The Annals of Applied Probability
2001, Vol. 11, No. 3, 964–1002

COMPOUND POISSON APPROXIMATION: A USER'S GUIDE

BY A. D. BARBOUR[1, 2] AND O. CHRYSSAPHINOU[2]

Universität Zürich and University of Athens

Compound Poisson approximation is a useful tool in a variety of applications, including insurance mathematics, reliability theory, and molecular sequence analysis. In this paper, we review the ways in which Stein's method can currently be used to derive bounds on the error in such approximations. The theoretical basis for the construction of error bounds is systematically discussed, and a number of specific examples are used for illustration. We give no numerical comparisons in this paper, contenting ourselves with references to the literature, where compound Poisson approximations derived using Stein's method are shown frequently to improve upon bounds obtained from problem specific, ad hoc methods.

1. Motivation. Many probability models [Aldous 1989] involve rare, isolated and weakly dependent clumps of interesting occurrences. A typical example is that of clusters of extreme events, such as earthquakes of magnitude exceeding 4.0; when one event occurs, several more may follow in quick succession, before normality returns. Clusters of events can then be expected to take place almost "at random," according to a Poisson process, in which case the number of clusters occurring in a given time interval would have a distribution close to a Poisson distribution $\mathrm{Po}(\lambda)$ with some mean λ, and the sizes of the individual clumps might well also be assumed to be approximately independent and identically distributed with some distribution $\boldsymbol{\mu}$. If these assumptions were precisely true, the total number W of occurrences would then have a compound Poisson distribution $\mathrm{CP}(\lambda, \boldsymbol{\mu})$, the distribution of the sum of a random $\mathrm{Po}(\lambda)$-distributed number of independent random variables, each with distribution $\boldsymbol{\mu}$: more formally, $\mathrm{CP}(\lambda, \boldsymbol{\mu})$ is defined by

$$\mathrm{CP}(\lambda, \boldsymbol{\mu}) = \mathscr{L}\left(\sum_{j=1}^{M} Y_j\right) = \mathscr{L}\left(\sum_{i\geq 1} iM_i\right),$$

for any $\lambda > 0$ and any probability distribution $\boldsymbol{\mu}$ on $\mathbb{N}$, where $(Y_j, j \geq 1)$ are independent, have distribution $\boldsymbol{\mu}$ and are independent also of $M \sim \mathrm{Po}(\lambda)$ and where $(M_i, i \geq 1)$ are independent, with $M_i \sim \mathrm{Po}(\lambda\mu_i)$. The former representation is that which ties in with the description above. The latter is an equivalent definition, which emphasizes that the number of clumps of each size $i \geq 1$ itself has a Poisson distribution and that these numbers are independent; this structure can also be useful.

Received December 1999; revised October 2000.

[1]Supported in part by Schweizer Nationalfonds Grant 20-50686.97.

[2]Supported in part by University of Athens Research Grant 70-4-2558.

AMS 2000 *subject classifications*. 6002, 60C05, 60E15, 60F05, 60K10, 62E17, 92D20.

Key words and phrases. Compound Poisson, Stein's method, total variation distance, Kolmogorov distance.

In practice, the assumptions above are usually not satisfied exactly; the clumps may occur only approximately as a Poisson process, and the clump sizes may not quite be independent. If this is the case, and if the CP $(\lambda, \boldsymbol{\mu})$ distribution is being used to approximate the distribution of W, it is important to know how accurate the approximation really is. In this paper, we are interested in showing how to quantify this closeness, when $W = \sum_{\gamma\in\Gamma} X_\gamma$ is a countable sum of nonnegative integer valued random variables $(X_\gamma, \gamma \in \Gamma)$. We review estimates for the distance between $\mathscr{L}(W)$ and CP $(\lambda, \boldsymbol{\mu})$, for suitably chosen λ and $\boldsymbol{\mu}$, with respect to the Kolmogorov distance d_K and the total variation distance d_{TV}, where, for probability distributions P and Q on $\mathbb{Z}_+$,

$$d_K(P, Q) := \sup_{j\geq 0} |P\{[0, j]\} - Q\{[0, j]\}|;$$

$$d_{TV}(P, Q) := \sup_{A\subset\mathbb{Z}_+} |P\{A\} - Q\{A\}|.$$

The X_γ are to be thought of as being generally weakly dependent, apart possibly from some strong local dependence, but we retain generality and flexibility by avoiding as far as possible making any specific assumptions in this respect.

1.1. *Insurance.* A simple model of an insurance portfolio assumes a finite number n of insured risks, each of which may lead to a claim with a small probability, independently of all the others. The distribution of the total number N of claims is then well approximated by the Poisson distribution $\mathrm{Po}(\lambda)$, with $\lambda = \sum_{j=1}^n p_j$, even if the claim probabilities p_j are not equal. Furthermore, as observed by Michel (1988) [see also Le Cam (1965)], if all claim amounts are independent and identically distributed, the difference in terms of total variation distance between the distribution of the aggregated claim amount and an appropriate compound Poisson distribution is no greater than the total variation distance $\Delta \leq \lambda^{-1} \sum_{i=1}^n p_i^2$ [Barbour and Hall (1984)] between the distribution of the total number of claims N and $\mathrm{Po}(\lambda)$. If the occurrences of claims are weakly dependent, but the claim amounts are still independent and identically distributed, Goovaerts and Dhaene (1996) have noted that Michel's observation can still be applied and that the new value of Δ, which will usually be larger than $\lambda^{-1} \sum_{i=1}^n p_i^2$, because of the dependence, can be estimated using the Stein–Chen method [Barbour, Holst and Janson (1992), hereafter BHJ (1992)].

In many insurance applications, however, there may be strong local dependence between claim occurrences. For instance, in storm damage insurance, the (rare) single occurrence of a tornado in a particular town may lead to some random number of claims on the portfolio. Since the distribution of this number of claims may well depend on the time of year, the preceding argument, even assuming independent and identically distributed individual claim amounts, cannot be applied, because the *aggregate* claim amounts resulting from occurrences of tornados are no longer independent and identically distributed. Despite this, it still seems reasonable to suppose that the distribution of the total number of claims is close to a *compound* Poisson distribution in

total variation, in which case, if once again the individual claim amounts are independent and identically distributed, the distribution of the aggregated claim amount is itself at least as close in total variation to an appropriate compound Poisson distribution, again by Michel's observation. To exploit this idea, if the possibility of substantial local dependence between the random claim numbers is also to be allowed, it is necessary to have an equivalent of the Stein–Chen method, which quantifies the error in total variation when approximating the distribution of a sum of nonnegative random variables, most of them taking the value zero with high probability, by a compound Poisson distribution. Once such an equivalent has been developed, there is the further advantage that one can dispense with Michel's observation and the assumption of independent and identically distributed individual claim amounts and prove compound Poisson approximations to the total claim amount directly.

1.2. *Reliability.* We consider a system of n linearly arranged components, having independent lifetimes with common distribution function F, connected in such a way that the system fails if at least k consecutive components fail. This reliability system is called the Consecutive-k-out-of-n:F [$C(k, n{:}F)$] system. Over the last two decades, the $C(k, n{:}F)$ and related systems have been extensively studied by many authors. One can find a rich literature in Chao, Fu and Koutras (1995). The advantages of using such a system are that it has higher reliability than a series system, but is less expensive to build than a parallel system. It has applications in telecommunication systems, oil pipelines, vacuum systems, computer ring networks, spacecraft relay stations and many other engineering systems.

The reliability of this system—the probability that it has not failed—has been exactly determined, but the explicit formula is quite complicated, especially if n and k are large. For this reason, upper and lower bounds for the reliability have been derived. In this context, the Stein–Chen method for Poisson approximation has proved a powerful tool. The approach is as follows. For any fixed time T, we associate independent Bernoulli Be(p) random variables $J_1, \ldots, J_n$ with the components, where $p = 1 - F(T)$; J_i takes the value 1(0) if the ith component works(fails), $i = 1, \ldots, n$. We then define the random variable $W = \sum_{i=1}^{n-k+1} I_i$, where $I_i = \prod_{j=i}^{i+k-1}(1 - J_j)$ takes the value 1 if all the components $i, i+1, \ldots, i+k-1$ fail and 0 in all other cases. Clearly, the $C(k, n{:}F)$ system fails if and only if $W > 0$. Although the components themselves work independently, the indicators I_i are locally dependent, tending to occur in clusters. Nonetheless, the random variable W is reasonably well approximated by a Poisson distribution Po(λ) with $\lambda = \sum_{j=1}^{n-k+1} q^k$, where $q = 1 - p$, provided that p is small. This has been established using the Stein–Chen "local" approach [BHJ (1992), (8.4.2), Chryssaphinou and Papastavridis (1990), Godbole (1991) and Roos (1993)], and the argument works equally well for more general situations, such as when the failure probabilities may differ from component to component. However, the probability $\mathbb{P}[W = 0]$ is more accurately estimated by using a Poisson approximation for the distribution of

the number of clusters, or, alternatively, a compound Poisson approximation to the distribution of W.

Some more complicated problems, which arise in the design of reliable electronic devices and in pattern detection, inspired Salvia and Lasher (1990) to consider a planar version of the above system. They introduce the *two-dimensional consecutive-k-out-of-n:F* $[C(k^2, n^2{:}F)]$ *system*, which consists of $n \times n$ components placed on a square grid; it fails if there is a square subgrid of size at least $k \times k$ with all its components failed. The exact reliability of this system is not known, and thus approximations to it are essential. Salvia and Lasher (1990) obtained bounds for the reliability of the system by relating it to certain $C(k, n{:}F)$ systems. Koutras, Papadopoulos and Papastavridis (1993) proposed bounds using results of Arratia, Goldstein and Gordon (1989, 1990), using the Stein–Chen method for Poisson approximation.

With the same assumptions about the operating behavior of the components as for the $C(k, n{:}F)$ system, we define the random variable $W = \sum_{i,\,j=1}^{n-k+1} I_{ij}$, where $I_{ij} = I$[all components in a $k \times k$ subgrid with left lowermost component (i, j) are failed]. Then the reliability of the system is just $\mathbb{P}[W = 0]$. The indicators I_{ij} and $I_{i'j'}$ are independent unless $|i - i'| \leq k - 1$ and $|j - j'| \leq k - 1$ but the local dependence between the $I_{i,\,j}$ is now frequently relatively strong. For example, the conditional probability that $I_{i+1,\,j} = 1$ given that $I_{i,\,j} = 1$ is q^k, as compared with the unconditional probability of q^{k^2}. Thus the indicators $I_{i,\,j}$ tend to occur in clusters, and the random variable W is best approximated by a compound Poisson distribution. The reliability $\mathbb{P}[W = 0]$ is then approximated by $e^{-\mathbb{E}N}$, where N is the number of clusters, rather than by the Poisson approximation $e^{-\mathbb{E}W}$, the two quantities differing inasmuch as $\mathbb{E}W = C\mathbb{E}N$, where C is the expected cluster size: see Barbour, Chryssaphinou and Roos (1995, 1996).

1.3. *Sequence matching.* Biological systematics has been revolutionized by the discovery of the double helix and the genetic code and by the development of cheap, fast and automatic sequencing machines. The degree of relationship between closely related species can now be assessed in terms of the similarity of their genomes. For more distant species, relationship at the DNA level may well have become obscured because too many random mutations have occurred since divergence, but functionally vital elements of the amino acid sequences composing their proteins are still likely to have been conserved. Thus unusual similarity between (parts of) sequences can be used as evidence for relationship. It then becomes important to be able to determine what measure of similarity can be interpreted as unusual.

The simplest model is to suppose that two finite sequences x and y of length n from the same finite alphabet $\mathscr{A}$ ($|\mathscr{A}| = 4$ for DNA, $|\mathscr{A}| = 20$ for amino acid sequences) are to be compared, which, on the null hypothesis of no relation, are each independently composed of letters drawn independently from $\mathscr{A}$ with the same probability distribution $\boldsymbol{\nu}$. A measure of relatedness

might be the count $W = \sum_{i,\,j=1}^{n-m} I_{ij}$, where

$$I_{ij} = \prod_{l=0}^{m} I[x_{i+l} = y_{j+l}],$$

the number of pairs of subsequences of length $m+1$ from the two sequences which exactly match, where m is to be suitably chosen. The indicators I_{ij} and $I_{i'j'}$ are independent unless either $|i - i'| \le m$ or $|j - j'| \le m$, so that dependence between the I_{ij} is in general weak; however, the conditional probability that $I_{i+1,\,j+1} = 1$ given that $I_{ij} = 1$ is typically substantial (at least $1/|\mathscr{A}|$), so that matching pairs tend to occur in local clusters. Hence a compound Poisson distribution is a suitable candidate for approximating the distribution of W. There are, of course, many generalizations of the model, the most important, for the purposes of practical algorithms, being to allow some degree of mismatching in the pairs of interest, through insertions and deletions of sequence segments and through the replacement of one amino acid by another similar one; see Neuhauser (1994, 1996).

2. Compound Poisson approximation.

2.1. *The Poisson process approach.* A first approach to compound Poisson approximation for sums of dependent indicators is to proceed by way of Poisson point process approximation. This is a very natural idea in the context of an underlying process consisting of rare, isolated and weakly dependent clumps of events. In such a system, the locations of the clumps, when suitably defined, occur more or less as a Poisson process on the index set Γ, and, if the sizes of the clumps are added to Γ as an extra index dimension, then the process of clump locations and sizes on $\Gamma \times \mathbb{N}$ is also almost a Poisson process. The typical strategy is to assign a location to each clump by using exactly one of the indices $\gamma \in \Gamma$ as the representative of each clump and to replace $W = \sum_{\gamma\in\Gamma} X_\gamma$ by a sum

$$W = \sum_{\gamma\in\Gamma} \sum_{l\ge 1} l I_{\gamma l}, \tag{2.1}$$

where $I_{\gamma l}$ now denotes the indicator of the event that γ is the index of the representative of a clump of size l; thus, for each clump, exactly one of the $I_{\gamma l}$ takes the value 1, and no index γ is representative of more than one clump. Poisson process approximation in total variation to the point process $\Xi = \sum_{\gamma\in\Gamma} \sum_{l\ge 1} I_{\gamma l} \boldsymbol{\delta}_l$, where $\boldsymbol{\delta}_l$ denotes the point mass at l, is then accomplished by using Stein's method for Poisson process approximation, and compound Poisson approximation in total variation to the random variable $W = \sum_{\gamma\in\Gamma} \sum_{l\ge 1} l I_{\gamma l}$, with exactly the same error estimate, follows as a consequence. There have been many successful applications of this approach, a number of which are given in Arratia, Goldstein and Gordon (1989, 1990).

To formulate the results, we introduce some notation. For each $(\gamma, l) \in \Gamma \times \mathbb{N}$, let $B(\gamma, l) \subset \Gamma \times \mathbb{N}$ be a set containing $\{\gamma\} \times \mathbb{N}$, and set

$$b_1 = \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \sum_{(\beta,j)\in B(\gamma,l)} \mathbb{E}I_{\gamma l}\mathbb{E}I_{\beta j}; \qquad b_2 = \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \sum_{\substack{(\beta,j)\in B(\gamma,l)\\(\beta,j)\neq(\gamma,l)}} \mathbb{E}(I_{\gamma l}I_{\beta j}); \tag{2.2}$$

$$b_3 = \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \mathbb{E}|\mathbb{E}\{I_{\gamma l} - \mathbb{E}I_{\gamma l}|\sigma(I_{\beta j};(\beta,j)\notin B(\gamma,l))\}|. \tag{2.3}$$

The set $B(\gamma, l)$ can be thought of as indexing the immediate neighborhood of (γ, l), and should be chosen so that the indicators $I_{\beta j}$, whose indices (β, j) do *not* belong to it, collectively have little or no influence on the value of $I_{\gamma l}$. The degree of influence from outside the immediate neighborhood, aggregated over all (γ, l), is measured by b_3. An alternative expression for it is

$$b_3 = \mathbb{E}N \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} w_{\gamma l}\mathbb{E}\left|\frac{\mathbb{P}[I_{\gamma l}=1 \mid \sigma(I_{\beta j};(\beta,j)\notin B(\gamma,l))]}{\mathbb{P}[I_{\gamma l}=1]} - 1\right|,$$

where $N := \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} I_{\gamma l}$ denotes the total number of clumps, and the $w_{\gamma l}$ are the weights $\mathbb{E}I_{\gamma l}/\mathbb{E}N$. This represents b_3 as the product of $\mathbb{E}N$ and a weighted average of a measure of the dependence of the distribution of $I_{\gamma l}$ on what happens outside its immediate neighborhood, the events in $\sigma(I_{\beta j};(\beta, j) \notin B(\gamma, l))$. Local dependence, which should also be weak for a Poisson process approximation to be good, is summarized in b_2. This quantity can be interpreted in a similar way, as the product of $\mathbb{E}N$ and a weighted average of the expected number of further clumps occurring in the immediate neighborhood of an index pair (γ, l), conditional on $I_{\gamma l} = 1$. Finally, b_1 has a similar interpretation, as $\mathbb{E}N$ times a weighted average, but now of the unconditional expected number of clumps with index pairs $(\beta, l) \in \{(\gamma, l)\} \cup B(\gamma, l)$; this is a measure of the "extent" of the neighborhoods, and should also be small if Poisson process approximation is to be accurate. Thus the choice of the sets $B(\gamma, l)$ is critical to the success of the approach.

With these definitions, the following compound Poisson estimate, derived by way of a Poisson process approximation, can be proved as in Arratia, Goldstein and Gordon [(1990), Section 4.2.1], with improved coefficients of b_1 and b_2 from BHJ [(1992), Theorem 10.A].

CPA PP. *If* $W = \sum_{\gamma\in\Gamma}\sum_{l\geq 1} lI_{\gamma l}$ *is as defined above, and* $\lambda := \sum_{\gamma\in\Gamma}\sum_{l\geq 1} \mathbb{E}I_{\gamma l} = \mathbb{E}N$ *and* $\mu_l := \lambda^{-1}\sum_{\gamma\in\Gamma}\mathbb{E}I_{\gamma l}$, $l \geq 1$, *then*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq b_1 + b_2 + b_3. \tag{2.4}$$

By choosing the sets $B(\gamma, l)$ carefully, very good results can be obtained, as long as λ is not too large.

There are two drawbacks to the point process approach. First, the identification of a unique representative for each clump ("declumping") is rarely natural and can pose difficulties. Second, if λ is large, the error estimates derived in this way are frequently far from accurate, because the error involved in

point process approximation in total variation is often much larger than that for compound Poisson approximation to $\mathscr{L}(W)$. However, the point process approach still provides flexible, useful and explicit error estimates for compound Poisson approximation.

EXAMPLE A. Let $X_{ij} = I_i I[Y_i \geq j]$, $1 \leq i \leq n$, $j \geq 1$, be a double array of indicators, in which the $I_i \sim \mathrm{Be}(p_i)$ are independent, and the $Y_i \sim \boldsymbol{\mu}^{(i)}$ are independent of each other and of the I_i. Set

$$W = \sum_{i=1}^{n} \sum_{j \geq 1} X_{ij} = \sum_{i=1}^{n} I_i Y_i.$$

"Declumping" is easily achieved by using representatives $\gamma \in \{1, 2, \ldots, n\} \times \{1\}$, denoted for short by i, and then defining $I_{il} = I_i I[Y_i = l]$ for each i, l. Then $\lambda = \sum_{i=1}^{n} p_i$ and $\mu_l = \lambda^{-1} \sum_{i=1}^{n} p_i \mu_l^{(i)}$, and, taking $B(i,l) = \{i\} \times \mathbb{N}$ for each i, it also follows that $b_1 = \sum_{i=1}^{n} p_i^2$ and $b_2 = b_3 = 0$. The Poisson process estimate CPA PP thus immediately implies that

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq \sum_{i=1}^{n} p_i^2. \tag{2.5}$$

To illustrate the implications of (2.5), let $p(n)$ be chosen in such a way that $p(n) \to 0$ and $np(n) \to \infty$ as $n \to \infty$, and consider three choices of the $p_i = p_i^{(n)}$ and $\boldsymbol{\mu}^{(i)} = \boldsymbol{\mu}^{(in)}$.

Case (a). Suppose that $p_i^{(n)} = p(n)$ and $\boldsymbol{\mu}^{(in)} = \boldsymbol{\mu}$ for all i. Then (2.5) gives a total variation error bound of $np(n)^2$ for approximation by CP $(np(n), \boldsymbol{\mu})$; however, the true error is actually much less, being at most $p(n)$.

Case (b). Suppose that the $p_i^{(n)}$ and $\boldsymbol{\mu}^{(in)}$ are as in Case (a) for $2 \leq i \leq n$, and that $\boldsymbol{\mu}$ is such that $\mu_1 > 0$; suppose also that $p_1^{(n)} = \frac{1}{2}$ and that $\boldsymbol{\mu}^{(1n)} = \boldsymbol{\delta}_1$ for all n. Then (2.5) gives an error estimate of at least $1/4$ for approximation by CP $(\lambda_n, \boldsymbol{\mu}_n)$, where

$$\lambda_n = (n-1)p(n) + \tfrac{1}{2} \quad \text{and} \quad \boldsymbol{\mu}_n = \lambda_n^{-1}\{(n-1)p(n)\boldsymbol{\mu} + \tfrac{1}{2}\boldsymbol{\delta}_1\};$$

here, the true error in fact tends to 0 with n, being of order $O(p(n) + [np(n)]^{-1})$.

Case (c). Suppose that everything is as in Case (b), except that now $\boldsymbol{\mu}\{2\mathbb{Z}_+\} = 1$, so that, in particular, $\mu_1 = 0$. In this case, the error estimate of order $O(1)$ furnished by (2.5) is of the correct order.

The contrast between Cases (b) and (c) indicates that improving upon the error estimates for compound Poisson approximation that are derived using the point process approach is likely to be a delicate matter.

2.2. *The direct approach.* If $\mathbb{E}W$ is large but finite, there is advantage to be gained by taking a direct approach by way of Stein's method for the compound Poisson distribution, introduced in Barbour, Chen and Loh (1992). Here, there is no need to rewrite $W = \sum_{\gamma\in\Gamma} X_\gamma$ in "declumped" form. Instead, for each γ, decompose W into nonnegative integer valued random variables in the form

$$W = W_\gamma + Z_\gamma + U_\gamma + X_\gamma, \tag{2.6}$$

where, for the representation to be useful, W_γ should be almost independent of (X_γ, U_γ), and U_γ and Z_γ should not be too large: the sense in which these requirements are to be interpreted becomes clear shortly. Such a decomposition is often realized by partitioning the indices $\{1, 2, \ldots, n\}$ into subsets $\{\gamma\}$, S_γ, N_γ and T_γ and setting

$$U_\gamma = \sum_{\beta\in S_\gamma} X_\beta \quad \text{and} \quad Z_\gamma = \sum_{\beta\in N_\gamma} X_\beta$$

[Roos (1994)]; S_γ contains those X_β, $\beta \neq \gamma$, which strongly influence X_γ, T_γ those X_β whose cumulative effect on (X_γ, U_γ) is negligible, and N_γ the remainder. This procedure is the analogue of the "local" approach to Poisson approximation [BHJ (1992), pages 9 and 10], which is recovered, for 0–1 valued X_γ, by taking $S_\gamma = \varnothing$ and hence $U_\gamma = 0$. Define the parameters λ and $\boldsymbol{\mu}$ of the canonical approximating compound Poisson distribution as follows.

Canonical parameters:

$$\begin{aligned} \lambda\mu_l &= \frac{1}{l}\sum_{\gamma\in\Gamma} \mathbb{E}\{X_\gamma I[X_\gamma + U_\gamma = l]\}, \qquad l \geq 1; \\ \lambda &= \sum_{l\geq 1} \lambda\mu_l = \sum_{\gamma\in\Gamma} \mathbb{E}\biggl\{\Bigl(\frac{X_\gamma}{X_\gamma + U_\gamma}\Bigr) I[X_\gamma + U_\gamma \geq 1]\biggr\}. \end{aligned} \tag{2.7}$$

Note that, if the X_γ are 0–1 valued and $S_\gamma = \varnothing$, then $\mu_1 = 1$ and all other μ_i are zero, and $\lambda = \mathbb{E}W$, all consistent with the simple Poisson approximation.

Then, setting $\pi_{jk}^{(\gamma)} = j\mathbb{P}[X_\gamma = j, U_\gamma = k]/m_{1\gamma}$, $j \geq 1$, $k \geq 0$, where $m_{1\gamma} = \mathbb{E}X_\gamma$, define the four following quantities which appear in the error estimates and which should be small for the estimates to be good:

$$\delta_1 = \sum_{\gamma\in\Gamma} m_{1\gamma} \sum_{j\geq 1}\sum_{k\geq 0} \pi_{jk}^{(\gamma)} \mathbb{E}\left|\frac{\mathbb{P}[X_\gamma = j, U_\gamma = k|W_\gamma]}{\mathbb{P}[X_\gamma = j, U_\gamma = k]} - 1\right|; \tag{2.8}$$

$$\delta_2 = 2\sum_{\gamma\in\Gamma} \mathbb{E}\bigl\{X_\gamma d_{\mathrm{TV}}(\mathscr{L}(W_\gamma|X_\gamma, U_\gamma), \mathscr{L}(W_\gamma))\bigr\}; \tag{2.9}$$

$$\delta_3 = \sum_{\gamma\in\Gamma} \mathbb{E}\bigl\{X_\gamma d_{W}(\mathscr{L}(W_\gamma|X_\gamma, U_\gamma), \mathscr{L}(W_\gamma))\bigr\}; \tag{2.10}$$

$$\delta_4 = \sum_{\gamma\in\Gamma} \bigl\{\mathbb{E}(X_\gamma Z_\gamma) + \mathbb{E}X_\gamma \mathbb{E}\{X_\gamma + U_\gamma + Z_\gamma\}\bigr\}. \tag{2.11}$$

In (2.10), the distance d_W is the Wasserstein L_1 metric on probability measures over $\mathbb{Z}_+$:

$$d_W(P, Q) = \sup_{\{f \in \mathrm{Lip}_1\}} \left| \int f\, dP - \int f\, dQ \right|,$$

where $\mathrm{Lip}_1 = \{f\colon |f(r) - f(s)| \le |r - s|, r, s \in \mathbb{Z}_+\}$.

The quantities δ_l, $1 \le l \le 4$, can be interpreted as follows. To start with, δ_4 is an analogue of $b_1 + b_2$ in (2.2), combining the effects of local dependence and neighborhood size, and it reduces to the corresponding element in the Poisson "local" bounds [BHJ (1992), (1.28)] if the X_γ are 0–1 valued and $S_\gamma = \emptyset$. Two points should be noted. First, the two weighted averages are now multiplied by the expectation $\sum_{\gamma \in \Gamma} \mathbb{E}X_\gamma$ of W, and not by the expected number λ of clumps. Second, in the term

$$\sum_{\gamma \in \Gamma} \mathbb{E}(X_\gamma Z_\gamma) = \mathbb{E}W \sum_{\gamma \in \Gamma} w'_\gamma \sum_{l \ge 1} \frac{l\mathbb{P}[X_\gamma = l]}{\mathbb{E}X_\gamma} \mathbb{E}(Z_\gamma | X_\gamma = l),$$

where the weights w'_γ are defined by $w'_\gamma := \mathbb{E}X_\gamma / \mathbb{E}W$, the average is over conditional expectations of Z_γ given the value of X_γ, and does not include any contribution from the strongly dependent U_γ; the effect of the U_γ is already accounted for in the definition of the canonical parameters (2.7).

Each of the quantities δ_1, δ_2 and δ_3 is a measure of the effect of any long range dependence on the joint distribution of (X_γ, U_γ), and can be contrasted with b_3 in (2.3). In δ_2 and δ_3, it is expressed in terms of the effect on the distribution of the "distant" W_γ exercised by the value of (X_γ, U_γ), measured either in terms of total variation or Wasserstein distance. In δ_1, the dependence is rewritten in terms of the effect on the distribution of (X_γ, U_γ) exercised by W_γ. All three can be viewed as weighted averages of measures of dependence at a distance, multiplied by $\mathbb{E}W$. For *independent* X_i, one can take $Z_i = U_i = 0$, for which choice $\delta_1 = \delta_2 = \delta_3 = 0$, and δ_4 reduces to $\sum_{i=1}^n (\mathbb{E}X_i)^2$.

REMARK. The distances $d(\mathcal{L}(W_\gamma | X_\gamma, U_\gamma), \mathcal{L}(W_\gamma))$ appearing in (2.9) and (2.10) are often bounded by constructing random variables $W_\gamma^{1,i,l}$ and $W_\gamma^{2,i,l}$ on the same probability space, for each $i \ge 1$ and $l \ge 0$, with $\mathcal{L}(W_\gamma^{1,i,l}) = \mathcal{L}(W_\gamma | (X_\gamma, U_\gamma) = (i, l))$ and $\mathcal{L}(W_\gamma^{2,i,l}) = \mathcal{L}(W_\gamma)$, in such a way that $W_\gamma^{1,i,l}$ and $W_\gamma^{2,i,l}$ are close—for instance, so that they coincide with high probability. In practice, it is often easier to make a coupling of $\mathcal{L}(W_\gamma | (X_\gamma, U_\gamma, Y_\gamma) = (i, l, y))$ and $\mathcal{L}(W_\gamma)$, where Y_γ summarizes additional information, for example the exact knowledge of $(I_\beta, \beta \in S_\gamma)$ rather than just the value of U_γ. This causes no extra difficulty, since it is always the case that

$$\mathbb{E}\{X_\gamma d(\mathcal{L}(W_\gamma | X_\gamma, U_\gamma), \mathcal{L}(W_\gamma))\} \le \mathbb{E}\{X_\gamma d(\mathcal{L}(W_\gamma | X_\gamma, U_\gamma, Y_\gamma), \mathcal{L}(W_\gamma))\}.$$

In terms of these quantities, the following estimate can be established: see Roos (1994), Barbour and Utev [(1999), Theorem 1.9], Barbour [(1999), Equations (5.13) and (5.14)].

CPA 1A. *There exist constants*

$$H_l^K = H_l^K(\lambda, \boldsymbol{\mu}) \le H_l^{\mathrm{TV}} = H_l^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}), \qquad l = 0, 1,$$

which do not involve W *in any way, such that, if* λ *and* $\boldsymbol{\mu}$ *are the canonical parameters, then*

$$\begin{aligned} d_K(\mathscr{L}(W),\ \mathrm{CP}\,(\lambda, \boldsymbol{\mu})) &\le \varepsilon_0 H_0^K + \varepsilon_1 H_1^K, \\ d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}\,(\lambda, \boldsymbol{\mu}) &\le \varepsilon_0 H_0^{\mathrm{TV}} + \varepsilon_1 H_1^{\mathrm{TV}}, \end{aligned} \tag{2.12}$$

in either bound, one can take (i) $\varepsilon_0 = \min(\delta_1, \delta_2)$ *and* $\varepsilon_1 = \delta_4$, *or* (ii) $\varepsilon_0 = 0$ *and* $\varepsilon_1 = \delta_3 + \delta_4$.

In general, when evaluating δ_2 and δ_3, it is often possible to compute the distances between distributions by means of couplings. Variant (ii), when applied with $Z_i = 0$, gives the analogue of the Poisson coupling estimate; variant (i) leads to the analogue of the Poisson local estimate [BHJ (1992), Theorems 1.B and 1.A, respectively].

If approximation by another compound Poisson distribution with parameters λ' and $\boldsymbol{\mu}'$ is preferred, a similar estimate is available [Barbour (1999), Compound Poisson Estimate 2]. One advantage of allowing distributions other than the canonical compound Poisson distribution as approximations is that the canonical distribution may be very complicated, whereas an approximation of the same order may be obtainable with a very simple compound Poisson distribution.

CPA 1B. *For any choices* λ' *and* $\boldsymbol{\mu}'$, *we have*

$$\begin{aligned} d_K(\mathscr{L}(W),\ \mathrm{CP}\,(\lambda', \boldsymbol{\mu}') &\le \varepsilon_0' H_0^{'K} + \varepsilon_1' H_1^{'K}; \\ d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}\,(\lambda', \boldsymbol{\mu}') &\le \varepsilon_0' H_0^{'\mathrm{TV}} + \varepsilon_1' H_1^{'\mathrm{TV}}, \end{aligned} \tag{2.13}$$

where $H_l^{'K} = H_l^K(\lambda', \boldsymbol{\mu}')$ *and* $H_l^{'\mathrm{TV}} = H_l^{\mathrm{TV}}(\lambda', \boldsymbol{\mu}')$ *for* $l = 0, 1$, *and where*

$$\varepsilon_0' = \varepsilon_0 + |\lambda m_1 - \lambda' m_1'| \quad \text{and} \quad \varepsilon_1' = \varepsilon_1 + \lambda m_1 d_W(Q', Q), \tag{2.14}$$

with ε_0 *and* ε_1 *as for* CPA 1A; *here* $m_1 = \sum_{l \ge 1} l\mu_l$, *and the probability measures* Q *and* Q' *on* $\mathbb{N}$ *are such that* $Q\{i\} = i\mu_i/m_1$ *and* $Q'\{i\} = i\mu_i'/m_1'$. *In particular, if* $\lambda' m_1' = \lambda m_1$, *then* $\varepsilon_0' = \varepsilon_0$.

REMARK. If $\lambda' m_1' = \lambda m_1$, then one can instead take

$$\varepsilon_1' = \varepsilon_1 + \sum_{l \ge 1} l(l-1)|\lambda' \mu_l' - \lambda \mu_l| \tag{2.15}$$

[Roos (1994), Theorem 3]. The formulas for the elements $\lambda\mu_l$ from the canonical parameters are easy to obtain from (2.7), and the alternative parameters λ' and $\boldsymbol{\mu}'$ are usually chosen for their simplicity, so that this quantity is easy to compute.

In order to exploit CPA 1A and CPA 1B, it thus remains to find suitable bounds for $H_l^K(\lambda, \boldsymbol{\mu})$ and $H_l^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})$, $l = 0, 1$. In the Poisson case, $\boldsymbol{\mu} = \boldsymbol{\delta}_1$, it is known that

$$H_0^{\mathrm{TV}}(\lambda, \boldsymbol{\delta}_1) \le \min\{1, \sqrt{2/e\lambda}\}, \qquad H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\delta}_1) \le \min\{1, \lambda^{-1}\}, \tag{2.16}$$

[BHJ (1992), Lemma 1.1.1 and Remark 10.2.4]. If bounds with similar λ-dependence could also be found for general compound Poisson distributions, the estimates of CPA 1A and CPA 1B would greatly improve upon the error estimates derived in CPA PP. The reason for this is quite simple. In the constituents δ_l, $1 \le l \le 4$, of the bounds given in CPA 1A and CPA 1B, as also in b_1, b_2 and b_3 of CPA PP, average measures of dependence at each location are multiplied by the mean of W or by λ, the mean number of clumps, each of which increases in proportion to the overall size of the system. Bounds of this magnitude are an intrinsic feature of total variation approximation for Poisson processes, and are thus unavoidable in CPA PP, but, as in the Poisson case, the same need not be true of total variation approximation to $\mathscr{L}(W)$. In particular, whenever $H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) = O(\lambda^{-1})$ is true, the elements in the estimate CPA 1A involving ε_1 can be made independent of the system size, since the factor λ^{-1} neutralizes the growth of the multiplying factor $\mathbb{E}W$ in δ_3 and δ_4: this is particularly advantageous for variant (ii) of the estimates. Unfortunately, the only known analogue of (2.16) for general $\boldsymbol{\mu}$ is the bound

$$H_l^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) \le \min\{1, (\lambda\mu_1)^{-1}\}e^{\lambda}, \quad l = 0, 1, \tag{2.17}$$

proved in Barbour, Chen and Loh (1992). This bound can be useful for small λ, but for larger λ the exponential factor rapidly makes itself felt. What is more, it is shown in Barbour and Utev (1998) that there can be no general analogue of (2.16) in which the bounds decrease with λ.

Placing some restrictions on $\boldsymbol{\mu}$, however, better bounds can be obtained. Under the condition

$$i\mu_i \ge (i+1)\mu_{i+1}, \qquad i \ge 1, \tag{2.18}$$

it follows that [Barbour, Chen and Loh(1992)]

$$\begin{aligned} H_0^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) &\le \min\left\{1, \frac{1}{\sqrt{\lambda(\mu_1 - 2\mu_2)}}\left(2 - \frac{1}{\sqrt{\lambda(\mu_1 - 2\mu_2)}}\right)\right\}, \\ H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) &\le \min\left\{1, \frac{1}{\lambda(\mu_1 - 2\mu_2)}\left(\frac{1}{4\lambda(\mu_1 - 2\mu_2)}\right.\right. \\ &\qquad \left.\left. + \log^+\{2\lambda(\mu_1 - 2\mu_2)\}\right)\right\}. \end{aligned} \tag{2.19}$$

Alternatively, if the condition

$$\theta := m_1^{-1}(m_2 - m_1) < 1/2 \tag{2.20}$$

holds, where $m_2 := \sum_{l \ge 1} l^2\mu_l$, then it follows that

$$H_0^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) \le \frac{1}{(1-2\theta)\sqrt{\lambda m_1}}, \qquad H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) \le \frac{1}{(1-2\theta)\lambda m_1}, \tag{2.21}$$

[Barbour and Xia (1999)], these latter bounds being of exactly the same order in λ as those of (2.16) for the Poisson distribution. Note that, for the canonical parameters λ and $\boldsymbol{\mu}$,

$$\theta = \sum_{\gamma\in\Gamma} \mathbb{E}(X_\gamma U_\gamma) \Big/ \sum_{\gamma\in\Gamma} \mathbb{E}X_\gamma \tag{2.22}$$

is a weighted average of the conditional expectations of the excess clump sizes U_γ at γ, given the possible positive values of X_γ; if the X_γ are indicators and the pairs (X_γ, U_γ) are identically distributed, then $\theta = \mathbb{E}(U_\gamma | X_\gamma = 1)$.

Neither of the conditions (2.18) and (2.20) allows the approximating compound Poisson to be too far from a Poisson distribution—indeed, in the latter case, it follows that $m_1 < 3/2$ and hence that $\mu_1 > 1/2$. Nonetheless, there are many applications, for example in the area of scan statistics, in which a Poisson approximation is a reasonable but crude first approximation, and approximation by a compound Poisson distribution which is not too far from the Poisson can be very much better. In such cases, the bounds (2.19) and (2.21) combined with the error estimates given in CPA 1A and CPA 1B can prove extremely effective; see, for example, (3.4), (3.7) and (3.20) below. For *Kolmogorov* distance, sharper bounds under Condition (2.18) are also available [Barbour and Xia (2000)]:

$$H_0^K(\lambda, \boldsymbol{\mu}) \le \min\left\{1, \sqrt{\frac{2}{e\lambda\mu_1}}\right\}; \qquad H_1^K(\lambda, \boldsymbol{\mu}) \le \min\left\{\frac{1}{2}, \frac{1}{\lambda\mu_1 + 1}\right\}. \tag{2.23}$$

If neither (2.18) nor (2.20) holds, there is as yet no simple fix, though the theorems in Section 2.3 frequently make it possible to obtain approximation errors of best asymptotic order, albeit with unpalatable constants.

EXAMPLE A (Continued). Define $\Gamma = \{(i, j)\colon 1 \le i \le n, j \ge 1\}$, and use the decomposition (2.6) with $U_{ij} = \sum_{l\ne j} X_{il}$ and $Z_{ij} = 0$. Then the pair (X_{ij}, U_{ij}) is independent of W_{ij}, so that $\delta_1 = \delta_2 = \delta_3 = \varepsilon_0 = 0$, and $\varepsilon_1 = \delta_4 = \sum_{i=1}^n p_i^2(\mathbb{E}Y_i)^2$; $\lambda = \sum_{i=1}^n p_i$ and $\mu_l = \lambda^{-1}\sum_{i=1}^n p_i \mu_l^{(i)}$ are as before. Then the direct estimate CPA 1A gives

$$d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}\ (\lambda, \boldsymbol{\mu})) \le H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) \sum_{i=1}^n p_i^2(\mathbb{E}Y_i)^2. \tag{2.24}$$

If condition (2.20) holds, then in Case (a) the bound (2.21) implies an error estimate of $(1-2\theta)^{-1}m_1 p(n)$, of the correct asymptotic order; in Case (b), the error estimate is less than

$$(1-2\theta)^{-1}\{m_1 p(n) + [4(n-1)m_1 p(n)]^{-1}\},$$

again of the correct asymptotic order; Case (c) is incompatible with condition (2.20).

If condition (2.18) holds with $\mu_1 > 2\mu_2$ and $m_1 < \infty$, then in Cases (a) and (b) the bound (2.19) leads to error estimates which are asymptotically slightly larger, the order being in both cases multiplied by a factor of $\log\{np(n)\}$;

again, Case (c) is impossible. If neither of conditions (2.18) and (2.20) are satisfied, the error estimate derived from (2.24) using (2.17) becomes rapidly worse as n increases and is useless if $m_1 = \infty$.

Comparison with the Poisson process estimate. If a "declumping" has been achieved, one can also use it in conjunction with CPA 1A and 1B. If $W = \sum_{\gamma\in\Gamma}\sum_{l\geq 1} lI_{\gamma l}$, as in (2.1), with $\sum_{l\geq 1} I_{\gamma l} \in \{0,1\}$ for each γ, decompose W as in (2.6), but with $\widehat{\Gamma} = \Gamma \times \mathbb{N}$ in place of Γ, taking

$$X_{\gamma l} = lI_{\gamma l}; \quad U_{\gamma l} = 0; \quad Z_{\gamma l} = \sum_{\substack{(\beta,\, j)\in B(\gamma,\, l)\\(\beta,\, j)\neq(\gamma,\, l)}} jI_{\beta j}; \quad W_{\gamma l} = \sum_{(\beta j)\notin B(\gamma,l)} jI_{\beta j}, \tag{2.25}$$

where $B(\gamma, l)$ is as for CPA PP. The canonical parameters λ and $\boldsymbol{\mu}$ defined using (2.7) are exactly as for CPA PP, and we can take $\varepsilon_0 = \delta_1 = b_3^*$ and $\varepsilon_1 = \delta_4 = b_1^* + b_2^*$ in CPA 1A and 1B, where

$$\begin{aligned} b_1^* &= \sum_{(\gamma,\, l)\in\Gamma\times\mathbb{N}} \sum_{(\beta,\, j)\in B(\gamma,l)} jl\mathbb{E}I_{\gamma l}\mathbb{E}I_{\beta j};\\ b_2^* &= \sum_{(\gamma,\, l)\in\Gamma\times\mathbb{N}} \sum_{\substack{(\beta,\, j)\in B(\gamma,l)\\(\beta,\, j)\neq(\gamma,\, l)}} jl\mathbb{E}(I_{\gamma l}I_{\beta j});\\ b_3^* &= \sum_{(\gamma,\, l)\in\Gamma\times\mathbb{N}} l\mathbb{E}|\mathbb{E}\{I_{\gamma l} - \mathbb{E}I_{\gamma l}|\sigma(I_{\beta j};\ (\beta,\, j)\notin B(\gamma,\, l))\}|. \end{aligned} \tag{2.26}$$

This gives the following estimate.

CPA 1C. *In the setting of CPA PP, for any choices λ' and $\boldsymbol{\mu}'$, we have*

$$\begin{aligned} d_K(\mathcal{L}(W),\ \mathrm{CP}(\lambda', \boldsymbol{\mu}')) &\leq \varepsilon_0' H_0^{'K} + \varepsilon_1' H_1^{'K};\\ d_{\mathrm{TV}}(\mathcal{L}(W),\ \mathrm{CP}(\lambda'\boldsymbol{\mu}')) &\leq \varepsilon_0' H_0^{'\mathrm{TV}} + \varepsilon_1' H_1^{'\mathrm{TV}}, \end{aligned} \tag{2.27}$$

where $H_l^{'K} = H_l^K(\lambda', \boldsymbol{\mu}')$ and $H_l^{'\mathrm{TV}} = H_l^{\mathrm{TV}}(\lambda', \boldsymbol{\mu}')$ for $l = 0, 1$, where b_1^*, b_2^* and b_3^* are as defined in (2.26), *and where, as in* (2.14),

$$\varepsilon_0' = b_3^* + |\lambda m_1 - \lambda' m_1'| \quad \textit{and} \quad \varepsilon_1' = b_1^* + b_2^* + \lambda m_1 d_W(Q', Q). \tag{2.28}$$

Comparing the error estimate in (2.27) to that of (2.4), note that the quantities b_t^* are larger than the corresponding b_t, because of the factors j and l. Thus CPA 1C is never better than CPA PP unless the $H_l^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})$ are small. This is the case under condition (2.20) as soon as λ becomes large, and then CPA 1C is substantially better than CPA PP; the same is typically true if condition (2.18) is satisfied. In other circumstances, CPA PP is normally preferable to CPA 1C, and the direct estimates CPA 1A and 1B are only competitive if a more advantageous decomposition of W as in (2.6) can be found, without using the declumping.

2.3. *Improved estimates.* The weakness of the estimates CPA 1A–1C when the $H_l^{\mathrm{TV}}(\lambda, \mathbf{\mu})$ are not small suggests that modification of the original Stein argument is needed. One such approach was exploited in Barbour and Utev (1998, 1999).

CPA 2A. *If W is decomposed as in* (2.6), *and if $\lambda > 2$, $\mathbf{\mu}$ and δ_l, $1 \le l \le 4$, are as in* (2.7)–(2.11), *then, for any $\lambda' > 2$ and $\mathbf{\mu}'$ satisfying $\lambda' m_1' = \lambda m_1$, and such that $\sum_{j=1} \mu_j' r^j < \infty$ for some $r > 1$ and that $\mathbf{\mu}'$ is aperiodic ($\mathbf{\mu}'\{l\mathbb{Z}_+\} < 1$ for all $l \ge 2$), we have*

$$d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}(\lambda', \mathbf{\mu}')) \le (\lambda')^{-1/2}\varepsilon_0 S_0(\mathbf{\mu}') + (\lambda')^{-1}\varepsilon_1' S_1(\mathbf{\mu}') + \mathbb{P}[W \le \phi(\mathbf{\mu}')\lambda m_1] S_2(\mathbf{\mu}'), \tag{2.29}$$

whenever $\lambda' m_1' \ge \{2(1 - \phi(\mathbf{\mu}'))\}^{-1}$, where $3/4 < \phi(\mathbf{\mu}') < 1$ and $S_l(\mathbf{\mu}') < \infty$, $0 \le l \le 2$, and where ε_0 and ε_1' are as given in (2.14) *and* CPA 1A.

The detailed way in which $\phi(\mathbf{\mu}')$ is to be chosen and in which the $S_l(\mathbf{\mu}')$ depend on the radius of convergence of the power series $\sum_{j\ge1} \mu_j' z^j$ and on the nearness of $\mathbf{\mu}'$ to being periodic are explicitly specified in Barbour and Utev (1999). The third term in (2.29) is a penalty incurred in modifying the straightforward Stein argument. Similar estimates for Kolmogorov distance are given in Barbour and Utev (1998), under less restrictive conditions on $\mathbf{\mu}'$.

Note that if a "declumping" has been achieved as in (2.1), then CPA 2A can be applied with $\varepsilon_0 = b_3^*$ and $\varepsilon_1 = b_1^* + b_2^*$, as defined in (2.26), giving the following estimate.

CPA 2B. *If W is declumped as in* (2.1) *and if $\lambda > 2$ and $\mathbf{\mu}$ are as for* CPA PP, *then, for any $\lambda' > 2$ and $\mathbf{\mu}'$ satisfying $\lambda' m_1' = \lambda m_1$, and such that $\sum_{j\ge1} \mu_j' r^j < \infty$ for some $r > 1$ and that $\mathbf{\mu}'$ is aperiodic ($\mathbf{\mu}'\{l\mathbb{Z}_+\} < 1$ for all $l \ge 2$), we have*

$$d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}(\lambda', \mathbf{\mu}')) \le (\lambda')^{-1/2}\varepsilon_0 S_0(\mathbf{\mu}') + (\lambda')^{-1}\varepsilon_1' S_1(\mathbf{\mu}') + \mathbb{P}[W \le \phi(\mathbf{\mu}')\lambda m_1] S_2(\mathbf{\mu}'), \tag{2.30}$$

whenever $\lambda' m_1' \ge \{2(1 - \phi(\mathbf{\mu}'))\}^{-1}$, where $3/4 < \phi(\mathbf{\mu}') < 1$ and $S_l(\mathbf{\mu}') < \infty$, $0 \le l \le 2$ and where $\varepsilon_0 = b_3^*$ and ε_1' is as given in (2.28).

The advantage of CPA 2A over CPA 1B is that good behavior as λ increases is obtained under much less restrictive conditions on $\mathbf{\mu}'$ than those of (2.18) and (2.20). Strong restrictions on the form of $\mathbf{\mu}'$ are replaced by requiring the existence of an exponential moment and an *aperiodic* $\mathbf{\mu}$, and the latter condition is essential. For fixed aperiodic $\mathbf{\mu}'$ having a finite exponential moment, the coefficients of ε_0 and ε_1' are of exactly the same satisfactory λ-order as in the Poisson case (2.16), and, provided that λ is reasonably large, the third term in (2.29) may be relatively unimportant. This makes for excellent asymp-

totic orders in the error estimates. The disadvantage is that the expressions for $S_l(\boldsymbol{\mu}')$ given in Barbour and Utev (1999), while explicit, are very complicated, and can be expected to be quite large in practical applications: thus error estimates of the correct asymptotic order may be obtained at the cost of unreasonably large constant factors. In the same way, CPA 2B frequently improves upon CPA PP in the asymptotic order of the error estimate, at the expense of introducing unwieldy constant factors.

EXAMPLE A (Continued). Suppose that $\sum_{j\geq 1}\mu_j z^j$ has radius of convergence greater than 1 and that $\boldsymbol{\mu}$ is aperiodic. Take λ' and $\boldsymbol{\mu}'$ to be the canonical parameters. In Cases (a) and (b), the $S_l(\boldsymbol{\mu}_n)$ are bounded uniformly in n, and $\phi(\boldsymbol{\mu}_n)$ is bounded away from 1. Hence, from (2.29), error estimates $O(p(n)+\exp\{-np(n)\alpha\})$ for some $\alpha>0$ are obtained, with Bernstein's inequality being used to derive the exponential bound for the third term in (2.29). This is of the ideal $O(p(n))$, except when $np(n)\to\infty$ very slowly with n. At first sight, it appears that the same error estimate should also follow in Case (c), contradicting the fact that the true distance in total variation is $O(1)$. The reason why this estimate is not obtained in case (c) is that the distribution $\boldsymbol{\mu}_n$ approaches a *periodic* limit $\boldsymbol{\mu}$ as $n\to\infty$, and the $S_l(\boldsymbol{\mu}_n)$ become unbounded.

In situations where asymptotic rates of approximation are of interest, both $\lambda=\lambda_n$ and $\boldsymbol{\mu}=\boldsymbol{\mu}_n$ typically vary with n. In such cases, the error estimate given in CPA 2A depends on n not only in the obvious way, through the quantities λ_n', $\varepsilon_0=\varepsilon_0^{(n)}$ and $\varepsilon_1'=\varepsilon_1'^{(n)}$, but also because the $S_l(\boldsymbol{\mu}_n')$, $l=0,1,2$, and $\phi(\boldsymbol{\mu}_n')$ depend on n through their dependence on $\boldsymbol{\mu}_n'$. This latter dependence is in general quite complicated. However, the following result, which is proved in Månsson [(1999), Proposition 2.3], is useful in showing that, for asymptotics, a single choice often suffices for the S_l and for ϕ.

PROPOSITION. *Suppose that $\boldsymbol{\nu}$ is an aperiodic probability measure on $\mathbb{N}$ and that the $\boldsymbol{\mu}_n'$ are such that, for some $r_0>1$ and $c>0$,*

(i) $$\sup_{n\geq 1}\sum_{j\geq 1}\mu_{jn}' r_0^j<\infty;$$

(ii) $$\inf_{n\geq 1}\mu_{jn}\geq c\nu_j \quad \textit{for each } j\geq 1.$$

Then

$$S_l^*:=\sup_{n\geq 1}S_l(\boldsymbol{\mu}_n')<\infty,\quad l=0,1,2 \quad\text{and}\quad 3/4<\phi^*:=\sup_{n\geq 1}\phi(\boldsymbol{\mu}_n')<1,$$

and CPA 2A *holds for each n, with S_l^* and ϕ^* in place of $S_l(\boldsymbol{\mu}_n')$ and $\phi(\boldsymbol{\mu}_n')$, whenever* $\inf_{n\geq 1}\lambda_n'>\{2(1-\phi^*)\}^{-1}$.

The proposition can then be combined with the estimate CPA 2A to give good asymptotic rates in a wide variety of problems.

3. Examples.

3.1. *Runs.*

A. *Success runs.* The problem of success runs is very well known in the literature, having already been considered by von Mises (1921) in the context of Poisson approximation. It is the simplest prototype for many problems in the general area of reliability and sequence analysis [Arratia, Goldstein and Gordon (1989, 1990), Arratia, Gordon and Waterman (1990)], and gives a good test of the effectiveness of the various compound Poisson approximations.

To formulate the problem, consider the independent identically distributed Bernoulli random variables $\xi_1, \ldots, \xi_n$, where $\mathbb{P}[\xi_i = 1] = p = 1 - \mathbb{P}[\xi_i = 0]$, $i = 1, \ldots, n$. We are interested in compound Poisson approximation to the number of k-runs of consecutive 1's. In order to avoid trivialities and edge effects, we assume that $n > 4k - 3$ and we identify all indices of the form $i + nj$ for $j \in Z$. We define $I_\gamma = \prod_{i=\gamma}^{\gamma+k-1} \xi_i$ and $W = \sum_{\gamma\in\Gamma} I_\gamma$, $\Gamma = \{1, \ldots, n\}$ and set $\psi := \mathbb{E}I_\gamma = p^k$. It is clear that the random variable W counts the number of locations among the first n at which a run of 1's of length at least k begins. It is also clear that runs of 1's occur in "clumps"; that is, if there is a run of 1's of length k beginning at position γ, then with probability p there will also be a run of 1's of length k at position $\gamma+1$, with probability p^2 a run of length k beginning at position $\gamma+2$ and so forth. This is an example, with average clump size $1 + p + p^2 + p^3 + \cdots$, of the "Poisson clumping heuristic" described by Aldous (1989).

We start by applying the Poisson process approach, as in Arratia, Goldstein and Gordon (1990). Defining the random variable $R = \sum_{\gamma\in\Gamma} X_\gamma$, where $X_\gamma = (1 - \xi_{\gamma-1})I_\gamma$, with $X_1 = I_1$ we observe that $R + \prod_{i=1}^n \xi_i$ counts the number of clumps and is approximately Poisson Po($\mathbb{E}R$) distributed with mean

$$\mathbb{E}R = \psi[(n-1)(1-p) + 1].$$

For interesting results, we want $\mathbb{E}R$ to be bounded away from 0, which is essentially the condition that $k \le \log_{1/p}(n(1-p))$. Note that if we are interested in the distribution M_n of the longest of these 1's runs, then we have $\mathbb{P}[M_n < k] = \mathbb{P}[R = 0]$.

The size of each clump minus 1 is the length by which the associated run of 1's exceeds k and is approximately distributed as a geometric random variable with parameter p. Furthermore, the clump sizes are almost independent of each other and of the total number R of clumps, so that the distribution of W is approximately Poisson Po($\mathbb{E}R$) compounded by geometric Ge(p), the Pólya–Aeppli distribution PA($\mathbb{E}R$, p): this is equivalently expressed as CP (λ, $\boldsymbol{\mu}$) with

$$\lambda = \mathbb{E}R; \qquad \mu_l = p^{l-1}(1-p), \qquad l \ge 1. \tag{3.1}$$

Although a clump size l could take any value between 1 and n, we shall simplify the calculation by considering only $l \in L = \{1, 2, \ldots, [n/2] - k - 1\}$, and declump by defining $I_{\gamma l} = (1 - \xi_{\gamma-1})\xi_\gamma\xi_{\gamma+1}\cdots\xi_{\gamma+k+l-2}(1 - \xi_{\gamma+k+l-1})$ for

$(\gamma, l) \in \Gamma \times L$; then the random variable

$$W^* = \sum_{\gamma\in\Gamma}\sum_{l\in L} lI_{\gamma l} = W - n\prod_{i=1}^{n}\xi_i - \sum_{\gamma\in\Gamma}\sum_{[n/2]-k}^{n-k} lI_{\gamma l}$$

satisfies

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathscr{L}(W^*)) \le p^n + np^{(n/2)-2}.$$

Now take

$$B(\gamma, l) = \{(\beta, j)\colon\ j \in L,\ \gamma - k - j \le \beta \le \gamma + k + l\}$$

and apply the estimate CPA PP. Simple calculations show that

$$b_1 \le n(1-p)\psi^2\{2 + (2k+1)(1-p)\};\quad b_2 \le 2n(1-p)\psi^2;\quad b_3 = 0; \tag{3.2}$$

and hence the error estimate that results from CPA PP is $O(nk\psi^2) = O(k\psi\mathbb{E}W)$.

We next turn to the direct approximations CPA 1A and 1B, as exemplified in the thesis of Roos (1993). We define neighborhoods of dependence for each $\gamma \in \Gamma$, taking into account the dependence structure of the problem:

$$\begin{aligned} S_\gamma &= \{\gamma - (k-1), \ldots, \gamma - 1, \gamma + 1, \ldots, \gamma + k - 1\},\\ N_\gamma &= \{\gamma - 2(k-1), \ldots, \gamma - k, \gamma + k, \ldots, \gamma + 2(k-1)\},\\ T_\gamma &= \Gamma\backslash\{\{\gamma\} \cup S_\gamma \cup N_\gamma\}. \end{aligned}$$

Under the above partioning of Γ, we have a decomposition of the random variable W as in (2.6), with

$$W_\gamma = \sum_{\beta\in T_\gamma} I_\beta,\quad U_\gamma = \sum_{\beta\in S_\gamma} I_\beta \quad\text{and}\quad Z_\gamma = \sum_{\beta\in N_\gamma} I_\beta.$$

We observe that $|S_\gamma| = |N_\gamma| = 2(k-1)$ and that $\{I_\beta\colon\ \beta \in S_\gamma \cup \{\gamma\}\}$ are independent of $\{I_\beta\colon\ \beta \in T_\gamma\}$ for each $\gamma \in \Gamma$, so that $\delta_2 = \delta_3 = \delta_4 = 0$. Furthermore, we have

$$\sum_{\gamma\in\Gamma}((\mathbb{E}I_\gamma)^2 + \mathbb{E}I_\gamma\mathbb{E}\{U_\gamma + Z_\gamma\}) = (4k-3)n\psi^2;$$

$$\sum_{\gamma\in\Gamma}\mathbb{E}\{I_\gamma Z_\gamma\} = 2(k-1)n\psi^2$$

and thus, by CPA 1A,

$$\begin{aligned} d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}\,(\lambda, \boldsymbol{\mu})) &\le H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})\sum_{\gamma\in\Gamma}((\mathbb{E}I_\gamma)^2\\ &\qquad + \mathbb{E}I_\gamma\mathbb{E}\{U_\gamma + Z_\gamma\} + \mathbb{E}\{I_\gamma Z_\gamma\})\\ &= H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})(6k-5)n\psi^2 \end{aligned}$$

and the canonical parameters λ and $\boldsymbol{\mu}$ are as given in (2.7). These can be explicitly computed:

$$\begin{aligned}\lambda\mu_i &= ni^{-1}\mathbb{E}\{I_k I[U_k + I_k] = i\}\\ &= np^k i^{-1}\mathbb{P}[I_1 + \cdots + I_{k-1} + I_{k+1} + \cdots + I_{2k-1} = i-1 | I_k = 1]\\ &= np^k i^{-1}\mathbb{P}[V_{k-1} + V'_{k-1} = i-1],\end{aligned}$$

where V_{k-1} and V'_{k-1} are independent and Ge(p) truncated at $k-1$ distributed random variables. Evaluating the last probability, we get

$$\lambda\mu_i = \begin{cases} n\psi p^{i-1}(1-p)^2, & \text{if } i = 1, \ldots, k-1;\\ i^{-1}n\psi\{2p^{i-1}(1-p)\\ \qquad +(2k-i-2)p^{i-1}(1-p)^2\}, & \text{if } i = k, \ldots, 2k-2;\\ (2k-1)^{-1}n\psi p^{2k-2}, & \text{if } i = 2k-1;\end{cases}$$

$$\lambda = \sum_{i=1}^{2k-1} \lambda\mu_i.$$

Finally, Roos (1993) showed that condition (2.18) is satisfied if $p \le 1/3$, so that $H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})$ can be bounded using (2.19). Writing $M = \lambda(\mu_1 - 2\mu_2) = np^k(1-p)^2(1-2p)$, we derive the error estimate

$$d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}\ (\lambda, \boldsymbol{\mu})) \le \left\{1 \wedge \frac{1}{M}\left[\frac{1}{4M} + \log^+ 2M\right]\right\}(6k-5)n\psi^2. \tag{3.3}$$

The estimate (3.3), $O(k\psi\log(n\psi))$, is a big improvement over the bound $O(nk\psi^2)$ obtained by using CPA PP and (3.2), whenever $\mathbb{E}W = n\psi$ is at all large. Furthermore, condition (2.20) is satisfied for $p < 1/5$, since $\theta \le 2p/(1-p)$, and (2.21) then gives the error estimate

$$d_{\mathrm{TV}}(\mathscr{L}(W),\ \mathrm{CP}\ (\lambda, \boldsymbol{\mu})) \le (6k-5)\psi(1-p)/(1-5p), \tag{3.4}$$

of even better asymptotic $O(k\psi)$. It is shown in Barbour, Chryssaphinou and Vaggelatou [(1999), (2.9)] that the canonical compound Poisson distribution can be replaced, using (2.15), by the Pólya–Aeppli distribution PA($n\psi(1-p)$, p) [see (3.1)], if $4k\psi(1-p)/(1-5p)$ is added to the error estimate. The Markov chain approach of Erhardsson (1999), outlined in Section 3.6, also gives an approximation of the same order when $p < 1/5$, but with a better constant factor.

If (essentially) no restriction on p is to be imposed, the estimate CPA 2A is still available; this has been applied in Eichelsbacher and Roos [(1999), Section 3.1] to give an error $O(k\psi + \exp(-\alpha_k n\psi))$ for some $\alpha_k > 0$, with an unspecified constant. The term involving the exponential comes from the third term in (2.29). Finally, Barbour and Xia (1999) examined the same problem in the special case where $k = 2$. Using a much less straightforward argument and a slightly different approximating compound Poisson distribution, they obtained an explicit error estimate $O(\psi(n\psi)^{-1/2})$, which is surprisingly even better than the best of the estimates above, which in this case is $O(\psi)$.

B. *Increasing sequences.* Consider a sequence $X_1, \ldots, X_n$ of independent random variables with the same continuous distribution F, and the events $\{X_{i-r+1} < \cdots < X_i\}$ of appearances of increasing sequences of length r, for $i = r, \ldots, n$. Pittel (1981) and Révész (1983) considered the limiting behavior of random variables closely related to the number W of appearances of the above event in a sequence of n such trials; we shall approximate its distribution by a compound Poisson distribution, using the estimate CPA 1C. This W could well be used as a test for local dependence in a supposedly random X-sequence, such as the innovations in a supposedly GARCH(1, 1) process, being used to model a financial time series, or in ARCH and ARMA processes, as well as in quality control; see Wolfowitz (1944) and Engle (1995).

We first define the indicators

$$I_i = I[X_{i-r+1} < \cdots < X_i] \text{ for } i = r, \ldots, n,$$

so that $W = \sum_{i=r}^n I_i$, with $\mathbb{E}(W) = (n - r + 1)\psi$ and $\psi := 1/r!$. Then, in order to achieve a declumping, we define the indicator random variable $I_{i,k}$ for the event "a k-clump occurs at the ith trial"; that is,

$$I_{i,k} = I[X_{i-r} > X_{i-r+1} < \cdots < X_i < \cdots < X_{i+k-1} > X_{i+k}], \qquad k \geq 1.$$

In this definition, it is assumed that there is a *doubly infinite* sequence of random variables X_i, $i \in \mathbb{Z}$, at our disposal, so that edge effects play no role. This simplifies the analysis, but introduces an error, in that we actually approximate the distribution of a new "declumped" random variable $\widehat{W} = \sum_{i=r}^n \sum_{k\geq 1} kI_{i,k}$, which is not quite the same as W; however, W differs from $\widehat{W}$ only when, in the infinite sequence, either $X_0 < X_1 < \cdots < X_r$ or $X_{n-r+1} < \cdots < X_n < X_{n+1}$, and hence

$$\mathbb{P}(W \neq \widehat{W}) \leq \frac{2}{(r+1)!} = \frac{2\psi}{r+1}. \tag{3.5}$$

The indicator $I_{i,k}$ is dependent only on the random variables $X_{i-r}, \ldots, X_{i+k}$, and is thus independent of all the indicators $I_{j,l}$ for which $j + l < i - r$ or $j - r > i + k$. This observation leads us to define the neighborhoods of dependence by

$$B(i, k) := \{(j, l): i - l - r \leq j \leq i + k + r\} \cap (\{1, \ldots, n\} \times \mathbb{N}), \tag{3.6}$$

ensuring that $b_3^* = 0$. The quantities b_1^* and b_2^*, given in (2.26), can then be bounded, and the canonical parameters λ and $\boldsymbol{\mu}$, as given for CPA PP, can be determined; taking into account the conditions (2.20) and (2.18), this leads to the following bounds:

1. If $r \geq 4$, then

$$d_{\mathrm{TV}}(\mathcal{L}(W),\ \mathrm{CP}\,(\lambda, \boldsymbol{\mu})) \leq \frac{4(r+1)^2}{((r-1)^2 - 8)}\Delta + \frac{2\psi}{r+1}; \tag{3.7}$$

2. If $r \geq 2$, then

$$(3.8)\quad d_K(\mathscr{L}(W),\ \mathrm{CP}\,(\lambda, \boldsymbol{\mu})) \leq \left\{2\mathbb{E}W \wedge \frac{4}{(\frac{r}{r+1} - \frac{1}{r+2} + (\mathbb{E}W)^{-1})}\right\}\Delta + \frac{2\psi}{r+1},$$

where

$$\Delta := r\psi\left\{\frac{r+1}{r-1}\right\}; \qquad \mathbb{E}W = (n-r+1)\psi;$$

$$\lambda := \frac{(n-r+1)r}{(r+1)!}; \qquad \mu_k := \frac{(r+1)!}{r}\left(\frac{k+r-1}{(k+r)!} - \frac{k+r}{(k+r+1)!}\right),\ k \geq 1.$$

We note that the above bounds are both $O(r\psi)$. For more details, further applications and the relevant literature; see Chryssaphinou and Vaggelatou (1999a).

3.2. *Reliability systems.*

A. *The two-dimensional consecutive-k-out-of-n:F system.* The system $C(k^2, n^2\colon F)$ consists of n^2 independent components, each with lifetime distribution F, placed on a square grid of size n. It fails if there is a square subgrid of side k with all its k^2 components failed. Let the set $\Gamma = ((i, j)\colon 1 \leq i, j \leq n-k+1)$ and let $A_\gamma \equiv A_{ij}$ denote the $k \times k$ subgrid with left lowermost component (i, j):

$$A_\gamma = \big((i+x-1, j+y-1) : x, y = 1, \ldots, k\big).$$

We fix a time T, and define the indicators $I_\gamma = I$ [all components in A_γ are failed at time T] for each $\gamma \in \Gamma$, setting $W = \sum_{\gamma\in\Gamma} I_\gamma$. The random variable W counts the number of possibly overlapping $k \times k$ squares with all components failed in the system. Clearly the reliability of the system is given by $\mathbb{P}[W = 0]$, and $\psi := \mathbb{E}I_\gamma = q^{k^2}$, where $q = 1 - F(T)$.

We first apply the estimate CPA 1A, as in Barbour, Chryssaphinou and Roos (1996), to approximate the distribution of W by an appropriate compound Poisson distribution $\mathrm{CP}\,(\lambda, \boldsymbol{\mu})$. Our first step is to define the neighborhoods of dependence S_γ, N_γ and T_γ for each $\gamma \in \Gamma$. Here, we take $S_\gamma = (\beta \in \Gamma\colon \beta \neq \gamma, |\beta \cap \gamma| = k^2 - k)$, consisting of the $k \times k$ subgrids $A_{i-1,\,j}$, $A_{i+1,\,j}$, $A_{i,\,j+1}$, $A_{i,\,j-1}$. We observe that there are $(n-k+1)^2$ possible positions of the $k \times k$ subgrid A_γ, of which 4 are corners, $4(n-k-1)$ are borders and the remaining $(n-k-1)^2$ are interior to Γ, and that then $|S_\gamma|$ is equal to 2, 3 and 4, respectively. Next we take $T_\gamma = (\gamma \in \Gamma\colon \gamma \cap \beta = \varnothing)$ for all $\beta \in S_\gamma \cup \{\gamma\}$ so that $\delta_2 = \delta_3 = \delta_4 = 0$, and assign the remaining $k \times k$ subgrids to N_γ. Since $|S_\gamma| \leq 4$, we find that $|S_\gamma| + |N_\gamma| \leq (2k+1)^2 - 1$. Then W can be written in the form

$$W = W_\gamma + Z_\gamma + U_\gamma + I_\gamma$$

as in (2.6), where $U_\gamma = \sum_{\gamma\in S_\gamma} I_\gamma$, $Z_\gamma = \sum_{\gamma\in N_\gamma} I_\gamma$ and $W_\gamma = \sum_{\gamma\in T_\gamma} I_\gamma$.

We now compute the value of δ_4 from (2.11), obtaining

$$\begin{aligned}
\sum_{\gamma\in\Gamma}(\mathbb{E}I_\gamma)^2 &= (n-k+1)^2\psi^2,\\
\sum_{\gamma\in\Gamma}\mathbb{E}I_\gamma\mathbb{E}\{U_\gamma+Z_\gamma\} &= \sum_{\gamma\in\Gamma}\sum_{\beta\in S_\gamma\cup N_\gamma}\mathbb{E}I_\gamma\mathbb{E}I_\beta\\
&\le (n-k+1)^2((2k+1)^2-1)\psi^2,\\
\sum_{\gamma\in\Gamma}\mathbb{E}\{I_\gamma Z_\gamma\} &= \sum_{\gamma\in\Gamma}\sum_{\beta\in\Gamma_\gamma^b}\mathbb{E}\{I_\gamma I_\beta\}\\
&\le (n-k+1)^2\psi^2\\
&\quad\times\left((8k-4)\psi+4\left(\sum_{r=1}^{k-1}\sum_{s=1}^{k-1}q^{k^2-rs}+\sum_{s=1}^{k-2}q^{k^2-ks}\right)\right),
\end{aligned}$$

where the complicated sums arise because of the differing overlaps possible between two $k\times k$ squares.

Next we compute the canonical parameters from (2.7), obtaining

$$\begin{aligned}
\lambda\mu_r &= \frac{1}{r}\sum_{\gamma\in\Gamma}\mathbb{E}\left\{I_\gamma I\left[I_\gamma+\sum_{\beta\in S_\gamma}I_\beta=r\right]\right\}\\
&= \frac{1}{r}\sum_{i=1}^{n-k+1}\sum_{j=1}^{n-k+1}\mathbb{E}\{I_{ij}I[I_{ij}+I_{i,\,j-1}+I_{i-1,\,j}+I_{i,\,j+1}+I_{i+1,\,j}=r]\}\\
&= r^{-1}\psi\{4\pi_1(r)+4(n-k-1)\pi_2(r)+(n-k-1)^2\pi_3(r)\}
\end{aligned}\tag{3.9}$$

for $r=1,\dots,5$, and $\lambda=\sum_{r=1}^5\lambda\mu_r$, where

$$\pi_1(r)=\mathbb{P}\left[\sum_{\beta\in S_\gamma}I_\beta=r-1|I_\gamma=1\right]=\mathbb{P}\big[\mathrm{Bi}(2,q^k)=r-1\big]$$

for the corner indicators,

$$\pi_2(r)=\mathbb{P}\left[\sum_{\beta\in S_\gamma}I_\beta=r-1|I_\gamma=1\right]=\mathbb{P}\big[\mathrm{Bi}(3,q^k)=r-1\big]$$

for the border indicators,

$$\pi_3(r)=\mathbb{P}\left[\sum_{\beta\in S_\gamma}I_\beta=r-1|I_\gamma=1\right]=\mathbb{P}\big[\mathrm{Bi}(4,q^k)=r-1\big]$$

for the interior indicators.

Hence, applying the approximation CPA 1A, we have

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \le H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})(n-k+1)^2 \times \psi\Bigg((4k^2+12k-3)\psi + 4\Bigg(\sum_{r=1}^{k-1}\sum_{s=1}^{k-1} q^{k^2-rs} + \sum_{s=1}^{k-2} q^{k^2-ks}\Bigg)\Bigg). \tag{3.10}$$

In reliability applications, it is usual to suppose that $\lambda \le \lambda m_1 = (n-k+1)^2\psi$ is small and the reliability high, in which case the bound $H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) \le e^{\lambda}$ from (2.17) is adequate.

In the above example, the computation of λ and $\boldsymbol{\mu}$ is not very complicated, since we have only to calculate five terms, and the resulting error estimate is uniformly $O(q^{2k-1})$ in $\lambda \le 1$. This provides an improvement on the $O(q^k)$ obtainable by using the Poisson "local" approach, by the factor $O(q^{k-1})$. Bigger improvements still could be obtained by expanding the set S_γ, but at the cost of more complicated calculations needed to determine λ and $\boldsymbol{\mu}$. The same approach is still valid for the case of unequal failure probabilities, when computer algebra can be used to evaluate the canonical parameters. From tables [see Barbour, Chryssaphinou and Roos (1996)], one can see that these error estimates are mostly comparable to or better than those presented by Fu and Koutras (1994), though, for $k = 2$, the Stein–Chen method is not good in either the Poisson or the compound Poisson approach.

For larger λ, relevant for reliability calculations if the system tolerates up to a large number m of failed $k \times k$ squares before it collapses, note that

$$\lambda(m_2 - m_1) = \sum_{r=1}^{5} r(r-1)\lambda\mu_r \le 4q^k\lambda m_1,$$

so that $\theta \le 4q^k$. Thus, from (2.21), we can take $(n-k+1)^2\psi H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) = (1-8q^k)^{-1}$, provided that $8q^k < 1$, yielding an error estimate which is still of asymptotic $O(q^{2k-1})$. If $8q^k \ge 1$ and λ is large, a larger neighborhood S_γ is needed, if accurate approximation is to be achieved, together with the asymptotically sharper estimate of CPA 2A. For the latter, a bound on $\mathbb{P}[W \le \phi\mathbb{E}W]$ for $\phi < 1$ is also required; Janson's (1990) inequality shows, for instance, that here

$$\mathbb{P}[W \le \phi\mathbb{E}W] \le \exp\{-\alpha\mathbb{E}W/(1+4k^2q^k)\}$$

for some $\alpha = \alpha(\phi) > 0$: see Eichelsbacher and Roos [(1999), Section 3.3].

B. *Multiple failure mode systems.* Reliability systems which are subject to more than one type of failure are also of interest; see Barlow, Hunter and Proschan (1963), Ross (1979), Satoh, Sasaki, Yuge and Yanasi (1993) and Koutras (1997). The system that we discuss here is also a generalization of consecutive-k-out-of-n: F system $C(k, n{:}\,F)$, but in a different direction from that of $C(k^2, n^2{:}\,F)$. We consider a system consisting of n linearly arranged components, in each of which any one of r defects may be present. The system

fails if, for any $1 \le i \le r$, at least k_i consecutive components have the defect of type i. We denote such a system by $C(k_1, \ldots, k_r; n\colon F)$. With the n components, we associate the sequence of random variables $X^{(n)} = \{X_1, \ldots, X_n\}$ taking values in $\mathscr{A} = \{0, 1, \ldots, r\}$, assuming it to be generated by a stationary Markov chain. The state $\{0\}$ denotes a correctly functioning component, while $\{i\}$ denotes a component defect of type i. Such a model can arise as the equilibrium distribution of a reversible nearest neighbor birth and death process, in which a correctly functioning component can become defective, or a defective component return to normal working (for example, after repair), with suitably chosen rates depending only on the states of the immediately neighbouring components: see, for example, Preston (1973). In Chryssaphinou and Vaggelatou (1999b), this model is successfully analyzed by using CPA 1C and a declumping. The techniques required are more complicated than those used above, because of the Markov dependence, and are similar in spirit to those that we use in Section 3.4, when counting the occurrences of copies of a word in a string of letters. The error bounds that they obtain are $O(\psi \log(1/\psi) \max_{1\le i\le r} k_i)$, where $\psi = \sum_{i=1}^{r} \psi_i$ and $\psi_i = \mathbb{P}[X_1 = X_2 = \cdots = X_{k_i} = i]$, and where the constants implicit in the order depend on the transition matrix of the underlying Markov chain.

C. *Connected-s systems.* Many reliability systems can be represented as graphs $G(V, E)$ with V a set of vertices (machines) and E a set of edges (connections) between them; see Chen, Hwang and Li (1993). Here, we suppose that the n vertices are independently subject to failure, with probabilities q_i, $1 \le i \le n$. The system has a collection of s-vertex "sensitive" subsets, often minimal cut sets in the underlying graph, and the system fails if there are at least m such subsets with all vertices failed. The two-dimensional consecutive-k-out-of-n:F system is a particular example of a connected-k^2 system.

Let Γ denote the set of all sensitive sets and $\gamma = (k_1, \ldots, k_s)$ its typical element, where $k_1, \ldots, k_s$ are the indices of the s vertices. Set $I_\gamma = I$ [all vertices in the set γ are failed], and $W = \sum_{\gamma\in\Gamma} I_\gamma$; then the reliability of the system is equal to $\mathbb{P}[W \le m-1]$, and $\mathbb{E}W = \sum_{\gamma\in\Gamma} \prod_{i\in\gamma} q_i$. To define neighborhoods of dependence, we first specify the minimum number R of items in common to α and γ when $\alpha \in S_\gamma$. The possible choices of R are $1, 2, \ldots, s-1$. Note that the choice of $R = s$ leads to the Poisson "local" approach. For $1 \le R \le s-1$, we define the sets:

$$
\begin{aligned}
S_\gamma(R) &= \{\alpha \in \Gamma\backslash\{\gamma\}\colon |\alpha \cap \gamma| = r,\ \ r = R, \ldots, s-1\};\\
N_\gamma(R) &= \{\beta\colon |\beta \cap \alpha| \ge 1 \ \text{ for some } \alpha \in \{\{\gamma\} \cup S_\gamma(R)\}\}\backslash\{\{\gamma\} \cup S_\gamma(R)\};\\
T_\gamma(R) &= \Gamma\backslash\{\{\gamma\} \cup \Gamma_\gamma^{vs}(R) \cup \Gamma_\gamma^{b}(R)\}.
\end{aligned}
$$

With the above definitions, and after some computations, CPA 1A implies that

$$
d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda(R), \boldsymbol{\mu}(R))) \le H_1^{\mathrm{TV}}(\lambda(R), \boldsymbol{\mu}(R)) Q(R) q_{\max}^{s-R+1} \mathbb{E}W, \tag{3.11}
$$

where

$$Q(R) = \{1 + \max_{\gamma\in\Gamma} |S_\gamma(R)| + 2 \max_{\gamma\in\Gamma} |N_\gamma(R)|\}$$

and $q_{\max}$ is the largest failure probability of an individual item.

Thus, for equal q's, if $|\Gamma|$ is large and q small, but $\mathbb{E}W = |\Gamma|\psi$ is of order 1, where $\psi := q^s$, compound Poisson approximation is reasonable with $R = 1$ if $Q(1)$ is much less than $|\Gamma|$. If, instead, $\mathbb{E}W$ is large, the approximation CPA 2A can be used to show that the error in compound Poisson approximation is $\psi Q(1) + \mathbb{P}[W \le \phi\mathbb{E}W]$ for suitable $\phi < 1$, and Janson's (1990) inequality can be used to bound the latter probability. However, if the structure of $S_\gamma(1)$ is complicated, it may be difficult to compute the canonical parameters, either numerically or by using computer algebra. In such cases, a larger value of R and correspondingly smaller $S_\gamma(R)$ can be chosen, which will however give an error estimate of higher order: we adopted this strategy in the $C(k^2, n^2\colon F)$ example, taking $R = k^2 - k$. Finally, we note that the Poisson estimate is typically poorer than the compound Poisson estimates, but that Poisson approximation may be much easier to achieve.

The above general approach is illustrated in Barbour, Chryssaphinou and Roos (1996), where the reliability of a model called "the double pipeline" is approximated by an appropriate compound Poisson using CPA 1A.

3.3. *Scan statistics.*

A. *Two-dimensional scan statistics.* As an example of the application of compound Poisson approximation to scan statistics, we take the two dimensional discrete scan statistic which was applied by Glaz (1996) to the problem of detecting minefields; see also Chen and Glaz (1996). Other applications are to be found in Glaz, Naus, Roos and Wallenstein (1994) and in Barbour and Månsson (2000).

A two-dimensional rectangular region $R = [0, L_1] \times [0, L_2]$ is inspected for the occurrence of certain events. Fix $n_1, n_2 \ge 1$, and divide R into $n_1 n_2$ subregions

$$J_{l_1,l_2} := [(l_1 - 1)h_1, l_1 h_1] \times [(l_2 - 1)h_2,\ l_2 h_2], \qquad 1 \le l_i \le n_i,$$

each of size $h_1 \times h_2$, where $h_i = L_i/n_i$; set $\Gamma = \{(i, j)\colon 1 \le i \le n_1 - k_1 + 1,\ 1 \le j \le n_2 - k_2 + 1\}$. For $\gamma \in \Gamma$, let the random variable Y_γ count the number of events that occur in the subregion J_γ. Set $B_{i,j} = \{(l_1, l_2)\colon i \le l_1 \le i + k_1 - 1,\ j \le l_2 \le j + k_2 - 1\}$ and define the random variable

$$S_\gamma = \sum_{\beta\in B_\gamma} Y_\beta, \qquad \gamma \in \Gamma,$$

which counts the total number of events in the rectangular subregion B_γ of R, which consists of $k_1 k_2$ adjacent smaller subregions. If S_γ exceeds a level m, then we will say that m events are clustered within the region. Finally, the two-dimensional discrete scan statistic is defined by

$$S_{\mathbf{k}} = \max\{S_\gamma;\ \gamma \in \Gamma\}, \qquad \mathbf{k} = (k_1, k_2).$$

It is of interest to test the null hypotheses of randomness under which it is assumed that the Y_β are independent distributed according to a binomial distribution with parameters N and p_0 versus the alternative that they are distributed according to a binomial with parameters N and p_1, with $p_1 > p_0$.

In order to deal with the testing procedure, an accurate approximation for the distribution $\mathbb{P}(S_{\mathbf{k}} \geq m)$ is necessary. A compound Poisson approximation is used for this. To derive one, we associate the random variable $S_{\mathbf{k}}$ with the random variable

$$W = \sum_{\gamma \in \Gamma} I_\gamma \quad \text{where} \quad I_\gamma = I[S_\gamma \geq m].$$

Clearly, the random variable W counts the number of sets of $k_1 \times k_2$ rectangular subregions in which at least m events occur. Thus we have

$$\mathbb{P}(W \geq 1) \ = \mathbb{P}(S_{\mathbf{k}} \geq m).$$

Assuming a Bernoulli model for the number of events in each $h_1 \times h_2$ subregion, we observe that the problem of approximating $\mathscr{L}(W)$ by a compound Poisson distribution can be approached in very much the same way as the two-dimensional consecutive k-out-of-n:F system $C(k^2, n^2{:}F)$. In fact, $C(k^2, n^2{:}F)$ is then a particular case of the two-dimensional discrete scan statistic, with $k_1 = k_2 = k$, $n_1 = n_2 = n$ and $m = k^2$. The extra generality in the choices of dimensions k and n makes essentially no difference to the argument, but allowing $m < k_1 k_2$ is a significant change, since, as m becomes smaller, the dependence between neighboring $k_1 \times k_2$ sets decreases more slowly with decreasing degree of overlap. Thus, in the notation of Section 3.2(C), a large choice of R such as $k_1 k_2 - \max\{k_1, k_2\}$ may no longer give an adequate approximation. In principle, the choice $R = 1$ would be good, but the computation of $\boldsymbol{\mu}(R)$ may only be feasible using a computer. If the Bernoulli model were replaced by the more general binomial model, an analogous approach could still be used, though the computation of $\boldsymbol{\mu}(R)$ would become still more involved.

B. *Linear conditional scan statistics.* The next example concerns one-dimensional scan statistics, used when testing for clustering among a fixed number n of points. This problem has been studied by many authors, and has been applied in a variety of fields, including geology, medicine and nuclear physics. For references, see Glaz and Balakrishnan (1999).

Let $X_1, \ldots, X_n$ be independent and identically distributed observations from the uniform distribution on the interval (0,1], and let $Y_t(w)$ be the number of X_i's contained in the scanning interval $(t, t + w]$. The scan statistic S_w, also called the linear conditional scan statistic, is defined by $S_w = \max_{0<t<1-w} Y_t(w)$. Because the points are on the line, the events $\{S_w \geq m\}$ can simply be rewritten as $\{W_{w,m} \geq 1\}$, where

$$W_{w,m} := \sum_{i=1}^{n-m+1} I_i \quad \text{and} \quad I_i := I[X_{(i+m-1)} - X_{(i)} < w],$$

with $X_{(j)}$ denoting the j'th order statistic. Thus the problem is rephrased in terms of counting short $(m-1)$-spacings. Even so, exact evaluation of the tail probabilities of the scan statistic is a complicated problem, and as the sample size n increases, even with moderate values of m and small values of w, it becomes practically impossible.

When nw is small, BHJ [(1992), Corollary 7.C.2] proved a Poisson approximation to $\mathscr{L}(\widehat{W}_{w,m})$, where $\widehat{W}_{w,m} := \sum_{i=1}^{n} \widehat{I}_i$ and

$$\widehat{I}_i := I_i,\ 1 \le i \le n-m+1;$$
$$\widehat{I}_i = I[1 + X_{(i+m-1-n)} - X_{(i)} < w];\ n-m+2 \le i \le n.$$

For these random variables, $\psi := \mathbb{E}\widehat{I}_i$ is the same for all i, being given by

$$\psi = \mathbb{P}[\mathrm{Bi}(n-1, w) \ge m-1] \sim (nw)^{m-1}/(m-1)!,$$

and thus $\mathbb{P}[W_{w,m} \ne \widehat{W}_{w,m}] \le (m-1)\psi$. The error that they obtained is $\psi^{1/(m-1)}$ and is thus unlikely to be accurate for $m \ge 3$. The reason for this is that the indicators I_i are dependent. However, indicators I_i and I_j are close to being independent if $|i-j| \ge m-1$, for m fixed and sufficiently large n. Since the indicators I_i and I_j for $|i-j| \le m-1$ are positively correlated, we expect that the 1's tend to occur in clusters, while the number of such clusters approximately follows a Poisson distribution. Thus the approximation of the distribution of $W_{w,m}$ by a suitable $\mathrm{CP}(\lambda, \boldsymbol{\mu})$ distribution arises in a natural way, giving

$$\mathbb{P}[S_w \ge m] = \mathbb{P}(W_{w,m} \ge 1) \approx 1 - \exp\{-\lambda\}.$$

Glaz, Naus, Roos and Wallenstein (1994), in work closely related to Roos (1993), use an approach based on CPA 1A to obtain a compound Poisson approximation, with error $O(\psi(1+\log^{+}\lambda))$, using $2m-1$ nonzero μ_j's; they also give a simpler approximation, involving only m nonzero μ_j's, with the same order of error, using CPA 1B. The evaluation of λ and the μ_j was accomplished using the clumping heuristic of Aldous (1989). Huffer and Lin (1997) suggested an alternative compound Poisson approximation.

3.4. *Occurrences of a word in a sequence of letters.* Let $\{\Xi_i,\ i \ge 0\}$ be independent and identically distributed random variables taking values in a set (alphabet) $\Omega = \{\omega_1, \ldots, \omega_q\}$, $q \ge 2$, with probabilities $p_s = \mathbb{P}[\Xi_i = \omega_s]$, $s = 1, \cdots, q$, $i \ge 0$. Let $A = \alpha_1 \ldots \alpha_k$ be a fixed string or word of length k. We define the random variable W which counts the overlapping appearances of the word A in the sequence $\Xi_1, \ldots, \Xi_n$. Chryssaphinou and Papastavridis (1988), using generating functions and combinatorial arguments, proved that the random variable W converges in distribution to a compound Poisson distribution under quite general conditions. BHJ (1992) examined the accuracy of Poisson approximation for $\mathscr{L}(W)$ taking into account the set of periods (see Theorem 8.F). In Example 10.4.2 of the same reference, the case of small periods was examined using a compound Poisson approximation, the accuracy of which is

always much better. Applying CPA PP, together with combinatorial arguments based on the set of principal periods of A, Chryssaphinou, Papastavridis and Vaggelatou (1999) obtained an upper bound on the total variation distance between $\mathscr{L}(W)$ and an appropriate compound Poisson distribution, but of the same accuracy as that obtained by BHJ (1992). As we shall see later, we can obtain sharper error estimates.

This model has been used to solve problems which arise in many areas. In particular, for a finite sequence $\Xi_1, \ldots, \Xi_n$ of letters taken from the alphabet $\{A, C, G, T\}$, the above mentioned results have proved useful for determining critical values for test statistics in the analysis of DNA sequences. For more literature see Arratia, Goldstein and Gordon (1989, 1990), Arratia, Martin, Reinert and Waterman (1996), BHJ (1992) and Waterman (1995).

The assumption of independent Ξ_i is not a good one for DNA sequences. In what follows, we show how to derive compound Poisson approximation for sequences modelled by Markov chains, as in Schbath (1995). We assume that the finite sequence $\Xi^n = \{\Xi_1, \ldots, \Xi_n\}$ of random variables, taking values in the alphabet $\mathscr{A} = \{A, C, G, T\}$, arises from a stationary realization of an irreducible, aperiodic, homogeneous Markov chain on the finite state space $\mathscr{A}$; the more general case of an m-order chain can be treated as a first-order chain on $\mathscr{A}^m$. Let $\Pi(\alpha_i, \alpha_{i+1})$ denote the first order transition probabilities of the chain, and $\pi(\alpha_i)$ the invariant probability of α_i.

Define the indicator random variables I_γ, with $\gamma \in \Gamma = \{1, \ldots, n-k+1\}$, by $I_\gamma = I[\Xi_\gamma = \alpha_1, \ldots, \Xi_{\gamma+k-1} = \alpha_k]$; then $W = \sum_{\gamma\in\Gamma} I_\gamma$. We say that A has period p if $\alpha_i = \alpha_{i+p}$ for all $i = 1, \ldots, k-p$, and let $\mathscr{P}(A)$ denote the set of periods of A. We also define the set of "principal periods" $\mathscr{P}'(A)$ of A, consisting of the minimal period and of those which are not a multiple of it.

Schbath (1995), applying CPA PP, proved that if the $\mathbb{E}W$ are bounded as $n \to \infty$, then $\mathscr{L}(W)$ can be approximated by a compound Poisson distribution. Her argument runs much as follows. First, replace W by a random variable W^* which is easier to analyze, defined using the entire doubly infinite stationary sequence $\{\Xi_i\}_{-\infty\le i\le\infty}$. Set

$$I_{\gamma l} := I[\text{an } l\text{-clump of A' s starts at } \gamma \text{ in the infinite sequence}];$$
$$W^* := \sum_{\gamma\in\Gamma}\sum_{l\ge 1} l I_{\gamma l}.$$

An l-clump consists of exactly l overlapping occurrences of A, whose union does not overlap any preceding or subsequent occurrence of A. Only the sequence $\Xi^{(n)}$ is observable in practice, and the definition of $I_{\gamma l}$ may involve ξ_i for indices $i \notin \{1, 2, \ldots, n\}$, so that we are actually interested in the random variable W rather than W^*. However, $W \neq W^*$ only if a copy of A in the infinite sequence overlaps one of the ends of the interval $\{1, 2, \ldots, n\}$, so that

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathscr{L}(W^*)) \le \mathbb{P}[W \neq W^*] \le 2(k-1)\psi, \tag{3.12}$$

where

$$\psi := \mathbb{E}I_\gamma = \pi(A) = \pi(\alpha_1)\prod_{i=1}^{k-1}\Pi(\alpha_i, \alpha_{i+1}).$$

The bound (3.12), of order $k\psi$, is no larger than other terms in the later estimates. Furthermore,

$$\mathbb{E}I_{\gamma l} = (1-L)^2 L^{l-1}\psi \quad \text{where} \quad L = \sum_{p\in\mathscr{P}'(\mathscr{A})}\prod_{i=1}^{p}\Pi(\alpha_i, \alpha_{i+1}), \tag{3.13}$$

so that $\mathbb{E}W^* = (n-k+1)\psi = \mathbb{E}W$.

In order to apply CPA PP to W^*, we begin by defining the neighborhoods $B(\gamma, l)$. In doing so, we introduce an integer r which determines their size and which can then be chosen to minimize the estimates obtained. Denote by $Z(\gamma, l)$ the set which contains the positions of the letters defining the random variable $I_{\gamma l}$. We say that the indices (γ, l) and (β, j) are not neighbors if the respective $Z(\gamma, l)$ and $Z(\beta, j)$ are separated by at least r positions, $r > 0$. Then the neighborhood $B_r(\gamma, l)$ is given by

$$B_r(\gamma, l) = \{(\beta, j):\ -(j+2)k - r \le \beta - \gamma \le (l+2)k + r\} \cap \Gamma \times \mathbb{N}.$$

With the above neighborhoods, computing the quantities (2.2), one obtains [Schbath (1995)]

$$b_1^{(r)} \le 2(n-k+1)(k\psi + (2k+r+1)\psi^*)\psi^*$$

and

$$b_2^{(r)} \le \frac{2}{\pi_{\inf}}(n-k+1)(k\psi + (r-k+1)\psi^*) + \frac{4}{\pi(\alpha_1)}(n-k+1)k\psi^2,$$

where

$$\psi^* = \sum_{l\ge 1}\mathbb{E}I_{\gamma l} = \psi(1-L)$$

and $\pi_{\inf} := \min_{a\in\mathscr{A}}\pi(a)$. The long range dependence term b_3 of (2.3) can be bounded using the geometric ergodicity of the Markov chain $\{\xi_i\}$ [Schbath (1995)], showing that $b_3^{(r)} = O(n^2\rho^r)$ for some $0 \le \rho < 1$. Hence, combining CPA PP with (3.12) through the triangle inequality, it follows that

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \le C_1 n(k+r+1)\psi^2 + C_2 n^2\rho^r + 2k\psi, \tag{3.14}$$

where the quantities C_1 and C_2 are $O(1)$, the parameters λ and $\boldsymbol{\mu}$ are given by

$$\lambda\mu_l = (n-k+1)\mathbb{E}I_{\gamma l}; \qquad \lambda = (n-k+1)\psi(1-L)$$

and r can be chosen at will.

To illustrate the possible asymptotics, let $n = n_m$, $A = A_m$ and $k = k_m$ all depend on $m \ge 1$, in such a way that $\psi_m := \pi(A_m) \to 0$ as $m \to \infty$. Note that the quantities ψ and ψ^* are typically geometrically small as functions of k,

so that if the mean $\mathbb{E}(W_m) = n_m\psi_m$ is kept bounded then $k_m/\log n_m$ stays bounded away from 0 and ∞. Taking $r = r_m = 3\log n_m/\log(\rho^{-1})$, the error estimate (3.14) is then $O(\psi_m \log(\psi_m^{-1}))$.

For the same sequence of words A_m, one could instead count the number of copies of A_m appearing in a sequence of letters from the Markov chain of length growing faster with m; for example, $n_m = [\psi_m^{-c}]$, for any fixed $c > 1$. In this case, the error estimate (3.14) is $O(\psi_m^{2-c} \log(\psi_m^{-1}))$, which is not even small with m if $c \geq 2$. Here we see the advantage of the improved compound Poisson estimates. If we again take $r_m = 3\log n_m/\log(\rho^{-1})$, the estimate (2.30) can be computed to be of order

$$\psi_m \log(\psi_m^{-1}) + \mathbb{P}[W \leq \phi\mathbb{E}W], \tag{3.15}$$

for some $\phi < 1$, and, for $c > 1$, the regenerative structure of the finite Markov chain can be used to show that the latter term is of smaller order. Hence the improved estimate CPA 2A yields a much better asymptotic order for $c > 1$ than does CPA PP, by a factor of ψ_m^{c-1}.

For small enough values of L, conditions (2.18) and (2.20) are satisfied: condition (2.18) if $L \leq 1/2$, allowing (2.27) to be applied using the bounds (2.19) and (2.23), and condition (2.20) if $L < 1/5$, in which case (2.27) can be applied with the bounds (2.21). Details and some numerical examples are given in Barbour, Chryssaphinou and Vaggelatou [(2000, Section 3]. The Markov chain approach of Erhardsson (1999) (see Section 3.6) can also be applied. The error estimates obtained from his theorems are not quite so explicit as those derived from (2.27), but they are of comparable accuracy when $\mathbb{E}W$ is small, and of similar $O(\psi_m \log(\psi_m^{-1}))$ in the asymptotic setting considered above. Erhardsson (1997) also considers the number of appearances of words from a prescribed set of words $A^{(i)}$, $1 \leq i \leq l$; see also Chryssaphinou, Papastavridis and Vaggelatou (2000).

Reinert and Schbath (1998) consider the joint distribution of the numbers of copies of each of a finite set of words in a sequence of n letters. Their approximations are expressed in terms of *independent* compound Poisson distributions, and are valid only under a hypothesis which restricts the possible overlapping of clumps of a single word with clumps of another. They do so by applying CPA PP. This is an example of the generality and usefulness of the point process approach; as yet, there is no bivariate analogue of the direct compound Poisson approach to use as an alternative.

3.5. *Sequence matching.* Let $\xi_1, \ldots, \xi_m$ and $\eta_1, \ldots, \eta_n$ be two independent sequences of independently chosen letters from a finite alphabet $\mathscr{A}$, the ξ_i chosen according to a distribution σ and the η_j according to ν. Fix k and set

$$I_{ij} := I[\xi_i = \eta_j,\ \xi_{i+1} = \eta_{j+1}, \ldots, \xi_{i+k-1} = \eta_{j+k-1}],$$

so that

$$W := \sum_{i=1}^{m-k+1} \sum_{j=1}^{n-k+1} I_{ij}$$

counts the number of times that pairs of matching strings of length k can be found in the two sequences. In molecular sequence applications, an observed value of W higher than that expected according to the above model would indicate evolutionary relationship between the two sequences. Previous work [Arratia, Goldstein and Gordon (1989), Neuhauser (1996)] has largely concentrated on approximating $\mathbb{P}[W = 0]$, which, by then varying k, translates into a statement about the length of the longest matching run; with this in mind, the strategy is typically to replace W by a random variable which counts distinct clumps of k-runs, and to approximate its distribution by a Poisson random variable. Here, as also in Månsson (1999), we use compound Poisson approximation to treat the whole distribution of W and to provide rather explicit estimates for the accuracy of the approximations obtained. Our approach is based on that of Månsson (1999) and also uses some refinements from Neuhauser (1996).

In order to simplify the canonical parameters (2.7) of the approximating compound Poisson distribution, we work instead with the random variable

$$W' := \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij},$$

derived from the ξ- and η-sequences by using the "torus convention" $\xi_{i+m} = \xi_i$, $\eta_{j+n} = \xi_j$ for all $i, j \in \mathbb{Z}$. Since

$$0 \le W' - W = \sum_{i=1}^{m} \sum_{j=n-k+2}^{n} I_{ij} + \sum_{i=m-k+2}^{m} \sum_{j=1}^{n-k+1} I_{ij},$$

it is immediate that

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathscr{L}(W')) \le (m + n - k + 1)(k - 1)\psi, \tag{3.16}$$

where $\psi := p^k$ and we assume that $0 < p := \sum_{\alpha \in \mathscr{A}} \sigma_\alpha \nu_\alpha < 1$. The random variable W' has expectation

$$\mathbb{E}W' = mnp^k,$$

and we are typically interested in values of k less than, say, $2\log(mn)/\log(1/p)$, so that $\mathbb{E}W'$ is not extremely small. In order to construct a suitable decomposition of the form $W' = I_{ij} + U_{ij} + Z_{ij} + W_{ij}$, as in (2.6), we note that the indicators most strongly dependent on I_{ij} are those of the form $I_{i+l,j+l}$ with $|l| \le k - 1$, so we take

$$U_{ij} = \sum_{1 \le |l| \le k-1} I_{i+l,j+l}$$

and

$$Z_{ij} = \left(\sum_{(r,s) \in N_{ij}} I_{rs} \right) - I_{ij} - U_{ij},$$

where

$$N_{ij} = \{(r,\ s);\ \min\{|r-i|, |s-j|\} \le 2(k-1)\}.$$

This yields $W' = I_{ij} + U_{ij} + Z_{ij} + W_{ij}$, in such a way that W_{ij} is independent of the pair (I_{ij}, U_{ij}), so that $\varepsilon_0 = 0$ in CPA 1A(i).

The canonical parameters for compound Poisson approximation are essentially the same as those for success runs, but with the new definition of p and with n replaced by mn:

$$\lambda\mu_i = \begin{cases} mn\psi p^{i-1}(1-p)^2, & \text{if } i = 1, \dots, k-1; \\ i^{-1}mn\psi\{2p^{i-1}(1-p) \\ \qquad +(2k-i-2)p^{i-1}(1-p)^2\}, & \text{if } i = k, \dots, 2k-2; \\ \{2k-1\}^{-1}mn\psi p^{2k-2}, & \text{if } i = 2k-1; \end{cases}$$

$$\lambda = \sum_{i=1}^{2k-1} \lambda\mu_i.$$

As before, condition (2.18) is satisfied if $p \le 1/3$, and condition (2.20) is satisfied with $(1-2\theta)^{-1} = (1-p)/(1-5p)$ if $p < 1/5$; once again, a Pólya–Aeppli PA $(mn\psi(1-p),\ p)$ approximation would contribute at most an extra $4k\psi(1-p)/(1-5p)$ to the total variation error estimate given below. To compute ε_1 of CPA 1A(i), it is immediate that

$$\mathbb{E}I_{ij}\mathbb{E}(I_{ij} + U_{ij} + Z_{ij}) \le (4k-3)(m+n)\psi^2, \tag{3.17}$$

and all that remains is to bound $\mathbb{E}(I_{ij}Z_{ij})$.

In order to express the result, we define three further quantities:

$$\begin{aligned} &q_1 := \sum_{\alpha\in\mathscr{A}} \sigma_\alpha^2\nu_\alpha; \quad q_2 := \sum_{\alpha\in\mathscr{A}} \sigma_\alpha\nu_\alpha^2; \\ &\gamma_+ := \max_{\alpha\in\mathscr{A}} \gamma_\alpha, \qquad \text{where } \gamma_\alpha := p^{-1}\sigma_\alpha\nu_\alpha, \end{aligned} \tag{3.18}$$

noting that $p\gamma_+ \ge q_i \ge p^2$, $i = 1, 2$. We then observe that there are at most $2kn$ pairs $(r,s) \in N_{ij}$ such that $|r-i| \le k-1$ and $|s-j| \ge k$, for each of which $\mathbb{E}(I_{ij}I_{rs}) \le q_2^k$, that there are at most $2kn$ pairs $(r,s) \in N_{ij}$ such that $|r-i| \ge k$ and $|s-j| \ge k$, for each of which $\mathbb{E}(I_{ij}I_{rs}) = p^{2k}$ and that there are at most $4k^2$ pairs $(r,s) \in N_{ij}$ such that $|r-i| \le k-1$ and $|s-j| \le k-1$, for each of which

$$\mathbb{E}(I_{ij}I_{rs}) \le \max\left\{ \sum_{\alpha\in\mathscr{A}} \sigma_\alpha(\sigma_\alpha\nu_\alpha)^k,\ \sum_{\alpha\in\mathscr{A}} \nu_\alpha(\sigma_\alpha\nu_\alpha)^k \right\} \le (p\gamma_+)^k.$$

Swapping the roles of r and s in the first two cases, this finally gives the bound

$$\mathbb{E}(I_{ij}Z_{ij}) \le 2k(mq_1^k + nq_2^k) + 2k(m+n)p^{2k} + 4k^2(p\gamma_+)^k. \tag{3.19}$$

Hence, in particular, adding over the mn possible pairs (i, j) and using the improved bounds (2.23) and (2.21), we deduce from CPA 1A(i) and (3.16) that

$$\begin{aligned} d_K(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) &\le \frac{1}{(1-p)^2 + \lambda^{-1}}\Delta_1 + \Delta_2, \qquad p \le 1/3; \\ d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) &\le \frac{1-p}{1-5p}\Delta_1 + \Delta_2, \qquad p < 1/5, \end{aligned} \tag{3.20}$$

where

$$\begin{aligned} \Delta_1 &:= 2k\{m(q_1/p)^k + n(q_2/p)^k + 3(m+n)p^k + 2k\gamma_+^k\}; \\ \Delta_2 &:= k(m+n)p^k. \end{aligned} \tag{3.21}$$

In the case when $p \ge 1/5$, as would be the case in the important application to the four-letter DNA alphabet with (almost) uniform distributions over the letters, the total variation estimate above cannot be used. However, it is once more possible to apply CPA 2A, to get error estimates of almost the same asymptotic order. The only essential difficulty lies in showing that the tail probability in (2.29) is small. Here, one can apply the exponential lower tail bounds of Janson [(1998, Theorem 10], which extend Suen's (1990) inequality to cover the event appearing in (2.29). This leads to the result that, whatever the value of $p < 1$, one has

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) = O(\Delta_1 + e^{-\alpha \mathbb{E}W'}), \tag{3.22}$$

uniformly in $\Delta_1 \le 1$, for some $\alpha > 0$. This makes for very good asymptotics whenever $\mathbb{E}W \to \infty$ at all fast; when this is not the case, the error estimate derived from CPA PP is usually close to best possible.

Most emphasis has previously been placed on asymptotics in which m and n tend to infinity in such a way that both $\log m/\log n$ and λ converge to finite, nonzero limits. Using (3.20) and (3.21), we can make more precise statements about how well the distribution of W is then being approximated, under less restrictive conditions (provided that p is in the permitted ranges). For m and n given, set

$$k = k_{mn} := \log(mn)/\log(1/p) - c_{mn}, \tag{3.23}$$

for any $c = c_{mn} \ge 0$, and define

$$l_1 = l_1(m, n) := \log m/\log(mn); \qquad l_2 = l_2(m, n) := 1 - l_1; \tag{3.24}$$

note also that then $\mathbb{E}W = p^{-c}$. Considering the elements of (3.21), we immediately have

$$(m+n)p^k = (m^{-1} + n^{-1})p^{-c} \tag{3.25}$$

and

$$2k\gamma_+^k \le \left\{\frac{2\log(mn)}{\log(1/p)}\right\}(mnp^c)^{-\delta_0}, \tag{3.26}$$

where $\delta_0 := \log(1/\gamma_+)/\log(1/p)$. Then, defining

$$\delta_i = \delta_i(m,n) := \left\{ \frac{\log(p/q_i)}{l_i \log(1/p)} - 1 \right\}, \qquad i = 1, 2, \tag{3.27}$$

we also have

$$(q_i/p)^k = p^{k(1+\delta_i)l_i} = p^{-c(1+\delta_i)l_i} \exp\{-l_i \log(mn)(1+\delta_i)\},$$

$i = 1, 2$, which, from the definition of l_i, and because $(1+\delta_i)l_i \le 1$ in view of (3.27) and $q_i \ge p^2$, implies that

$$m(q_1/p)^k \le m^{-\delta_1} p^{-c}; \qquad n(q_2/p)^k \le n^{-\delta_2} p^{-c}. \tag{3.28}$$

Thus Δ_1 is bounded, uniformly in $c \ge 0$, by the rather explicit formula

$$\begin{aligned} &\left\{ \frac{2\log(mn)}{\log(1/p)} \right\} \Bigg(p^{-c}[m^{-\delta_1} + n^{-\delta_2} + 3(m^{-1} + n^{-1})] \\ &\qquad + \left\{ \frac{2\log(mn)}{\log(1/p)} \right\} (mnp^c)^{-\delta_0} \Bigg), \end{aligned} \tag{3.29}$$

which is small so long as $\delta_1(m,n)$ and $\delta_2(m,n)$ are sufficiently positive and c is not too large; Δ_2 is small if Δ_1 is. In asymptotic terms, one would require that

$$l_i^* := \limsup_{m,\, n \to \infty} l_i(m,n) < \frac{\log(p/q_i)}{\log(1/p)}, \tag{3.30}$$

since then

$$\liminf_{m,n\to\infty} \delta_i \ge \frac{\log(p/q_i)}{l_i^* \log(1/p)} - 1 > 0, \qquad i = 1, 2,$$

and then ensure that c was at most some suitably small multiple of $\log n$, corresponding to a growth in $\mathbb{E}W$ of order at most $(mn)^\delta$, for some small $\delta > 0$. Previous asymptotics have mostly assumed that $\mathbb{E}W$ remains fixed, so that this last condition was automatically satisfied.

As already observed in Neuhauser (1996), condition (3.30) is stronger than is actually necessary for the approximation to be accurate asymptotically. To see why this is, write I_{ij} in the form

$$\sum_{\mathbf{k} \in S_k} I_{ij;\mathbf{k}}, \tag{3.31}$$

where

$$S_k := \left\{ \mathbf{k} \in \mathbb{Z}_+^{|\mathscr{A}|} : \sum_{\alpha \in \mathscr{A}} k_\alpha = k \right\}; \qquad I_{ij;\mathbf{k}} := \prod_{\alpha \in \mathscr{A}} I[K_{ij;\alpha} = k_\alpha],$$

and where

$$K_{ij;\alpha} := \#\{l\colon 0 \le l \le k-1,\ \xi_{i+l} = \eta_{j+l} = \alpha\}, \qquad \alpha \in \mathscr{A}.$$

Then it is possible for $\mathbb{E}(Z_{ij} \mid I_{ij;\mathbf{k}} = 1)$ to be very large for values of $\mathbf{k}$ for which $\mathbb{E}I_{ij;\mathbf{k}}$ is small, with a significant contribution to $\mathbb{E}(I_{ij}Z_{ij})$ resulting. In

such circumstances, the element $m(q_1/p)^k + n(q_2/p)^k$ in Δ_1 may be unduly pessimistic. This element derives from the inequality

$$\begin{aligned} &|g(l + Z_{ij} + W_{ij}) - g(l + W_{ij})| I_{ij} I[U_{ij} = l - 1] \\ &\quad \le \|\Delta g\| Z_{ij} I_{ij} I[U_{ij} = l - 1], \end{aligned} \tag{3.32}$$

used in deriving the basic compound Poisson estimates, which may give a larger result than the alternative inequality

$$\begin{aligned} &|g(l + Z_{ij} + W_{ij}) - g(l + W_{ij})| I_{ij} I[U_{ij} = l - 1] \\ &\quad \le 2\|g\| I[Z_{ij} \ge 1] I_{ij} I[U_{ij} = l - 1] \end{aligned}$$

if Z_{ij} is likely to be large when $I_{ij} = 1$. So instead, using (3.31), we can use both inequalities to arrive at the bound

$$\begin{aligned} &\left| \mathbb{E}\left\{ I_{ij} g(W) - \sum_{l \ge 1} I_{ij} I[U_{ij} = l - 1] g(l + W_{ij}) \right\} \right| \\ &\quad \le \|\Delta g\| \sum_{\mathbf{k} \in B} \mathbb{E}(I_{ij;\mathbf{k}} Z_{ij}) + 2\|g\| \sum_{\mathbf{k} \notin B} \mathbb{E} I_{ij;\mathbf{k}}, \end{aligned} \tag{3.33}$$

where $B \subset \mathbb{Z}_+^{|\mathscr{A}|}$ can be chosen at will; we take

$$B := \{\mathbf{k} \in S_k \colon |k_\alpha - k\gamma_\alpha| \le \varepsilon k \gamma_\alpha,\ \alpha \in \mathscr{A}\},$$

for a suitably chosen $\varepsilon > 0$.

The random vector $\mathbf{K}_{ij}$, conditional on $I_{ij} = 1$, has the multinomial distribution $\mathrm{MN}(k; \gamma_\alpha,\ \alpha \in \mathscr{A})$, and hence

$$\sum_{\mathbf{k} \notin B} \mathbb{E}_{ij;\mathbf{k}} \le \psi \sum_{\substack{\alpha \in \mathscr{A} \\ \gamma_\alpha \ne 0}} \mathbb{P}[|K_{ij;\alpha} - k\gamma_\alpha| > \varepsilon k \gamma_\alpha],$$

thus giving a contribution to (3.33) of at most

$$2\|g\| \psi \sum_{\substack{\alpha \in \mathscr{A} \\ \gamma_\alpha \ne 0}} \left\{ \varepsilon^{-1}(1+\varepsilon) \exp\{-\lambda \varepsilon^2 / 2(2+\varepsilon)\} + 2\varepsilon^{-1} \exp\{-\lambda \varepsilon^2 / 2(2-\varepsilon)\} \right\}, \tag{3.34}$$

if $k\varepsilon \ge 2$, this last from the tail bounds for the binomial distribution given in BHJ [(1992), Proposition A.2.5]. For $\mathbf{k} \in B$, splitting N_{ij} as before, only the $2kn$ pairs (r, s) with $|r - i| \le k - 1$ and $|s - j| \ge k$ and the $2km$ pairs with $|r - i| \ge k$ and $|s - j| \le k - 1$ need to be estimated differently. For the first set, we have

$$\mathbb{E}(I_{rs} \mid I_{ij;\mathbf{k}} = 1) \le \prod_{\alpha \in \mathscr{A}} \nu_\alpha^{k\gamma_\alpha(1-\varepsilon)}$$

for each (r, s), giving a total contribution to (3.33) of at most

$$\|\Delta g\|\, 2kn\, \psi\, \exp\{-k(1-\varepsilon) v_2 \log(1/p)\}, \tag{3.35}$$

where

$$v_1 := 1 - H(\sigma \mid \gamma)/\log(1/p), \qquad v_2 := 1 - H(\nu \mid \gamma)/\log(1/p)$$

and

$$H(\rho \mid \gamma) := \sum_{\alpha \in \mathscr{A}} \gamma_\alpha \log(\gamma_\alpha/\rho_\alpha) \geq 0.$$

The second set contributes at most

$$\|\Delta g\| 2km\psi \exp\{-k(1-\varepsilon)v_1 \log(1/p)\} \tag{3.36}$$

to (3.33). The sum of the three contributions (3.34)–(3.36) is then added over the $(m-k+1)(n-k+1)$ possible pairs (i, j), and the improved bounds for $\|g\|$ and $\|\Delta g\|$ are applied. This yields alternative error estimates in place of (3.20):

$$\begin{aligned} d_K(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) &\leq \frac{1}{(1-p)^2 + \lambda^{-1}} \Delta_1' + \Delta_2 \\ &\quad + C_K(p)\{mn\psi\}^{1/2}\Delta_3, \qquad p \leq 1/3; \\ d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) &\leq \frac{1-p}{1-5p}\{\Delta_1' + \Delta_2 + 2\{mn\psi\}^{1/2}\Delta_3\}, \\ &\qquad\qquad p < 1/5, \end{aligned} \tag{3.37}$$

where

$$\begin{aligned} \Delta_1' &:= 2k\{m \exp\{-k(1-\varepsilon)v_1 \log(1/p)\} \\ &\qquad + n \exp\{-k(1-\varepsilon)v_2 \log(1/p)\} + 3(m+n)p^k + 2k\gamma_+^k\}; \\ \Delta_2 &:= \sum_{\substack{\alpha \in \mathscr{A} \\ \gamma_\alpha \neq 0}} \big\{\varepsilon^{-1}(1+\varepsilon)\exp\{-k\gamma_\alpha\varepsilon^2/2(2+\varepsilon)\} \\ &\qquad + 2\varepsilon^{-1}\exp\{-k\gamma_\alpha\varepsilon^2/2(2-\varepsilon)\}\big\} \end{aligned} \tag{3.38}$$

Δ_2 is as in (3.21) and $C_K(p) := 2^{3/2}\{e((1-p)^2 + \lambda^{-1})\}^{-1/2}$.

In the circumstances illustrated in (3.23), provided that $l_i < v_i$, $i = 1, 2$, a suitable choice for ε is given by

$$\varepsilon = 1 - \max\left\{\sqrt{\frac{l_1}{v_1}}, \sqrt{\frac{l_2}{v_2}}\right\}. \tag{3.39}$$

This has the effect of replacing the exponents δ_1 and δ_2 in (3.29) by $\varepsilon/(1-\varepsilon)$, and adding the element involving Δ_3; this alternative bound can then be used whenever $k\varepsilon \geq 2$. In asymptotic terms, it is now enough that $l_i^* < v_i$, $i = 1, 2$, which is a less restrictive condition than (3.30), to allow the choice of

$$\varepsilon^* = 1 - \max\left\{\sqrt{\frac{l_1^*}{v_1}}, \sqrt{\frac{l_2^*}{v_2}}\right\} \tag{3.40}$$

for ε, in which case the overall error estimate is $O(p^{-c}n^{-\delta}) = O(\mathbb{E}Wn^{-\delta})$ for some $\delta > 0$; this once again converges to zero, as long as $\mathbb{E}W$ only grows with n at most as fast as a small power of n.

3.6. *Markov chains.* Counting both success runs and occurrences of a word can be rephrased as particular cases of the more general problem of counting the number of visits W to a "rare" set S_1 in n steps of a recurrent Markov chain X. For success runs, as in Section 3.1, the state space $\mathscr{S}$ can be taken to be $\{0, 1, \ldots, k\}$, by setting $X_i = j$ if $\xi_i = \xi_{i-1} = \cdots = \xi_{i-j+1}$ and $\xi_{i-j} = 0$, $1 \le j \le k-1$ and $X_i = k$ otherwise; then, apart from edge effects, the number of success runs of length k becomes the number of visits of X to $S_1 = \{k\}$. For copies of a single word $A = a_1 a_2 \cdots a_k$ of length k in the Markov model of Section 3.4, set $X_i = j$ if

$$\max(0, \{l\colon \xi_{i-l+1}, \ldots, \xi_i = a_1 a_2 \cdots a_l\}) = j \in \{1, 2, \ldots, k\},$$

and set $X_i = (\xi_i, 0)$ otherwise; the number of copies of A is then the number of visits to $S_1 = \{k\}$. When visits to S_1 generally occur singly, a Poisson approximation to $\mathscr{L}(W)$ is appropriate [BHJ (1992), Section 8.5], but it is more often the case that there is a tendency for visits to occur in clumps, and then compound Poisson approximation gives much sharper results.

This problem has been studied in some generality by Erhardsson (1999), exploiting the regenerative structure of a recurrent Markov chain to derive very pleasing approximation theorems. The Markov chain X is assumed to be stationary and Harris recurrent on $\mathscr{S}$, having a unique stationary distribution ν, and $W = \sum_{i=1}^{n} 1_{\{X_i \in S_1\}}$ counts the number of visits to $S_1 \in \mathscr{S}$: define $\psi := \nu(S_1)$. The approximating compound Poisson distribution $\mathrm{CP}(\lambda, \boldsymbol{\mu})$ is defined in terms of regeneration cycles; λ is the expected number of cycles in $1, 2, \ldots, n$ which contain at least one visit to S_1, and $\boldsymbol{\mu}$ is the conditional distribution of the number of visits to S_1 in a cycle, given that at least one occurs. In particular, if S_1 is an atom—for example, a single state—then $\boldsymbol{\mu}$ is geometric, and $\mathrm{CP}(\lambda, \boldsymbol{\mu})$ is a Pólya–Aeppli distribution. A formal definition of λ and $\boldsymbol{\mu}$ is given in Erhardsson [(1999), Definition 3.2].

The simplest theorem is obtained when regeneration is defined in terms of visits to an atom S_0 such that $\nu(S_0) > 0$ and $S_0 \cap S_1 = \varnothing$. In the example of success runs, the choice $S_0 = \{0\}$ is appropriate; for the occurrence of a word, any singleton of the form $\{(a, 0)\}$ could be used, or one could take $S_0 = \cup_{a \in A'} \{(a, 0)\}$, for any collection $A' \subset \mathscr{A}$ such that $\pi(a, \cdot)$ is the same for all $a \in A'$. Let τ_{S_0} and τ_{S_1} denote the first times that X hits S_0 and S_1, $\tau^R_{S_0}$ the first time that the *reversed* chain hits S_0. Then Erhardsson uses CPA 1A combined with a coupling argument to prove that

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \\ &\quad \le 2H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu}) n \psi^2 \{\mathbb{E}_{\nu|S_1}(\tau_{S_0} + \tau^R_{S_0}) + \nu^{-1}(S_0)\mathbb{E}_\nu(\tau_{S_0})\} \\ &\qquad + 2\mathbb{P}_\nu[\tau_{S_1} < \tau_{S_0}]. \end{aligned} \tag{3.41}$$

Good bounds for $H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})$ are currently only known under either of conditions (2.18) and (2.20). In the examples of word counts and success runs, when condition (2.20) holds, the results obtained for total variation approximation are nonetheless of the best asymptotic order generally known. Erhardsson

(1999) shows that (2.18) holds whenever

$$\nu \operatorname*{ess\,sup}_{x \in S_1} \mathbb{P}_x[\tau_{S_1} < \tau_{S_0}] \le \{3 - 2\nu \operatorname*{ess\,inf}_{x \in S_1} \mathbb{P}_x[\tau_{S_1} < \tau_{S_0}]\}^{-1},$$

and that then

$$\lambda(\mu_1 - 2\mu_2) \ge n\psi\{1 - 4\mathbb{P}_{\nu|S_1}[\tau_{S_1} < \tau_{S_0}]\},$$

enabling $H_1^{\mathrm{TV}}(\lambda, \boldsymbol{\mu})$ to be effectively bounded using (2.19) in these circumstances. More particularly, if S_1 is an atom, as is the case in the examples of success runs and occurrences of a word, then $\mu_j = (1-p)p^{j-1}$ for $j \ge 1$, where $p = \mathbb{P}_{S_1}[\tau_{S_1} < \tau_{S_0}]$, and thus condition (2.18) holds for $p \le 1/2$ with $\mu_1 - 2\mu_2 = (1-p)(1-2p)$, and the approximating distribution is then a Pólya–Aeppli distribution. However, the regenerative structure of such a Markov chain also lends itself well to proving good bounds for the quantity $\mathbb{P}[W \le \phi\lambda m_1]$ in CPA 2A, so that it is to be expected that the better asymptotic order normally given by CPA 2A could be achieved here, too, when condition (2.18) fails to hold.

Acknowledgment. ADB thanks the Department of Mathematics and Statistics at Monash University for their warm hospitality while part of this paper was being written.

REFERENCES

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

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

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5** 403–34.

ARRATIA, R., GORDON, L. and WATERMAN, M. S. (1990). The Erdős–Rényi law in distribution, for coin tossing and sequence matching. *Ann. Statist.* **18** 539–570.

ARRATIA, R., MARTIN, D., REINERT, G. and WATERMAN, M. S. (1996). Poisson process approximation for sequence repeats and sequencing by hybridization. *J. Comp. Biol*. **3** 425–463.

BARBOUR, A. D. (1999). Topics in Poisson approximation. *Handbook of Statistics*. To appear.

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., CHRYSSAPHINOU, O. and ROOS, M. (1995). Compound Poisson approximation in reliability theory. *IEEE Trans. Reliability* **44** 398–402.

BARBOUR, A. D., CHRYSSAPHINOU, O. and ROOS, M. (1996). Compound Poisson approximation in systems reliability. *Naval Res. Logist.* **43** 251–264.

BARBOUR, A. D., CHRYSSAPHINOU, O. and VAGGELATOU, E. (2000). Applications of compound Poisson approximation. In *Statistical Distribution Theory and Inference* (N. Balakrishnan, M. V. Koutras and C. Charalambides, eds.) CRC Press, Boca Raton, FL.

BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.

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

BARBOUR, A. D. and MÅNSSON, M. (2000). Compound Poisson approximation and the clustering of random points. *Adv. in Appl. Probab.* **32** 19–38.

BARBOUR, A. D. and UTEV, S. (1998). Solving the Stein equation in compound Poisson approximation. *Adv. in Appl. Probab.* **30** 449–475.

BARBOUR, A. D. and UTEV, S. (1999). Compound Poisson approximation in total variation. *Stochastic Process. Appl.* **82** 89–125.

BARBOUR, A. D. and XIA, A. (2000). Estimating Stein's constants for compound Poisson approximation. *Bernoulli*. **6** 581–590.

BARBOUR, A. D. and XIA, A. (1999). Poisson perturbations. *ESAIM: P&S* **3** 131–150.

BARLOW, R. E., HUNTER, L. C. and PROSCHAN, F. (1963). Optimum redundancy when components are subject to two kinds of failure. *J. Soc. Indust. Appl. Math.* **11** 64–73.

CHAO, M. T., FU, J. C. and KOUTRAS, M. V. (1995). A survey of the reliability studies of consecutive-k-out-of-n:F system and related systems. *IEEE Trans. Reliability* **44** 120–127.

CHEN, J. and GLAZ, J. (1996). Two-dimensional discrete scan statistics. *Statist. Probab. Let.* **31** 59–68.

CHEN, R. W., HWANG, F. K. and LI, W.-C. W. (1993). Consecutive-2-out-of-n-:F systems with node and link failures. *IEEE Trans. Reliability* **42** 497–502.

CHRYSSAPHINOU, O. and PAPASTAVRIDIS, S. (1988). A limit theorem for the number of overlapping appearances of a pattern in a sequence of independent trials. *Probab. Theory Related Fields* **79** 129–143.

CHRYSSAPHINOU, O. and PAPASTAVRIDIS, S. (1990). Limit distribution for a consecutive-k-out-of-n:F system. *Adv. in Appl. Probab.* **22** 491–493.

CHRYSSAPHINOU, O., PAPASTAVRIDIS, S. and VAGGELATOU, E. (1999). On the number of appearances of a word in a sequence of i.i.d. trials. *Meth. Comp. Appl. Probab.* **1** 329–348.

CHRYSSAPHINOU, O., PAPASTAVRIDIS, S. AND VAGGELATOU, E. (2000). Poisson approximation for the nonoverlapping appearances of several words in Markov chains. *Combin. Probab. Comput.* To appear.

CHRYSSAPHINOU, O. and VAGGELATOU, E. (1999a). Compound Poisson approximation for the numbers of long increasing sequences. *J. Appl. Probab.* To appear.

CHRYSSAPHINOU, O. and VAGGELATOU, E. (1999b). Compound Poisson approximation for multiple runs in a Markov chain. *Ann. Inst. Statist. Math.* To appear.

EICHELSBACHER, P. and ROOS, M. (1999). Compound Poisson approximation for dissociated random variables via Stein's method. *Combin. Probab. Computing* **8** 335–346.

ENGLE, R. (1995). *ARCH*. Oxford Univ. Press.

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. Appl. Probab.* **27** 565–596.

FU, J. C. and KOUTRAS, M. V. (1994). Poisson approximations for two-dimensional patterns. *Ann. Inst. Statist. Math.* **46** 179–192.

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

GLAZ, J. (1996). Discrete scan statistics with applications to minefield detection. In *SPIE* **2765** 420–429.

GLAZ, J. and BALAKRISHNAN, N. (1999). *Scan Statistics and Applications.* Birkhäuser, Boston.

GLAZ, J., NAUS, J., ROOS, M. and WALLENSTEIN, S. (1994). Poisson approximations for the distribution and moments of ordered m-spacings. *J. Appl. Probab.* **31** 271–281.

GODBOLE, A. (1991). Poisson approximations for runs and patterns of rare events. *Adv. in Appl. Probab.* **23** 851–865.

GODBOLE, A. P. and SCHAFFNER, A. A. (1993). Improved Poisson approximations for word patterns. *Adv. in Appl. Probab.* **25** 334–347.

GOOVAERTS, M. J. and DHAENE, J. (1996). The compound Poisson approximation for a portfolio of dependent risks. *Insurance Math. Econom.* **18** 81–85.

HUFFER, F. and LIN, C. T. (1997). Approximating the distribution of the scan statistic using moments of the number of clumps. *J. Amer. Statist. Soc.* **92** 1466–1475.

JANSON, S. (1990). Poisson approximation for large deviations. *Random Structures Algorithms* **1** 221–229.

JANSON, S. (1998). New versions of Suen's correlation inequality. *Random Structures Algorithms* **13** 467–483.

KOUTRAS, M. V. (1997). Consecutive-k, r-out-of-n:DFM systems. *Microelectron. Reliab.* **37** 597–603.

KOUTRAS, M. V., PAPADOPOULOS, G. K. and PAPASTAVRIDIS, S. G. (1993). Reliability of two-Dimensional consecutive-k-out-of-n:F systems. *IEEE Trans. Reliability* **44** 658-661.

LE CAM, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197.

MÅNSSON, M. (1999). On compound Poisson approximation for sequence matching. Preprint.

MICHEL, R. (1988). An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. *ASTIN Bulletin* **17** 165–169.

NEUHAUSER, C. (1994). A Poisson approximation for sequence comparisons with insertions and deletions. *Ann. Statist.* **22** 1603–1629.

NEUHAUSER, C. (1996). A phase transition for the distribution of matching blocks. *Combin. Probab. Comput.* **5** 139–159.

PITTEL, B. G. (1981). Limiting behavior of a process of runs. *Ann. Probab.* **9** 119–129.

PRESTON, C. J. (1973). Generalized Gibbs states and Markov random fields. *Adv. in Appl. Probab.* **5** 242–261.

REINERT, G., and SCHBATH, S. (1998). Compound Poisson and Poisson process approximations for occurrences of multiple words in Markov chains. *J. Comp. Biol.* **5** 223–253.

RÉVÉSZ, P. (1983). Three problems on the lengths of increasing runs. *Stochastic Process. Appl.* **15** 169–179.

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.

ROSS, S. (1979). Multivalued state component systems. *Ann. Probab.* **7** 379–383.

SALVIA, A. A., and LASHER, W. C. (1990). Two-dimensional consecutive-k-out-of-n:F models. *IEEE Trans. Reliability* **39** 382–385.

SATOH, N., SASAKI, M., YUGE, T. and YANASI, S. (1993). Reliability of three state device systems. *IEEE Trans. Reliability* **42** 470–477.

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

SUEN, W. C. S. (1990). A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures Algorithms* **1** 231–242.

VON MISES, R. (1921). Das Problem der Iterationen. *Z. Angew. Math. Mech.* **1** 298–307.

WATERMAN, M. S. (1995). *Introduction to Computational Biology.* Chapman and Hall, London.

WOLFOWITZ, J. (1944). Asymptotic distribution of runs up and down. *Ann. Math. Statist.* **15** 163–172.

APPLIED MATHEMATICS
UNIVERSITY OF ZÜRICH
WINTERTHURERSTRASSE 190
CH-8057 ZÜRICH
SWITZERLAND
E-MAIL: adb@amath.unizh.ch

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ATHENS PANEPISTIMIOUPOLIS
GR-157 84, ATHENS
GREECE
E-MAIL: ocrysaf@cc.uoa.gr