Solving the Stein Equation in compound Poisson approximation.

A. D. Barbour*

Abteilung für Angewandte Mathematik, Universität Zürich-Irchel, Winterthurerstrasse 190, CH 8057 Zürich, Switzerland.

Sergey Utev*

Institute of Mathematics, Novosibirsk University, Universitetsky pr. 4, Novosibirsk 630090, Russia.

ABSTRACT

The accuracy of compound Poisson approximation can be estimated using Stein's method in terms of quantities similar to those which must be calculated for Poisson approximation. However, the solutions of the relevant Stein equation may in general grow exponentially fast with the mean number of 'clumps', leading to many applications in which the bounds are of little use. In this paper, we introduce a method for circumventing the difficulty. We prove good bounds for those solutions of the Stein equation which are needed to measure the accuracy of approximation with respect to Kolmogorov distance, but only in a restricted range of the argument; the restriction on the range is then compensated by a truncation argument. Examples are given to show that the method clearly outperforms its competitors, as soon as the mean number of clumps is even moderately large.

AMS subject classification: 62E17, 60F05

Keywords: Stein's method, compound Poisson, distributional approximation

* Supported in part by Schweizerischer Nationalfonds Projekte Nr. 20-31262.91 and 21-37354.93. ADB acknowledges the warm hospitality of the Statistics Section at Monash University, where part of this work was undertaken.

1. Introduction

The Stein–Chen method is a powerful tool for quantifying the accuracy of a Poisson approximation to the distribution of a sum W of Bernoulli random variables. All that is needed is to show that

$$|\mathbb{E}\{\lambda g(W+1) - Wg(W)\}| \leq \varepsilon_0 M_0(g) + \varepsilon_1 M_1(g) \tag{1.1}$$

for all bounded functions $g : \mathbb{N} \to \mathbb{R}$ and for some (small) ε_0 and ε_1, where

$$M_0(g) = \sup_{j\geq 1} |g(j)|; \quad M_1(g) = \sup_{j\geq 1} |g(j+1) - g(j)|; \tag{1.2}$$

this turns out in many applications to be quite easy (see Arratia, Goldstein and Gordon (1990) and Barbour, Holst and Janson (1992)). It then follows that

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}\,(\lambda)) \leq \varepsilon_0 H_0 + \varepsilon_1 H_1, \tag{1.3}$$

where the constants H_0 and H_1 which multiply the small quantities ε_0 and ε_1 are defined by $H_l = \sup_{A\subset\mathbb{Z}_+} M_l(g_A)$, $l = 0, 1$, g_A being the solution of the Stein Equation

$$\lambda g(w+1) - wg(w) = 1_{\{w\in A\}} - \mathrm{Po}\,(\lambda)\{A\}. \tag{1.4}$$

Now H_0 and H_1 only depend on λ, and can be computed once and for all: the bounds

$$H_0 \leq \min(1, \lambda^{-1/2}) \quad \text{and} \quad H_1 \leq \lambda^{-1}(1 - e^{-\lambda}) \tag{1.5}$$

are proved in Barbour, Holst and Janson (1992), Lemma 1.1.1. The fact that these bounds on H_0 and H_1 decrease as λ increases greatly improves their usefulness when λ is large. In particular, in the simplest possible setting, the Poisson $\mathrm{Po}\,(np)$ approximation of the binomial distribution $\mathrm{Bi}\,(n, p)$, it is easy to show that (1.1) is satisfied with $\varepsilon_0 = 0$ and $\varepsilon_1 = np^2$, leading in (1.3) to an error bound of $np^2 H_1$. If H_1 did not vary with np, the error would only have been proved to be small for $p = o(n^{-1/2})$. Since (1.5) shows that $H_1 \leq 1/np$, we actually have an error bound of p, which is small whenever $p = o(1)$, and which is sharp up to a constant multiple.

For many sums W of dependent indicator random variables, there is sufficient local dependence to make a Poisson approximation unrealistic, because of the presence of clumps of 1's. A wide variety of stochastic models which can loosely be described by means of

Poisson clump processes are discussed in Aldous (1989), and counts deriving from these models are far more likely to exhibit an approximately compound Poisson distribution. One way of trying to estimate the error in such an approximation is then to proceed by way of Stein's method for Poisson point processes approximation (Arratia, Goldstein and Gordon (1990), Barbour and Brown (1992)), since a bound on the total variation distance from Poisson of the point process describing the positions and sizes of the clumps implies the same bound for the total variation distance from Poisson of the distribution of W. This turns out to be quite useful if $\mathbb{E}W$ is small, despite the awkwardness which may be involved in defining a clump and its position. However, with this approach, the analogues of the H_l do not decrease as $\mathbb{E}W$ becomes large, and if a single clump has a non trivial probability of occurring then the whole estimate becomes very weak. Simple examples of these drawbacks are given in Examples 1.1 and 1.2.

A better idea is to use Stein's method directly, as introduced in compound Poisson approximation in Barbour, Chen and Loh (1992). If the distribution of a random variable W is to be approximated by the compound Poisson distribution $\mathrm{CP}(\boldsymbol{\lambda})$, the distribution of $\sum_{i\geq 1} iZ_i$ when the Z_i are independent $\mathrm{Po}(\lambda_i)$ random variables and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots)$, it is merely necessary to show that

$$\Big|\mathbb{E}\Big\{\sum_{i\geq 1} i\lambda_i g(W+i) - Wg(W)\Big\}\Big| \leq \varepsilon_0 M_0(g) + \varepsilon_1 M_1(g) \tag{1.6}$$

for all bounded functions $g : \mathbb{N} \to \mathbb{R}$, much as for Poisson approximation, and Roos (1994) has shown how to achieve this in practice, for suitably chosen $\boldsymbol{\lambda}$. It then follows that

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}(\boldsymbol{\lambda})) \leq \varepsilon_0 H_0 + \varepsilon_1 H_1, \tag{1.7}$$

where $H_l = \sup_{A\subset\mathbb{Z}_+} M_l(g_A)$, $l = 0, 1$, and g_A is now the solution of the Stein Equation

$$\sum_{i\geq 1} i\lambda_i g(w+i) - wg(w) = 1_A(w) - \mathrm{CP}(\boldsymbol{\lambda})\{A\}, \qquad w \geq 0. \tag{1.8}$$

Now, for compound Poisson approximation, the general estimates of the form

$$H_0, H_1 \leq Ce^{\lambda}\min\{1, \lambda^{-1}\}; \qquad \lambda = \sum_{i\geq 1}\lambda_i, \tag{1.9}$$

given in Barbour, Chen and Loh, behave very differently from the Poisson estimates (1.5), and usually lead to useless results as soon as λ is at all big. A better pair of bounds,

having the right λ–dependence (apart from a logarithmic term in the numerator of the H_1–bound)

$$H_0 \leq \min\Big(1, \frac{2}{\sqrt{\lambda\nu_1}}\Big); \quad H_1 \leq \min\Big\{1, \frac{1}{\lambda\nu_1}\Big(\frac{1}{4\lambda\nu_1} + \log^+(2\lambda\nu_1)\Big)\Big\}, \tag{1.10}$$

where $\mu_i = \lambda_i/\lambda$ and $\nu_i = i\mu_i - (i+1)\mu_{i+1}$, were also established in Barbour, Chen and Loh, but these only apply when the additional condition

$$\nu_i \geq 0, \quad i \geq 1 \tag{1.11}$$

is satisfied. Condition (1.11) does hold in many applications, especially those in which the compound Poisson distribution under consideration is not too far from a Poisson distribution, and compound Poisson approximation using Stein's method then works well. However, it is unsatisfactory for the method only to work for a restricted set of distributions, especially since the argument which led to the restrictions, involving an auxiliary probabilistic construction, does not indicate that they are of an essential nature. The aim of this paper is to try to remedy this situation, making direct use of Stein's method for compound Poisson approximation generally feasible.

In order to emphasize the natural comparison with (1.5), we formulate all our results with the idea of keeping the μ_i fixed while letting $\lambda \to \infty$, although the main theorems are actually always proved for any particular fixed choice of λ and the μ_i, and are thus generally valid. The first results that we obtain, in Section 2, show that no universal bound with λ–behaviour similar to that of (1.10) is possible, even when the constants are allowed to depend on the μ_i, as they also do to some extent in (1.10). More surprisingly, if one restricts to the case where only μ_1 and μ_2 are non–zero, it turns out that Condition (1.11) is in fact sharp, and that, when it does not hold, H_0 grows exponentially fast with λ. In Section 3, we prove bounds similar in spirit to those of (1.10) under a condition, (3.2), which is somewhat less restrictive than (1.11), and reduces (as it should) to (1.11) when only μ_1 and μ_2 are non–zero. The main difference is that we replace the quantities H_0 and H_1 by

$$J_l = \sup_{m \geq 0} M_l(g_{[0,m]}), \qquad l = 0, 1 \tag{1.12}$$

which play the same rôle as H_0 and H_1 if one considers only the smaller family of sets $A = [0, m]$, $m \geq 0$, on the right hand side of (1.8). The corresponding estimate is then for approximation in terms of Kolmogorov distance, instead of in total variation, the analogue of (1.7) being

$$d_K(\mathcal{L}(W), \mathrm{CP}\,(\boldsymbol{\lambda})) \leq \varepsilon_0 J_0 + \varepsilon_1 J_1. \tag{1.13}$$

In Section 4, a rather different approach is taken. Here, in view of the impossibility of obtaining a bound like (1.10) for a general choice of the μ_i, we show that better bounds can be found for quantities $J_l^{(a)}$, $l = 0, 1$. These are obtained in the same way as the J_l, but with M_l replaced by $M_l^{(a)}$, where

$$M_0^{(a)}(g) = \sup_{j \geq a} |g(j)|; \qquad M_1^{(a)}(g) = \sup_{j \geq a} |g(j+1) - g(j)|, \tag{1.14}$$

and where $a \geq 1$ may be chosen to suit. The bounds can then be used, in conjunction with a more complicated analogue of (1.13) (see Equation (4.4)), to express closeness in Kolmogorov distance in terms of ε_0, ε_1 and any bound on the lower tail of the distribution of W, the main result being presented in Theorem 4.3.

There are two problems with this formulation. The first is that total variation distance is replaced by the weaker Kolmogorov distance, for which central limit theorems offer an alternative route to approximation, one which is effective precisely when λ is large, the setting for which we have specifically designed our improvements. What progress have we really made? A simple answer is to take Poisson approximation to the binomial as analogy, where the uncorrected central limit theorem has an error of order $(np)^{-1/2}$, as compared to the error of order p for Poisson approximation. The choice between the two approximations thus depends on the size of p in relation to n, and Poisson approximation is better so long as the value of p is small enough. Reassuringly, that part of our compound Poisson estimate (4.4) which corresponds to (1.13) is of the correct order p, so that it does as well as it can; similar considerations also apply in the settings of Examples 1.1 and 1.2.

The second problem concerns the additional truncation component present in (4.4), which is not needed in the restricted setting in which (1.13) is valid. Here, however, a second moment Chebyshev estimate is enough to show that the extra component is typically of order at most λ^{-1}, and an exponential concentration inequality could make it much smaller. Using the binomial–Poisson example to get a feel for its importance, even λ^{-1} is of order $(np)^{-1}$, an order of magnitude smaller than the error in the normal approximation. Thus the truncation element is of small importance when compared with the error in a central limit approximation, even though, for moderate λ, it can still dominate the 'main' component in the compound Poisson error estimate (4.4). Once again, Examples 1.1 and 1.2 illustrate the points clearly.

The bounds of Section 4, unlike (1.9), are still not valid in complete generality, though the conditions are enormously weaker than (1.11). We assume that the $\boldsymbol{\mu}$ distribution has a

fourth moment, and, more importantly, that it satisfies an aperiodicity condition, expressed precisely in Assumption A of Section 4. The significance of this latter condition emerges already in Example 1.2.

A particularly attractive feature of the bound derived for $J_1^{(a)}$ is that $\log^+ \lambda$ no longer appears in the numerator, so that its λ–behaviour is exactly the same as (1.5) for Poisson approximation. What is more, the relevant part of the proof can also be used under the conditions of Section 3, implying there, too, a bound for J_1 which does not involve $\log^+ \lambda$. Getting rid of similar $\log^+ \lambda$ factors in the solution of Stein Equations, for instance when considering Poisson process approximation, has so far proved to be very difficult, and indeed, in some circumstances, impossible (Brown and Xia, 1994). Yet were such a logarithmic factor present in the Poisson bound (1.5) for H_1, the correct order p of approximation would not be obtained in the $\mathrm{Bi}\,(n,p)$ example, and for $\lambda \geq \exp\{1/p\}$ the error estimate would be useless.

We now give two examples illustrating approximation in the compound Poisson setting. These examples are deliberately chosen to be as simple as possible, and they show that the problems being addressed in this paper arise even for sums of independent random variables. Example 5.1, taken from the theory of random graphs, exhibits more of the power and flexibility of the technique, and shows that our approach can also work well when dependence is involved, and even when typical values from the $\boldsymbol{\mu}$ distribution become large with λ.

Example 1.1

Suppose that $(I_l, Y_l;\ 1 \leq l \leq n)$ are all independent, with $I_l \sim \mathrm{Be}\,(p)$ and $\mathbb{P}[Y_l = i] = \mu_i$, $i \geq 1$, for each l, and set $W = \sum_{l=1}^n I_l Y_l$. Since W has the distribution of $\sum_{k=1}^N Y_k$ and $\mathrm{CP}\,(\lambda\boldsymbol{\mu})$ that of $\sum_{k=1}^{N'} Y_k$, where $N \sim \mathrm{Bi}\,(n,p)$ and $N' \sim \mathrm{Po}\,(np)$ are independent of the Y's, it follows that $d_{TV}(\mathcal{L}(W), \mathrm{CP}\,(\lambda\boldsymbol{\mu})) \leq p$ (Michel, 1988), just because of the closeness of $\mathrm{Po}\,(np)$ and $\mathrm{Bi}\,(n,p)$. The result is of best possible order, but the proof depends on the strong symmetry present in the example. The approach by way of point process approximation considers the sequence as a marked point process, with a point for each l such that $I_l = 1$, carrying the mark Y_l. Total variation approximation of the point process by a Poisson point process then implies a corresponding bound for the compound Poisson approximation of W. The method of Section 3 of Arratia, Goldstein and Gordon (1990) yields a bound of order $O(np^2)$, which is not at all good if np is large. A better result is obtained from Theorem 10.H in Barbour, Holst and Janson (1992), which gives an order

of $O(p[1+\log^+(np)])$. However, the bound still suffers from an undesirable logarithmic factor; the theorem also relies on the independence of the Y's from one another and from the I's, which drastically limits its usefulness in more complicated applications.

To test out the efficacy of compound Poisson approximation using Stein's method, evaluate the left hand side of (1.6) directly; by independence, we have

$$\mathbb{E}I_l Y_l g(W) = p\sum_{i\geq 1} i\mu_i \mathbb{E}g(W_l+i),$$

where $W_l = \sum_{k\neq l} I_k Y_k$, and hence it follows that

$$\mathbb{E}\Big\{\sum_{i\geq 1} i\lambda_i g(W+i) - Wg(W)\Big\} = p\sum_{i\geq 1} i\mu_i \sum_{l=1}^{n} \mathbb{E}\{g(W+i)-g(W_l+i)\},$$

showing that we can take $\varepsilon_0 = 0$ and $\varepsilon_1 = np^2\Big\{\sum_{i\geq 1} i\mu_i\Big\}^2$. Hence, if the μ_i are such that Condition (3.2) holds, the corresponding bound in Kolmogorov distance following from our approach is of order $J_1 np^2$, now with $J_1 = O(1/np)$, giving an error estimate of the optimal order p as $n\to\infty$. If, instead, only the much weaker conditions of Section 4 are satisfied by the μ_i, we have to add a term of the form $\mathbb{P}[W \leq cnp\mathbb{E}Y]$ for some fixed $c<1$, but this becomes negligible at rate $e^{-\alpha np}$ for some $\alpha > 0$, if Y_1 has an exponential moment, because of the Chernoff bounds. Thus the current approach yields very encouraging results, even though there is still some slight restriction on the choices of the μ_i which are allowed. Michel's argument requires no such restrictions; nor does Theorem 1 of Zaitsev (1988), which, when specialized to this problem, also gives order p, but, like Theorem 10.H of Barbour, Holst and Janson, requires independence of the Y's and I's.

Example 1.2

This example shows that the aperiodicity assumption of Section 4 is really necessary. Suppose that an extra independent pair (I_0, Y_0) is added to the setup in Example 1, with $I_0 \sim \mathrm{Be}(1/2)$ and $Y_0 = 3$ with probability one. This increases the mean number of clumps of size 3 by 1/2, so we increase the intensity λ_3 in the approximating compound Poisson distribution by 1/2; this corresponds to the prescription given in Roos (1994). Symmetry is broken, so that Michel's argument cannot be used here, and both Zaitsev's theorem and the approach in Arratia, Goldstein and Gordon (1990) yield errors of order $O(1)$, which are unhelpful. Theorem 10.H of Barbour, Holst and Janson (1992) is still applicable, and

does better, yielding a total variation bound of order

$$O\Big(p[1+\log^+(np)] + \min\Big\{1, \frac{1}{np\mu_3}[1+\log^+(np\mu_3)]\Big\}\Big),$$

looking much as before when np is large, unless, for instance, $\mu_3 = 0$. However, this latter condition is by no means spurious: if $\mu_i = 0$ for all *odd* i, the true total variation difference between the distributions is at least

$$\mathrm{Po}\,(1/2)\{2\mathbb{Z}_+\} - 1/2 > 0.183,$$

and not of order p, so that the conclusion appears to be quite good.

For direct compound Poisson approximation using Stein's method, we evaluate (1.6) once more, concluding that we can take $\varepsilon_0 = 0$ and

$$\varepsilon_1 = np^2\Big\{\sum_{i\geq 1} i\mu_i\Big\}^2 + 9/4.$$

This leads to an excellent bound of order $O(p + 1/np)$ for Kolmogorov distance, under the conditions of Sections 3 and 4. Note that the aperiodicity condition of Section 4 allows many choices of the μ_i in which $\mu_3 = 0$, but properly enters in the case where all odd μ_i are zero. However, it also enters in any case where l.c.m. $\{i : \mu_i > 0\} = 3$, though here good approximation can in fact also be obtained (for example, by dividing each of the Y_l by 3, and then starting afresh...)

It may seem curious that the values of H_l and J_l obtainable for the solutions of the Stein Equation (1.8) can turn out to be large. The phenomenon is related to the fact that (1.8) is not, like most other commonly used Stein Equations, derived from the Elementary Stein Identity (Barbour, 1995), but makes use of the General Stein Identity. There is then no automatic interpretation of the Stein Operator as the generator of an associated Markov process, enabling the H_l to be estimated by probabilistic means. Only under the additional Condition (1.11) has such an interpretation as yet been found.

Throughout the remainder of the paper, λ_i, $i \geq 1$, are non–negative real numbers satisfying $\sum_{i\geq 1} i\lambda_i < \infty$. λ is used to denote $\sum_{i\geq 1}\lambda_i$, $\mu_i = \lambda_i/\lambda$, $m_l = \sum_{i\geq 1} i^l\mu_i$, $l \geq 1$, and $\sigma^2 = m_2 - m_1^2$. We always assume that $m_1 < \infty$, and we usually assume in addition that $m_l < \infty$ for some larger value of l.

2. Solving the Stein Equation

An explicit expression for the solution g to the Stein equation

$$\sum_{i\geq 1} i\lambda_i g(w+i) - wg(w) = f(w) - \mathbb{E}f(z), \quad w \in \mathbb{Z}_+, \tag{2.1}$$

for f a given bounded function and Z a random variable with distribution CP $(\boldsymbol{\lambda})$, is given in Barbour, Chen and Loh (1992) as

$$wg(w) = -\sum_{k\geq 0} \lambda^k \mathbb{E}\left\{ \frac{\left(\prod_{l=1}^k Y_l\right)[f(w+S_k) - \mathbb{E}f(Z)]}{\prod_{l=1}^k (w+S_l)} \right\}, \quad w \geq 1, \tag{2.2}$$

where the Y_l, $l \geq 1$ are independent and satisfy $\mathbb{P}[Y_l = j] = \mu_j$, $j \geq 1$, and where $S_m = \sum_{l=1}^m Y_l$. The bounds given in (1.9) are obtained by analyzing a recursive scheme derived from (2.1), but a slightly weaker bound follows directly from (2.2) by observing that

$$\begin{aligned} &\sum_{k\geq 0} \lambda^k \mathbb{E}\Big\{\prod_{l=1}^k Y_l \Big/ \prod_{l=1}^k (w+S_l)\Big\} \\ &\qquad \leq \sum_{k\geq 0} \lambda^k \mathbb{E}\Big\{\prod_{l=1}^k Y_l \Big/ \prod_{l=1}^k S_l\Big\} = \sum_{k\geq 0} \lambda^k/k! = e^\lambda. \end{aligned}$$

This argument also shows that, if g_A denotes the solution of (2.1) for the choice $f = 1_A$, then

$$g_A(w) = \sum_{j\in A} g_{\{j\}}(w) \text{ for all } A \subset \mathbb{Z}_+, \tag{2.3}$$

with absolute convergence in the j sum.

As an alternative to representation (2.2), we use a Fourier approach, replacing the general f on the right hand side of (2.1) by

$$f(w) = f^t(w) = t^w - \mathbb{E}t^Z, \quad w \geq 0, \tag{2.4}$$

for any complex t such that $|t| \leq 1$. The use of complex valued functions f actually involves nothing new, since real and imaginary parts can be separated, the left hand side of (2.1) being linear in g. If (2.1) can now be solved for each f^t, with solution g^t, say, then complex inversion can be used to derive a solution for any bounded f. This is because f^t can be written as the (functional) power series

$$f^t = \sum_{j\geq 0} t^j \{1_{\{j\}} - \mathbb{P}[Z = j]\}, \tag{2.5}$$

invertible to

$$1_{\{j\}} - \mathbb{P}[Z = j] = \frac{1}{2\pi i} \int_{|t|=1} t^{-j-1} f^t \, dt. \tag{2.6}$$

Applying Fubini's theorem in (2.2) with f as in (2.6), it then follows that

$$g_{\{j\}} = \frac{1}{2\pi i} \int_{|t|=1} t^{-j-1} g^t \, dt; \tag{2.7}$$

for general bounded f, we have

$$g = \sum_{j \geq 0} f(j) g_{\{j\}} = \frac{1}{2\pi i} \sum_{j \geq 0} f(j) \int_{|t|=1} t^{-j-1} g^t \, dt. \tag{2.8}$$

Theorem 2.1. *Write $\mu(t) = \sum_{i \geq 1} \mu_i t^i$. Then Equation (2.1) with right hand side as in (2.4) is solved by*

$$g^t(w) = e^{\lambda \mu(t)} \int_t^1 u^{w-1} e^{-\lambda \mu(u)} \, du, \quad w \geq 1, \tag{2.9}$$

where the integral is taken along any contour in the unit disc from t to 1.

Proof. Take g^t as defined in (2.9). Then, integrating by parts and using Fubini's theorem, we obtain

$$\begin{aligned}
\sum_{i \geq 1} i \lambda \mu_i g^t(w+i) &= e^{\lambda \mu(t)} \sum_{i \geq 1} i \lambda \mu_i \int_t^1 u^{w-1+i} e^{-\lambda \mu(u)} \, du \\
&= e^{\lambda \mu(t)} \int_t^1 u^w \lambda \mu'(u) e^{-\lambda \mu(u)} \, du \\
&= e^{\lambda \mu(t)} \left\{ [-u^w e^{-\lambda \mu(u)}]_t^1 + \int_t^1 w u^{w-1} e^{-\lambda \mu(u)} \, du \right\} \\
&= t^w - e^{\lambda[\mu(t) - \mu(1)]} + w g^t(w),
\end{aligned}$$

as required.

Example 2.2

Suppose that $\mu_i = 0$, $i \geq 3$, so that $\mu(t) = \mu_1 t + \mu_2 t^2$. Then

$$g^{-1}(1) = e^{\lambda \mu(-1)} \int_{-1}^1 e^{-\lambda \mu(u)} \, du = \int_{-1}^1 \exp\{-\lambda(1+u)(\mu_1 - \mu_2 + \mu_2 u)\} \, du. \tag{2.10}$$

The factor $(1+u)$ is always positive in $-1 \leq u \leq 1$, but, if $2\mu_2 > \mu_1$, the factor $(\mu_1 - \mu_2 + \mu_2 u)$ is negative near enough to $u = -1$, with the result that $g^{-1}(1)$ receives a contribution

from the range of u near -1 which is exponentially large with λ. In fact, the saddle point method shows that, for large λ and $\mu_1 < 2\mu_2$,

$$g^{-1}(1) \sim \sqrt{\frac{\pi}{\lambda}} \exp\{\lambda\mu_2(1-\mu_1/2\mu_2)^2\}. \tag{2.11}$$

Thus, in this example, if condition (1.11) is violated, the function $f(w) = (-1)^w$, for which $M_0(f) = 1$ and $M_1(f) = 2$, gives rise to a solution g of (2.1) for which $M_0(g)$ grows exponentially fast with λ.

When applying Stein's method for compound Poisson approximation as in Roos (1995), one needs estimates of both $M_0(g)$ and $M_1(g)$, for the solutions g of (2.1) corresponding to the chosen family of test functions f. Theorem 2.1 and (2.7) provide the basis on which bounds for M_0 and M_1 can be constructed, for the functions $g_{\{j\}}$, $j \geq 0$. Although (2.8) can in principle then be used to extend the results to arbitrary test functions f, the extra sum coming into the expression can lead to a very weak bound, if absolute values are introduced too soon in the argument, even for the bounded test functions used in total variation approximation. However, if one restricts attention to Kolmogorov distance, it is enough to consider the test functions $1_{[0,j]}$ for $j \in \mathbb{Z}_+$, for which (2.8) yields the explicit sum

$$g_{[0,j]} = \frac{1}{2\pi i} \int_{|t|=1} \Big(\frac{1-t^{-j-1}}{1-t^{-1}}\Big) t^{-1} g^t \, dt. \tag{2.12}$$

This can be further simplified to

$$g_{[0,j]} = -\frac{1}{2\pi i} \int_{|t|=1} \Big(\frac{t^{-j-1}}{1-t^{-1}}\Big) t^{-1} g^t \, dt, \tag{2.13}$$

by letting $j \to \infty$ in (2.12), using (2.3) and the Riemann–Lebesgue lemma (since, in view of (2.9), $g^t/(t-1)$ is bounded near 0) to give

$$g_{Z_+} = \frac{1}{2\pi i} \int_{|t|=1} \Big(\frac{t^{-1}}{1-t^{-1}}\Big) g^t \, dt;$$

but then, from (2.2), one has $g_{Z_+} = 0$ for all w. Thus, in what follows, we shall mainly be

concerned with finding bounds for the quantities

$$
\begin{aligned}
I_0^{j,w} &= \Big|\frac{1}{2\pi i}\int_{|t|=1} t^{-j-1} g^t(w)\,dt\Big| = |g_{\{j\}}(w)|;\\
I_1^{j,w} &= \Big|\frac{1}{2\pi i}\int_{|t|=1} t^{-j-1}\{g^t(w+1)-g^t(w)\}\,dt\Big| = |g_{\{j\}}(w+1)-g_{\{j\}}(w)|;\\
J_0^{j,w} &= \Big|\frac{1}{2\pi i}\int_{|t|=1} \frac{t^{-j-2}}{1-t^{-1}} g^t(w)\,dt\Big| = |g_{[0,j]}(w)|;\\
J_1^{j,w} &= \Big|\frac{1}{2\pi i}\int_{|t|=1} \frac{t^{-j-2}}{1-t^{-1}}\{g^t(w+1)-g^t(w)\}\,dt\Big| = |g_{[0,j]}(w+1)-g_{[0,j]}(w)|,
\end{aligned}
\tag{2.14}
$$

for $j \in \mathbb{Z}_+$ and $w \in \mathbb{N}$, where g^t is given by (2.9).

3. Extending the 'good' region

In this section, we show that we can still find useful bounds for the quantities in (2.14), even when Condition (1.8) is relaxed somewhat. Define

$$\nu_i = i\mu_i - (i+1)\mu_{i+1}, \qquad \nu_i^- = \min(0,\nu_i), \tag{3.1}$$

and note that Condition (1.8) is the same as requiring that $\nu_i^- = 0$ for all i. Here, we shall assume the weaker condition

$$\nu_1 + \sum_{i\geq 2} i^2 \nu_i^- := \delta \geq 0, \tag{3.2}$$

which allows some negativity among the ν_i, but not too much. For $l = 0,1$, let I_l (J_l) be defined to be $\sup_{j\geq 0, w\geq 1} I_l^{j,w}$ ($J_l^{j,w}$): note that J_0 and J_1 are as defined in (1.9).

Theorem 3.1. *If $m_1 < \infty$ and (3.2) holds, then*

$$
\begin{aligned}
I_0 &\leq \frac{2\sqrt{2}}{\pi} \wedge \frac{1}{2\lambda\delta}\{1+\log^+(\pi\lambda\delta)\}; \qquad I_1 \leq \Big[\frac{2}{\pi}(\pi-1)\Big] \wedge \frac{2}{\lambda\delta};\\
J_0 &\leq 1 \wedge 2\sqrt{\frac{2}{\lambda\delta}}; \qquad J_1 \leq \frac{4\sqrt{2}}{\pi} \wedge \frac{1}{\lambda\delta}\{1+\log^+(\pi\lambda\delta)\}.
\end{aligned}
$$

Remark. The bounds on J_0 and J_1 are very similar in structure to those given in (1.7) for H_0 and H_1.

Proof. Integrating along the chord joining t and 1, g^t can be expressed as

$$g^t(w) = e^{\lambda\mu(t)} \int_0^1 (t+p(1-t))^{w-1} e^{-\lambda\mu(t+p(1-t))}(1-t)\,dp, \tag{3.3}$$

from which it immediately follows that

$$|g^t(w)| \le |1-t| \int_0^1 \exp\{\lambda \mathcal{R}e\,[\mu(t) - \mu(t+p(1-t))]\}\, dp. \tag{3.4}$$

Now it can be checked, using Fubini's theorem, that

$$\mu(x) - \mu(z) = -\sum_{i\ge 1} \nu_i \int_x^z \frac{1-s^i}{1-s}\, ds = -\nu_1(z-x) - \sum_{i\ge 2} \nu_i \int_x^z \frac{1-s^i}{1-s}\, ds, \tag{3.5}$$

for $|z|, |x| \le 1$. Thus, from Lemmas 6.6 and 6.7 and (3.5), we have

$$\mathcal{R}e\,[\mu(t) - \mu(t+p(1-t))] \le -\{\nu_1 + \sum_{i\ge 2} i^2 \nu_i^-\} p(1-\mathcal{R}e\, t) = -\delta p(1-\mathcal{R}e\, t);$$

substituting this into (3.4), it follows that

$$|g^t(w)| \le |1-t| \min\Big\{1, \frac{1}{\lambda\delta(1-\mathcal{R}e\, t)}\Big\}. \tag{3.6}$$

A similar argument also gives

$$\begin{aligned} |g^t(w+1) - g^t(w)| &\le |1-t|^2 \int_0^1 (1-p) \exp\{\lambda \mathcal{R}e\,[\mu(t) - \mu(t+p(1-t))]\}\, dp \\ &\le |1-t|^2 \min\Big\{1, \frac{1}{\lambda\delta(1-\mathcal{R}e\, t)}\Big\} = \min\{|1-t|^2, 2/\lambda\delta\}. \end{aligned} \tag{3.7}$$

With (3.6) and (3.7), we are now in a position to make estimates of the quantities in (2.14). Starting with (3.6) and $I_0^{j,w}$, we always have

$$I_0^{j,w} \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{2}\sin(\theta/2)\, d\theta = 2\sqrt{2}/\pi.$$

On the other hand, for $\pi\lambda\delta \ge 1$, we can write

$$\begin{aligned} I_0^{j,w} &\le \frac{1}{2\pi} \int_{-\pi}^{\pi} \Big|g^{e^{i\theta}}(w)\Big|\, d\theta \le \frac{\sqrt{2}}{\pi} \int_0^{\pi} \min\Big\{\sqrt{1-\cos\theta}, \frac{1}{\lambda\delta\sqrt{1-\cos\theta}}\Big\}\, d\theta \\ &\le \frac{\sqrt{2}}{\pi}\Big\{\int_0^{\sqrt{\pi/\lambda\delta}} \frac{\theta}{\sqrt{2}}\, d\theta + \frac{1}{\lambda\delta}\int_{\sqrt{\pi/\lambda\delta}}^{\pi} \frac{\pi}{\theta\sqrt{2}}\, d\theta\Big\} = \frac{1}{2\lambda\delta}\{1 + \log^+(\pi\lambda\delta)\}. \end{aligned}$$

Since the estimates do not depend on j or w, the bound required for I_0 is thus proved. The bound for I_1 is easier, following directly from (3.7), which gives

$$I_1^{j,w} \le \min\Big\{\frac{2}{\lambda\delta}, \frac{1}{2\pi}\int_{-\pi}^{\pi} 2(1-\cos\theta)\, d\theta\Big\}.$$

Then, for J_0, we use (3.6) to give

$$J_0^{j,w} \le \frac{2}{\pi}\int_0^\pi \min\Big\{1, \frac{\pi^2}{2\lambda\delta\theta^2}\Big\}\,d\theta,$$

bounded by 2 if $2\lambda\delta \le 1$, and by

$$\frac{2}{\pi}\Big\{\frac{\pi}{\sqrt{2\lambda\delta}} + \frac{\pi^2}{2\lambda\delta}\Big[\frac{\sqrt{2\lambda\delta}}{\pi} - \frac{1}{\pi}\Big]\Big\} = 2\sqrt{\frac{2}{\lambda\delta}} - \frac{1}{\lambda\delta}$$

otherwise. Finally, from (3.7),

$$J_1^{j,w} \le \frac{2\sqrt{2}}{\pi}\int_0^\pi \min\Big\{\sqrt{1-\cos\theta}, \frac{1}{\lambda\delta\sqrt{1-\cos\theta}}\Big\}\,d\theta,$$

which is estimated in the same way as in the argument for I_0.

Remark. There is an undesirable logarithmic factor in the estimates of both I_0 and J_1, as there is also in (1.10). This factor arises because the simple estimates (3.6) and (3.7) are a little too coarse near $\theta = 0$. However, if $m_4 < \infty$, Lemma 4.2 can be used to bound the contributions to the integrals in (2.14) in the range $|\theta| \le \sqrt{m_2/m_4}$ by quantities of order $1/\lambda m_2$, as a result of which one obtains bounds for I_0 and J_1 of the correct order λ^{-1}. It should, however, be noted that the constants which appear are quite large, and that an improvement over Theorem 3.1 can only be expected when $\frac{m_2}{\delta}\log^+(\pi\lambda\delta)$ is at least 15.

4. The general case

As is clear from Example 2.2, we cannot in general expect H_0 and H_1 to be bounded by quantities which decrease with λ. However, in order to use the Stein equation (1.4) to demonstrate the closeness in distribution of a random variable W and $\mathrm{CP}(\boldsymbol{\lambda})$, one can often get by with less. For instance, one can use (1.4) to write

$$\begin{aligned}
&1_A(w) - \mathrm{CP}(\boldsymbol{\lambda})\{A\}\\
&= \{1_A(w) - \mathrm{CP}(\boldsymbol{\lambda})\{A\}\}(1-u_{ab}(w)) + u_{ab}(w)\Big\{\sum_{i\ge1} i\lambda_i g(w+i) - wg(w)\Big\}\\
&\quad = \{1_A(w) - \mathrm{CP}(\boldsymbol{\lambda})\{A\}\}(1-u_{ab}(w)) + \Big\{\sum_{i\ge1} i\lambda_i g(w+i)u_{ab}(w+i) - wg(w)u_{ab}(w)\Big\}\\
&\qquad\qquad - \sum_{i\ge1} i\lambda_i g(w+i)(u_{ab}(w+i) - u_{ab}(w)),
\end{aligned} \tag{4.1}$$

where, for $0 < a < b < \infty$,

$$u_{ab}(w) = \begin{cases} 0 & \text{if } w \le a; \\ \frac{w-a}{b-a} & \text{if } a \le w \le b; \\ 1 & \text{if } w \ge b, \end{cases}$$

giving

$$|u_{ab}(w+i) - u_{ab}(w)| \le \Big(1 \wedge \frac{i}{b-a}\Big) 1_{[0,b]}(w) 1_{(a,\infty)}(w+i).$$

Putting W for w and taking expectations, this yields

$$\begin{aligned} &|\mathbb{P}[W \in A] - \mathrm{CP}\,(\boldsymbol{\lambda})\{A\}| \\ &\qquad \le \mathbb{P}[W \le b] + \Big|\mathbb{E}\Big\{\sum_{i \ge 1} i\lambda_i g(W+i) u_{ab}(W+i) - W g(W) u_{ab}(W)\Big\}\Big| \\ &\qquad\qquad + \frac{\lambda m_2 M_0^{(a)}(g)}{b-a} \mathbb{P}[W \le b], \end{aligned} \tag{4.2}$$

where $M_l^{(a)}(g) = M_l(g(\cdot + a))$, $l = 1, 2$. Now, writing $g_{ab}(w) = g(w) u_{ab}(w)$, we have

$$\begin{aligned} M_0(g_{ab}) &\le \sup_{w \ge a} |g(w)| = M_0^{(a)}(g); \\ M_1(g_{ab}) &\le \sup_{w \ge a} |g(w+1) - g(w)| + \frac{1}{b-a} \sup_{w \ge a} |g(w)| = M_1^{(a)}(g) + \frac{1}{b-a} M_0^{(a)}(g), \end{aligned} \tag{4.3}$$

and hence the middle contribution to the error in (4.2) can be bounded by

$$\varepsilon_0 M_0^{(a)}(g) + \varepsilon_1 \Big(\frac{M_0^{(a)}(g)}{b-a} + M_1^{(a)}(g)\Big),$$

if (1.6) holds. The remaining contribution

$$\mathbb{P}[W \le b]\Big\{1 + \frac{\lambda m_2 M_0^{(a)}(g)}{b-a}\Big\}$$

has then to be bounded by other means, for instance by using Chebyshev's inequality. This, for example, leads to estimates

$$\begin{aligned} d_{TV}(\mathcal{L}(W), \mathrm{CP}\,(\boldsymbol{\lambda})) &\le \varepsilon_0 H_0^{(a)} + \varepsilon_1 \Big(\frac{H_0^{(a)}}{b-a} + H_1^{(a)}\Big) + \mathbb{P}[W \le b]\Big\{1 + \frac{\lambda m_2 H_0^{(a)}}{b-a}\Big\}; \\ d_K(\mathcal{L}(W), \mathrm{CP}\,(\boldsymbol{\lambda})) &\le \varepsilon_0 J_0^{(a)} + \varepsilon_1 \Big(\frac{J_0^{(a)}}{b-a} + J_1^{(a)}\Big) + \mathbb{P}[W \le b]\Big\{1 + \frac{\lambda m_2 J_0^{(a)}}{b-a}\Big\}, \end{aligned} \tag{4.4}$$

to be compared with (1.7) and (1.13), where

$$H_l^{(a)} = \sup_{A \subset \mathbb{Z}_+} M_l^{(a)}(g_A); \quad J_l^{(a)} = \sup_{j \ge 0} M_l^{(a)}(g_{[0,j]}).$$

The values of a and b are still free to be chosen, with a not too small (so that the $H_l^{(a)}$ and $J_l^{(a)}$ remain small enough) and b not too large (so that $\mathbb{P}[W \le b]$ is small), while keeping $b-a$ large (so that $\lambda/(b-a)$ is not too big). For instance, one can take $a = c_1\lambda m_1$ and $b = c_2\lambda m_1$, for $0 < c_1 < c_2 < 1$. Then, if the mean and variance of W are close to those of $\mathrm{CP}(\boldsymbol{\lambda})$, which it is natural to expect, Chebyshev's inequality implies that

$$\mathbb{P}[W \le c_2\lambda m_1] = O\Big(\frac{\lambda^{-1}m_2}{(1-c_2)^2 m_1^2}\Big); \qquad \frac{\lambda m_2}{b-a} = \frac{m_2}{(c_2-c_1)m_1}, \tag{4.5}$$

so that the 'truncation error' introduced by replacing the usual Stein estimate by (4.2) is of order λ^{-1} if $m_2 < \infty$, and is thus small for large λ. Using Chebyshev's inequality with a higher moment of W may often reduce this error further.

The upshot of these considerations is that, to be able to apply (4.4) effectively, all that is now required is to find bounds on $|g(w)|$ and $|g(w+1)-g(w)|$ in the restricted range $w \ge c_1\lambda m_1$, the hope being that these bounds will be tighter than those where w may take any value in $\mathbb{Z}_+$: as always, g solves (2.1) for f in the relevant class of test functions. Our technique for deriving the improved bounds uses complex variable arguments in conjunction with (2.7) – (2.9). For the first of the results, we need an aperiodicity condition on the $\boldsymbol{\mu}$–distribution. Define

$$\rho_1(\theta) = 1 - \sum_{i\ge1}\mu_i\cos i\theta; \qquad \rho_2(\theta) = 1 - \frac{1}{m_1}\sum_{i\ge1} i\mu_i\cos i\theta, \tag{4.6}$$

noting that $0 \le \rho_l(\theta) \le 2$, $l = 1,2$; set $\rho_l^*(\zeta) = \inf_{\zeta\le\theta\le\pi}\rho_l(\theta)$.

Assumption A. *For any $0 < \zeta \le \pi$ and for $l = 1,2$, $\rho_l^*(\zeta) > 0$.*

Remark. If $\mu_1 > 0$, Assumption A is satisfied, since, for $0 < \zeta < \pi$, $\rho_1^*(\zeta) \ge \mu_1(1-\cos\zeta)$ and $\rho_2^*(\zeta) \ge (\mu_1/m_1)(1-\cos\zeta)$.

Lemma 4.1. *Suppose that Assumption A holds, and define $\rho^*(\zeta) = \min(\rho_1^*(\zeta), \frac{1}{2}\rho_2^*(\zeta), 1)$. Fix any $0 < \zeta < \pi$. Then, for any $t = e^{i\theta}$ such that $|\theta| \ge \zeta$, we have*

$$|g^t(w)| \le 3e\Big(1 \wedge \frac{2}{\lambda m_1\rho^*(\zeta)}\Big);$$

$$|g^t(w+1) - g^t(w)| \le e\Big(1 \wedge \frac{2}{\lambda m_1\rho^*(\zeta)}\Big)\sqrt{2(1-\cos\zeta)} + e\Big(1 \wedge \frac{4}{\lambda m_1^2\rho^*(\zeta)}\Big\{1 + \frac{1}{\lambda\rho^*(\zeta)}\Big\}\Big),$$

uniformly in $w \ge \lambda m_1(1-\rho^*(\zeta)/2)$.

Proof. We use the formula (2.9), integrating along the rays from t to 0 and from 0 to 1, giving

$$\begin{aligned} g^t(w) &= -t^w e^{\lambda\mu(t)} \int_0^1 s^{w-1} e^{-\lambda\mu(st)}\, ds + e^{\lambda\mu(t)} \int_0^1 s^{w-1} e^{-\lambda\mu(s)}\, ds \\ &= G_1 + G_2, \end{aligned} \tag{4.7}$$

say. For G_1, we have

$$\begin{aligned} |G_1| &= \Big| \int_0^1 (1-v)^{w-1} \exp\{\lambda[\mu(t) - \mu((1-v)t)]\}\, dv \Big| \\ &\le \int_0^1 \exp\Big\{-(w-1)v + \lambda \sum_{i\ge 1} \mu_i \mathcal{R}e\,(t^i)\{1-(1-v)^i\}\Big\}\, dv \\ &\le \int_0^1 \exp\{-(w-1)v + \lambda m_1(1-\rho_2(\theta))v\}\, dv, \end{aligned}$$

since

$$\mathcal{R}e\,(t^i)[1-(1-v)^i] \le iv \max(\cos i\theta, 0) \le (iv/2)[1+\cos i\theta].$$

From Assumption A and the lower bound on the value of w, it now follows that

$$|G_1| \le e \int_0^1 \exp\{-\lambda m_1 \rho^*(\zeta) v/2\}\, dv \le e\Big(1 \wedge \frac{2}{\lambda m_1 \rho^*(\zeta)}\Big).$$

For G_2, we write

$$\begin{aligned} |G_2| &= \Big| e^{\lambda(\mu(t)-\mu(1))} \int_0^1 (1-v)^{w-1} e^{\lambda(\mu(1)-\mu(1-v))}\, dv \Big| \\ &\le \exp\Big\{\lambda \sum_{i\ge 1} \mu_i [\mathcal{R}e\,(t^i) - 1]\Big\} \int_0^1 \exp\Big\{-(w-1)v + \lambda \sum_{i\ge 1} \mu_i [1-(1-v)^i]\Big\}\, dv. \end{aligned}$$

Thus it follows that

$$\begin{aligned} |G_2| &\le \exp\{-\lambda\rho_1(\theta)\} \int_0^1 \exp\{-(w-1)v + \lambda(1 \wedge m_1 v)\}\, dv \\ &= \exp\{1-\lambda\rho_1(\theta)\}\Big\{\int_0^{1/m_1} e^{(\lambda m_1 - w)v} dv + e^{\lambda} \int_{1/m_1}^1 e^{-wv} dv\Big\}. \end{aligned}$$

Now, because $w \ge \lambda m_1(1-\rho^*(\zeta)/2)$, we have

$$\begin{aligned} |G_2| &\le \exp\{1-\lambda\rho^*(\zeta)/2\}\Big\{\Big(\frac{2}{\lambda m_1 \rho^*(\zeta)} \wedge \frac{1}{m_1}\Big) + \Big[\Big(1-\frac{1}{m_1}\Big) \wedge \frac{1}{\lambda m_1(1-\rho^*(\zeta)/2)}\Big]\Big\} \\ &\le e\Big(1 \wedge \frac{4}{\lambda m_1 \rho^*(\zeta)}\Big) \exp\{-\lambda\rho^*(\zeta)/2\}, \end{aligned}$$

proving the first estimate. For the second, proved in similar fashion, we have $|g^t(w+1) - g^t(w)| \le |G_1'| + |G_2'|$, with

$$\begin{aligned}|G_1'| &\le \int_0^1 \{v + |t-1|\} \exp\Big\{-(w-1)v + \lambda \sum_{i\ge 1} \mu_i \mathcal{R}e\,(t^i)\{1-(1-v)^i\}\Big\}\, dv \\ &\le e\Big(1 \wedge \frac{2}{\lambda m_1 \rho^*(\zeta)}\Big)|t-1| + e\Big(\frac{1}{2} \wedge \frac{4}{[\lambda m_1 \rho^*(\zeta)]^2}\Big)\end{aligned}$$

and

$$\begin{aligned}|G_2'| &\le \exp\{-\lambda\rho_1(\theta)\} \int_0^1 v \exp\{-(w-1)v + \lambda(1 \wedge m_1 v)\}\, dv \\ &\le \Big(\frac{1}{2} \wedge \frac{4}{\lambda m_1^2 \rho^*(\zeta)}\Big) \exp\{-\lambda\rho^*(\zeta)/2\}.\end{aligned}$$

As a consequence of Lemma 4.1, we have the bounds

$$\begin{aligned}J_0(\zeta) &= \sup_{j\ge 1} \sup_{w \ge w_0} \Big| \frac{1}{2\pi} \int_{\zeta \le |\theta| \le \pi} \frac{e^{-ij\theta}}{1-e^{-i\theta}}\, g^{e^{i\theta}}(w)\, d\theta \Big| \\ &\le \frac{1}{\pi} \int_\zeta^\pi \frac{3e}{\sqrt{2(1-\cos\theta)}} \Big(1 \wedge \frac{2}{\lambda m_1 \rho^*(\theta)}\Big)\, d\theta \le \frac{3e}{\sqrt{2(1-\cos\zeta)}} \Big(1 - \frac{\zeta}{\pi}\Big)\Big(1 \wedge \frac{2}{\lambda m_1 \rho^*(\zeta)}\Big)\end{aligned} \tag{4.8}$$

and

$$\begin{aligned}J_1(\zeta) &= \sup_{j\ge 1} \sup_{w \ge w_0} \Big| \frac{1}{2\pi} \int_{\zeta \le |\theta| \le \pi} \frac{e^{-ij\theta}}{1-e^{-i\theta}} \Big\{ g^{e^{i\theta}}(w+1) - g^{e^{i\theta}}(w) \Big\}\, d\theta \Big| \\ &\le \frac{e}{\pi} \int_\zeta^\pi \Big\{ \Big(1 \wedge \frac{2}{\lambda m_1 \rho^*(\theta)}\Big) + \frac{1}{\sqrt{2(1-\cos\theta)}} \Big[1 \wedge \frac{4}{\lambda m_1^2 \rho^*(\theta)} \Big(1 + \frac{1}{\lambda \rho^*(\theta)}\Big)\Big]\Big\}\, d\theta \\ &\le e\Big(1 - \frac{\zeta}{\pi}\Big) \Big\{ \Big(1 \wedge \frac{2}{\lambda m_1 \rho^*(\zeta)}\Big) + \frac{1}{\sqrt{2(1-\cos\zeta)}} \Big[1 \wedge \frac{4}{\lambda m_1^2 \rho^*(\zeta)} \Big(1 + \frac{1}{\lambda \rho^*(\zeta)}\Big)\Big]\Big\},\end{aligned} \tag{4.9}$$

whenever $w_0 \ge \lambda m_1(1 - \rho^*(\zeta)/2)$. These in turn lead to bounds of order λ^{-1} for the contributions to the integrals $I_l^{j,w}$ and $J_l^{j,w}$ in (2.14) from the values of t away from 1 (corresponding to θ away from 0), whenever $w \ge \lambda m_1(1 - \rho^*(\zeta)/2)$.

Bounding the remaining contribution from the range $|\theta| \le \zeta$ turns out to be much more delicate, and we do so only under the further assumption that $m_4 < \infty$. On the other hand, the estimate that we obtain for it remains valid for all $w \ge 0$. We state the result here, deferring the proof until Section 6.

Lemma 4.2. *If $m_4 < \infty$, $\zeta \le \sqrt{m_2/m_4}$ and $\zeta(\lambda m_2)^{1/2} \ge 1$, we have*

$$(a) \quad \Big|\frac{1}{2\pi}\int_{|\theta|\le\zeta} e^{-ij\theta} g^{e^{i\theta}}(w)\,d\theta\Big| \le \frac{67}{2\pi\lambda m_2};$$

$$(b) \quad \Big|\frac{1}{2\pi}\int_{|\theta|\le\zeta} e^{-ij\theta}\Big\{g^{e^{i\theta}}(w+1) - g^{e^{i\theta}}(w)\Big\}\,d\theta\Big| \le \frac{5\zeta}{2\pi\lambda m_2};$$

$$(c) \quad \Big|\frac{1}{2\pi}\int_{|\theta|\le\zeta} \frac{e^{-ij\theta}}{1-e^{-i\theta}}\, g^{e^{i\theta}}(w)\,d\theta\Big| \le \frac{9}{2\pi\sqrt{\lambda m_2}},$$

and

$$(d) \quad \Big|\frac{1}{2\pi}\int_{|\theta|\le\zeta} \frac{e^{-ij\theta}}{1-e^{-i\theta}}\Big\{g^{e^{i\theta}}(w+1) - g^{e^{i\theta}}(w)\Big\}\,d\theta\Big| \le \frac{44}{\pi\lambda m_2}$$

The estimates (4.8) and (4.9) of $J_0(\zeta)$ and $J_1(\zeta)$ can now be combined with Lemma 4.2 to give bounds on $J_0^{(a)}$ and $J_1^{(a)}$ for $a = c_1\lambda m_1$, when $c_1 = 1-\rho^*(\zeta_0)/2$ and $\zeta_0 = \sqrt{m_2/m_4}$. These can in turn be used in (4.4) and (4.5), where we also take $c_2 = 1-\rho^*(\zeta_0)/4$. The result is given in the following theorem.

Theorem 4.3. *Suppose that Assumption A is satisfied and that $m_4 < \infty$. Let W be a random variable such that (1.6) holds for some quantities ε_0 and ε_1 and for all bounded g. Then it follows that*

$$\begin{aligned} &d_K(\mathcal{L}(W), \mathrm{CP}(\boldsymbol{\lambda})) \\ &\quad \le \varepsilon_0 J_0^* + \varepsilon_1\Big(J_1^* + \frac{4J_0^*}{\lambda m_1\rho^*(\zeta_0)}\Big) + \mathbb{P}[W \le \lambda m_1(1-\rho^*(\zeta_0)/4)]\Big\{1 + \frac{4m_2 J_0^*}{m_1\rho^*(\zeta_0)}\Big\}, \end{aligned}$$

where $\zeta_0 = \sqrt{m_2/m_4}$, ρ^* is as defined in Lemma 4.1,

$$J_0^* = J_0(\zeta_0) + \frac{9}{2\pi\sqrt{\lambda m_2}}; \qquad J_1^* = J_1(\zeta_0) + \frac{44}{\pi\lambda m_2},$$

and J_0 and J_1 are as given in (4.8) and (4.9).

Corollary 4.4. *Under the assumptions of Theorem 4.3, and if also $\lambda^{-1}\mathbb{E}W > m_1(1-\rho^*(\zeta_0)/4)$, it follows that*

$$\begin{aligned} d_K(\mathcal{L}(W), \mathrm{CP}(\boldsymbol{\lambda})) &\le \varepsilon_0 J_0^* + \varepsilon_1\Big(J_1^* + \frac{4J_0^*}{\lambda m_1\rho^*(\zeta_0)}\Big) \\ &\quad + \frac{1}{\lambda}\Big\{1 + \frac{4m_2 J_0^*}{m_1\rho^*(\zeta_0)}\Big\}\Big\{\frac{\lambda^{-1}\mathrm{Var}\,W}{(\lambda^{-1}\mathbb{E}W - m_1 + m_1\rho^*(\zeta_0)/4)^2}\Big\}. \end{aligned}$$

In particular, for a sequence W_n in which $\lambda_n \to \infty$ but $\boldsymbol{\mu}$ remains fixed and satisfies Assumption A, and in which $\lambda_n^{-1}\mathrm{Var}\,W_n$ is bounded and $\lambda_n^{-1}\mathbb{E}W_n \to m_1$, the bound in Corollary 4.4 is of order

$$\lambda_n^{-1/2}\varepsilon_0^{(n)} + \lambda_n^{-1}\varepsilon_1^{(n)} + \lambda_n^{-1}, \tag{4.10}$$

and better results can be expected from Theorem 4.3 if a better bound on the lower tail probability of W is available than that given by Chebyshev's inequality.

The quantities J_0^* and J_1^* are of the ideal orders $\lambda^{-1/2}$ and λ^{-1} respectively, if Assumption A is satisfied, $\boldsymbol{\mu}$ is held fixed and $\lambda \to \infty$. However, if $\boldsymbol{\mu}$ varies with λ, and in particular in such a way that $m_1 \to \infty$ with λ, it is useful to be able to assess their size in relation also to the moments of the $\boldsymbol{\mu}$–distribution. This is not entirely straightforward, because the expressions (4.8) and (4.9) for $J_0(\zeta_0)$ and $J_1(\zeta_0)$ contain the factor $1/\rho^*(\zeta)$, which can be large, since $\zeta_0 = \sqrt{m_2/m_4} = O(m_1^{-1}) \to 0$. Nevertheless, if the $\boldsymbol{\mu}$–distributions remain well spread out and essentially aperiodic, in a sense which is made precise below, then reasonable estimates of J_0^* and J_1^* can be determined, using the integral forms of (4.8) and (4.9).

To make such estimates of J_0^* and J_1^* when $m_1 \to \infty$ with λ, observe first that $\rho_1(\theta) = 1 - \mathbb{E}\cos(Y_1\theta)$ and $\rho_2(\theta) = 1 - \mathbb{E}\cos(Y_2\theta)$, where Y_1 denotes a random variable with distribution $\boldsymbol{\mu}$, and Y_2 has the size–biassed distribution with $\mathbb{P}[Y_2 = j] = j\mu_j/m_1$. Now finding useful lower bounds for $\mathbb{E}\{1 - \cos X\}$ is difficult for general random variables X, because the function $1 - \cos x$ is close to zero whenever x is near to $2k\pi$ for some integer k. However, for bounded random variables, the situation can be somewhat better, and we exploit this fact to make progress.

If $|X| \le 1$ a.s., the inequality $1 - \cos x \ge \alpha x^2$ in $|x| \le 2$, where $\alpha = (1 - \cos 2)/4 > 1/3$, implies that

$$\mathbb{E}\{1 - \cos(a + X)\} \ge \tfrac{1}{3}\operatorname{Var} X \tag{4.11}$$

for any $a \in \mathbb{R}$, because $1 - \cos(a + x) \ge 1 - \cos(b + x)$, uniformly for all $|x| \le 1$, for some b with $|b| \le 1$. So let Y be a random variable with mean m and variance s^2, and let Y^* have the distribution of Y conditional on $|Y - m| \le 2s$; write $c_1 = \operatorname{Var} Y^*/4s^2$, and note that $p = \mathbb{P}[|Y - m| \le 2s] \ge 3/4$. Then, taking $X = \theta(Y^* - m)$, satisfying $|X| \le 1$ in $|\theta| \le 1/2s$, it follows from (4.11) that

$$\begin{aligned} \mathbb{E}\{1 - \cos(\theta Y)\} &\ge p\mathbb{E}\{1 - \cos(\theta Y^*)\} \\ &\ge \tfrac{3}{4}\mathbb{E}\{1 - \cos(\theta m + X)\} \ge \tfrac{1}{4}\theta^2 \operatorname{Var} Y^* = \theta^2 c_1 s^2 \end{aligned} \tag{4.12}$$

in $|\theta| \le 1/2s$, giving a useful lower bound in this range.

On the other hand, for any y such that $|y - m| \le 2s$ and any $a > 0$, we have

$$\begin{aligned} 4c_1 s^2 = \operatorname{Var} Y^* &\le \mathbb{E}\{(Y^* - y)^2\} = p^{-1}\mathbb{E}\{(Y - y)^2 I[|Y - m| \le 2s]\} \\ &\le \tfrac{4}{3}\{a^2 s^2 + 16 s^2 \mathbb{P}[|Y - y| > as]\}; \end{aligned}$$

thus, for any $a \le \sqrt{3c_1/2}$, we find that

$$\mathbb{P}[|Y - y| > as] \ge 3c_1/32,$$

for any y such that $|y - m| \le 2s$. Hence, if c_1 is not small, the distribution of θY for $|\theta| \ge 1/2s$ cannot be concentrated close to $2k\pi$ for any single integer k — it is well spread out on its natural scale s — and so, in these circumstances,

$$c_2 = \inf_{(2s)^{-1} \le |\theta| \le \pi} \mathbb{E}\{1 - \cos(\theta Y)\} \tag{4.13}$$

can only be small if the distribution of Y is close to being periodic.

So, applying these considerations to Y_1 and Y_2, define

$$c_{i1} = \frac{\operatorname{Var} Y_i^*}{4s_i^2}; \qquad c_{i2} = \inf_{(2s_i)^{-1} \le |\theta| \le \pi} \rho_i(\theta); \qquad i = 1, 2, \tag{4.14}$$

where $s_1^2 = \operatorname{Var} Y_1 = \sigma^2 = m_2 - m_1^2$ and $s_2^2 = \operatorname{Var} Y_2 = m_1^{-1}(m_3 m_1 - m_2^2)$. We define a sequence of $\boldsymbol{\mu}$–distributions to be well spread out and essentially aperiodic if each of the c_{ij}, $1 \le i, j \le 2$, remains bounded away from 0. Then, from (4.12) and (4.14), it follows that

$$\rho_i(\theta) \ge \begin{cases} c_{i1} s_i^2 \theta^2, & \text{if } |\theta| \le 1/2s_i; \\ c_{i2}, & \text{if } 1/2s_i < |\theta| \le \pi. \end{cases} \tag{4.15}$$

Observing that

$$\frac{1}{\rho^*(\theta)} \le \frac{1}{\rho_1(\theta)} + \frac{2}{\rho_2(\theta)} + 1, \tag{4.16}$$

the bounds (4.15) can be combined with the integral forms of (4.8) and (4.9) to give estimates of J_0^* and J_1^*, since

$$J_0(\zeta_0) = O\Big(\frac{1}{\lambda m_1} \int_{\zeta_0}^{\pi} \frac{d\theta}{\theta \rho^*(\theta)}\Big) \tag{4.17}$$

and

$$J_1(\zeta_0) = O\Big(\frac{1}{\lambda m_1} \int_{\zeta_0}^{\pi} d\theta \Big\{\frac{1}{\rho^*(\theta)} + \frac{1}{m_1 \theta \rho^*(\theta)} + \frac{1}{\lambda m_1 \theta [\rho^*(\theta)]^2}\Big\}\Big). \tag{4.18}$$

In particular, writing

$$s_+^2 = \max\{m_2 - m_1^2, (m_3 m_1 - m_2^2)/m_1\}; \qquad s_-^2 = \min\{m_2 - m_1^2, (m_3 m_1 - m_2^2)/m_1\},$$

in any sequence of compound Poisson distributions in which $m_1 \to \infty$ as $\lambda \to \infty$ in such a way that

$$m_1^{-4} m_4 \quad \text{and} \quad \lambda^{-1/2} s_-^{-2} m_1^2 (1 + \log^+(2\pi s_+))$$

remain uniformly bounded and $\min_{1\le i,j\le 2} c_{ij}$ remains uniformly bounded away from 0, then

$$J_0^* = O\Big(\frac{1}{\sqrt{\lambda m_2}}\Big); \quad J_1^* = O\Big(\frac{1}{\lambda}\Big(\frac{1}{m_1} + \frac{1}{s_-^2}\Big)\Big). \tag{4.19}$$

5. Applications

Theorem 4.3 can be used anywhere that an inequality of the form (1.6) can be established, and it applies in particular to those examples of compound Poisson approximation already studied in Roos (1993,1994) and Roos and Stark (1996). In exchange for restricting consideration to Kolmogorov distance, one obtains error estimates with the right behaviour as $\lambda \to \infty$; no unwanted $\log \lambda$ factor when (1.11) holds, no unwanted e^λ factor when (1.11) is violated, so long as (4.16) holds and C_*/c_* is bounded above. So, for example, when approximating the distribution of the number of overlapping k–runs of heads in an independent sequence of coin tosses with $\mathbb{P}[\text{head}] = p$, Arratia, Goldstein and Gordon (1990), Section 4.1.2, give a bound on the total variation distance from a compound Poisson approximation of order $nkp^{2k}(1-p)$, and Roos (1993, Section 3.3) improves this to order $kp^k \log(np^k)$ when $p < 1/2$. Taking her estimate of ε_1 and using Janson's inequality to bound $\mathbb{P}[W \le \lambda m_1(1-\rho^*(\zeta_0)/4)]$ (a more difficult example is treated in detail in Example 5.1), Theorem 4.3 leads to a bound on the Kolmogorov distance from the same compound Poisson distribution of order

$$kp^k + e^{-cnp^k(1-p)}, \tag{5.1}$$

for some $c > 0$, uniformly in the range of n and p in which the mean number $\lambda = np^k(1-p)$ of runs of length at least k satisfies $\lambda \ge 1$. This new bound becomes rapidly smaller than the previous estimates when λ increases, and does not deteriorate, as the other bounds do, when $n \to \infty$ for k and p fixed: nor is there any restriction on the value of p, other than that implied by the condition $\lambda \ge 1$.

Theorem 4.3 can also be applied in the setting used for compound Poisson approximation by way of Poisson point process approximation. For instance, using the notation of Arratia, Goldstein and Gordon (1990), Section 3.1, W is expressed in the form $W = \sum_{\alpha\in I}\sum_{i\ge 1} iX_{\alpha i}$, where $X_{\alpha i}$ is the indicator that an event which adds exactly i to W occurs at index α; for a problem such as the k–runs discussed above, α would denote

the first index in a consecutive sequence of exactly $k+i-1$ heads. For each $\alpha \in I$, let $B(\alpha) \subset I$ contain α, and set

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B(\alpha)} \mathbb{E}X_\alpha \mathbb{E}X_\beta; \quad b_2 = \sum_{\alpha \in I} \sum_{\substack{\beta \in B(\alpha) \\ \beta \neq \alpha}} \mathbb{E}(X_\alpha X_\beta); \\ b_3 &= \sum_{\alpha \in I} \sum_{i \geq 1} \mathbb{E}|\mathbb{E}\{X_{\alpha i} - \mathbb{E}X_{\alpha i} \,|\, \sigma(X_{\beta j};\ \beta \notin B(\alpha), j \geq 1)\}|, \end{aligned} \tag{5.2}$$

where $X_\alpha = \sum_{i \geq 1} X_{\alpha i}$. Then, arguing as in Arratia, Goldstein and Gordon (1990), Section 4.2.1, we conclude that

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}\,(\boldsymbol{\lambda})) \leq b_1 + b_2 + b_3, \tag{5.3}$$

where $\lambda_i = \sum_{\alpha \in I} \mathbb{E}X_{\alpha i}$; the improved coefficients of b_1 and b_2 are from Barbour, Holst and Janson (1992, Theorem 10.A).

Instead of using the estimate (5.3), which is typically inaccurate if λ is large, one can verify (1.6) directly, using the same structure. Write $Y_\alpha = \sum_{i \geq 1} iX_{\alpha i}$, and set

$$Z_\alpha = \sum_{\substack{\beta \in B(\alpha) \\ \beta \neq \alpha}} Y_\beta; \quad W_\alpha = \sum_{\beta \notin B(\alpha)} Y_\beta.$$

Then, since

$$\mathbb{E}(X_{\alpha i} g(W)) = \mathbb{E}\{X_{\alpha i} g(i + Z_\alpha + W_\alpha)\},$$

it follows that

$$|\mathbb{E}(X_{\alpha i} g(W)) - \mathbb{E}\{X_{\alpha i} g(i + W_\alpha)\}| \leq M_1(g) \mathbb{E}(X_{\alpha i} Z_\alpha);$$

similar considerations yield

$$|\mathbb{E}X_{\alpha i}\{\mathbb{E}g(i+W) - \mathbb{E}g(i+W_\alpha)\}| \leq M_1(g)\mathbb{E}X_{\alpha i}\mathbb{E}\{Y_\alpha + Z_\alpha\}$$

and

$$|\mathbb{E}\{X_{\alpha i} g(i+W_\alpha)\} - \mathbb{E}X_{\alpha i}\mathbb{E}g(i+W_\alpha)| \leq M_0(g)\mathbb{E}|\mathbb{E}\{X_{\alpha i} - \mathbb{E}X_{\alpha i} \,|\, \sigma(X_{\beta j};\ \beta \notin B(\alpha), j \geq 1)\}|. \blacksquare$$

Hence, with the λ_i defined as above, we have

$$\left| \mathbb{E}\Big\{ \sum_{i \geq 1} i\lambda_i g(W+i) - Wg(W) \Big\} \right| \leq M_1(g)\{b_1^* + b_2^*\} + M_0(g) b_3^*,$$

which is (1.6) with $\varepsilon_0 = b_3^*$ and $\varepsilon_1 = b_1^* + b_2^*$, where

$$\begin{aligned} b_1^* &= \sum_{\alpha\in I}\sum_{\beta\in B(\alpha)} \mathbb{E}Y_\alpha \mathbb{E}Y_\beta; \quad b_2^* = \sum_{\alpha\in I}\sum_{\substack{\beta\in B(\alpha)\\ \beta\neq\alpha}} \mathbb{E}(Y_\alpha Y_\beta); \\ b_3^* &= \sum_{\alpha\in I}\sum_{i\geq 1} i\mathbb{E}|\mathbb{E}\{X_{\alpha i} - \mathbb{E}X_{\alpha i}\,|\,\sigma(X_{\beta j};\ \beta\notin B(\alpha), j\geq 1)\}|. \end{aligned} \tag{5.4}$$

Theorem 4.3 can now be used to give a bound on the accuracy of approximation in Kolmogorov distance,

$$\begin{aligned} d_K(\mathcal{L}(W), \mathrm{CP}\,(\boldsymbol{\lambda})) \leq (b_1^*+b_2^*)\Big(J_1^* + \frac{4J_0^*}{\lambda m_1 \rho^*(\zeta_0)}\Big) + b_3^* J_0^* \\ + \Big\{1 + \frac{4m_2 J_0^*}{m_1\rho^*(\zeta_0)}\Big\}\mathbb{P}[W \leq \lambda m_1(1-\rho^*(\zeta_0)/4)], \end{aligned} \tag{5.5}$$

where the quantities J_0^*, J_1^*, m_1, m_2 and $\rho^*(\zeta_0)$ are determined from the λ_i's.

Comparing the error bound in (5.5) to that of (5.3), note that the quantities b_l^* are larger than the corresponding b_l, because $Y_\alpha \geq X_\alpha$; in the example of k–runs, by a factor of $(1-p)^{-2}$. On the other hand, their coefficients become small as λ increases and also as m_1 increases, in view of (4.19); for k–runs, $J_1^* = O(1/np^k)$, giving an overall improvement in the order of the error by a factor of $1/\{np^k(1-p)^2\}$, since $b_3^* = b_3 = 0$. Thus (5.4) and (5.5) can be used wherever (5.3) has previously been applied, and the error bounds obtained, although now with respect to Kolmogorov distance, are typically much tighter as soon as λ becomes at all large.

We conclude the section with an example from the theory of random graphs, in which we use the methods of Roos (1994) to establish (1.6). The example is also interesting, in that the structure of the problem dictates that $m_1 \to \infty$ with λ.

Example 5.1. *The number of copies of the triangle and whisker graph $H = \{(1,2),(2,3),(1,3),(1,4)\}$ in the Bernoulli random graph $K_{n,p}$.*

The graph H is balanced, but not strictly balanced (Bollobás 1985, Chapter IV), and so the number of copies of H near the threshold does not have an approximately Poisson distribution. A suitably chosen compound Poisson distribution offers a good approximation, as we now indicate.

Let $[n] = \{1,2,\ldots,n\}$, and define $\Gamma = \{(\gamma,i,j);\ \gamma\in[n]^3, i\in\gamma, j\in[n]\setminus\gamma\}$, where $[n]^r$ denotes all r–subsets of $[n]$. Let the independent $\mathrm{Be}\,(p)$ distributed edge indicator random variables in $K_{n,p}$ be denoted by E_{lm}, $\{l,m\}\in[n]^2$, and define

$$I_\alpha = E_{ij}\prod_{\{l,m\}\subset\gamma} E_{lm} \quad \text{for } \alpha = (\gamma,i,j)\in\Gamma. \tag{5.6}$$

Then $I_\alpha = 1$ exactly when there is a copy of H in $K_{n,p}$ with the triangle on γ and with whisker $\{i,j\}$. Set $W = \sum_{\alpha\in\Gamma} I_\alpha$, and note that $\mathbb{E}W = \frac{1}{2}n(n-1)(n-2)(n-3)p^4$. If $np \ll 1$, it thus follows that $W = 0$ with high probability. We investigate what happens when np is moderately large; for simplicity, we shall henceforth assume that $(3n-11)p \geq 1$ and that $p \leq 1/2$.

We start by applying Roos (1994, Theorem 2) to establish (1.6) with appropriate choices of ε_0, ε_1 and $(\lambda_j,\ j \geq 1)$. Using her notation, define

$$\Gamma^{vs}_{(\gamma,i,j)} = \{(\gamma, i', j');\ i' \in \gamma, j' \in [n] \setminus \gamma\} \setminus \{(\gamma, i, j)\}$$

and

$$\Gamma^{vw}_{(\gamma,i,j)} = \{(\gamma', i', j');\ (\gamma' \cup \{j'\}) \cap \gamma = \emptyset\},$$

so that the collections $\{I_\beta,\ \beta \in \Gamma^{vs}_\alpha\}$ and $\{I_\beta,\ \beta \in \Gamma^{vw}_\alpha\}$ are independent of one another. Set

$$U_\alpha = \sum_{\beta\in\Gamma^{vs}_\alpha} I_\beta; \quad Z_\alpha = U_\alpha + I_\alpha; \quad X_\alpha = \sum_{\beta\in\Gamma\setminus(\Gamma^{vs}_\alpha\cup\Gamma^{vw}_\alpha\cup\{\alpha\})} I_\beta.$$

Then it follows from Roos's theorem that (1.6) holds with

$$\varepsilon_0 = 0; \quad \varepsilon_1 = \sum_{\alpha\in\Gamma}\{(\mathbb{E}I_\alpha)^2 + \mathbb{E}I_\alpha\mathbb{E}\{U_\alpha + X_\alpha\} + \mathbb{E}(I_\alpha X_\alpha)\}, \tag{5.7}$$

if $j\lambda_j = \sum_{\alpha\in\Gamma}\mathbb{E}\{I_\alpha I[Z_\alpha = j]\}$. Direct calculation then shows that

$$\varepsilon_1 = O(n^7p^8); \quad \lambda_j = j^{-1}\mathbb{E}W\ \mathrm{Bi}\,(3n-10,p)\{j-1\},\ j \geq 1, \tag{5.8}$$

where the order estimate is uniform in the range of n and p under consideration. Thus the approximating distribution in this case is a Poisson compounded Binomial distribution, with mean which may be large. Note also that, for np at all big, Condition (3.2) is not satisfied.

In order to apply Theorem 4.3, observe that it follows from (5.8) that $\lambda = \sum_{j\geq 1}\lambda_j \asymp \mathbb{E}W/(np) \asymp (np)^3$ and that $m_r \asymp (np)^r$, $r \geq 1$, and $\sigma^2 \asymp np$. Thus $\zeta_0 \asymp (np)^{-1}$ can be very small in our range of n, p, as can $\rho^*(\zeta_0) \asymp (np)^{-1}$ also. However, it is easy to show that the conditions preceding (4.19) are satisfied in our range of n and p, implying that

$$J_0^* = O((np)^{-5/2}); \quad J_1^* = O((np)^{-4}).$$

Hence, and from (5.8), we have

$$\varepsilon_0 J_0^* + \varepsilon_1\{J_1^* + 4J_0^*/(\lambda m_1 \rho^*(\zeta_0))\} = O(n^3 p^4), \tag{5.9}$$

uniformly in our range of n, p.

The remaining element in Theorem 4.3 is of order $\mathbb{P}[W \leq \lambda m_1(1 - \rho^*(\zeta_0)/4)]$. We estimate the probability by using Janson's inequality (Barbour, Holst and Janson (1992), Theorem 2.S), which here gives

$$\mathbb{P}[W \leq (1-\eta)\mathbb{E}W] \leq \exp\Big\{-\frac{\eta^2 \mathbb{E}W}{2(1+\delta)}\Big\},$$

with

$$\delta = \{\mathbb{E}W\}^{-1} \sum_{\alpha \in \Gamma} \mathbb{E}\{I_\alpha(U_\alpha + X_\alpha)\} = O(np)$$

and $\eta = \rho^*(\zeta_0)/4 \asymp (np)^{-1}$. Thus this last contribution to the error bound in Theorem 4.3 is of order $\exp\{-cnp\}$ for some $c > 0$, again uniformly in our n, p range. Hence Theorem 4.3 implies that

$$d_K(\mathcal{L}(W), \mathrm{CP}\,(\boldsymbol{\lambda})) = O(n^3 p^4 + e^{-cnp}), \tag{5.10}$$

where the λ_j are as defined in (5.8) and the error is uniform in $p \leq 1/2$ and $(3n-11)p \geq 1$.

Note that the bound is useful asymptotically if $n^{-1} \ll p \ll n^{-3/4}$, but that the second contribution stays big if np remains bounded. However, in the latter case, we can instead use (1.7) with (1.9) to conclude that

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}\,(\boldsymbol{\lambda})) = O(n^2 p^3 \exp\{\mathbb{E}W/(np)\}) = O(n^{-1}),$$

so that compound Poisson approximation is also accurate at the threshold. The worst rate obtained by our approach is in the intermediate range where $np \asymp \{\log n\}^{1/3}$. There is no reason to suppose that the approximation here is really bad; it is rather the truncation method which is pessimistic when $\mathbb{E}W$ is only moderately large. Indeed, (5.8) suggests that Theorem 10.B of Barbour, Holst and Janson (1992) might be applied to give a distance of order $O(n^7 p^8)$, an order which is inferior to that in (5.10) as soon as $np \geq c \log n$ for some suitable c, but is better for the very small values of np. Theorem 10.H is inapplicable because of the dependence between clump sizes, as is also Zaitsev's (1988) Theorem 1.

6. Technicalities

The main part of this section is concerned with the proof of Lemma 4.2. Two lemmas used in the proof of Theorem 3.1 are proved at the end.

For Lemma 4.2, we need bounds on integrals of the form

$$I(f,h) = \int_{e^{-i\zeta}}^{e^{i\zeta}} t^{-1} f(t) e^{\lambda\mu(t)} \int_t^1 h(v) v^{-1} e^{-\lambda\mu(v)}\, dv\, dt, \tag{6.1}$$

where the integrals are taken along the unit circle; the choices of f and h in (6.1) are

$$f_1(t) = t^{-j} \quad \text{and} \quad f_2(t) = t^{-j}/(1-t^{-1}), \quad j \in \mathbb{Z}_+,$$

and

$$h_1(v) = v^w \quad \text{and} \quad h_2(v) = v^w(v-1), \quad w \in \mathbb{Z}_+.$$

Matching the contributions to (6.1) from t and its conjugate $\bar{t}$, we find that

$$\begin{aligned} I(f,h) &= 2\int_0^\zeta d\phi \int_0^\phi d\theta\, \mathcal{I}m \left\{ f(e^{i\phi}) e^{\lambda\mu(e^{i\phi})} h(e^{i\theta}) e^{-\lambda\mu(e^{i\theta})} \right\} \\ &= \frac{2}{\lambda} \int_0^\zeta d\phi \int_0^{\lambda\phi} du\, \exp\{-C_\lambda(\phi,u)\} D_\lambda(f,h,\phi,u), \end{aligned} \tag{6.2}$$

where

$$C_\lambda(\phi,u) = \lambda \sum_{j\geq 1} \mu_j [\cos\{j(\phi - u/\lambda)\} - \cos j\phi] \tag{6.3}$$

and

$$D_\lambda(f,h,\phi,u) = \mathcal{I}m \left\{ f(e^{i\phi}) h(e^{i(\phi-u/\lambda)}) \exp\{-i\lambda \sum_{j\geq 1} \mu_j [\sin\{j(\phi-u/\lambda)\} - \sin j\phi]\} \right\}. \tag{6.4}$$

In order to make further headway with (6.2), we need a number of preliminary results.

Lemma 6.1. *For $0 \leq u \leq \lambda\phi$ and $\phi \leq \zeta \leq \sqrt{m_2/m_4}$, we have*

(a) $\quad |C_\lambda(\phi,u) - m_2(u\phi - u^2/2\lambda)| \leq \dfrac{m_4}{24}\{\lambda^{-2}u^3\phi + 4u\phi^3\}$;

(b) $\quad |\lambda \sum_{j\geq 1} \mu_j [\sin\{j(\phi - u/\lambda)\} - \sin j\phi] + m_1 u| \leq \dfrac{m_3}{6}\{3u\phi^2 + 3\lambda^{-1}u^2\phi + \lambda^{-2}u^3\}$;

(c) $\quad C_\lambda(\phi,u) \geq m_2 u\phi / K_\zeta$,

where

$$K_\zeta^{-1} = \frac{1}{2}\left\{1 - \frac{m_4\zeta^2}{6m_2}\right\} \geq \frac{5}{12}.$$

Proof. For part (a), we have

$$\begin{aligned} C_\lambda(\phi,u) &= 2\lambda \sum_{j\ge 1} \mu_j \sin\{j(\phi - u/2\lambda)\} \sin\{ju/2\lambda\} \\ &= 2\lambda \sum_{j\ge 1} \mu_j \sin\{j(\phi - u/2\lambda)\}\{ju/2\lambda\} + E_1 \\ &= m_2(u\phi - u^2/2\lambda) + E_1 + E_2, \end{aligned}$$

where, by Taylor's expansion,

$$|E_1| \le 2\lambda \sum_{j\ge 1} \mu_j j\phi \frac{j^3u^3}{48\lambda^3} \quad \text{and} \quad |E_2| \le 2\lambda \sum_{j\ge 1} \mu_j \frac{ju}{2\lambda}\frac{j^3\phi^3}{6}.$$

On the other hand, for part (c), we use the inequality $\sin x \ge x(1 - x^2/6)$ for $x \ge 0$, yielding

$$\begin{aligned} C_\lambda(\phi,u) &\ge \sum_{j\ge 1} \mu_j j^2 u(\phi - u/2\lambda)\{1 - j^2\phi^2/6\} \\ &\ge \frac{m_2 u\phi}{2}\Big\{1 - \frac{m_4\zeta^2}{6m_2}\Big\}. \end{aligned}$$

Finally, for part (b),

$$\begin{aligned} \lambda \sum_{j\ge 1} \mu_j[\sin j\phi - \sin\{j(\phi - u/\lambda)\}] &= 2\lambda \sum_{j\ge 1} \mu_j \cos\{j(\phi + u/2\lambda)\} \sin\{ju/2\lambda\} \\ &= 2\lambda \sum_{j\ge 1} \mu_j \cos j\phi \sin\{ju/2\lambda\} + E_3 \\ &= m_1 u + E_3 + E_4, \end{aligned}$$

where

$$|E_3| \le 4\lambda \sum_{j\ge 1} \mu_j \frac{ju}{2\lambda}\frac{ju}{4\lambda}\Big[j\phi + \frac{ju}{4\lambda}\Big]$$

and

$$|E_4| \le 2\lambda \sum_{j\ge 1} \mu_j \Big\{\frac{ju}{2\lambda}\frac{j^2\phi^2}{2} + \frac{j^3u^3}{48\lambda^3}\Big\}.$$

The next three lemmas provide routine estimates of some integrals which appear frequently in what follows. Hereafter, K denotes the K_ζ of Lemma 6.1(c).

Lemma 6.2. *Integral estimates of the first kind:*

$$(a)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\phi^{-1} \le 4\Big(\frac{K}{\lambda m_2}\Big)^{1/2};$$

$$(b)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\frac{2u}{\lambda\phi} \le 2\sqrt{2}\Big(\frac{K}{\lambda m_2}\Big);$$

$$(c)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\frac{u^2}{\lambda^2\phi} \le \frac{4\sqrt{2}}{3}\Big(\frac{K}{\lambda m_2}\Big)^{3/2};$$

$$(d)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\frac{u^2\phi}{\lambda} \le \frac{4K}{m_2}\Big(\frac{K}{\lambda m_2}\Big)^{3/2};$$

$$(e)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\frac{u^3}{3\lambda^2} \le \frac{2K}{3m_2}\Big(\frac{K}{\lambda m_2}\Big)^{3/2};$$

$$(f)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\frac{u^3\phi}{\lambda^2} \le \frac{4K}{m_2}\Big(\frac{K}{\lambda m_2}\Big)^{2}.$$

Proof. We only give the proof of part (a): the others are similar. Observe that the ϕ–integral can be split into the ranges $0<\phi<c(\lambda m_2)^{-1/2}$ and $\phi\ge c(\lambda m_2)^{-1/2}$, where $c>0$ can be chosen at will, and that different estimates of the u–integral can be used in each range. For (a), $\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}$ can be estimated by either $\lambda\phi$ or by $K/m_2\phi$. Use the first of these for $0<\phi<c(\lambda m_2)^{-1/2}$ and the second for larger values of ϕ, giving the bound

$$\frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\phi^{-1} \le \frac{2c}{\sqrt{\lambda m_2}} + 2\Big(\frac{K}{\lambda m_2}\Big)\frac{\sqrt{\lambda m_2}}{c},$$

optimized by taking $c=\sqrt{K}$.

Lemma 6.3. *Integral estimates of the second kind:*

$$(a)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\phi \le 2\zeta\Big(\frac{K}{\lambda m_2}\Big);$$

$$(b)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}u\phi^2 \le 2\frac{K\zeta}{m_2}\Big(\frac{K}{\lambda m_2}\Big);$$

$$(c)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}4u\phi^3 \le 4\frac{K\zeta^2}{m_2}\Big(\frac{K}{\lambda m_2}\Big).$$

Proof. In each case, the bound is derived by taking the u–integral out to ∞.

Lemma 6.4. *Integral estimates of the third kind:*

$$(a)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}(u/\lambda) \le \sqrt{\pi}\Big(\frac{K}{\lambda m_2}\Big)^{3/2};$$

$$(b)\qquad \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}(u^2/2\lambda^2) \le \Big(\frac{K}{\lambda m_2}\Big)^{2}.$$

Proof. In each case, first switch the order of integration. For (a), we then obtain

$$\frac{2}{\lambda}\int_0^{\zeta} d\phi \int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}(u/\lambda) = \frac{2}{\lambda}\int_0^{\lambda\zeta} du \int_{u/\lambda}^{\zeta} d\phi\, e^{-m_2 u\phi/K}(u/\lambda)$$
$$\leq \frac{2}{\lambda}\int_0^{\lambda\zeta} du\, \frac{u}{\lambda}\, e^{-m_2 u^2/K\lambda}\, \frac{K}{m_2 u},$$

from which the result follows.

There are two cases in which (6.2) is easy to bound. For $I(f_1, h_2)$, we have

$$|D_\lambda(f_1, h_2, \phi, u)| \leq |e^{i(\phi - u/\lambda)} - 1| \leq \phi$$

in the range of the integrals in (6.2), and hence, by Lemmas 6.1(c) and 6.3(a), it follows that

$$|I(f_1, h_2)| \leq 2\zeta\Big(\frac{K}{\lambda m_2}\Big). \tag{6.5}$$

This and the fact that $K \leq 12/5$ proves Lemma 4.2(b). Then, for $I(f_2, h_1)$, we have

$$|D_\lambda(f_2, h_1, \phi, u)| \leq \frac{1}{\sqrt{2(1-\cos\phi)}} \leq \frac{\sqrt{2}}{\phi},$$

since in $0 < \phi \leq \zeta \leq \sqrt{(m_2/m_4)}$ it follows that

$$1 - \cos\phi \geq \phi^2(1 - \zeta^2/2)/2 \geq \phi^2/4. \tag{6.6}$$

Hence, from Lemmas 6.1(c) and 6.2(a), we conclude that

$$|I(f_2, h_1)| \leq 4\sqrt{2}\Big(\frac{K}{\lambda m_2}\Big)^{1/2}, \tag{6.7}$$

proving Lemma 4.2(c).

The remaining pair require more thought. For $I(f_1, h_1)$, we simply have

$$D_\lambda(f_1, h_1, \phi, u) = \sin S_{\lambda,j,w}(\phi, u), \tag{6.8}$$

where

$$S_{\lambda,j,w}(\phi, u) = -j\phi + w(\phi - u/\lambda) - \lambda\sum_{j\geq 1}\mu_j[\sin\{j(\phi - u/\lambda)\} - \sin j\phi]. \tag{6.9}$$

For $I(f_2, h_2)$, the expression is somewhat more complicated:

$$D_\lambda(f_2, h_2, \phi, u) = -\sin S_{\lambda,j,w}(\phi, u) + E_\lambda(u, \phi), \tag{6.10}$$

where

$$E_\lambda(u,\phi) = \frac{1}{2}\sin S_{\lambda,j,w}(\phi,u)\Big\{[1-\cos(\lambda^{-1}u)] + \frac{\sin\phi\sin(\lambda^{-1}u)}{1-\cos\phi}\Big\} \\ + \frac{1}{2}\cos S_{\lambda,j,w}(\phi,u)\Big\{\sin(\lambda^{-1}u) - \frac{(1-\cos(\lambda^{-1}u))\sin\phi}{1-\cos\phi}\Big\}. \tag{6.11}$$

The contribution to $I(f_2,h_2)$ from (6.11) can now be directly bounded, using Lemma 6.1(c) as before, combined with Lemma 6.4(b) for the term with $[1-\cos(\lambda^{-1}u)]$; with (6.6) and Lemma 6.2(b) for that with $\sin\phi\sin(\lambda^{-1}u)/(1-\cos\phi)$; with Lemma 6.4(a) for that with $\sin(\lambda^{-1}u)$; and with Lemma 6.2(c) for that with $(1-\cos(\lambda^{-1}u))\sin\phi/(1-\cos\phi)$. Putting these together, the result is no larger in modulus than

$$\begin{aligned} &\frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\ \exp\{-C_\lambda(\phi,u)\}|E_\lambda(\phi,u)| \\ &\le \frac{1}{2}\Big(\frac{K}{\lambda m_2}\Big)^2 + 2\sqrt{2}\Big(\frac{K}{\lambda m_2}\Big) + \frac{1}{2}\sqrt{\pi}\Big(\frac{K}{\lambda m_2}\Big)^{3/2} + \frac{4\sqrt{2}}{3}\Big(\frac{K}{\lambda m_2}\Big)^{3/2} \\ &\le \frac{20}{\lambda m_2}, \end{aligned} \tag{6.12}$$

when $K \le 12/5$ and $\lambda m_2 \ge 1$. Hence, to estimate both $I(f_1,h_1)$ and $I(f_2,h_2)$ and thus complete the proof of Lemma 4.2, all that remains is to prove the following lemma.

Lemma 6.5. *For $0 < \zeta \le \sqrt{m_2/m_4}$ and $\zeta(\lambda m_2)^{1/2} \ge 1$, we have*

$$\Big|\frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\ \exp\{-C_\lambda(\phi,u)\}\sin S_{\lambda,j,w}(\phi,u)\Big| \le \frac{67}{\lambda m_2}.$$

Remark. Note that the conditions of the lemma imply that $\lambda \ge m_4/m_2^2$, and that this is exactly the condition required of λ if ζ is taken equal to its upper limit.

Proof. We start by reducing the integral in two steps, to a form in which we can explicitly make computations. First, we use Lemmas 6.1(b) and (c) to prove that

$$\begin{aligned} &\Big|\frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\ \exp\{-C_\lambda(\phi,u)\}[\sin S_{\lambda,j,w}(\phi,u) - \sin\{-j\phi + w(\phi-u/\lambda) + m_1 u\}] \\ &\qquad \le \frac{2}{\lambda}\int_0^\zeta d\phi\int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\,\frac{m_3}{6}\{3u\phi^2 + 3\lambda^{-1}u^2\phi + \lambda^{-2}u^3\} \\ &\qquad \le \frac{Km_3}{m_2}\Big(\frac{K}{\lambda m_2}\Big)\Big\{\zeta + \frac{7}{3}\Big(\frac{K}{\lambda m_2}\Big)^{1/2}\Big\}, \end{aligned} \tag{6.13}$$

where the last line follows by applying Lemmas 6.3(b) and 6.2(d,e): $K = K_\zeta$ as before. Then, since $|e^{-a} - e^{-b}| \le e^{-(a\wedge b)}|a-b|$ for $a, b > 0$, it follows from Lemmas 6.1(a,c) that

$$\Big|\frac{2}{\lambda}\int_0^\zeta d\phi \int_0^{\lambda\phi} du\,\big[\exp\{-C_\lambda(\phi,u)\} - \exp\{-m_2(u\phi - u^2/2\lambda)\}\big] \sin\{-j\phi + w(\phi - u/\lambda) + m_1 u\}\Big|$$
$$\le \frac{2}{\lambda}\int_0^\zeta d\phi \int_0^{\lambda\phi} du\, e^{-m_2 u\phi/K}\,\frac{m_4}{24}\{4u\phi^3 + \lambda^{-2}u^3\phi\} \le \frac{Km_4}{6m_2}\Big(\frac{K}{\lambda m_2}\Big)\Big\{\zeta^2 + \Big(\frac{K}{\lambda m_2}\Big)\Big\}, \tag{6.14}$$

where the last line follows now from Lemmas 6.3(c) and 6.2(f). So it now remains, after changing the order of integration, to consider

$$\frac{2}{\lambda}\int_0^{\lambda\zeta} du \int_{u/\lambda}^{\zeta} d\phi\, \exp\{-m_2 u\phi + m_2 u^2/2\lambda\}\{\sin k\phi \cos su + \cos k\phi \sin su\}, \tag{6.15}$$

where $s = m_1 - w/\lambda \in \mathbb{R}$ and $k = w - j \in \mathbb{Z}$.

The ϕ–integrations can be explicitly evaluated:

$$\int_{u/\lambda}^{\zeta} d\phi\, e^{-m_2 u\phi}\sin k\phi$$
$$= \frac{e^{-m_2u^2/\lambda}}{m_2^2u^2 + k^2}\{m_2 u\sin(ku/\lambda) + k\cos(ku/\lambda)\} - \frac{e^{-m_2u\zeta}}{m_2^2u^2+k^2}\{m_2u\sin k\zeta + k\cos k\zeta\}, \tag{6.16}$$

(zero if $k = 0$) and

$$\int_{u/\lambda}^{\zeta} d\phi\, e^{-m_2 u\phi}\cos k\phi$$
$$= \frac{e^{-m_2u^2/\lambda}}{m_2^2u^2 + k^2}\{m_2 u\cos(ku/\lambda) - k\sin(ku/\lambda)\} - \frac{e^{-m_2u\zeta}}{m_2^2u^2+k^2}\{m_2u\cos k\zeta - k\sin k\zeta\}, \tag{6.17}$$

leaving eight u–integrals in all to consider, in order to estimate (6.15). The first four, obtained when (6.16) is multiplied by $\frac{2}{\lambda}\exp\{m_2u^2/2\lambda\}\cos su$ and integrated from 0 to $\lambda\zeta$, are relatively simply bounded by

$$\frac{1}{\lambda m_2} + \frac{\pi}{\lambda m_2} + \frac{2}{\lambda m_2 e} + \frac{\pi}{\lambda m_2}; \tag{6.18}$$

for the first and third terms, the AM–GM inequality gives $2m_2uk \le m_2^2u^2 + k^2$, giving

$$\frac{2}{\lambda}\int_0^{\lambda\zeta} du\, e^{-m_2u^2/(2\lambda)} m_2 u\sin(ku/\lambda) \le \lambda^2\int_0^\infty u e^{-m_2u^2/(2\lambda)}\,du = \frac{1}{\lambda m_2}$$

and

$$\Big|\frac{2}{\lambda}\int_0^{\lambda\zeta} du\,\frac{e^{-m_2u(\zeta-(u/2\lambda))}}{m_2^2u^2+k^2}\,m_2u\sin k\zeta\cos su\Big|$$
$$\le \frac{2}{\lambda}\int_0^{\infty} du\,km_2u\zeta\frac{e^{-m_2u\zeta/2}}{m_2^2u^2+k^2}\le\frac{\zeta}{\lambda}\int_0^{\infty}e^{-m_2u\zeta/2}\,du=\frac{2}{\lambda m_2};$$

the second and fourth are bounded by

$$\frac{2}{\lambda}\int_0^{\infty}\frac{k}{m_2^2u^2+k^2}\,du=\frac{\pi}{\lambda m_2}.$$

Two of the last four, obtained from (6.17) by multiplying by $\frac{2}{\lambda}\exp\{m_2u^2/(2\lambda)\}\sin su$ and integrating from 0 to $\lambda\zeta$, pose more of a problem:

$$\begin{aligned}\Big|\frac{2}{\lambda}\int_0^{\lambda\zeta} du\,\frac{e^{-m_2u^2/2\lambda}}{m_2^2u^2+k^2}\,m_2u\cos(ku/\lambda)\sin su\Big|\\ \le A_1+\frac{2}{\lambda}\int_0^{\lambda\zeta}du\,\frac{e^{-m_2u^2/2\lambda}}{m_2^2u^2+k^2}\frac{m_2u^3}{2\lambda^2},\end{aligned}\tag{6.19}$$

where

$$A_1=\Big|\frac{2}{\lambda}\int_0^{\lambda\zeta}du\,\frac{e^{-m_2u^2/2\lambda}}{m_2^2u^2+k^2}\,m_2u\sin su\Big|,\tag{6.20}$$

and

$$\begin{aligned}\Big|\frac{2}{\lambda}\int_0^{\lambda\zeta} du\,\frac{e^{-m_2u(\zeta-(u/2\lambda))}}{m_2^2u^2+k^2}\,m_2u\cos k\zeta\sin su\Big|\\ \le A_2+\frac{2}{\lambda}\int_0^{\lambda\zeta}du\,\frac{e^{-m_2u\zeta/2}}{m_2^2u^2+k^2}m_2u\frac{k^2\zeta^2}{2},\end{aligned}\tag{6.21}$$

where

$$A_2=\Big|\frac{2}{\lambda}\int_0^{\lambda\zeta}du\,\frac{e^{-m_2u(\zeta-(u/2\lambda))}}{m_2^2u^2+k^2}\,m_2u\sin su\Big|.\tag{6.22}$$

Apart from A_1 and A_2, the remaining four contributions to (6.15) from (6.17) can be seen to give at most

$$\frac{2}{\lambda m_2}+\frac{2}{\lambda m_2}+\frac{4}{\lambda m_2}+\frac{4}{\lambda m_2},\tag{6.23}$$

all obtained using the inequality $k^2\le m_2^2u^2+k^2$. For both A_1 and A_2, an integration by parts is needed.

First, if $|s|\le\sqrt{\frac{m_2\pi}{2\lambda}}$, both A_1 and A_2 are bounded by

$$\begin{aligned}\frac{2}{\lambda}\int_0^{\infty}du\,\frac{|s|}{m_2}\frac{m_2^2u^2}{m_2^2u^2+k^2}\exp\{-m_2u^2/(2\lambda)\}\\ \le\frac{2}{\lambda m_2}\int_0^{\infty}du\,\sqrt{\frac{m_2\pi}{2\lambda}}\exp\{-m_2u^2/(2\lambda)\}\le\frac{\pi}{\lambda m_2}.\end{aligned}\tag{6.24}$$

Then, if $|s| > \sqrt{\frac{m_2\pi}{2\lambda}}$, the integrals defining A_1 and A_2 can be separated into the ranges $u \le \pi/(2|s|)$ and $u > \pi/(2|s|)$. In the former, the bound $|s|/m_2$ for the integrand in either integral is enough to show that the contribution to each is no larger than

$$\frac{2|s|}{\lambda m_2}\frac{\pi}{2|s|} = \frac{\pi}{\lambda m_2}. \tag{6.25}$$

For the latter range, taking A_1, we have

$$\begin{aligned}\frac{2}{\lambda}\int_{\pi/(2|s|)}^{\lambda\zeta} du \frac{e^{-m_2u^2/2\lambda}}{m_2^2u^2+k^2} m_2 u \sin su &= \frac{2}{\lambda}\Big[-\frac{m_2 u\cos su}{s}\frac{e^{-m_2u^2/2\lambda}}{m_2^2u^2+k^2}\Big]_{\pi/(2|s|)}^{\lambda\zeta} \\ &+ \frac{2}{\lambda}\int_{\pi/(2|s|)}^{\lambda\zeta} du \frac{m_2\cos su}{s}\frac{e^{-m_2u^2/2\lambda}}{m_2^2u^2+k^2}\Big\{1-\frac{2m_2^2u^2}{m_2^2u^2+k^2}-\frac{m_2u^2}{\lambda}\Big\},\end{aligned} \tag{6.26}$$

with the first part no greater in modulus than

$$\frac{2}{\lambda}\frac{e^{-m_2\lambda\zeta^2/2}}{m_2\lambda\zeta|s|} \le \frac{2\sqrt{2/\pi}}{\lambda m_2\zeta\sqrt{\lambda m_2}}e^{-m_2\lambda\zeta^2/2}, \tag{6.27}$$

and the second bounded by

$$\frac{2}{\lambda}\int_{\pi/(2|s|)}^{\infty} du\Big\{\frac{1}{m_2u^2|s|}+\frac{1}{\lambda|s|}e^{-m_2u\pi/4\lambda|s|}\Big\} \le \frac{4}{\lambda m_2\pi}+\frac{8}{\lambda m_2\pi} = \frac{12}{\lambda m_2\pi}. \tag{6.28}$$

Combining (6.24) – (6.28), we thus find that

$$A_1 \le \frac{1}{\lambda m_2}\Big\{\frac{1}{\pi}(12+\pi^2)+\frac{2\sqrt{2/\pi}}{\zeta\sqrt{\lambda m_2}}e^{-m_2\lambda\zeta^2/2}\Big\}. \tag{6.29}$$

Finally, to compute the contribution to A_2 from the range $u > \pi/(2|s|)$ when also $|s| > \sqrt{\frac{m_2\pi}{2\lambda}}$, we have

$$\begin{aligned}\frac{2}{\lambda}\int_{\pi/(2|s|)}^{\lambda\zeta} du \frac{e^{-m_2u(\zeta-(u/2\lambda))}}{m_2^2u^2+k^2} m_2 u \sin su &= \frac{2}{\lambda}\Big[-\frac{m_2 u\cos su}{s}\frac{e^{-m_2u(\zeta-(u/2\lambda))}}{m_2^2u^2+k^2}\Big]_{\pi/(2|s|)}^{\lambda\zeta} \\ &+ \frac{2}{\lambda}\int_{\pi/(2|s|)}^{\lambda\zeta} du \frac{m_2\cos su}{s}\frac{e^{-m_2u(\zeta-(u/2\lambda))}}{m_2^2u^2+k^2}\Big\{1-\frac{2m_2^2u^2}{m_2^2u^2+k^2}-m_2u(\zeta-u/\lambda)\Big\},\end{aligned} \tag{6.30}$$

with the first part bounded by

$$\frac{2\sqrt{2/\pi}}{\lambda m_2\zeta\sqrt{\lambda m_2}}e^{-m_2\lambda\zeta^2/2}, \tag{6.31}$$

as before, and the second by

$$\frac{2}{\lambda}\int_{\pi/(2|s|)}^{\infty} du\,\Big\{\frac{1}{m_2u^2|s|}+\frac{\zeta}{u|s|}e^{-m_2\zeta u/2}\Big\}\le\frac{12}{\lambda m_2\pi}, \tag{6.32}$$

since $u|s|\ge\pi/2$ in the range of integration; thus, for A_2 also, from (6.24) – (6.25) and (6.31) – (6.32), we have the bound

$$A_2\le\frac{1}{\lambda m_2}\Big\{\frac{1}{\pi}(12+\pi^2)+\frac{2\sqrt{2/\pi}}{\zeta\sqrt{\lambda m_2}}e^{-m_2\lambda\zeta^2/2}\Big\}. \tag{6.33}$$

Combining (6.23), (6.29) and (6.33), we thus have the bound

$$\frac{1}{\lambda m_2}\Big\{12+\frac{2}{\pi}(12+\pi^2)+\frac{4\sqrt{2/\pi}}{\zeta\sqrt{\lambda m_2}}e^{-m_2\lambda\zeta^2/2}\Big\}$$

for the four integrals contributing to (6.15) by way of (6.17), and hence, adding (6.18) as well to complete the bound on (6.15), and then incorporating (6.13) and (6.14), we finally arrive at the estimate

$$\begin{aligned}\Big|\frac{2}{\lambda}\int_0^{\zeta}d\phi\int_0^{\lambda\phi}du\,\exp\{-C_\lambda(\phi,u)\}\sin S_{\lambda,j,w}(\phi,u)\Big|&\\ \le\frac{1}{\lambda m_2}\Big\{36+\frac{4\sqrt{2/\pi}}{\zeta\sqrt{\lambda m_2}}e^{-m_2\lambda\zeta^2/2}&\\ +K^2\Big[\frac{m_3}{m_2}\Big\{\zeta+\frac{7}{3}\Big(\frac{K}{\lambda m_2}\Big)^{1/2}\Big\}+\frac{m_4}{6m_2}\Big\{\zeta^2+\Big(\frac{K}{\lambda m_2}\Big)\Big\}\Big]\Big\}.&\end{aligned} \tag{6.34}$$

In particular, if $\zeta\le\sqrt{m_2/m_4}\le 1$ and $\zeta\sqrt{\lambda m_2}\ge 1$, it follows that $K\le 12/5$, and the conclusion of the lemma is implied by simple estimates of (6.34), using Hölder's inequality.

We need two further technical lemmas, used in the proof of Theorem 3.1.

Lemma 6.6. *For all x on the unit circle, all $z=x+t(1-x)$ on the line segment $[x,1]$ and all $j=1,2,\ldots$, we have*

$$\mathcal{R}e\int_{[x,z]}\frac{1-s^j}{1-s}\,ds\ge 0.$$

Proof. We observe that for any z with $|z|\le 1$ we have

$$\mathcal{R}e\,z^j\le|z|^j\le 1.$$

However, substituting $x + u(1-x)$ for s in the integral, with $0 \le u \le t$, we obtain

$$\int_{[x,z]} \frac{1-s^j}{1-s}\, ds = \int_0^t \frac{1-\{x+u(1-x)\}^j}{(1-x)(1-u)}\,(1-x)\,du$$
$$= \int_0^t \frac{1-\{x+u(1-x)\}^j}{(1-u)}\,du,$$

from which it now follows that

$$\mathcal{R}e \int_{[x,z]} \frac{1-s^j}{1-s}\, ds = \int_0^t \frac{1-\mathcal{R}e\,\{[x+u(1-x)]^j\}}{(1-u)}\,du \ge 0.$$

Lemma 6.7. *For all x on the unit circle, all $z = x + t(1-x) \in [x,1]$ and all $j = 1, 2, \ldots$, we have*

$$\mathcal{R}e \int_{[x,z]} \frac{1-s^j}{1-s} ds \le j^2 t(1-\mathcal{R}e\,x).$$

Proof. We begin by observing that

$$1 - a^j = j(1-a) + \sum_{k=1}^{j-1} (1-a)(a^k - 1),$$

and that therefore, if $|a| \le 1$,

$$1 - \mathcal{R}e\,a^j \le j(1-\mathcal{R}e\,a) + \sum_{k=1}^{j-1} |1-a||a^k - 1|$$
$$\le j(1-\mathcal{R}e\,a) + \sum_{k=1}^{j-1} k|1-a|^2 = j(1-\mathcal{R}e\,a) + \frac{j(j-1)}{2}|1-a|^2.$$

Hence, for x on the unit circle and $0 \le u \le 1$, we deduce that

$$1 - \mathcal{R}e\,\{(x+u(1-x))^j\} \le j(1-\mathcal{R}e\,[(x+u(1-x)]) + \frac{j(j-1)}{2}|1-[x+u(1-x)]|^2$$
$$= (1-u)j(1-\mathcal{R}e\,x) + j(j-1)(1-\mathcal{R}e\,x)(1-u)^2 \le j^2(1-u)(1-\mathcal{R}e\,x).$$

Thus it follows that

$$\mathcal{R}e \int_{[x,z]} \frac{1-s^j}{1-s}\, ds = \int_0^t \frac{1-\mathcal{R}e\,(x+u(1-x))^j}{1-u}\,du$$
$$\le \int_0^t j^2(1-\mathcal{R}e\,x)\,du = j^2 t(1-\mathcal{R}e\,x),$$

as claimed.

References

[1] D. J. Aldous (1989) *Probability approximations via the Poisson clumping heuristic.* Springer, New York.

[2] R. Arratia, L. Goldstein and L. Gordon (1990) Poisson approximation and the Chen–Stein method. *Stat. Science* **5,** 403–434.

[3] A. D. Barbour (1995) Stein's method. In *Encyclopaedia of Statistical Science,* 2nd Edn, Wiley, New York.

[4] A. D. Barbour and T. C. Brown (1992) Stein's method and point process approximation. *Stoch. Procs Applics* **43,** 9–31.

[5] A. D. Barbour, L. H. Y. Chen and W.–L. Loh (1992) Compound Poisson approximation for nonnegative random variables using Stein's method. *Ann. Probab.* **20,** 1843–1866.

[6] A. D. Barbour, L. Holst and S. Janson (1992) *Poisson approximation.* Oxford University Press.

[7] B. Bollobás (1985) *Random graphs.* Academic Press, London.

[8] T. C. Brown and A. Xia (1995) On Stein–Chen factors for Poisson approximation. *Stat. Prob. Letters* **23,** 327–332.

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

[10] M. Roos (1993) Stein–Chen method for compound Poisson approximation. Ph. D. Thesis, University of Zürich.

[11] M. Roos (1994) Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Prob.* **4,** 1177–1187.

[12] M. Roos and D. Stark (1996) Compound poisson approximation for visits to small sets in a Markov chain. Preprint.

[13] A. Yu. Zaitsev (1988) On the uniform approximation of distributions of sums of independent random variables. *Th. Prob. Applics* **32,** 40–48.