\mathscr{L}(\sum_{i=1}^{M} T_i)$, where the random 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 $M \sim \mathrm{Po}(\lambda)$. $\mathscr{L}(T_1)$ is called the *compounding* distribution. In the case when the compounding distribution is *geometric* with parameter θ [i.e., $\lambda_k/\lambda = (1-\theta)^{k-1}\theta$ for each $k \in Z'_+$], $\mathrm{CP}(\lambda_1, \lambda_2, \dots)$ is called the *Pólya–Aeppli*(λ, θ) distribution.

3. Counts of head runs.

THEOREM 3.1. *Let Φ be a stationary Markov chain on the state space $\{0, 1\}$ with transition matrix p, defined by*

$$p := \begin{pmatrix} 1-\beta & \beta \\ 1-\alpha & \alpha \end{pmatrix},$$

where α, $\beta \in (0,1)$. For each $n \geq r \geq 1$, define $M(n, r) := \sum_{i=r}^{n} I\{\Phi_{i-r+1} = \cdots = \Phi_i = 1\}$. Then,

$$\begin{aligned} &d_{\mathrm{TV}}\bigl(\mathscr{L}(M(n,r)), \text{Pólya–Aeppli}(\lambda, 1-\alpha)\bigr) \\ &\quad\leq H(\lambda, \theta)\biggl(2r + 2 + \frac{4\alpha}{1-\alpha} + \frac{2\beta}{(1-\alpha)^2}\biggr)(n - r + 1)\mu(r)^2 \\ &\qquad + \bigl(2(1-\alpha)r + 2\alpha\bigr)\mu(r), \end{aligned}$$

where $\mu(r) = P(\Phi_1 = \cdots = \Phi_r = 1) = (\alpha^{r-1}\beta/(1-\alpha+\beta))$, *and* $\lambda = (n-r+1)\mu(r)(1-\alpha)$. *Also,* $H(\lambda,\theta) := ((\lambda(1-\alpha))^{-1} \wedge 1)e^{\lambda}$, *unless* $\alpha \le \frac{1}{2}$, *in which case,*

$$H(\lambda,\theta) := \frac{1}{\lambda(1-\alpha)(1-2\alpha)}\left(\frac{1}{4\lambda(1-\alpha)(1-2\alpha)} + \log^{+}\big(2\lambda(1-\alpha)(1-2\alpha)\big)\right) \wedge 1.$$

PROOF. We define an embedded Markov chain η on the state space $S = \{0,\dots,r\}$ in the following way:

$$\eta_t := \min\{i \in Z_+; X_{t-i} = 0\} \wedge r \qquad \forall\, t \in Z.$$

It is easily shown that η has transition matrix p', defined by

$$p' := \begin{pmatrix} 1-\beta & \beta & 0 & \cdots & 0 & 0 & 0 \\ 1-\alpha & 0 & \alpha & \cdots & 0 & 0 & 0 \\ 1-\alpha & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1-\alpha & 0 & 0 & \cdots & 0 & \alpha & 0 \\ 1-\alpha & 0 & 0 & \cdots & 0 & 0 & \alpha \\ 1-\alpha & 0 & 0 & \cdots & 0 & 0 & \alpha \end{pmatrix}.$$

The reason for introducing η is that the quantity we are interested in, $M(n,r)$, can be expressed as $\sum_{i=r}^{n} I\{\eta_i = r\}$, that is, as the number of visits by η to the "rare" set $S_1 := \{r\}$ during the time $\{r,\dots,n\}$. Then η is a stationary irreducible Markov chain on a finite state space, and the stationary distribution μ is easily calculated:

$$\mu(i) = \begin{cases} \dfrac{1-\alpha}{1-\alpha+\beta}, & \text{if } i = 0, \\ \alpha^{i-1}\beta\left(\dfrac{1-\alpha}{1-\alpha+\beta}\right), & \text{if } i \in \{1,\dots,r-1\}, \\ \dfrac{\alpha^{r-1}\beta}{1-\alpha+\beta}, & \text{if } i = r. \end{cases}$$

Hence, η is Harris recurrent, so Theorem 4.3 in Erhardsson (1999) can be applied. This theorem gives a bound for $d_{\mathrm{TV}}(\mathscr{L}(M(n,r)), \mathrm{CP}(\lambda_1,\lambda_2,\dots))$, where $\lambda_k := (n-r+1)P(Z_0 = k)$ for each $k \in Z'_+$. Here,

$$Z_0 := I\{\eta_0 \in S_0\} \sum_{i=1}^{\tau_{S_0}-1} I\{\eta_i \in S_1\},$$

where $S_0 \subset S_1^c$ is a singleton. We choose $S_0 = \{0\}$, which is natural and convenient in the present situation. Since $S_1 = \{r\}$ is also a singleton, it follows as in Remark 3.1 in Erhardsson (1999) that $\mathrm{CP}(\lambda_1, \lambda_2, \ldots)$ is the Pólya–Aeppli(λ, θ) distribution, with parameters

$$\theta := P_r(\tau_0 < \tau_r) = 1 - \alpha,$$

$$\lambda := (n - r + 1)\mu(r)\theta = (n - r + 1)\frac{\alpha^{r-1}\beta}{1 - \alpha + \beta}(1 - \alpha),$$

or equivalently, that

$$\lambda_k := (1 - \theta)^{k-1}\theta\lambda = (n - r + 1)\frac{(1 - \alpha)^2\beta\alpha^{r+k-2}}{1 - \alpha + \beta} \qquad \forall\, k \in Z'_+.$$

The bound given in Theorem 4.3 in Erhardsson (1999) is the following:

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathscr{L}(M(n, r)), \text{Pólya–Aeppli}(\lambda, 1 - \alpha)\big) \\ &\quad\le 2H(\lambda, \theta)\left(E(\tau_{S_0} Z_0) + \mu(S_1)\frac{E(\bar{\tau}_{S_0})}{\mu(S_0)}\right)(n - r + 1)\mu(S_1) \\ &\qquad + 2P(\tau_{S_1} < \tau_{S_0}), \end{aligned}$$

where $H(\lambda, \theta)$ is defined as above; see also Proposition 3.2 and Theorem 3.1 in Erhardsson (1999). All quantities appearing in the bound can be explicitly computed. From Theorem 4.4 in Erhardsson (1999) we get, since S_1 is a singleton,

$$\begin{aligned} E(\tau_{S_0} Z_0) &= \frac{E(\tau_{S_1} I\{\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}\}) + 2\mu(S_1)E_{S_1}(\tau_{S_1} I\{\tau_{S_1} < \tau_{S_0}\})}{P_{S_1}(\tau_{S_0} < \tau_{S_1})} \\ &\quad + \frac{\mu(S_1)E_{S_1}(\tau_{S_0} I\{\tau_{S_0} < \tau_{S_1}\})}{P_{S_1}(\tau_{S_0} < \tau_{S_1})} \\ &= \frac{r\mu(r)(1 - \alpha) + 2\mu(r)\alpha + \mu(r)(1 - \alpha)}{1 - \alpha} = \left(r + \frac{2\alpha}{1 - \alpha} + 1\right)\mu(r). \end{aligned}$$

Also,

$$E(\bar{\tau}_{S_0}) = \sum_{i=1}^{r} \mu(i)E_i(\tau_0) = \sum_{i=1}^{r} \mu(i)\frac{1}{1 - \alpha} = \frac{\beta}{(1 - \alpha)(1 - \alpha + \beta)}$$

and for the last term in the bound we get

$$P(\tau_{S_1} < \tau_{S_0}) = \mu(0)\beta\alpha^{r-1} + \cdots + \mu(r)\alpha = \big((1 - \alpha)r + \alpha\big)\mu(r). \qquad \square$$

REMARK 3.1. As mentioned in Section 1, GGSSW (1995) also derive an approximating compound Poisson distribution for $\mathscr{L}(M(n, r))$, and a total variation distance error bound. They use the Stein–Chen method for discrete Poisson process approximation to find a bound for the total variation distance between the distribution of a certain sequence of indicator variables and that of a sequence of independent Poisson distributed random variables; from this an error bound for a compound Poisson approximation can be deduced. It is worth noting that, although the approach of the present paper is different from the one of GGSSW we still end up proposing exactly the same approximating compound Poisson distribution as them. Moreover, GGSSW arrive at the following total variation distance error bound (their Theorem 2.1):

$$d_{\mathrm{TV}}\big(\mathscr{L}(M(n,r)), \text{Pólya–Aeppli}(\lambda, 1-\alpha)\big) \le (6r+16)n\mu(r)^2 + 2\mu(r).$$

As is pointed out in GGSSW (1995), this bound has the drawback of being unnecessarily large for large values of $n\mu(r)$. They also write, "It would be interesting to see whether our results could be further improved upon using Stein's method directly, along the lines of the development in Barbour, Chen and Loh (1992) or Roos (1994)." Not considering the values of the constants, a comparison to the bound in Theorem 3.1 shows that the latter bound contains the factor $H(\lambda, \theta)$ in the first term. As a consequence, the bound is much smaller than the bound in GGSSW (1995) for large values of $n\mu(r)$ (provided that $\alpha \le \frac{1}{2}$).

REMARK 3.2. If the parameters α, β and r depend on n in such a way that $\limsup_{n\to\infty} \alpha < 1$ and

$$0 < \liminf_{n\to\infty} n\mu(r) \le \limsup_{n\to\infty} n\mu(r) < \infty,$$

then Theorem 3.1 implies that for some explicit $C < \infty$ and n large enough,

$$d_{\mathrm{TV}}\big(\mathscr{L}(M(n,r)), \text{Pólya–Aeppli}(\lambda, 1-\alpha)\big) \le Cr\mu(r),$$

where C is particularly small if in fact $\limsup_{n\to\infty} \alpha < \frac{1}{2}$. If, in addition to the above, $\lim_{n\to\infty} r\mu(r)/\alpha = 0$, then it follows from Remark 6.1 in Erhardsson (1999) that for some explicit $0 < C \le C' < \infty$ and n large enough,

$$C\alpha \le d_{\mathrm{TV}}\big(\mathscr{L}(M(n,r)), \mathrm{Po}((n-r+1)\mu(r))\big) \le C'\alpha.$$

In other words, the error in the simple Poisson approximation for $\mathscr{L}(M(n,r))$ converges to 0 no faster and no slower than α.

REMARK 3.3. If α, β and r depend on n in such a way that $\limsup_{n\to\infty} \alpha < \frac{1}{2}$ and $\lim_{n\to\infty} n\mu(r) = \infty$, then, for some explicit $C < \infty$ and n large enough,

$$d_{\mathrm{TV}}\big(\mathscr{L}(M(n,r)), \text{Pólya–Aeppli}(\lambda, 1-\alpha)\big) \le C\log\big(n\mu(r)\big)r\mu(r).$$

In this situation it is natural to consider a normal approximation for $W \sim$ Pólya–Aeppli$(\lambda, 1-\alpha)$. The first three moments of the geometric distribution with parameter $1-\alpha$ are easily calculated:

$$m_1 = \frac{1}{1-\alpha}, \qquad m_2 = \frac{1+\alpha}{(1-\alpha)^2}, \qquad m_3 = \frac{1+4\alpha+\alpha^2}{(1-\alpha)^3},$$

so we get, as in Theorem 6.2 in Erhardsson (1999),

$$\sup_{x\in R}\left| P\left(\frac{W-(n-r+1)\mu(r)}{\sqrt{(1+\alpha)/(1-\alpha)(n-r+1)\mu(r)}} < x\right) - \Phi(x)\right| \le \frac{0.8(1+4\alpha+\alpha^2)}{\sqrt{(1+\alpha)^3(1-\alpha)(n-r+1)\mu(r)}}.$$

4. The first and the longest head run.

THEOREM 4.1. *Let* Φ *be the stationary Markov chain defined in Theorem* 3.1. *For each* $n \geq r \geq 1$, *define* $N(n,r) := \sum_{i=r}^{n} I\{\Phi_i = 0, \Phi_{i+1} = \cdots = \Phi_{i+r} = 1\}$. *Then*

$$d_{\mathrm{TV}}\big(\mathscr{L}(N(n,r)), \mathrm{Po}((n-r+1)(1-\alpha)\mu(r))\big)$$
$$\leq \big(1 - \exp(-(n-r+1)(1-\alpha)\mu(r))\big)\left(2(1-\alpha)r + 2 + \frac{2\beta}{1-\alpha}\right)\mu(r),$$

where $\mu(r) := \alpha^{r-1}\beta/(1-\alpha+\beta)$.

PROOF. Let η be the embedded Markov chain defined in the proof of Theorem 3.1. Clearly, $N(n,r) = \sum_{i=r}^{n} I\{\eta_i = 0, \tau_r^i < \tau_0^i\}$. Moreover, since η is stationary and regenerative (it regenerates at the times of visits to $S_0 = \{0\}$), it contains an embedded stationary renewal reward process [see Section 3 in Erhardsson (2000)], for which the renewals $\{X_i; i \in Z\}$ are the times of visits by η to $\{0\}$, ordered in such a way that $\cdots < X_{-1} < X_0 \leq 0 < X_1 < \cdots$, and the corresponding rewards $\{Y_i; i \in Z\}$ are defined by $Y_i := I\{\tau_r^{X_i} < \tau_0^{X_i}\}$ for each $i \in Z$. Hence, Theorem 5.1 in Erhardsson (2000) is applicable, and gives the following bound:

$$d_{\mathrm{TV}}\left(\mathscr{L}(N(n,r)), \mathrm{Po}\left((n-r+1)\frac{E(Y_0^{\circ})}{E(T_0^{\circ})}\right)\right)$$
$$\leq \left(1 - \exp\left(-(n-r+1)\frac{E(Y_0^{\circ})}{E(T_0^{\circ})}\right)\right)\left(2\frac{E(T_0^{\circ}Y_0^{\circ})}{E(T_0^{\circ})} + \frac{E(T_0^{\circ}(T_0^{\circ}-1))E(Y_0^{\circ})}{E(T_0^{\circ})}\right),$$

where $P((T_0^{\circ}, Y_0^{\circ}) \in \cdot) = P((X_1, Y_0) \in \cdot \mid X_0 = 0) = P((\tau_0, I\{\tau_r < \tau_0\}) \in \cdot \mid \eta_0 = 0)$. Clearly $E(Y_0^{\circ}) = P_0(\tau_r < \tau_0) = \beta\alpha^{r-1}$, and it follows from the Palm inversion formula for regenerative random sequences that $E(T_0^{\circ}) = 1/\mu(0) = (1-\alpha+\beta)/(1-\alpha)$ and that

$$\frac{E(T_0^{\circ}(T_0^{\circ}-1))}{2E(T_0^{\circ})} = E(\bar{\tau}_0) = \frac{\beta}{(1-\alpha)(1-\alpha+\beta)}.$$

Finally, recalling the definition of Z_0 from the proof of Theorem 3.1, it follows from the proof of Theorem 4.4 in Erhardsson (1999) that

$$\frac{E(T_0^{\circ}Y_0^{\circ})}{E(T_0^{\circ})} = E\big(\tau_0 I\{Z_0 > 0\}]\big) = E\big(\tau_{S_1} I\{\eta_0 \in S_0, \tau_{S_1} < \tau_{S_0}\}\big)$$
$$+ \mu(S_1)E_{S_1}\big(\tau_{S_1} I\{\tau_{S_1} < \tau_{S_0}\}\big) + \mu(S_1)E_{S_1}\big(\tau_{S_0} I\{\tau_{S_0} < \tau_{S_1}\}\big)$$
$$= r\mu(r)(1-\alpha) + \mu(r)\alpha + \mu(r)(1-\alpha) = ((1-\alpha)r + 1)\mu(r). \qquad \square$$

COROLLARY 4.1. *Let Φ be the stationary Markov chain defined in Theorem* 3.1. *Define* $T_r := \min\{t \in Z_+; \Phi_{t+1} = \cdots = \Phi_{t+r} = 1\}$. *Then,*

$$\sup_{x \in R_+} \left| P\big((1-\alpha)\mu(r)T_r > x\big) - e^{-x} \right| \le \left(4(1-\alpha)r + 3 + \alpha + \frac{2\beta}{1-\alpha}\right)\mu(r).$$

PROOF. Let η be the embedded Markov chain defined in the proof of Theorem 3.1. Clearly, $T_r = \bar{\tau}_r^r$, and $P(\bar{\tau}_r^r > k) = P(\sum_{i=r}^{r+k} I\{\eta_i = r\} = 0)$ for each $k \in Z_+$, implying that

$$\begin{aligned}
&\left|P\big((1-\alpha)\mu(r)T_r > x\big) - e^{-x}\right| \\
&\quad\le \left|P\left(N\left(\left[\frac{x}{(1-\alpha)\mu(r)}\right] + r, r\right) = 0\right) - e^{-x}\right| \\
&\qquad + \left|P\left(\bar{\tau}_r^r > \left[\frac{x}{(1-\alpha)\mu(r)}\right]\right) - P\left(N\left(\left[\frac{x}{(1-\alpha)\mu(r)}\right] + r, r\right) = 0\right)\right| \\
&\quad\le d_{\mathrm{TV}}\left(\mathscr{L}\left(N\left(\left[\frac{x}{\mu(r)(1-\alpha)}\right] + r, r\right)\right),\right. \\
&\qquad\left.\mathrm{Po}\left(\left(\left[\frac{x}{\mu(r)(1-\alpha)}\right] + 1\right)(1-\alpha)\mu(r)\right)\right) \\
&\qquad + (1-\alpha)\mu(r)e^{-x} + 2P(\tau_r < \tau_0) \\
&\quad\le \left(4(1-\alpha)r + 3 + \alpha + \frac{2\beta}{1-\alpha}\right)\mu(r) \qquad \forall\, x \in R_+.
\end{aligned}$$

□

COROLLARY 4.2. *Let Φ be the stationary Markov chain defined in Theorem* 3.1. *Define* $M_n := \max\{r \in Z_+; \sum_{i=r}^{n} I\{\Phi_{i-r+1} = \cdots = \Phi_i = 1\} > 0\}$, *and let* $a_n := \log(\beta(1-\alpha)n/\alpha(1-\alpha+\beta))/\log(1/\alpha)$ *for each* $n \in Z'_+$. *Then,*

$$\begin{aligned}
&\left|P(M_n - a_n \le x) - P\left(\left[\frac{W}{\log(1/\alpha)} + \rho(a_n)\right] - \rho(a_n) \le x\right)\right| \\
&\qquad \le \left(5(1-\alpha)[x + a_n] + 6 - 2\alpha + \frac{2\beta}{1-\alpha}\right)\mu([x + a_n] + 1) \\
&\qquad\qquad \forall\, x \in [-a_n, n - a_n),
\end{aligned}$$

where W is Gumbel *distributed, that is,* $P(W \le x) = \exp(-e^{-x})$ *for each* $x \in R$, *and* $\rho\colon R \to [0, 1)$ *is defined by* $\rho(x) := x - [x]$.

PROOF. This corollary complements Theorem 1 in Gordon, Schilling and Waterman (1986) in the case of (uninterrupted) head runs, by giving a bound for the error in the approximation of $\mathscr{L}(M_n)$ with their "integerized extreme value" distribution. Let η be the embedded Markov chain defined in the proof of Theorem 3.1. Clearly, $M_n = \max\{r \in Z_+; \sum_{i=r}^n I\{\eta_i = r\} > 0\}$, and $P(M_n \le k) = P(\sum_{i=k+1}^n I\{\eta_i = k+1\} = 0)$ for each $k \in Z_+$, implying that

$$\begin{aligned}
&\left|P(M_n - a_n \le x) - P\left(\left[\frac{W}{\log(1/\alpha)} + \rho(a_n)\right] - \rho(a_n) \le x\right)\right| \\
&\quad \le \left|P(N(n, [x+a_n]+1) = 0) - P\left(\left[\frac{W}{\log(1/\alpha)} + \rho(a_n)\right] - \rho(a_n) \le x\right)\right| \\
&\qquad + \left|P(M_n \le [x+a_n]) - P(N(n, [x+a_n]+1) = 0)\right| \\
&\quad \le d_{\mathrm{TV}}\big(\mathscr{L}(N(n, [x+a_n]+1)), \\
&\qquad\qquad \mathrm{Po}((n - [x+a_n])(1-\alpha)\mu([x+a_n]+1))\big) \\
&\qquad + [x+a_n](1-\alpha)\mu([x+a_n]+1) + 2P(\tau_{[x+a_n]+1} < \tau_0) \\
&\quad \le \left(5(1-\alpha)[x+a_n] + 6 - 2\alpha + \frac{2\beta}{1-\alpha}\right)\mu([x+a_n]+1) \\
&\qquad\qquad\qquad \forall\, x \in [-a_n, n-a_n). \qquad \square
\end{aligned}$$

5. Counts of general patterns. Let Φ be a stationary irreducible Markov chain on a finite state space E, with transition matrix p and stationary distribution μ. For each $r \in Z_+$, let S_r be the space of sequences $e_1e_2\cdots e_r$ of length r of elements in E which are such that $\mu(e_1)p(e_1, e_2)\cdots p(e_{r-1}, e_r) > 0$. For each $s = e_1e_2\cdots e_r \in S_r$ and each $0 \le l \le r$, define the *initial part* of length l as $e_1e_2\cdots e_l$, and the *end part* of length l as $e_{r-l+1}e_{r-l+2}\cdots e_r$. Let $\omega := \{s_1, \dots, s_N\}$, where $s_i := e_{i,1}e_{i,2}\cdots e_{i,r_i} \in S_{r_i}$ for each $i \in \{1, \dots, N\}$, and assume that s_i is not an end part of s_j for any $i \ne j$, $i, j \in \{1, \dots, N\}$. For each $n \ge \max_{1\le i\le N} r_i$, define $M(n, \omega) := \sum_{i=1}^N \sum_{j=r_i}^n I\{\Phi_{j-r_i+1} = e_{i,1}, \dots, \Phi_j = e_{i,r_i}\}$. Can we find an approximating compound Poisson distribution for $\mathscr{L}(M(n, \omega))$, and an error bound for this approximation?

This problem can be dealt with, at least in principle, in the same manner as the problem of compound Poisson approximation for the number of head runs treated in Section 3. We introduce the embedded Markov chain η, taking values in the space S_r, where $r := \max_{1\le i\le N} r_i$, defined by $\eta_i := (\Phi_{i-r+1}, \Phi_{i-r+2}, \dots, \Phi_i)$ for each $i \in Z$. We modify S_r into S_r^{mod} by "lumping" together certain subsets of S_r into singletons. We lump together a set into a singleton if it belongs to one of the following two types:

1. A set containing those sequences in S_r which *end* with a particular initial part of one (or several) of the sequences in ω, but not with any other *strictly longer* initial part.
2. A set containing those sequences in S_r which end with a particular element in E, and which do not belong to any set of type 1.

Since Φ is stationary and irreducible, the same holds for η, so Theorem 4.3 in Erhardsson (1999) can again be used. In this case, the theorem gives

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathscr{L}(M(n,\omega)), \mathrm{CP}(\lambda_1,\lambda_2,\dots)\big) \\ &\quad \le 2H(\lambda_1,\lambda_2,\dots)\bigg(E_{S_1}(\tau_{S_0}) + E_{S_1}(\tau_{S_0}(\eta^R)) + \frac{E(\bar{\tau}_{S_0})}{\mu'(S_0)}\bigg) n\mu'(S_1)^2 \\ &\quad\quad + 2P(\tau_{S_1} < \tau_{S_0}), \end{aligned}$$

where μ' is the stationary distribution of η, the set S_1 consists of the elements in S_r^{mod} which end with sequences in ω, and the set S_0 is a single element in $S_1^c \subset S_r^{\mathrm{mod}}$. Moreover, $\lambda_k = nP(I\{\eta_0 \in S_0\}\sum_{i=1}^{\tau_{S_0}-1} I\{\eta_i \in S_1\} = k)$ for each $k \in Z'_+$, and $H(\lambda_1,\lambda_2,\dots) := ((\lambda_1)^{-1} \wedge 1)\exp(\sum_{k=1}^{\infty}\lambda_k)$, unless $\{k\lambda_k; k \in Z'_+\}$ is monotonically decreasing towards 0, in which case

$$H(\lambda_1,\lambda_2,\dots) := \frac{1}{\lambda_1 - 2\lambda_2}\bigg(\frac{1}{4(\lambda_1-2\lambda_2)} + \log^+\big(2(\lambda_1-2\lambda_2)\big)\bigg) \wedge 1.$$

If S_1 is a singleton, then $\mathrm{CP}(\lambda_1,\lambda_2,\dots) = \text{Pólya–Aeppli}(\lambda,\theta)$, where $\theta = P_{S_1}(\tau_{S_0} < \tau_{S_1})$ and $\lambda = \sum_{k=1}^{\infty}\lambda_k = n\mu'(S_1)\theta$. The generating function for the parameters $\{\lambda_k; k \in Z'_+\}$, and the quantities appearing in the bound, can be found as solutions to linear equation systems with dimensions at most $\mathrm{card}(S_r^{\mathrm{mod}}) - 1 \le \mathrm{card}(E) + \sum_{i=1}^{N} r_i$; see Section 5 in Erhardsson (1999).

EXAMPLE 5.1. Let the state space of Φ be $E = \{A, C, G, T\}$ and let the set of patterns for which we want to count the number of occurrences be $\omega = \{ACACA\}$ (thus, ω contains only one pattern). The state space S_r^{mod} of the embedded Markov chain η then consists of the following eight elements: ACACA, ACAC, ACA (not preceded by AC), AC (not preceded by AC), A (not preceded by AC), C (not preceded by A), G, and T. The transition matrix p' of η is

$$p' := \begin{pmatrix} 0 & p_{A,C} & 0 & 0 & p_{A,A} & 0 & p_{A,G} & p_{A,T} \\ p_{C,A} & 0 & 0 & 0 & 0 & p_{C,C} & p_{C,G} & p_{C,T} \\ 0 & p_{A,C} & 0 & 0 & p_{A,A} & 0 & p_{A,G} & p_{A,T} \\ 0 & 0 & p_{C,A} & 0 & 0 & p_{C,C} & p_{C,G} & p_{C,T} \\ 0 & 0 & 0 & p_{A,C} & p_{A,A} & 0 & p_{A,G} & p_{A,T} \\ 0 & 0 & 0 & 0 & p_{C,A} & p_{C,C} & p_{C,G} & p_{C,T} \\ 0 & 0 & 0 & 0 & p_{G,A} & p_{G,C} & p_{G,G} & p_{G,T} \\ 0 & 0 & 0 & 0 & p_{T,A} & p_{T,C} & p_{T,G} & p_{T,T} \end{pmatrix}.$$

In this situation, the "rare" set is of course $S_1 = \{ACACA\}$. S_0 could be chosen as any singleton in S_r^{mod} except $\{ACACA\}$; the optimal choice depends on p.

We finally give a more explicit (but less sharp) result for counts of general patterns, which has some resemblance to Theorem 2.3 in GGSSW (1995) and Theorem 4 in Schbath (1995). For simplicity, we assume that S_0 is chosen as the singleton in S_r^{mod} formed by lumping together those sequences in S_r which end with s_0, where $s_0 = e_{0,1}e_{0,2}\cdots e_{0,m} \in S_m$ $(1 \le m \le r)$ is not an end part of an initial part of any sequence in ω and has no sequence in ω as an end part. (We can always augment ω with s_0 before constructing S_r^{mod} so that this assumption can be satisfied.)

THEOREM 5.1. *Let Φ be a stationary irreducible* aperiodic *Markov chain on a finite state space E, with transition matrix p and stationary distribution μ. Let $\omega := \{s_1, \dots, s_N\}$, where $s_i := e_{i,1}e_{i,2}\cdots e_{i,r_i} \in S_{r_i}$ for each $i \in \{1, \dots, N\}$, and assume that s_i is not an end part of s_j for any $i \ne j$, $i, j \in \{1, \dots, N\}$. For each $n \ge r := \max_{1\le i\le N} r_i$, define $M(n, \omega) := \sum_{i=1}^{N}\sum_{j=r_i}^{n} I\{\Phi_{j-r_i+1} = e_{i,1}, \dots, \Phi_j = e_{i,r_i}\}$. Define the space S_r^{mod}, the Markov chain η, the set $S_1 \subset S_r^{\mathrm{mod}}$, the singleton $S_0 \subset S_r^{\mathrm{mod}}$, the parameters $\{\lambda_k; k \in Z'_+\}$, and the constant $H(\lambda_1, \lambda_2, \dots)$ as in the previous paragraphs. Then, for each $a \in (0,1)$,*

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathscr{L}(M(n,\omega)), \mathrm{CP}(\lambda_1, \lambda_2, \dots)\big)\\ &\quad\le \frac{2H(\lambda_1, \lambda_2, \dots)}{(1-a)\mu'(S_0)}\Bigg(\frac{1}{\mu'(S_0)}\Bigg(\Bigg\lceil\frac{\log\big(4a^2\mu(e_{0,1})^2\check{\mu}\big)}{\log\beta_1}\Bigg\rceil + m\Bigg)\\ &\qquad\qquad + \Bigg\lceil\frac{\log\big(4a^2\mu(e_{0,m})^2\check{\mu}\big)}{\log\beta_{1,R}}\Bigg\rceil + r\Bigg)n\mu'(S_1)^2\\ &\qquad + \frac{2}{(1-a)\mu'(S_0)}\Bigg(\Bigg\lceil\frac{\log\big(4a^2\mu(e_{0,m})^2\check{\mu}\big)}{\log\beta_{1,R}}\Bigg\rceil + r\Bigg)\mu'(S_1), \end{aligned}$$

where μ' is the stationary distribution of η, $\check{\mu} := \min_{y\in E}\mu(y)$, β_1 is the second largest eigenvalue of the matrix pp^R and $\beta_{1,R}$ is the second largest eigenvalue of the matrix $p^R p$.

PROOF. Theorem 4.3 in Erhardsson (1999) applied to the Markov chain η gives

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathscr{L}(M(n,\omega)), \mathrm{CP}(\lambda_1, \lambda_2, \dots)\big)\\ &\quad\le 2H(\lambda_1, \lambda_2, \dots)\\ &\qquad\times\Bigg(E_{S_1}(\tau_{S_0}) + E_{S_1}(\tau_{S_0}(\eta^R)) + \frac{\mu'(S_0^c)}{\mu'(S_0)}E_{S_0^c}(\tau_{S_0})\Bigg)n\mu'(S_1)^2 \qquad (5.1)\\ &\qquad + 2E_{S_1}(\tau_{S_0}(\eta^R))\mu'(S_1), \end{aligned}$$

where we also used the fact that

$$P(\tau_{S_1} < \tau_{S_0}) = P(\bar{\tau}_{S_1} < \bar{\tau}_{S_0}) \le \sum_{i=0}^{\infty} P(\eta_i \in S_1, \bar{\tau}_{S_0} > i)$$
$$= \sum_{i=0}^{\infty} P\big(\eta_0^R \in S_1, \tau_{S_0}(\eta^R) > i\big) = E_{S_1}(\tau_{S_0}(\eta^R))\mu'(S_1).$$

To find bounds for the expectations in (5.1), we note that if there exists constants $n_1 \in Z'_+$ and $C < 1$ such that $\sup_{x \in S_0^c} P_x(\bar{\tau}_{S_0} > n_1) \le C$, then, for any $B \subset S_0^c$,

$$E_B(\bar{\tau}_{S_0}) = \sum_{i=0}^{\infty} P_B(\bar{\tau}_{S_0} > i) \le \sum_{i=0}^{\infty} \Big(\sup_{x \in S_0^c} P_x(\bar{\tau}_{S_0} > n_1)\Big)^{[i/n_1]}$$
$$\le \sum_{i=0}^{\infty} C^{[i/n_1]} = \frac{n_1}{1-C}.$$

Theorem 2.1 in Fill (1991) tells us that for any irreducible aperiodic Markov chain Φ on a finite state space E with stationary distribution μ, it holds that

$$d_{\mathrm{TV}}\big(P_x(\Phi_n \in \cdot), \mu\big)^2 \le \frac{1}{4\mu(x)} \beta_1^n \qquad \forall\, x \in E, n \in Z_+.$$

This implies that, for each $x \in S_0^c$ and $n \ge m$,

$$P_x(\bar{\tau}_{S_0} > n) \le 1 - P_x(\eta_n \in S_0)$$
$$\le 1 - \left(\mu(e_{0,1}) - \frac{1}{2}\sqrt{\frac{\beta_1^{n-m+1}}{\check{\mu}}}\right)^+ \prod_{i=1}^{m-1} p(e_{0,i}, e_{0,i+1}),$$

so that we may choose the constants $n_1 \in Z'_+$ and $C < 1$ above as

$$n_1 := \left\lceil \frac{\log\big(4a^2\mu(e_{0,1})^2\check{\mu}\big)}{\log \beta_1} \right\rceil + m; \qquad C := 1 - (1-a)\mu'(S_0).$$

Analogous calculations for the reverse chain η^R give the remaining bounds. □

REMARK 5.1. If ω depends on n in such a way that $\limsup_{n\to\infty} n\mu'(S_1) < \infty$, and if E, p and S_0 do not depend on n, then Theorem 5.1 implies that for some $C < \infty$ and n large enough,

$$d_{\mathrm{TV}}\big(\mathscr{L}(M(n,\omega)), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\big) \le Cr\mu'(S_1).$$

To compute C explicitly we must bound the eigenvalues β_1 and $\beta_{1,R}$. For further information on this topic, see Fill (1991).

6. Birth–death chains. By a *birth–death* chain (or a *general random walk* on Z_+), we mean an irreducible Markov chain on the state space Z_+ with a transition probability p defined by

$$p(j,k) := \begin{cases} p_j, & \text{if } k = j+1, \\ r_j, & \text{if } k = j, \\ q_j, & \text{if } k = j-1, \\ 0, & \text{otherwise.} \end{cases}$$

It is convenient to define, for each $i, k \in Z_+$,

$$\pi_{i,k} := \begin{cases} \dfrac{p_i}{q_k} \displaystyle\prod_{j=i+1}^{k-1} \dfrac{p_j}{q_j}, & \text{if } i < k, \\ \dfrac{q_i}{p_k} \displaystyle\prod_{j=k+1}^{i-1} \dfrac{q_j}{p_j}, & \text{if } i > k, \\ 1, & \text{if } i = k. \end{cases}$$

It is easily checked that $\pi_{i,k} = \pi_{i,j}\pi_{j,k}$ for each i, j, $k \in Z_+$ such that $i \le j \le k$, and that $\{\pi_{0,k}; k \in Z_+\}$ defines a reversible measure on Z_+. If $\sum_{i \in Z_+} \pi_{0,i} < \infty$, then the birth–death chain is positive recurrent with the unique stationary distribution μ, defined by

$$\mu(k) := \frac{\pi_{0,k}}{\sum_{i \in Z_+} \pi_{0,i}} \qquad \forall\, k \in Z_+.$$

For any birth–death chain η it is well known that, for each $a, k, b \in Z_+$ such that $a \le k \le b$,

$$P_k(\tau_a < \tau_b) = \left(\sum_{s=a}^{b-1} \frac{1}{p_s \pi_{0,s}} \right)^{-1} \sum_{r=k}^{b-1} \frac{1}{p_r \pi_{0,r}}. \tag{6.1}$$

Likewise, for any positive recurrent birth–death chain η,

$$E_k(\tau_i) = \begin{cases} \displaystyle\sum_{r=i}^{k-1} \sum_{j=r+1}^{\infty} \frac{\pi_{r,j}}{p_r}, & \text{if } i < k, \\ \displaystyle\sum_{r=k+1}^{i} \sum_{j=0}^{r-1} \frac{\pi_{r,j}}{q_r}, & \text{if } i > k. \end{cases} \tag{6.2}$$

Equation (6.1) is Theorem 3.7 in Chapter 5 of Durrett (1991). It can also be verified that (6.1) is the unique bounded solution of Poisson's equation, as in Proposition 5.2(i) in Erhardsson (1999). Similarly, (6.2) satisfies Poisson's equation as in Proposition 5.2(iii) in Erhardsson (1999), but it is not necessarily a bounded solution. However, in this case a coupling argument can be used; see Section 3 in Chapter 4 of Erhardsson (1997).

THEOREM 6.1. *Let η be a stationary birth–death chain. Let $S_1 \subset Z'_+$, and define $z_{\min} := \min\{z \in S_1\}$. Then,*

$$d_{\mathrm{TV}}\biggl(\mathscr{L}\biggl(\sum_{i=1}^{n} I\{\eta_i \in S_1\}\biggr), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\biggr)$$

$$\le 2H(\lambda_1, \lambda_2, \ldots)$$

$$\times \biggl(2 \sum_{k \in S_1} \frac{\mu(k)}{\mu(S_1)} \sum_{r=0}^{k-1} \sum_{j=r+1}^{\infty} \frac{\pi_{r,j}}{p_r} + \sum_{k=1}^{\infty} \frac{\mu(k)}{\mu(0)} \sum_{r=0}^{k-1} \sum_{j=r+1}^{\infty} \frac{\pi_{r,j}}{p_r}\biggr) n\mu(S_1)^2 \tag{6.3}$$

$$+ 2\biggl(\sum_{k=z_{\min}}^{\infty} \mu(k) + \sum_{s=0}^{z_{\min}-1} \frac{1}{p_s \pi_{0,s}}^{-1} \sum_{k=1}^{z_{\min}-1} \mu(k) \sum_{r=0}^{k-1} \frac{1}{p_r \pi_{0,r}}\biggr),$$

where $\lambda_k = nP(I\{\eta_0 = 0\} \sum_{i=1}^{\tau_0 - 1} I\{\eta_i \in S_1\} = k)$ for each $k \in Z'_+$. If $S_1 = \{z\}$ $(z \in Z'_+)$, then $\mathrm{CP}(\lambda_1, \lambda_2, \ldots) =$ Pólya–Aeppli(λ, θ), where $\theta = P_z(\tau_0 < \tau_z)$ and $\lambda := \sum_{k=1}^{\infty} \lambda_k = n\mu(0)P_0(\tau_z < \tau_0) = n\mu(z)P_z(\tau_0 < \tau_z)$. Moreover, $H(\lambda_1, \lambda_2, \ldots) := ((\lambda_1)^{-1} \wedge 1)\exp(\lambda)$, unless $\{k\lambda_k; k \in Z'_+\}$ is monotonically decreasing towards 0, *in which case*

$$H(\lambda_1, \lambda_2, \ldots) := \frac{1}{\lambda_1 - 2\lambda_2}\biggl(\frac{1}{4(\lambda_1 - 2\lambda_2)} + \log^+\bigl(2(\lambda_1 - 2\lambda_2)\bigr)\biggr) \wedge 1.$$

PROOF. The proof follows from Theorem 4.3, Remark 3.1 and Proposition 3.2 in Erhardsson (1999), using also the reversibility of η, (6.1) and (6.2).

EXAMPLE 6.1. Berman (1986) considers a stationary birth–death chain η with transition probabilities satisfying the following conditions: $p_0 = 1$, $\lim_{j \to \infty} p_j = p < \frac{1}{2}$ and $p_j + q_j = 1$ for each $j \in Z_+$. He claims (but does not explicitly prove) the following compound Poisson limit theorem (his Theorem 6.1):

$$\mathscr{L}\biggl(\sum_{i=1}^{[B_z]} I\{\eta_i = z\}\biggr) \xrightarrow{d} \text{Pólya–Aeppli}(\mu(0), q) \quad \text{as } z \to \infty,$$

where $B_z := \sum_{r=0}^{z} (1/p_r \pi_{0,r})$. However, our Theorem 6.1 yields a slightly different result. Clearly, for each p' such that $p < p' < \frac{1}{2}$, there exists a $N \in Z_+$ such that $p_j \le p'$ and $q_j \ge q' := 1 - p'$ for each $j > N$, and also two constants $1 \le C_p < \infty$ and $0 < C_q \le 1$ such that $p_j \le C_p p'$ for each $j \in Z_+$ and $q_j \ge C_q q'$ for each $j \in Z'_+$. Therefore, $\mu(k)/\mu(0) = \pi_{0,k} \le (C_p/C_q)^N (p'/q')^k$ for each $k \in Z'_+$, and

$$\sum_{r=0}^{k-1} \sum_{j=r+1}^{\infty} \frac{\pi_{r,j}}{p_r} \le \biggl(\frac{C_p}{C_q}\biggr)^N \frac{1}{q'} \sum_{r=0}^{k-1} \frac{1}{1 - p'/q'} = \biggl(\frac{C_p}{C_q}\biggr)^N \frac{k}{q' - p'} \qquad \forall\, k \in Z'_+.$$

Also,

$$B_k \frac{\mu(k)}{\mu(0)} = \sum_{r=0}^{k} \frac{\pi_{0,k}}{p_r \pi_{0,r}} = \sum_{r=0}^{k-1} \frac{\pi_{r,k}}{p_r} + \frac{1}{p_k} \qquad \forall\, k \in Z'_+,$$

which implies

$$1 \le B_k \frac{\mu(k)}{\mu(0)} \le \left(\frac{C_p}{C_q}\right)^N \frac{1}{q' - p'} + \frac{1}{\inf_{j \in Z_+} p_j} \qquad \forall\, k \in Z'_+.$$

Hence, if $p > 0$ then, for some $C < \infty$ and $z \in Z_+$ large enough,

$$d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{i=1}^{[B_z]} I\{\eta_i = z\}\right), \text{Pólya–Aeppli}(\lambda, \theta)\right) \le Cz\left(\frac{p'}{q'}\right)^z,$$

where

$$\theta = q_z\left(1 - \frac{B_{z-2}}{B_{z-1}}\right); \qquad \lambda = [B_z]\mu(0)P_0(\tau_z < \tau_0) = \frac{[B_z]}{B_{z-1}}\mu(0).$$

Since $\lim_{z\to\infty} B_z = \infty$ and $\lim_{z\to\infty} B_z/B_{z+1} = p/q$ [both results according to Lemma 2.1 in Berman (1986)], it holds that

$$\text{Pólya–Aeppli}(\lambda, \theta) \xrightarrow{d} \text{Pólya–Aeppli}\left(\frac{q}{p}\mu(0), q - p\right) \quad \text{as } z \to \infty,$$

a limiting distribution which is not completely identical to the one given in (the unproven) Theorem 6.1 in Berman (1986).

EXAMPLE 6.2. Serfozo (1980) considers a stationary birth–death chain η with the following transition probabilities: $p_0 = 1$, $p_j = p < \frac{1}{2}$ for each $j \in Z'_+$, and $q_j = 1 - p =: q$ for each $j \in Z'_+$. He proves the following compound Poisson limit theorem (in his Corollary 3.1):

$$\mathscr{L}\left(\sum_{i=1}^{[a_z]} I\{\eta_i = z\}\right) \xrightarrow{d} \text{Pólya–Aeppli}\left(\frac{q-p}{2q}, q - p\right) \quad \text{as } z \to \infty,$$

where $a_z := (q/p - 1)^{-1}(q/p)^z$. This result can be extended, using Theorem 6.1. First,

$$\pi_{j,k} = \begin{cases} \left(\dfrac{p}{q}\right)^{k-j}, & \text{if } 0 < j < k, \\[2ex] \dfrac{1}{p}\left(\dfrac{p}{q}\right)^{k}, & \text{if } 0 = j < k, \end{cases}$$

which implies that

$$\mu(k) = \frac{\pi_{0,k}}{\sum_{j=0}^{\infty} \pi_{0,j}} = \begin{cases} \dfrac{q-p}{2q}, & \text{if } k = 0, \\[2ex] \dfrac{q-p}{2pq}\left(\dfrac{p}{q}\right)^{k}, & \text{if } k > 0, \end{cases}$$

and that

$$\sum_{k=z}^{\infty} \mu(k) = \frac{q-p}{2pq} \sum_{k=z}^{\infty} \left(\frac{p}{q}\right)^k = \frac{q-p}{2pq}\left(\frac{p}{q}\right)^z \frac{1}{1-p/q} = \frac{1}{2p}\left(\frac{p}{q}\right)^z.$$

Moreover,

$$\begin{aligned}\sum_{r=0}^{k-1} \sum_{j=r+1}^{\infty} \frac{\pi_{r,j}}{p_r} &= \frac{1}{p} \sum_{r=0}^{k-1} \sum_{j=r+1}^{\infty} \left(\frac{p}{q}\right)^{j-r} \\ &= \frac{1}{p}\left(\frac{p/q}{1-p/q}\right)k = \frac{k}{q-p} \qquad \forall\, k \in Z'_+;\end{aligned}$$

$$a_k \mu(k) = \frac{1}{q/p-1}\left(\frac{q}{p}\right)^k \frac{q-p}{2pq}\left(\frac{p}{q}\right)^k = \frac{1}{2q} \quad \forall\, k \in Z'_+;$$

$$\begin{aligned}&\left(\sum_{s=0}^{z-1} \frac{1}{p_s \pi_{0,s}}\right)^{-1} \sum_{k=1}^{z-1} \mu(k) \sum_{r=0}^{k-1} \frac{1}{p_r \pi_{0,r}} \\ &\quad = \frac{q/p-1}{(q/p)^z - 1} \sum_{k=1}^{z-1} \frac{1}{2q}\left(1 - \left(\frac{p}{q}\right)^k\right) \\ &\quad = \frac{q-p}{2pq(1-(p/q)^z)}\left(\frac{p}{q}\right)^z \left(z - 1 - \frac{p}{q-p}\left(1 - \left(\frac{p}{q}\right)^{z-1}\right)\right).\end{aligned}$$

Hence,

$$\begin{aligned}&d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{i=1}^{[a_z]} I\{\eta_i = z\}\right), \text{Pólya–Aeppli}(\lambda, \theta)\right) \\ &\quad \le 2H(\lambda_1, \lambda_2, \ldots)\left(\frac{2z}{q-p} + \frac{q}{(q-p)^3}\right)[a_z]\left(\frac{q-p}{2pq}\right)^2 \left(\frac{p}{q}\right)^{2z} \\ &\qquad + \frac{1}{p}\left(\frac{p}{q}\right)^z + \frac{(q-p)(z-1) - p(1-(p/q)^{z-1})}{1-(p/q)^z} \frac{1}{pq}\left(\frac{p}{q}\right)^z,\end{aligned}$$

where

$$\theta = \frac{q-p}{1-(p/q)^z}; \qquad \lambda = [a_z] \frac{(q-p)^2}{2pq(1-(p/q)^z)}\left(\frac{p}{q}\right)^z.$$

The bound converges to 0 as $z \to \infty$. Moreover, $\lim_{z\to\infty} \theta = q - p$ and $\lim_{z\to\infty} \lambda = (q-p)/(2q)$.

EXAMPLE 6.3. Serfozo (1980) proves (also in his Corollary 3.1), for the same birth–death chain as in Example 6.2, that

$$\mathscr{L}\left(\sum_{i=1}^{[a_z]} I\{\eta_i \ge z\}\right) \xrightarrow{d} \mathrm{CP}(\lambda_1, \lambda_2, \ldots) \quad \text{as } z \to \infty,$$

where $\lambda = (q-p)/(2q)$ and

$$\frac{\lambda_k}{\lambda} = \sum_{i=1}^{k} g^{i*}(k)\left(\frac{p}{q}\right)^{i-1}\left(1-\frac{p}{q}\right) \qquad \forall\, k \in Z'_+.$$

Here, $g^{i*}(\cdot)$ denotes g convoluted with itself i times, and the function $g\colon Z'_+ \to [0,1]$ is defined by

$$\text{(6.4)} \qquad \begin{aligned} g(2i-1) &= \frac{(-1)^{i-1}}{2p}\binom{1/2}{i}(4pq)^i \qquad \forall\, i \in Z'_+;\\ g(2i) &= 0 \qquad \forall\, i \in Z'_+. \end{aligned}$$

This result can also be extended using Theorem 6.1. It holds that

$$\begin{aligned} \sum_{k=z}^{\infty} k\mu(k) &= \frac{q-p}{2pq}\left(\frac{p}{q}\right)\sum_{k=z}^{\infty} k\left(\frac{p}{q}\right)^{k-1} = \frac{q-p}{2pq}\left(\frac{p}{q}\right)\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{x^z}{1-x}\right]_{x=p/q}\\ &= \frac{q-p}{2pq}\left(\frac{p}{q}\right)\frac{(1-p/q)z+p/q}{(1-p/q)^2}\left(\frac{p}{q}\right)^{z-1} = \frac{(q-p)z+p}{2p(q-p)}\left(\frac{p}{q}\right)^{z}. \end{aligned}$$

Hence,

$$\begin{aligned} &d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{i=1}^{[a_z]} I\{\eta_i \ge z\}\right)\mathrm{CP}(\lambda_1, \lambda_2, \ldots)\right)\\ &\quad \le 2H(\lambda_1, \lambda_2, \ldots)\left(\frac{2z}{q-p} + \frac{2p(q-p)+q}{(q-p)^3}\right)[a_z]\left(\frac{1}{2p}\right)^2\left(\frac{p}{q}\right)^{2z}\\ &\qquad + \frac{1}{p}\left(\frac{p}{q}\right)^z + \frac{(q-p)(z-1)-p\big(1-(p/q)^{z-1}\big)}{1-(p/q)^z}\frac{1}{pq}\left(\frac{p}{q}\right)^z, \end{aligned}$$

where

$$\begin{aligned} \lambda &= [a_z]\mu(0)P_0(\tau_z < \tau_0) = [a_z]\mu(z)P_z(\tau_0 < \tau_z)\\ &= [a_z]\frac{(q-p)^2}{2pq\big(1-(p/q)^z\big)}\left(\frac{p}{q}\right)^z, \end{aligned}$$

and it follows as in the proof of Corollary 3.1 in Serfozo (1980) that

$$\begin{aligned} \frac{\lambda_k}{\lambda} &= P_z\left(\sum_{i=1}^{\tau_0-1} I\{\eta_i \ge z\} = k\right)\\ &= \sum_{i=1}^{k} g^{i*}(k)\left(1-\frac{1-p/q}{1-(p/q)^z}\right)^{i-1}\left(\frac{1-p/q}{1-(p/q)^z}\right) \qquad \forall\, k \in Z'_+. \end{aligned}$$

The bound converges to 0 as $z \to \infty$. Moreover, $\lim_{z\to\infty}\lambda = (q-p)/(2q)$, and

$$\lim_{z\to\infty}\frac{\lambda_k}{\lambda} = \sum_{i=1}^{k} g^{i*}(k)\left(\frac{p}{q}\right)^{i-1}\left(1-\frac{p}{q}\right) \qquad \forall\, k \in Z'_+.$$

REFERENCES

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

BARBOUR, A. D. (1997). Stein's method. In *Encyclopedia of Statistical Science* **1**. Wiley, New York.

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.

BERMAN, S. M. (1986). Extreme sojourns for random walks and birth-and-death processes. *Comm. Statist. Stochastic Models* **2** 393–408.

DURRETT, R. (1991). *Probability: Theory and Examples.* Duxbury Press, Belmont, CA.

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

ERHARDSSON, T. (1999). Compound Poisson approximation for Markov chains using Stein's method. *Ann. Probab.* **27** 565–596.

ERHARDSSON, T. (2000). On stationary renewal reward processes where most rewards are zero. *Probab. Theory Related Fields*. To appear.

FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.

GESKE, M. X., GODBOLE, A. P., SCHAFFNER, A. A., SKOLNICK, A. M. and WALLSTROM, G. L. (1995). Compound Poisson approximations for word patterns under Markovian hypotheses. *J. Appl. Probab.* **32** 877–892.

GODBOLE, A. P. and PAPASTAVRIDIS, S. G., eds. (1994). *Runs and Patterns in Probability: Selected Papers.* Kluwer, Dordrecht.

GORDON, L., SCHILLING, M. F. and WATERMAN, M. S. (1986). An extreme value theory for long head runs. *Probab. Theory Related Fields* **72** 279–287.

ROOS, M. (1994). Stein–Chen method for compound Poisson approximation: the coupling approach. In *Probab. Theory Math. Statist. Proceedings of the Sixth Vilnius Conference* 645–660.

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

SCHBATH, S. (1995). Compound Poisson approximation of word counts in DNA sequences. *ESAIM Probab. Statist.* **1** 1–16.

SERFOZO, R. (1980). High-level exceedances of regenerative and semi-stationary processes. *J. Appl. Probab.* **17** 423–431.

STEFANOV, V. T. and PAKES, A. G. (1997). Explicit distributional results in pattern formation. *Ann. Appl. Probab.* **7** 666–678.

WANG, Y. H. and JI, S. (1995). Limit theorems for the number of occurrences of consecutive k successes in n Markovian trials. *J. Appl. Probab.* **32** 727–735.

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