The Annals of Probability
2001, Vol. 29, No. 3, 1373–1403

STEIN'S METHOD AND BIRTH-DEATH PROCESSES[1]

BY TIMOTHY C. BROWN AND AIHUA XIA

University of Melbourne

Barbour introduced a probabilistic view of Stein's method for estimating the error in probability approximations. However, in the case of approximations by general distributions on the integers, there have been no purely probabilistic proofs of Stein bounds till this paper. Furthermore, the methods introduced here apply to a very large class of approximating distributions on the non-negative integers, among which there is a natural class for higher-order approximations by probability distributions rather than signed measures (as previously). The methods also produce Stein magic factors for process approximations which do not increase with the window of observation and which are simpler to apply than those in Brown, Weinberg and Xia.

1. Introduction. Stein's method first appeared in Stein (1971) and has proved successful in estimating the error in normal approximation to the sum of dependent random variables. Stein's method has been adapted for various distributions, including Poisson in Chen (1975) and Barbour and Hall (1984), the Poisson process in Barbour (1988) and Barbour and Brown (1992), binomial in Ehm (1991), multinomial in Loh (1992), compound Poisson in Barbour, Chen and Loh (1992) and negative binomial in Brown and Phillips (1999), etc. [see also Barbour, Holst and Janson (1992) and references therein]. To adapt Stein's method for a particular distribution is to establish an identity for the distribution (often called the Stein identity), and from this establish a Stein equation which is solved.

Stein's method has been spectacularly successful with the Poisson distribution (in this case it is often appropriately called the Stein-Chen method). For example, Barbour and Hall (1984) establish upper and lower bounds of the same, and therefore, correct order for the error in approximating the distribution of the sum of independent 0-1 random variables. This contrasts with simpler coupling methods which yield upper bounds of the wrong order.

The example of the sum of independent 0-1 random variables is a prototype for many others. Coupling methods produce upper bounds which grow indefinitely as the mean of the distribution increases, whilst Stein's method produces upper bounds which are at worst constant in the size of the mean. This behavior in the upper bound has an important practical consequence: if the distribution being approximated is from a stochastic process in time or

Received April 2000; revised January 2001.

[1]Supported by ARF Research grant RP 3982719 from the National University of Singapore.

AMS 2000 *subject classifications*. Primary 60E05; secondary 60E15, 60F05, 60G55.

Key words and phrases. Stein's method, birth-death process, distributional approximation, total variation distance, Poisson process approximation, Wasserstein distance, compound Poisson distribution, negative binomial distribution, polynomial birth-death distribution.

space, the order of the upper bound from Stein's method typically does not depend on how long or over what volume the process is observed. That is, with Stein's method the order of error does not increase with the size of the window of observation. Typically, the order of error depends instead on a crucial system parameter; in the case of the sum of independent 0-1 random variables the parameter (and bound) is the maximum probability of 1. Someone wanting to use the approximation therefore need only be concerned about the size of the system parameter rather than the size of the window of observation.

Implementing Stein's method often involves two parts: the first part obtains estimates on the Stein equation solution and the second part uses these properties for analysing different problems. The lack of dependence on the size of the window of observation, mentioned in the previous paragraph, depends crucially on the the first part. In the case of discrete distributions, it is the differences of the solution of the Stein equation which are estimated in the first part, and the maximum size of these is often called the Stein "magic factor." For the Poisson case, the Stein magic factor is essentially the reciprocal of the mean of the distribution. The Stein magic factor depends only on the approximating distribution and not on the distribution being approximated. Thus good estimates of the magic factor can then be applied to many different problems.

Barbour (1988) introduced an important new view of Stein's method using reversible Markov processes. In this view, the distribution used for approximation is the equilibrium distribution of a Markov process, and the Stein identity links the equilibrium distribution to the generator of the Markov process. For the case of the Poisson distribution, and the Poisson process, the Markov process is an immigration-death process. The introduction of Markov processes permits probability theory to *generate* new Stein identities for new approximating distributions. Furthermore, it has been hoped that knowledge in probability theory would illuminate Stein's magic factors.

To date, the potential advantages of the probabilistic approach in Stein's method have been limited in two ways. Probabilists would hope and expect that the probabilistic approach would yield elegant and intuitive derivations of Stein magic factors. Moreover, probabilists would hope that the approach would also work well with process approximation as well as distribution approximation. On the other hand, till now there has been no elegant probabilistic derivation of the Stein magic factor, except for the Poisson random variable case in Xia (1999). Furthermore, although the probabilistic approach was crucial in developing bounds for Poisson process approximation, until recently the Stein magic factor for process approximation grew with the logarithm of the size of the window of observation. Brown, Weinberg and Xia (2000) remedied this but at the expense of bounds which are very complicated to compute for particular problems.

In Section 2 of this paper, we give a neat probabilistic derivation of Stein magic factors, not only for the Poisson distribution but for a very large class of distributions on the non-negative integers. This occurs when the Stein equation comes from a birth-death process on the integers. A key point is that the solution to Stein equation is an explicit linear combination of mean upward and downward transition times of the birth-death process (Lemma 2.1). There is a very pleasing probabilistic derivation of explicit formulae for these means (Lemma 2.2) (essentially three pictures). In many cases, all differences of the solution of the Stein equation are negative except one. This particular structure of signs of the differences is another key point in the derivation of the Poisson Stein magic factors. Necessary and sufficient conditions are given for this structure if the approximating distribution is on the non-negative integers and the Stein equation comes from a birth-death process. These conditions include the sufficient conditions in Barbour, Holst and Jansen [(1992), Lemma 9.2.1].

The general theory suggests a wide class of distributions on the integers for which probabilities are very simple to compute and which can have an arbitrary number of parameters. The distributions are called *polynomial birth-death distributions* and are introduced in section 3. They include Poisson, negative binomial, binomial and hypergeometric distributions. In the case of two parameters, the new distributions are perhaps more comparable to the normal distribution than the Poisson distribution since the normal distribution is determined by two parameters while the Poisson distribution is determined by one. It is shown in section 3 that a polynomial birth-death distribution with two parameters can approximate the sum of independent Bernoulli trials with order of accuracy as good as the compound Poisson *signed* measure approximation in Barbour and Xia (1999). The benefit of approximation by a probability distribution rather than a signed measure is that all of the tools of probability theory are available for the approximand, and the meaning of moments of the approximand is clear. Unlike the binomial, the new distribution does not require truncation in approximation and the accuracy of approximation is usually also higher. Numerical examples are provided to compare the performance of Poisson, binomial and polynomial birth-death approximations.

Section 4 shows that the negative binomial distribution approximates the number of 2-runs of successes in Bernoulli trials with accuracy at worst $p^2/\sqrt{\lambda}$, where p is the success probability and λ is the mean number of success runs. The natural approximation to the runs distribution would be compound Poisson with a geometric summand (since the length of a run of successes is geometric). However, the order of approximation here is p^2. This reveals [as did approximation with a more complicated compound Poisson distribution in Barbour and Xia (1999)] a very surprising fact from Stein's method - the order of approximation can even improve as the window of observation increases! It should also be noted that negative binomial approximation is desirable since this distribution is widely available in computer packages such as EXCEL.

The reason for including these applications is that they are relatively straightforward and illustrate the power of the general results. In the case of the negative binomial approximation, the results do not follow from the results in Barbour, Holst and Janson (1992).

The techniques used in Sections 3 and 4 for calculating bounds use the Stein magic factors of Section 2: this is part one from the fourth paragraph of the introduction. In Section 4, the technique for part 2 is similar to that in Barbour and Hall (1984) in the way in which independence is exploited, although the calculations are more complicated. This technique amounts to use of the Palm distribution of a point process on a discrete space, and thus extension to dependence is possible. It is not done here to avoid obscuring the beauty and power of the techniques. Whenever there are unit per capita death rates, Palm process calculations can be done for part 2. In Section 3, the death rates are quadratic, and the appropriate point process calculations use second order Palm distributions. Again, it should be possible, with more complicated calculations, to incorporate dependence. Furthermore, with cubic or higher order death rates, Palm distributions of higher order could be used, and the result would be bounds of arbitrary higher order. Again, it should be stressed that in each case the approximand would be a probability distribution not a signed measure.

Poisson process approximation is a natural step forward from Poisson random variable approximation. Barbour (1988) and Arratia, Goldstein and Gordon (1989) extended Stein's method to the approximation in distribution by a Poisson process of discrete sums of the form $\Xi = \sum_{i=1}^{n} X_i \delta_{Y_i}$, where Y_i is a (possibly random) mark associated with X_i. The extension to general point processes was accomplished in Barbour and Brown (1992). There, a point process is regarded as a random configuration in a compact metric space and the approximations were based on either Janossy densities (the "local" approach) or Palm distributions (the "coupling" approach). It has been shown that Stein's magic factors (interpreted as the maximum over all states) for Poisson process approximation are by no means as good as those for Poisson random variable approximation [see Brown and Xia (1995b)], and thus the bounds of errors based on these uniform magic factors in applications are not as good as would be hoped. A non-uniform bound is suggested in Brown, Weinberg and Xia (2000) and has been proved successful in applications, namely, the error bounds are of optimal order. However, it is necessary to compute a large number of quantities which are specific to particular applications and this limits the usefulness of the bounds in Brown, Weinberg and Xia (2000). This defect, using some of the ideas from section 2 on integer distributions, is remedied in section 5.

For random variable approximation, we use the total variation distance to compare the difference between two probability measures $\mathbf{Q}_1$ and $\mathbf{Q}_2$ on $\mathbf{Z}_+$:

$$\begin{aligned} d_{TV}(\mathbf{Q}_1, \mathbf{Q}_2) &= \sup_{f \in \mathscr{F}} \left| \int f d\mathbf{Q}_1 - \int f d\mathbf{Q}_2 \right| \\ &= \tfrac{1}{2} \sum_{i=0}^{\infty} |\mathbf{Q}_1\{i\} - \mathbf{Q}_2\{i\}| \\ &= \sup_{A \subset \mathbf{Z}_+} |\mathbf{Q}_1(A) - \mathbf{Q}_2(A)|, \end{aligned}$$

where $\mathscr{F} := \{f:\ \mathbf{Z}_+ \mapsto [0,1]\}$.

2. Stein identities and birth-death processes. As explained in the Introduction, implementations of Stein's method often involve two parts. This section is concerned only with the first part: calculation of Stein magic factors. The result is Theorem 2.10. The main departures from previous work are that only probabilistic concepts are involved, there is no recursion and necessary and sufficient conditions are given for the existence of very simple Stein magic factors.

Suppose π is a distribution on $\mathbf{Z}_+ := \{0, 1, 2, \ldots\}$ or $\{0, 1, 2, \ldots, m\}$ for some finite m. Suppose that π attributes positive probability to each integer in the range - if not, relabel the states so it does so. Consider a birth-death process that has π as its stationary distribution. From now on we take the infinite state space but note that everything works in the same way for the finite state space.

There are infinitely many such birth-death processes and they are all positive recurrent: if the parameters process are α_i for births and β_i for deaths, then we suppose that the parameters are designed so that the detailed balance equations are satisfied:

$$\pi_i \alpha_i = \pi_{i+1}\beta_{i+1}, \ \ i \in \mathbf{Z}_+. \tag{2.1}$$

Accordingly, any such process is time-reversible [see Keilson (1979)]. Once α_0 or β_1 is specified (and this can be done in an arbitrary fashion as any non-negative number), the other is determined by (2.1). Similarly once α_1 or β_2 is specified, the other is determined by (2.1) and so on. The positive recurrence follows from lemma 2.2 and the fact that the transition from any state to another is the sum of transition times from adjoining states.

Equation (2.1), if multiplied by any bounded function g on $\mathbf{Z}_+$ and summed over i, gives

$$\sum_{i=0}^{\infty} g(i+1)\alpha_i \pi_i = \sum_{i=0}^{\infty} g(i+1)\pi_{i+1}\beta_{i+1},$$

which, on introducing a random variable X with distribution π, may be written as

$$\mathbf{E}(\alpha_X g(X+1) - \beta_X g(X)) = 0. \tag{2.2}$$

This equation is the *Stein identity* for π, $\{\alpha_i\}$, $\{\beta_i\}$. The Stein identity suggests the *Stein equation*: for any bounded function f on $\mathbf{Z}_+$ let g be the solution to

$$\mathscr{B}g(i) := \alpha_i g(i+1) - \beta_i g(i) = f(i) - \pi(f), \tag{2.3}$$

[conventionally $g(0) = 0$]. Taking g to be the indicator of $\{i\}$ gives (2.1) from (2.2). Thus, the reversibility criterion (2.1) is equivalent to the Stein identity (2.2), and either leads to the Stein equation (2.3).

Suppose that for $i \in \mathbf{Z}_+$, Z_i is a birth-death process satisfying (2.1) and started in state i. For $i, j \in \mathbf{Z}_+$, define

$$\tau_{ij} = \inf\{t : Z_i(t) = j\}, \qquad \tau_j^+ = \tau_{j,j+1}, \ \ \tau_j^- = \tau_{j,j-1} \tag{2.4}$$

and

$$\overline{\tau_j^+} = \mathbf{E}\left(\tau_j^+\right); \qquad \overline{\tau_j^-} = \mathbf{E}\left(\tau_j^-\right).$$

LEMMA 2.1. *If g_A is the solution for $f = 1_A$ in (2.3) with $A \subset \mathbf{Z}_+$, then, for $i \geq 1$,*

$$g_A(i) = \overline{\tau_i^-}\pi(A \cap [0, i-1]) - \overline{\tau_{i-1}^+}\pi(A \cap [i, \infty)). \tag{2.5}$$

PROOF. Let $g(i) = h(i) - h(i-1)$, and define

$$\mathscr{A}_{rv}h(i) = \alpha_i[h(i+1) - h(i)] + \beta_i[h(i-1) - h(i)], \qquad i \in \mathbf{Z}_+. \tag{2.6}$$

Then $\mathscr{A}_{rv}$ is the generator of the birth-death process Z (rv for random variable) and we can rewrite the Stein equation (2.3) as

$$\mathscr{A}_{rv}h(i) = f(i) - \pi(f), \qquad i \in \mathbf{Z}_+. \tag{2.7}$$

The solution of (2.7) [see Barbour and Brown (1992), noting that Lemma 2.2 shows the required positive recurrence] is

$$h(f, i) = -\int_0^\infty \mathbf{E}[f(Z_i(t)) - \pi(f)]dt. \tag{2.8}$$

In the case that $A = \{j\}$, we simply write g_A as g_j. If $i \leq j$, it follows from (2.8) and the strong Markov property that

$$\begin{aligned} h(1_{\{j\}}, i-1) &= -\mathbf{E}\int_0^{\tau_{i-1,i}}[1_{\{j\}}(Z_{i-1}(t)) - \pi_j]dt - \mathbf{E}\int_{\tau_{i-1,i}}^\infty [1_{\{j\}}(Z_{i-1}(t)) - \pi_j]dt \\ &= \pi_j \mathbf{E}\tau_{i-1,i} + h(1_{\{j\}}, i), \end{aligned}$$

giving

$$g_j(i) = -\pi_j\overline{\tau_{i-1}^+}. \tag{2.9}$$

Similarly, for $i \geq j+1$,

$$g_j(i) = \pi_j\overline{\tau_i^-}. \tag{2.10}$$

Now, (2.5) follows from summing up (2.9) and (2.10) over $j \in A$. □

LEMMA 2.2. *For $j \in \mathbf{Z}_+$,*

$$\overline{\tau_j^+} = \frac{F(j)}{\alpha_j\pi_j} \quad \text{and} \quad \overline{\tau_j^-} = \frac{\bar{F}(j)}{\beta_j\pi_j}, \tag{2.11}$$

where

$$F(j) = \sum_{i=0}^{j}\pi_i; \quad \bar{F}(j) = \sum_{i=j}^{\infty}\pi_i.$$

PROOF. These can be easily proved by conditioning on the time of the first jump after leaving the starting state, and producing a recurrence relation using the fact that the time of transition between states which are two apart is the sum of the times of transition between two adjoining states. It is not necessary to solve the recurrence relation, just to multiply by π_j and add. This is the proof that is given in Keilson [(1979), page 61] for the first formula in (2.11). Alternatively, there is a direct probabilistic proof that reveals the probability behind these pleasingly simple formulae.

We consider only the case of downward transitions because the other case is entirely similar. Consider a stationary process Z^* on $\{j, j+1, \dots, \}$ which has the same birth and death parameters as the process on $\mathbf{Z}_+$. The distribution of $Z^*(t)$ for each t is that of π conditioned to be on $\{j, j+1, \dots\}$. Define a point process N using Z^* by inserting a point at an independent exponential(β_j) time after each transition into state j, provided the process is still in state j after the exponential time. Continue the same probabilistic mechanism until the process leaves j and record no further points until the process next enters j. The inter-point times are therefore independent and identically distributed, apart from the first one, as after the first one, each inter-point time is the exponential(β_j) with probability $\frac{\beta_j}{\alpha_j+\beta_j}$ and otherwise is exponential$(\alpha_j + \beta_j)$ plus a geometric number of independent upwards excursion times from j back to j followed by an exponential(β_j). The point process N is stationary, because Z^* is stationary. Let the mean number of points be κ per unit time. Since the compensator of N is $\{\int_0^t \beta_j 1_{[Z^*(s)=j]} ds\}_{t\geq 0}$,

$$\kappa = \mathbf{E}(N(1)) = \int_0^1 \beta_j \mathbf{P}(Z^*(s) = j) ds = \frac{\beta_j \pi_j}{\bar{F}(j)}. \tag{2.12}$$

But κ is then the reciprocal of the mean time between points in the stationary renewal process N and this is the right hand side of (2.11). The proof is complete if we can construct a stationary version of Z in which the times of transition from j to $j-1$ are the inter-point times of N.

Independent of the process Z^*, realize a coin toss with probability of heads $F(j)$. If the coin is tails, then define Z to coincide with Z^* up to the first point of N. If the coin is heads, construct an independent fragment of the unconditional chain which starts in a state distributed according to the conditional distribution of π on $\{0, 1, 2, \dots, j-1\}$ and finishes at the first entrance to j. Omit, if necessary, the first part of Z^* up till the time that Z^* enters j and use the fragment of Z^* up to the first point of N. The process Z then has been defined to start with the stationary distribution π, and it evolves according to the right laws up to the first point of N. Let $Z_1, Z_2, \dots$ be independent realisations of fragments of the Markov chain each one consisting of an excursion of the chain starting in $j-1$ and finishing in j. Insert one of these fragments after each point of N. The process Z is stationary, Markov and the times between points in Z^* are precisely the times of transition from j to $j-1$, with the possible exception of the first time. However the strong

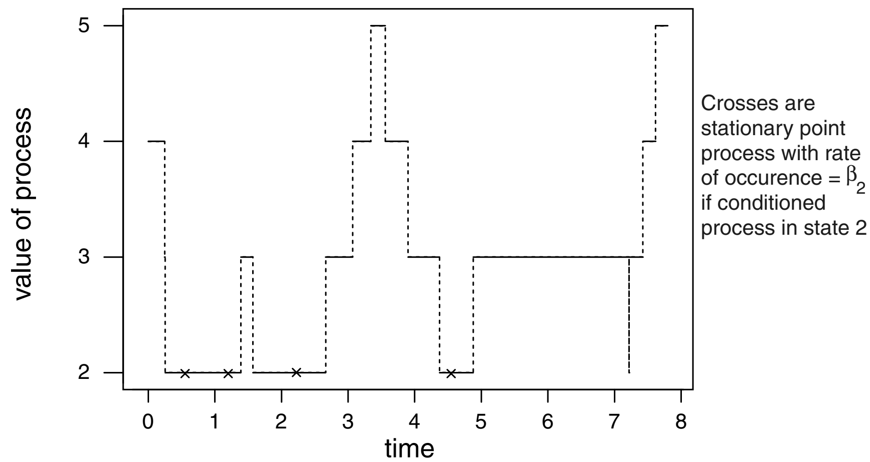

FIG. 1. *Birth-death process conditioned to be on* $\{2, 3\ldots\}$

law still shows that the expected time between points is the limiting sample average of the times between points in N, as required.

The process of construction is illustrated in Figures 1–3. □

Combining Lemma 2.2 with (2.9) and (2.10) gives the following lemma.

LEMMA 2.3. *For* $i \le j$,

$$g_j(i) = \frac{-\pi_j F(i-1)}{\beta_i \pi_i};$$

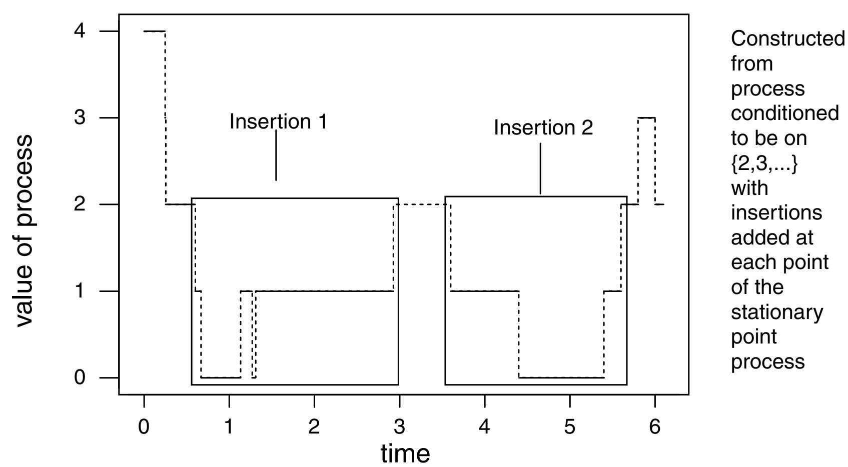

FIG. 2. *Stationary process on whole state space*

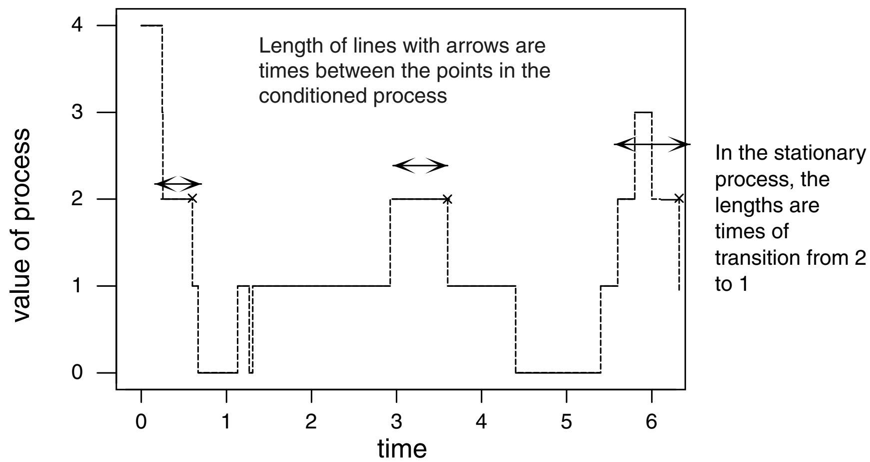

FIG. 3. *Stationary process on whole state space*

and for $i \geq j+1$,

$$g_j(i) = \frac{\pi_j \bar{F}(i)}{\alpha_{i-1}\pi_{i-1}}.$$

LEMMA 2.4. Let $\Delta g_j(i) = g_j(i+1) - g_j(i)$ and let $\delta_i = (\beta_{i+1} - \beta_i) - (\alpha_{i+1} - \alpha_i)$. The following are equivalent:

(C0) For each $j \in \mathbf{Z}_+$, $\Delta g_j(j)$ is the only non-negative difference for Δg_j.

(C1) $\overline{\tau_j^+}$ is increasing in j and $\overline{\tau_j^-}$ is decreasing in j. Here and in the sequel, we use increasing to mean non-decreasing and decreasing to mean non-increasing.

(C2) For each $k = 1, 2, \ldots,$

$$\frac{F(k)}{F(k-1)} \geq \frac{\alpha_k}{\beta_k} \geq \frac{\bar{F}(k+1)}{\bar{F}(k)}. \tag{2.13}$$

(C3) For each $k = 1, 2, \ldots,$

$$\delta_{k-1} \geq \frac{-\sum_{l=0}^{k-2} \delta_l F(l)}{F(k-1)} \tag{2.14}$$

and

$$\delta_k \geq \frac{-\sum_{l=k+1}^{\infty} \delta_l \bar{F}(l+1)}{\bar{F}(k+1)}. \tag{2.15}$$

In particular, a sufficient condition for any of (C0), (C1), (C2) or (C3) is

(C4) For each $k = 1, 2, \ldots,$ defining $\beta_0 = 0$,

$$\alpha_k - \alpha_{k-1} \leq \beta_k - \beta_{k-1}.$$

We defer the proof of Lemma 2.4 to the end of this section.

REMARK 2.5. If (2.14) holds for all the values up to $k-1$, then the right hand side is negative and the condition specifies that δ_k must be at least a certain negative number determined by $\delta_0, \dots, \delta_{k-1}$ and $\pi_0, \dots, \pi_k$. Similarly for (2.15).

REMARK 2.6. Given any probability distribution there is a range of possible values of the birth and death parameters for these conditions to hold. It is easy to construct artificial examples where the conditions do not hold. For example, take $\pi_0 = \frac{1}{8}$, $\pi_1 = \frac{1}{2}$, $\alpha_0 = 1$, $\alpha_1 = 10$. Then detailed balance in equation (2.1) gives $\beta_1 = \frac{1}{4}$ and equation (2.14) is not satisfied for $k = 1$.

REMARK 2.7. For the Poisson distribution, as in Barbour, Holst and Janson (1992), (C0) is satisfied if the birth rates are constant at the parameter of the distribution. This follows immediately from (C4).

EXAMPLE 2.8. If the birth parameters are decreasing and the death parameters are increasing, then (C4) is clearly satisfied and (C0) holds. This example is given as Lemma 9.2.1 in Barbour, Holst and Janson (1992). This includes the Binomial(n, p) distribution with $\alpha_j = (n-j)p$, $\beta_j = j(1-p)$ as noted in Barbour, Holst and Janson (1992). This also includes the Hypergeometric distribution with parameters $R, B, n \in \mathbf{Z}_+$ $(n \le R + B)$:

$$\pi_i = \binom{R}{i}\binom{B}{n-i}\Big/\binom{R+B}{n}, \qquad \max(0, n-B) \le i \le \min(n, R),$$

by taking $\alpha_j = (n-j)(R-j)$, $\beta_j = j(B-n+j)$.

EXAMPLE 2.9. For the negative binomial distribution with parameters $r > 0$ and $0 < q < 1$:

$$\pi_i = \frac{\Gamma(r+i)}{\Gamma(r)i!} q^r (1-q)^i, \qquad i \in \mathbf{Z}_+,$$

and taking $\alpha_j = a + bj$ and $\beta_j = j$ with $a = r(1-q)$ and $b = 1-q$, hence (C4) is satisfied [Brown and Phillips (1999)].

The following proposition provides non-uniform bounds for the solution g to Stein equation (2.3).

THEOREM 2.10. *Any of* (C0)–(C4) *implies that the solution* g *to Stein equation* (2.3) *satisfies*

$$\sup_{f \in \mathscr{F}} |\Delta g(i)| = \frac{\bar{F}(i+1)}{\alpha_i} + \frac{F(i-1)}{\beta_i} \tag{2.16}$$

$$\le \frac{1}{\alpha_i} \wedge \frac{1}{\beta_i} \qquad \forall i \in \mathbf{Z}_+. \tag{2.17}$$

PROOF. Replacing f by $1-f$ if necessary, it suffices to give an upper bound for $\Delta g(i)$. Using (2.19) gives

$$\Delta g(i) = \sum_{j=0}^{\infty} f(j)\Delta g_j(i). \tag{2.18}$$

Lemma 2.4 ensures that the only positive term in the right hand side of (2.18) is for $j = i$, leading to

$$\Delta g(i) \le \Delta g_i(i).$$

Now, (2.16) follows immediately from Lemma 2.3.

Finally, since (C2) implies that $F(i-1)/\beta_i \le F(i)/\alpha_i$ and $\bar{F}(i+1)/\alpha_i \le \bar{F}(i)/\beta_i$, the claim (2.17) is evident. $\square$

REMARK 2.11. One may ask, as happened in the Poisson approximation case, whether the bound (2.16) is decreasing in i so that a uniform bound could be achieved at $i = 0$ from conditions (C0)–(C4). Unfortunately, the answer is generally negative. For example, if we take π as Binomial(n, p) so that $\alpha_i = (n-i)p$ and $\beta_i = (1-p)i$ for $0 \le i \le n$, then $\sup_{f\in\mathscr{F}} |\Delta g(0)| = [1-(1-p)^n]/(np)$ and $\sup_{f\in\mathscr{F}} |\Delta g(n)| = (1-p^n)/[n(1-p)]$, hence $\sup_{f\in\mathscr{F}} |\Delta g(i)|$ is not decreasing if $p > 1/2$.

However, in the special case where the α's are a constant and the β's are increasing, the case $i = 0$ is the maximum. This includes Poisson approximation [see Barbour and Eagleson (1983)] and certain other polynomial birth-death distributions as shown in section 3.

COROLLARY 2.12. *If $\alpha_j = \alpha$ for all $j \in \mathbf{Z}_+$ and β_j is increasing, then*

$$\sup_{f\in\mathscr{F}} |\Delta g(i)| \le \frac{1-\pi_0}{\alpha} \qquad \forall\ i \in \mathbf{Z}_+.$$

PROOF [cf. Xia (1999)]. In fact,

$$\begin{aligned}
\sup_{f\in\mathscr{F}} |\Delta g(i)| &= \frac{1-\pi_0}{\alpha} + \frac{\pi_0 + \dots + \pi_{i-1}}{\beta_i} - \frac{\pi_1 + \dots + \pi_i}{\alpha} \\
&= \frac{1-\pi_0}{\alpha} + \sum_{j=1}^{i} \frac{\pi_j}{\alpha}\left(\frac{\beta_j}{\beta_i} - 1\right) \\
&\le \frac{1-\pi_0}{\alpha},
\end{aligned}$$

completing the proof. $\square$

To end this section, we give a proof for Lemma 2.4.

PROOF OF LEMMA 2.4. Noting that the solution g to (2.3) satisfies

$$g(k) = \sum_{i=0}^{\infty} f(i) g_i(k), \tag{2.19}$$

and if f is 1, the solution is $g = 0$, we have

$$\Delta g_j(j) + \sum_{i \neq j} \Delta g_i(j) = 0.$$

Condition (C0) is equivalent to all differences except $\Delta g_j(j)$ being negative.

For $i \leq j-1$, from Lemma 2.1 and (2.11) we have

$$\Delta g_j(i) = -\pi_j \left(\overline{\tau_i^+} - \overline{\tau_{i-1}^+} \right). \tag{2.20}$$

Similarly, for $i \geq j+1$,

$$\Delta g_j(i) = \pi_j \left(\overline{\tau_{i+1}^-} - \overline{\tau_i^-} \right). \tag{2.21}$$

Considering equations (2.20) and (2.21) and allowing j to vary shows that (C0) is equivalent to (C1).

On the other hand, it follows from (2.11) that, for $k = 1, 2, 3, \ldots$, $\overline{\tau_{k-1}^+} \leq \overline{\tau_k^+}$ and $\overline{\tau_k^-} \geq \overline{\tau_{k+1}^-}$ are equivalent to

$$\beta_k F(k) - \alpha_k F(k-1) \geq 0 \tag{2.22}$$

and

$$\beta_k \bar{F}(k+1) - \alpha_k \bar{F}(k) \leq 0, \tag{2.23}$$

using the detailed balance relations (2.1). Rearrangement gives the equivalence of (C1) and (C2).

But the left side of inequality (2.22) is, using the detailed balance condition $\beta_k\pi_k = \alpha_{k-1}\pi_{k-1}$,

$$\begin{aligned}
\beta_k F(k) - \alpha_k F(k-1) &= \beta_k\pi_k + \sum_{l=0}^{k-1}(\beta_k - \alpha_k)\pi_l \\
&= [(\beta_k - \beta_{k-1}) - (\alpha_k - \alpha_{k-1})]\,\pi_{k-1} \\
&\quad + \beta_{k-1}\pi_{k-1} + \sum_{l=0}^{k-2}(\beta_k - \alpha_k)\pi_l \\
&= \delta_{k-1}\pi_{k-1} + [(\beta_k - \beta_{k-2}) - (\alpha_k - \alpha_{k-2})]\pi_{k-2} \\
&\quad + \beta_{k-2}\pi_{k-2} + \sum_{l=0}^{k-3}(\beta_k - \alpha_k)\pi_l \\
&= \delta_{k-1}\pi_{k-1} + (\delta_{k-1} + \delta_{k-2})\pi_{k-2} \\
&\quad + \beta_{k-2}\pi_{k-2} + \sum_{l=0}^{k-3}(\beta_k - \alpha_k)\pi_l \\
&= \cdots \\
&= \sum_{l=0}^{k-1}(\delta_{k-1} + \cdots + \delta_l)\pi_l = \sum_{l=0}^{k-1}\delta_l F(l)
\end{aligned} \tag{2.24}$$

showing that (2.22) is the same as (2.14). We have used the definition of β_0 as 0 in the second last step of these equalities and collected all the terms with the same δ in the last. Likewise,

$$\begin{aligned}
\beta_k\bar{F}(k+1) - \alpha_k\bar{F}(k) &= -\alpha_k\pi_k + \sum_{l=k+1}^{\infty}(\beta_k - \alpha_k)\pi_l \\
&= -\delta_k\pi_{k+1} - \alpha_{k+1}\pi_{k+1} + \sum_{l=k+2}^{\infty}(\beta_k - \alpha_k)\pi_l \\
&= \cdots \\
&= -\sum_{l=k+1}^{\infty}(\delta_k + \ldots + \delta_{l-1})\pi_l \\
&= -\sum_{l=k+1}^{\infty}\delta_{l-1}\bar{F}(l)
\end{aligned} \tag{2.25}$$

showing that (2.23) is the same as (2.15). This gives the equivalence of (C2) and (C3).

The sufficiency of (C4) follows from the equivalences proved and the fact that (C4) is the same as all the δ's being non-negative. □

3. Approximation to the sum of independent Bernoulli trials. Let X_i, $1 \le i \le n$ be independent indicator random variables with distribution

$$\mathbf{P}(X_i = 1) = 1 - \mathbf{P}(X_i = 0) = p_i, \qquad 1 \le i \le n.$$

Set $W = \sum_{i=1}^n X_i$, $\lambda_l = \sum_{i=1}^n p_i^l$, and $\theta_l = \lambda_l/\lambda_1$, $l = 1, 2, \ldots$ We write $\lambda = \lambda_1$. Define

$$\sigma_k = \sqrt{\sum_{i=k+1}^{n} \rho_i}, \tag{3.1}$$

where ρ_i is the ith largest number of $p_1(1-p_1)$, $p_2(1-p_2)$, $\cdots$, $p_n(1-p_n)$. We use $\mathscr{L}W$ to denote the distribution of W.

It is well-known that the Poisson distribution provides a good approximation to $\mathscr{L}W$ if all the p_i's are small [see Barbour, Holst and Janson (1992) and references therein]. Barbour and Hall (1984) show, using the Stein-Chen method, that

$$\frac{1}{32}\min\left\{\frac{1}{\lambda}, 1\right\}\sum_{i=1}^n p_i^2 \le d_{TV}(\mathscr{L}W, \mathrm{Po}(\lambda)) \le \frac{1-e^{-\lambda}}{\lambda}\sum_{i=1}^n p_i^2, \tag{3.2}$$

where $\mathrm{Po}(\nu)$ stands for the Poisson distribution with mean ν. The order of Poisson approximation error in the upper and lower bounds is the same.

On the other hand, it is known that compound Poisson signed measures can improve the approximation precision [see Barbour and Xia (1999) and references therein]. It is not attractive to approximate any non-negative quantity by a negative one, and signed measures lack standard interpretations of moments. Thus, it is desirable to find an easily calculable probability measure which decreases the order of approximation error. Section 2 provides an algorithm for finding such approximating probability measures.

Recall that, in estimating Poisson approximation to $\mathscr{L}W$, we take $\alpha_j = \lambda$ and $\beta_j = j$, for $j \in \mathbf{Z}_+$. Intuitively, if we aim at higher precision than Poisson approximation, what we need to do is to reduce the variance of the approximating distribution so that the 'tailored' stationary distribution fits $\mathscr{L}W$ better. One way of doing this is to keep the death rates and reduce the birth rates to $\alpha_j = c(m-j)$, where c is a constant: this results in binomial approximation [see Ehm (1991) and Barbour, Holst and Janson (1992)]. Another way is to keep the birth rates as a constant and increase the death rates. More precisely, let

$$\alpha_j = \alpha,\ \ \beta_j = \beta j + j(j-1),$$

$$\pi_j = \frac{\alpha^j}{\Pi_{i=1}^j[\beta i + i(i-1)]}\left\{1 + \sum_{k=1}^{\infty}\frac{\alpha^k}{\Pi_{i=1}^k[\beta i + i(i-1)]}\right\}^{-1}, \qquad j \in \mathbf{Z}_+.$$

Since both birth and death rates are polynomial functions, we call this distribution a *polynomial birth-death* distribution with birth rates determined by α and death rates $0 + \beta \times j + 1 \times j(j-1)$, abbreviated as $PBD(\alpha; 0, \beta, 1)$. Accordingly, the Poisson distribution is $\mathrm{PBD}(\alpha; 0, 1)$, the binomial distribution is $\mathrm{PBD}(np, -p; 0, 1-p)$, the negative binomial is $\mathrm{PBD}(r(1-q), 1-q; 0, 1)$ and hypergeometric is $\mathrm{PBD}(nR, -R-n+1, 1; 0, B-n+1, 1)$. Note that the parameters of the PBD distributions are not uniquely determined.

THEOREM 3.1. *With the above setup, if*

$$\beta = \lambda^2\lambda_2^{-1} - 1 - 2\lambda + 2\lambda_3\lambda_2^{-1}, \qquad \alpha = \beta\lambda + \lambda^2 - \lambda_2, \tag{3.3}$$

then

$$d_{TV}(\mathscr{L}W, PBD(\alpha; 0, \beta, 1)) \le \frac{\beta\lambda_3}{\alpha\sigma_1} + \frac{2\lambda\lambda_2}{\alpha\sigma_2} \tag{3.4}$$

$$\le \frac{\theta_3}{\sigma_1} + \frac{2\theta_2^2}{\sigma_2(1-\theta_2-\theta_2/\lambda)}, \tag{3.5}$$

where (3.5) *is valid provided* $\theta_2 + \theta_2/\lambda < 1$.

REMARK. When λ is large and p_i's are small, σ_1 and σ_2 are close to $\sqrt{\lambda}$, so the upper bound in (3.5) is asymptotically $(\theta_3 + 2\theta_2^2)/\sqrt{\lambda}$. The order of the bound is as good as that of compound Poisson signed measure approximation [see Barbour and Xia (1999)].

REMARK. The parameters were chosen in such a way as to make the bound as small as possible. The error expression is rearranged in such a way that it involves only second differences of the solution g to the Stein equation. The parameter choice is quite crucial and delicate in this: equations (3.6) and (3.7) ensure that various terms have multipliers that match. Interestingly, the exact choice given here arose from numerical work as well as algebra. A first version had a cruder choice of the β parameter in that the term λ_2 was left out of the right hand side of (3.7). The resulting error expression involved a first difference as well as a second difference. The resulting calculated exact total variation distance between the law of W and the polynomial birth-death distribution came to about the same as for binomial. However, computer exploration of parameter choice showed that subtracting one from the β parameter yielded a much lower total variation distance. The algebraic reason for this then became apparent by re-examining the error expression and giving the results here.

REMARK. Although the particular implementation here uses the independence of the X's, a similar expression holds for dependent trials but the random variables W^i (resp. W^{ij}) need in general to have the distribution of $W - X_i$ (resp. $W - X_i - X_j$) conditional on $X_i = 1$ (resp. $X_i = 1$ and $X_j = 1$). In point process terms, the calculations use the first and second order reduced Palm distributions. It is difficult but not conceptually impossible to extend the analysis to higher order Palm distributions and to dependent trials. This would involve some degree of case-by-case analysis, with a careful choice of parameters to match the multipliers in the various terms.

Before we prove the theorem, we present an example to illustrate the performance of the PBD approximation.

EXAMPLE 3.2. Suppose

$$SX_1 \sim \text{Binomial}(70, 0.1), \ \ SX_2 \sim \text{Binomial}(9, 1/3)$$
$$\text{and } SX_3 \sim \text{Binomial}(2, 1/\sqrt{2})$$

are independent random variables and let $W = SX_1 + SX_2 + SX_3$, then $\mathbf{E}W = 11.4142$, Var(W)=8.7142. We apply (3.3) to get $\alpha = 415.765255$ and $\beta = 25.247555$, then the mean and variance of PBD($\alpha; 0, \beta, 1$) are 11.4142 and 8.71390 respectively, and an exact calculation gives

$$d_{TV}(\mathscr{L}W, \text{PBD}(\alpha; 0, \beta, 1)) = 0.00040$$

[the bound of (3.4) is 0.074692]. However, an exact calculation gives

$$d_{TV}(\mathscr{L}W, \text{Po}(\mathbf{E}W)) = 0.066$$

(the upper bound of (3.2) is 0.236545) and

$$d_{TV}(\mathscr{L}W, \text{Binomial}(n, p)) = 0.0048,$$

[the upper bound of Theorem 9.E of Barbour, Holst and Janson (1992) is 0.218542] with $n = 48$ and $p = 0.237796$ so that $np = \mathbf{E}W$ and $np(1-p) \approx$ Var(W). The PBD($\alpha; 0, \beta, 1$) approximation is more than 10 times better than the binomial approximation, and the binomial approximation does more than 10 times better than the Poisson approximation. It is worth noting that the computation of PBD distribution is as quick as Poisson and Binomial (in EXCEL or Minitab, logarithms of PBD probabilities are differences of logarithms of polynomials; knowledge of means and variances of the actual distribution makes it easy to pick the range where the probabilities are not essentially zero), but it is much harder to calculate the actual probabilities of W (in general, computation of the probabilities of W is thought to be NP-complete). Figures 4 and 5 provide details of the comparison among the three approximations.

PROOF OF THEOREM 3.1. Set $W^i = W - X_i$ and $W^{ij} = W - X_i - X_j$, for $i, j \in \mathbf{Z}_+$, $i \neq j$, so that

$$\begin{aligned}\mathbf{E}\mathscr{B}g(W) &= \alpha\mathbf{E}g(W+1) - \beta\mathbf{E}\sum_{i=1}^{n} p_i g(W^i+1) - \sum_{j\neq i} p_i p_j \mathbf{E}g(W^{ij}+2) \\ &= \beta\sum_{i=1}^{n} p_i^2 \mathbf{E}\Delta g(W^i+1) + (\alpha - \beta\lambda)\mathbf{E}g(W+1) - \sum_{j\neq i} p_i p_j \mathbf{E}g(W^{ij}+2).\end{aligned}$$

Taking

$$\alpha - \beta\lambda = \sum_{j\neq i} p_i p_j = \lambda^2 - \lambda_2, \tag{3.6}$$

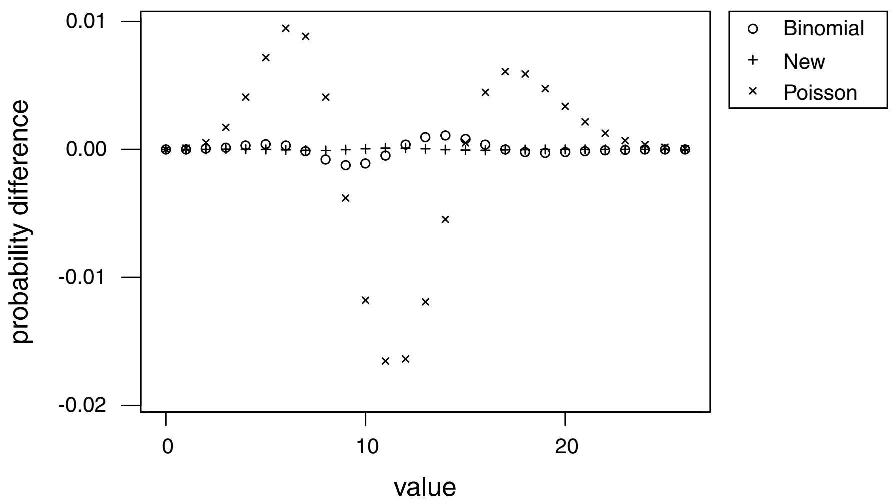

FIG. 4. *Approximate minus actual probability versus value*

we get

$$\begin{aligned}\mathbf{E}\mathscr{B}g(W) &= \beta\sum_{i=1}^{n} p_i^2\mathbf{E}\Delta g(W^i+1) + \sum_{j\neq i} p_i p_j[\mathbf{E}g(W+1) - \mathbf{E}g(W^{ij}+2)]\\ &= \beta\sum_{i=1}^{n} p_i^2\mathbf{E}\Delta g(W^i+1) + \sum_{j\neq i} p_i^2 p_j^2\mathbf{E}\Delta g(W^{ij}+2)\\ &\quad - \sum_{j\neq i} p_i p_j(1-p_i)(1-p_j)\mathbf{E}\Delta g(W^{ij}+1)\end{aligned}$$

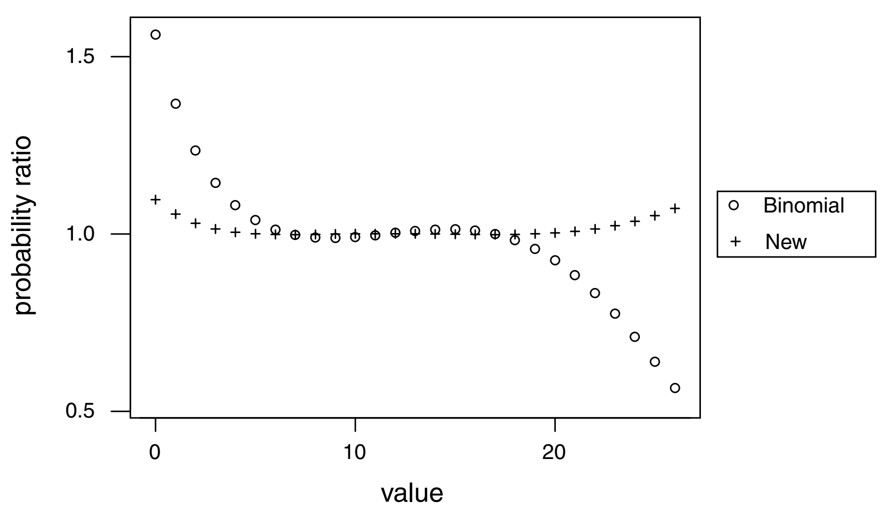

FIG. 5. *Approximate divided actual probability versus value*

$$= \beta \sum_{i=1}^{n} p_i^2 \mathbf{E}\Delta g(W^i + 1) + \sum_{j\neq i} p_i^2 p_j^2 \mathbf{E}\Delta^2 g(W^{ij} + 1)$$
$$- \sum_{j\neq i} p_i p_j (1 - p_i - p_j)\mathbf{E}\Delta g(W^{ij} + 1),$$

where $\Delta^2 g(k) = \Delta g(k+1) - \Delta g(k),\ \ k \in \mathbf{Z}_+$. Let

$$\beta\lambda_2 = \sum_{j\neq i} p_i p_j (1 - p_i - p_j) = \lambda^2 - \lambda_2 - 2\lambda\lambda_2 + 2\lambda_3. \tag{3.7}$$

It follows that

$$\begin{aligned}
\mathbf{E}\mathscr{B}g(W) &= \beta \sum_{i=1}^{n} p_i^2 \mathbf{E}[\Delta g(W^i + 1) - \Delta g(W + 1)] + \sum_{j\neq i} p_i^2 p_j^2 \mathbf{E}\Delta^2 g(W^{ij} + 1) \\
&\quad + \sum_{j\neq i} p_i p_j (1 - p_i - p_j)\mathbf{E}[\Delta g(W + 1) - \Delta g(W^{ij} + 1)] \\
&= -\beta \sum_{i=1}^{n} p_i^3 \mathbf{E}\Delta^2 g(W^i + 1) + \sum_{j\neq i} p_i^2 p_j^2 \mathbf{E}\Delta^2 g(W^{ij} + 1) \\
&\quad + \sum_{j\neq i} p_i p_j (1 - p_i - p_j)[p_i(1 - p_j) + p_j(1 - p_i)]\mathbf{E}\Delta^2 g(W^{ij} + 1) \\
&\quad + \sum_{j\neq i} p_i^2 p_j^2 (1 - p_i - p_j)\mathbf{E}[\Delta^2 g(W^{ij} + 1) + \Delta^2 g(W^{ij} + 2)] \\
&= -\beta \sum_{i=1}^{n} p_i^3 \mathbf{E}\Delta^2 g(W^i + 1) \\
&\quad + \sum_{j\neq i} p_i p_j (p_i + p_j)(1 - p_i)(1 - p_j)\mathbf{E}\Delta^2 g(W^{ij} + 1) \\
&\quad + \sum_{j\neq i} p_i^2 p_j^2 (1 - p_i - p_j)\mathbf{E}\Delta^2 g(W^{ij} + 2).
\end{aligned}$$

Noting that

$$|\mathbf{E}\Delta^2 g(W^i + 1)| \le 2\|\Delta g\| d_{TV}(\mathscr{L}W^i, \mathscr{L}(W^i + 1)) \le \frac{\|\Delta g\|}{\sigma_1},$$

where $\|\Delta g\| := \sup_k |\Delta g(k)|$ [see Barbour and Jensen (1989)], and correspondingly

$$|\mathbf{E}\Delta^2 g(W^{ij} + k)| \le \frac{\|\Delta g\|}{\sigma_2} \qquad \forall k \in \mathbf{Z}_+,$$

we obtain

$$\begin{aligned}
|\mathbf{E}\mathscr{B}g(W)| &\le \|\Delta g\| \left\{ \frac{\beta\lambda_3}{\sigma_1} + \frac{\sum_{j\neq i}[p_i p_j (p_i + p_j)(1 - p_i)(1 - p_j) + p_i^2 p_j^2]}{\sigma_2} \right\} \\
&\le \|\Delta g\| \left[\frac{\beta\lambda_3}{\sigma_1} + \frac{2\lambda\lambda_2}{\sigma_2} \right],
\end{aligned}$$

so (3.4) follows from (2.17). Equation (3.5) is due to the facts that $\alpha \geq \beta\lambda$ and $\alpha \geq \lambda^3\lambda_2^{-1} - \lambda - \lambda^2$.

Finally, (3.3) comes from solving the equations (3.6) and (3.7). $\square$

4. Negative binomial approximation to the number of 2-runs. Let $J_1, J_2, \ldots, J_n$ be independent identically distributed Bernoulli random variables with $\mathbf{P}[J_i = 1] = p$, $1 \leq i \leq n$. To avoid edge effects, we treat $i + nj$ as i for $1 \leq i \leq n$, $j = 0, \pm 1, \pm 2, \ldots$. Define $I_i = J_i J_{i-1}$ and $W = \sum_{i=1}^n I_i$. Our random variable of interest is W, which counts the number of 2-runs of J_i, $1 \leq i \leq n$.

The approximation to the number of 2-runs has been well studied previously, see, for example, compound Poisson approximation in total variation in Arratia, Goldstein and Gordon (1990), Roos (1993), Eichelsbacher and Roos (1999) and Barbour and Xia (1999). In this section, we will estimate the accuracy of negative binomial approximation to $\mathscr{L}W$.

It is easy to establish $\mathbf{E}I_i = p^2$, $\mathbf{E}W = np^2$ and $\mathrm{Var}(W) = np^2(1+2p-3p^2)$. If $p < 2/3$, then $\mathrm{Var}(W) > \mathbf{E}W$, indicating that, in comparison with Poisson approximation, we need to either reduce the death rate or increase the birth rate. For simplicity, we increase the birth rate by taking $\alpha_j = a + bj$ with $0 \leq b < 1$ and $\beta_j = j$ in (2.1). The equilibrium distribution is then a negative binomial distribution [see Example 2.9]. To keep our notation consistent, we denote the equilibrium distribution as $\mathrm{PBD}(a, b; 0, 1)$.

Our argument is similar to that in Barbour and Xia (1999). The following result can be found as Lemma 5.1 of Barbour and Xia (1999).

LEMMA 4.1. *Let* $(\eta_m,\ m \geq 1)$ *be independent indicator random variables with* $\mathbf{P}(\eta_m = 1) = \gamma_m$, $m \geq 1$, *and set* $\eta_0 = 0$, *that is*, $\gamma_0 = 0$, *and* $Y_m = \sum_{i=1}^m \eta_i\eta_{i-1}$. *Then, for each* $n \geq 2$,

$$b_n(\gamma_1, \gamma_2, \ldots, \gamma_n) := 2d_{TV}(\mathscr{L}(Y_n), \mathscr{L}(Y_n + 1)) \leq \frac{4.6}{\sqrt{\sum_{i=2}^n (1-\gamma_{i-2})^2\gamma_{i-1}(1-\gamma_{i-1})\gamma_i}}.$$

Now, we state the main result of this section.

THEOREM 4.2. *Let* $b = \frac{2p-3p^2}{1+2p-3p^2}$ *and* $a = (1-b)np^2$, *if* $p < 2/3$, *then*

$$d_{TV}(\mathscr{L}(W), PBD(a, b; 0, 1)) \leq \frac{32.2p^2}{\sqrt{(n-1)p^2(1-p)^3}}.$$

PROOF. Let $W_1 = W - I_1 - I_2 - I_3$. Using the fact that W_1, J_1 and J_2 are independent, we have

$$\begin{aligned}
&\mathbf{E}g(W+1) \\
&\quad = \mathbf{E}\{g(W+1)[J_1J_2 + (1-J_2)J_1 + (1-J_1)J_2 + (1-J_1)(1-J_2)\}
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{E}\{g(W_1+J_n+J_3+2)J_1J_2\} + \mathbf{E}\{g(W_1+J_n+1)(1-J_2)J_1\}\\
&\quad +\mathbf{E}\{g(W_1+J_3+1)(1-J_1)J_2\} + \mathbf{E}\{g(W_1+1)(1-J_1)(1-J_2)\}\\
&= \mathbf{E}\{[g(W_1+J_n+J_3+2) - g(W_1+J_n+2) - g(W_1+J_3+2)\\
&\quad +g(W_1+2)]J_1J_2\}\\
&\quad +\mathbf{E}\{[g(W_1+J_n+2) - g(W_1+J_n+1) - g(W_1+2) + g(W_1+1)]J_1J_2\}\\
&\quad +\mathbf{E}\{[g(W_1+J_3+2) - g(W_1+J_3+1) - g(W_1+2) + g(W_1+1)]J_1J_2\}\\
&\quad +\mathbf{E}\{[g(W_1+J_n+1) - g(W_1+1)]J_1\}\\
&\quad +\mathbf{E}\{[g(W_1+J_3+1) - g(W_1+1)]J_2\}\\
&\quad +\mathbf{E}\{\Delta g(W_1+1)J_1J_2\} + \mathbf{E}g(W_1+1)\\
&= p^2\mathbf{E}\{\Delta^2 g(W_1+2)J_3J_n\} + p^2\mathbf{E}\{\Delta^2 g(W_1+1)(J_3+J_n)\}\\
&\quad +p\mathbf{E}\{\Delta g(W_1+1)(J_3+J_n+p)\} + \mathbf{E}g(W_1+1)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbf{E}I_2 g(W)\\
&\quad = p^2\mathbf{E}\{g(W_1+J_n+J_3+1)[J_nJ_3+J_n(1-J_3)+J_3(1-J_n)+(1-J_n)(1-J_3)]\}\\
&\quad = p^2\mathbf{E}\{g(W_1+3)J_nJ_3+g(W_1+2)[J_n(1-J_3)+(1-J_n)J_3]\\
&\qquad +g(W_1+1)(1-J_n)(1-J_3)\}\\
&\quad = p^2\big\{\mathbf{E}\{\Delta^2 g(W_1+1)J_nJ_3\}+\mathbf{E}\{\Delta g(W_1+1)(J_n+J_3)\}+\mathbf{E}g(W_1+1)\big\}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbf{E}I_2 g(W+1) = {}& p^2\mathbf{E}\{\Delta^2 g(W_1+2)J_nJ_3\} + p^2\mathbf{E}\{\Delta^2 g(W_1+1)(J_n+J_3)\}\\
&+p^2\mathbf{E}\{\Delta g(W_1+1)(J_n+J_3+1)\} + p^2\mathbf{E}g(W_1+1).
\end{aligned}
$$

Combining the three expansions, we find that

$$
\begin{aligned}
&\mathbf{E}[(a+bW)g(W+1) - Wg(W)]\\
&\qquad = p^2(a+nb)\mathbf{E}\{\Delta^2 g(W_1+2)J_3J_n\}\\
&\qquad\quad + p^2\mathbf{E}\{\Delta^2 g(W_1+1)[-nJ_nJ_3 + (a+nb)J_n + (a+nb)J_3]\}\\
&\qquad\quad + \mathbf{E}\{\Delta g(W_1+1)[apJ_3 + apJ_n + ap^2 + nbp^2(J_n+J_3+1)\\
&\qquad\quad - np^2(J_n+J_3)]\} + (a - np^2 + nbp^2)\mathbf{E}g(W_1+1).
\end{aligned}
\tag{4.1}
$$

We choose

$$
a = np^2(1-b) \tag{4.2}
$$

so that the last term of (4.1) vanishes. Lemma 4.1 may now be applied to bound (4.1). The first term of (4.1) is bounded by

$$
\begin{aligned}
&p^2(a+nb)|\mathbf{E}\{\Delta^2 g(W_1+2)J_3J_n\}|\\
&\qquad \le p^4(a+nb)|\mathbf{E}\{\Delta^2 g(W_2+J_{n-1}+J_4+2)\}|\\
&\qquad \le p^4(a+nb)\|\Delta g\| b_{n-2}(1, p, \ldots, p, 1),
\end{aligned}
$$

where $W_2 = W_1 - I_n - I_4$. By symmetry, the second term of (4.1) can be bounded by

$$np^4|\mathbf{E}\{\Delta^2 g(W_2 + J_{n-1} + J_4 + 1)\}| + 2(a+nb)p^3|\mathbf{E}\{\Delta^2 g(W_3 + J_{n-1} + 1)\}|$$
$$\leq np^4\|\Delta g\| b_{n-2}(1, p, \ldots, p, 1) + 2(a+nb)p^3\|\Delta g\| b_{n-2}(p, \ldots, p, 1),$$

where $W_3 = W_1 - I_n$. Finally, by taking

$$2(a + nbp - np) + (a + nb) = 0, \tag{4.3}$$

the third term of (4.1) can be reduced to

$$|\mathbf{E}\{\Delta g(W_1+1)[p(a+nbp-np)(J_3+J_n)+(a+nb)p^2]\}|$$
$$= p^2|2[a+np(b-1)]\mathbf{E}\Delta g(W_3+J_{n-1}+1)+(a+nb)\mathbf{E}\Delta g(W_3+J_{n-1}J_n+1)|$$
$$= (a+nb)p^2|\mathbf{E}\{\Delta^2 g(W_3+1)J_{n-1}(1-J_n)\}|$$
$$\leq (a+nb)(1-p)p^3|\mathbf{E}\{\Delta^2 g(W_4+J_{n-2}+1)\}|$$
$$\leq (a+nb)(1-p)p^3\|\Delta g\| b_{n-3}(p,\ldots,p,1),$$

where $W_4 = W_3 - I_{n-1}$.

Now, we have from Lemma 4.1 that $b_{n-2}(1, p, \ldots, p, 1)$, $b_{n-2}(p, \ldots, p, 1)$ and $b_{n-3}(p, \ldots, p, 1)$ are all bounded by

$$\frac{4.6}{\sqrt{p^2(1-p) + p(1-p)^3 + (n-6)p^2(1-p)^3}} \leq \frac{4.6}{\sqrt{(n-1)p^2(1-p)^3}}$$

since

$$p^2(1-p) + p(1-p)^3 \geq 5p^2(1-p)^3, \quad 0 \leq p \leq 1.$$

The values of a and b follow from (4.2) and (4.3). It is easy to show that condition (C4) in Lemma 2.4 is satisfied so Theorem 2.10 gives $\|\Delta g\| \leq a^{-1}$. The proof is now complete by collecting the three estimates above and simplifying the bound. □

REMARK 4.3. Another way to tackle this problem is to declump the sequence of $J_1, \ldots, J_n$ into strings of 1's separated by 0's, as discussed in Arratia, Goldstein and Gordon (1990). As each string follows approximately a geometric distribution and the number of strings is roughly binomial, we can approximate the number of strings of at least two 1's by an appropriate Poisson distribution, giving us a bound of order p^2. This shows how impressive the bound in Theorem 4.2 is since it is of order $p^2/\sqrt{np^2}$.

The bound in Theorem 4.2 is as good as that of compound Poisson approximation obtained in Barbour and Xia (1999). As a matter of fact, a negative binomial can also be viewed as a Poisson sum of variables each with logarithmic distribution [see Johnson, Kotz and Kemp (1992), page 204], so the result here is also approximation by a compound Poisson distribution. It is surprising that the order of approximation seems to be better for this compound Poisson than the one in Arratia, Goldstein and Gordon (1990).

In general, let X_i ~Poisson(λ_i), $i \geq 1$ be independent and set $X = X_1 + 2X_2 + 3X_3 + \cdots$, then X has compound Poisson distribution, denoted as CP($\lambda_1, \lambda_2, \ldots$). Our PBD distribution approximation theory developed in section 2 includes a family of compound Poisson approximations provided λ_{i+1}/λ_i, $i \geq 1$ are all small. For example, the following proposition gives an estimate for the difference between the compound Poisson distribution with two free parameters and the negative binomial distribution [cf Corollary 4.8 of Barbour and Xia (1999)].

PROPOSITION 4.4. *Let* $\lambda_3 = \lambda_4 = \cdots = 0$ *and define*

$$b = \frac{2\lambda_2}{\lambda_1 + 4\lambda_2}, \quad a = \frac{(\lambda_1 + 2\lambda_2)^2}{\lambda_1 + 4\lambda_2},$$

then

$$d_{TV}(CP(\lambda_1, \lambda_2, 0, \ldots), PBD(a, b; 0, 1)) \leq \frac{8\lambda_2^2}{(\lambda_1 + 2\lambda_2)^2\sqrt{2e\lambda_1}}.$$

PROOF. In fact,

$$\begin{aligned}
&\mathbf{E}[ag(X+1) + bXg(X+1) - Xg(X)] \\
&\quad = \mathbf{E}[2b\lambda_2 g(X+3) + (b\lambda_1 - 2\lambda_2)g(X+2) + (a - \lambda_1)g(X+1)] \\
&\quad = 2b\lambda_2 \mathbf{E}[\Delta^2 g(X+1)] + (b\lambda_1 - 2\lambda_2 + 4b\lambda_2)\mathbf{E}[g(X+2)] \\
&\qquad +(a - \lambda_1 - 2b\lambda_2)\mathbf{E}[g(X+1)].
\end{aligned}$$

Take $b\lambda_1 - 2\lambda_2 + 4b\lambda_2 = 0$ and $a - \lambda_1 - 2b\lambda_2 = 0$, which are equivalent to $b = \frac{2\lambda_2}{\lambda_1+4\lambda_2}$ and $a = \frac{(\lambda_1+2\lambda_2)^2}{\lambda_1+4\lambda_2}$, so that the last two terms vanish. Thus,

$$\begin{aligned}
&|\mathbf{E}[ag(X+1) + bXg(X+1) - Xg(X)]| \\
&\quad \leq 4b\lambda_2 \|\Delta g\| d_{TV}(\mathscr{L}(X), \mathscr{L}(X+1)) \leq \frac{8\lambda_2^2}{(\lambda_1 + 2\lambda_2)^2\sqrt{2e\lambda_1}}
\end{aligned}$$

[see Barbour, Holst and Janson (1992), page 262], completing the proof. □

An immediate consequence of the Proposition is that the negative binomial distribution can approximate the number of 2-runs of $W = \sum_{i=1}^n J_{i-1}J_i$ with the same precision as compound Poisson approximation, even in the case that $J_1, \ldots, J_n$ are independent, but not identically distributed, indicators, as investigated in Barbour and Xia (1999). We omit the details of the analysis due to the complexity of the exercise.

5. Poisson process approximation. The idea presented in section 2 can also be extended to solve a counterpart problem in Poisson process approximation. To start with, we recall briefly the notation from Barbour and

Brown (1992) [see also Brown and Xia (1995a)]. Let Γ be a compact metric space with metric d_0 bounded by 1, and let $\mathscr{H}$ be the space of finite configurations on Γ, that is, the space of integer-valued finite measures on Γ. For each $\xi \in \mathscr{H}$, we use $|\xi|$ to denote the total number of points of ξ. The metric d_1 on $\mathscr{H}$ is 1 if the two configurations do not have the same number of points and is otherwise the average d_0-distance between the points of the configurations under the closest matching. The metric d_2 between two probability distributions on $\mathscr{H}$ is the infimum of expected values of the d_1 distance between any two realisations of the probability distributions. Define, for suitable function h on $\mathscr{H}$, the generator $\mathscr{A}_{pp}$ by

$$\mathscr{A}_{pp}h(\xi) = \int_\Gamma [h(\xi + \delta_x) - h(\xi)]\boldsymbol{\lambda}(dx) + \int_\Gamma [h(\xi - \delta_x) - h(\xi)]\xi(dx). \tag{5.1}$$

Then $\mathscr{A}_{pp}$ is the generator of an $\mathscr{H}$-valued immigration-death process $\mathbf{Z}_\xi(t)$ with immigration intensity $\boldsymbol{\lambda}$ and unit per capita death rate, where $\mathbf{Z}_\xi(0) = \xi$ and *pp* stands for Poisson process. Note that, when $\xi = 0$, the immigration-death process $\mathbf{Z}_\xi$ is written as $\mathbf{Z}_0$. The equilibrium distribution of $\mathbf{Z}_\xi$ is a Poisson process with mean measure $\boldsymbol{\lambda}$, denoted by $\mathrm{Po}(\boldsymbol{\lambda})$. Thus, as in Section 2, the generator $\mathscr{A}_{pp}$ can be used to establish a Stein equation for $\mathrm{Po}(\boldsymbol{\lambda})$ approximation:

$$\mathscr{A}_{pp}h(\xi) = f(\xi) - \mathrm{Po}(\boldsymbol{\lambda})(f), \tag{5.2}$$

for bounded function f on $\mathscr{H}$. The solution to (5.2) is given by

$$h(\xi) = -\int_0^\infty [\mathbf{E}f(\mathbf{Z}_\xi(t)) - \mathrm{Po}(\boldsymbol{\lambda})(f)]dt \tag{5.3}$$

[see Barbour and Brown (1992), Proposition 2.3].

Let $\mathscr{G}$ be the space of all d_1-Lipschitz functions f on $\mathscr{H}$, namely, $|f(\xi_1) - f(\xi_2)| \le d_1(\xi_1, \xi_2)$ for all $\xi_1, \xi_2 \in \mathscr{H}$ (so in particular, since d_1 is bounded by 1, f is bounded). Then for any point process Ξ on Γ,

$$d_2(\mathscr{L}\Xi, \mathrm{Po}(\boldsymbol{\lambda})) = \sup_{f \in \mathscr{G}} |\mathbf{E}f(\Xi) - \mathrm{Po}(\boldsymbol{\lambda})(f)|.$$

Hence, to use the Stein equation (5.2) in bounding the errors of $\mathrm{Po}(\boldsymbol{\lambda})$ approximation to $\mathscr{L}\Xi$, we need to estimate

$$\Delta^2 h(\xi; x, y) := h(\xi + \delta_x + \delta_y) - h(\xi + \delta_x) - h(\xi + \delta_y) + h(\xi),$$

where h is the solution (5.3) to Stein's equation (5.2), $f \in \mathscr{G}$, $x, y \in \Gamma$. Theorem 2.1 of Brown, Weinberg and Xia (2000) gives

$$|\Delta^2 h(\xi; x, y)| \le \frac{3}{\lambda} + \frac{1}{n+3} + \frac{n}{\lambda^2} + \frac{1.65}{\lambda^{1.5}}|n + 1 - \lambda|, \tag{5.4}$$

where $n = |\xi|$ and $\lambda = \boldsymbol{\lambda}(\Gamma)$.

The estimate (5.4) is indeed of optimal order, as demonstrated in a number of applications in the paper. However, there are four terms in the bound and extra effort is needed in applying the bound in concrete problems. Here, we prove a simplified version of the bound using the techniques from Section 2.

THEOREM 5.1. *For each $f \in \mathscr{G}$, $x, y \in \Gamma$ and $\xi \in \mathscr{H}$ with $|\xi| = n$, the solution h of* (5.3) *satisfies*

$$|\Delta^2 h(\xi; x, y)| \leq \frac{5}{\lambda} + \frac{3}{n+1}. \tag{5.5}$$

To prove Theorem 5.1, we need a few technical lemmas. Note that $|\mathbf{Z}_\xi|$ is a birth-death process with birth rates $\alpha_i = \lambda$ and death rates $\beta_i = i$, for $i \in \mathbf{Z}_+$, the notation in section 2 is hence in force, with $\pi_i = \mathrm{Po}(\lambda)\{i\}$.

LEMMA 5.2. *Let $e_n^+ = \mathbf{E}\exp(-\tau_n^+)$ and $e_n^- = \mathbf{E}\exp(-\tau_n^-)$, then*

$$e_n^+ = \frac{\lambda F(n)}{(n+1)F(n+1)}, \qquad e_n^- = 1 + \frac{n}{\lambda} - \frac{\bar{F}(n-1)}{\bar{F}(n)}. \tag{5.6}$$

PROOF. By conditioning on the time of the first jump after leaving the initial state, we can produce recurrence relations for e_n^+ and e_n^-:

$$e_n^+ = \frac{\lambda}{\lambda + n + 1 - n e_{n-1}^+}, \qquad n \geq 1;\ e_0^+ = \frac{\lambda}{\lambda+1}; \tag{5.7}$$

$$e_n^- = \frac{n+\lambda}{\lambda} - \frac{n-1}{\lambda e_{n-1}^-}, \qquad n \geq 2;\ e_1^- = \frac{\lambda+1}{\lambda} - \frac{1}{\lambda \int_0^\infty p_{00}(t) e^{-t} dt}, \tag{5.8}$$

where $p_{00}(t) = \mathbf{P}(|\mathbf{Z}_0(t)| = 0)$ [see Wang and Yang (1992), pages 154–156 and pages 174–176]. Now, the claim for e_n^+ follows from (5.7) and mathematical induction. On the other hand, it is well-known that $|\mathbf{Z}_0(t)| \sim \mathrm{Po}\,(\lambda(1-e^{-t}))$, so

$$e_1^- = \frac{\lambda+1}{\lambda} - \frac{1}{1-e^{-\lambda}}.$$

Using mathematical induction in (5.8) gives the claim for e_n^-. $\square$

LEMMA 5.3. *If $|\xi_1| = |\xi_2| = n$, then for each $f \in \mathscr{G}$, the solution h of* (5.3) *satisfies*

$$|h(\xi_1) - h(\xi_2)| \leq n d_1(\xi_1, \xi_2)\left(\frac{1}{\lambda} + \frac{1}{n+1}\right).$$

PROOF. Write $\xi_1 = \sum_{i=1}^n \delta_{x_i}$ and $\xi_2 = \sum_{i=1}^n \delta_{y_i}$. Without loss of generality, we may assume $d_1(\xi_1, \xi_2) = \frac{\sum_{i=1}^n d_0(x_i, y_i)}{n}$. Set $\eta_k = \sum_{i=1}^k \delta_{x_i} + \sum_{i=k+1}^n \delta_{y_i}$, $0 \leq k \leq n$, then $\eta_0 = \xi_2$ and $\eta_n = \xi_1$. By the triangle inequality, it suffices to show that

$$|h(\eta_k) - h(\eta_{k+1})| \leq d_0(x_{k+1}, y_{k+1})\left(\frac{1}{\lambda} + \frac{1}{n+1}\right), \qquad 0 \leq k \leq n-1. \tag{5.9}$$

Fix $0 \le k \le n-1$, and let $\psi_k = \sum_{i=1}^k \delta_{x_i} + \sum_{i=k+2}^n \delta_{y_i}$. Let $S \sim \exp(1)$ be independent of $\mathbf{Z}_{\psi_k}$ and construct $\mathbf{Z}_{\eta_k}(t) = \mathbf{Z}_{\psi_k}(t) + \delta_{y_{k+1}} 1_{S>t}$ and $\mathbf{Z}_{\eta_{k+1}}(t) = \mathbf{Z}_{\psi_k}(t) + \delta_{x_{k+1}} 1_{S>t}$, so that

$$\begin{aligned}
|h(\eta_k) - h(\eta_{k+1})| &= \left| \int_0^\infty \mathbf{E}[f(\mathbf{Z}_{\eta_k}(t)) - f(\mathbf{Z}_{\eta_{k+1}}(t))]dt \right| \\
&= \left| \int_0^\infty e^{-t}\mathbf{E}[f(\mathbf{Z}_{\psi_k}(t) + \delta_{y_{k+1}}) - f(\mathbf{Z}_{\psi_k}(t) + \delta_{x_{k+1}})]dt \right| \\
&\le d_0(x_{k+1}, y_{k+1}) \int_0^\infty e^{-t}\mathbf{E}\frac{1}{|\mathbf{Z}_{\psi_k}(t)| + 1} dt \\
&\le d_0(x_{k+1}, y_{k+1}) \left(\frac{1}{\lambda} + \frac{1}{n+1} \right),
\end{aligned}$$

where the last inequality is from Lemma 2.2 of Brown, Weinberg and Xia (2000). $\square$

The following lemma is taken from a statement in the proof of Lemma 3.2 of Brown, Weinberg and Xia (2000).

LEMMA 5.4. *If $|\xi_1| = |\xi_2| = n$, then for each $f \in \mathscr{G}$ and $x \in \Gamma$,*

$$|[h(\xi_1 + \delta_x) - h(\xi_1)] - [h(\xi_2 + \delta_x) - h(\xi_2)]| \le \frac{2n}{n+1} d_1(\xi_1, \xi_2).$$

PROOF OF THEOREM 5.1. By conditioning on the first jump time of leaving $\xi + \delta_x$, it follows from (5.3) that

$$\begin{aligned}
h(\xi + \delta_x) = &\frac{-[f(\xi + \delta_x) - \mathrm{Po}(\boldsymbol{\lambda})(f)]}{n + 1 + \lambda} \\
&+ \frac{\lambda}{n+1+\lambda}\mathbf{E}h(\xi + \delta_x + \delta_V) + \frac{n+1}{n+1+\lambda}\mathbf{E}h(\xi + \delta_x - \delta_{U_x}),
\end{aligned}$$

where $V \sim \boldsymbol{\lambda}/\lambda$ and U_x is uniformly distributed on $\{z_1, \dots, z_n, x\}$ with $z_1, \dots, z_n$ the atoms of ξ. Rearranging the equation gives

$$\begin{aligned}
\mathbf{E}h(\xi + \delta_x + \delta_V) = &\frac{f(\xi + \delta_x) - \mathrm{Po}(\boldsymbol{\lambda})(f)}{\lambda} \\
&+ \frac{n+1+\lambda}{\lambda} h(\xi + \delta_x) - \frac{n+1}{\lambda}\mathbf{E}h(\xi + \delta_x - \delta_{U_x}).
\end{aligned}$$

This in turn yields

$$\begin{aligned}
\Delta^2 h(\xi; x, y) = &[h(\xi + \delta_x + \delta_y) - \mathbf{E}h(\xi + \delta_x + \delta_V)] + [h(\xi + \delta_x) - h(\xi + \delta_y)] \\
&+ [h(\xi) - \mathbf{E}h(\xi + \delta_x - \delta_{U_x})] + \frac{f(\xi + \delta_x) - \mathrm{Po}(\boldsymbol{\lambda})(f)}{\lambda} \\
&+ \frac{n+1-\lambda}{\lambda}[h(\xi + \delta_x) - \mathbf{E}h(\xi + \delta_x - \delta_{U_x})].
\end{aligned} \tag{5.10}$$

Swap x and y to get

$$\begin{aligned}\Delta^2 h(\xi; x, y) = {} & [h(\xi+\delta_x+\delta_y) - \mathbf{E}h(\xi+\delta_y+\delta_V)] + [h(\xi+\delta_y) - h(\xi+\delta_x)] \\ & + [h(\xi) - \mathbf{E}h(\xi+\delta_y-\delta_{U_y})] + \frac{f(\xi+\delta_y) - \mathrm{Po}(\boldsymbol{\lambda})(f)}{\lambda} \\ & + \frac{n+1-\lambda}{\lambda}[h(\xi+\delta_y) - \mathbf{E}h(\xi+\delta_y-\delta_{U_y})],\end{aligned} \tag{5.11}$$

where U_y is uniformly distributed on $\{z_1, \ldots, z_n, y\}$.

Adding up (5.10) and (5.11) and then dividing by 2, we obtain

$$\begin{aligned}\left|\Delta^2 h(\xi; x, y)\right| \le {} & \max_{z=x,y} \left|h(\xi+\delta_x+\delta_y) - \mathbf{E}h(\xi+\delta_z+\delta_V)\right| \\ & + \max_{z=x,y}\left|h(\xi) - \mathbf{E}h(\xi+\delta_z-\delta_{U_z})\right| \\ & + \max_{z=x,y}\left|\frac{f(\xi+\delta_z) - \mathrm{Po}(\boldsymbol{\lambda})(f)}{\lambda}\right| \\ & + \max_{z=x,y}\left|\frac{n+1-\lambda}{\lambda}[h(\xi+\delta_z) - \mathbf{E}h(\xi+\delta_z-\delta_{U_z})]\right|.\end{aligned} \tag{5.12}$$

First, using the fact that f is Lipschitz, the third term in (5.12) is clearly bounded by $1/\lambda$. Next, we apply Lemma 5.3 to the first two terms of (5.12) to conclude that each is bounded by $\frac{1}{\lambda} + \frac{1}{n+1}$. Finally, for $z = x$ or y,

$$\left|\frac{n+1-\lambda}{\lambda}[h(\xi+\delta_z) - \mathbf{E}h(\xi+\delta_z-\delta_{U_z})]\right| \le \frac{2}{\lambda} + \frac{1}{n+1} \tag{5.13}$$

so that (5.5) follows from collecting the bounds for the terms in (5.12).

To prove (5.13), it is enough to consider the following two cases.

Case I: $n+1 > \lambda$. Let $\tau_{n+1}^- = \inf\{t : |\mathbf{Z}_{\xi+\delta_z}(t)| = n\}$. Then using strong Markov property gives

$$h(\xi+\delta_z) = -\mathbf{E}\int_0^{\tau_{n+1}^-}[f(\mathbf{Z}_{\xi+\delta_z}(t)) - \mathrm{Po}(\boldsymbol{\lambda})(f)]dt + \mathbf{E}h(\zeta),$$

where $\zeta = \mathbf{Z}_{\xi+\delta_z}(\tau_{n+1}^-)$. Consequently,

$$\begin{aligned}& \left|\frac{n+1-\lambda}{\lambda}[h(\xi+\delta_z) - \mathbf{E}h(\xi+\delta_z-\delta_{U_z})]\right| \\ & \quad \le \frac{n+1-\lambda}{\lambda}\mathbf{E}\tau_{n+1}^- + \frac{n+1-\lambda}{\lambda}|\mathbf{E}h(\zeta) - \mathbf{E}h(\xi+\delta_z-\delta_{U_z})|.\end{aligned} \tag{5.14}$$

However, by (2.11) and Proposition A.2.3 of Barbour, Holst and Janson (1992),

$$\frac{n+1-\lambda}{\lambda}\mathbf{E}\tau_{n+1}^- = \frac{(n+1-\lambda)\bar{F}(n+1)}{\lambda(n+1)\mathrm{Po}(\lambda)\{n+1\}} \le \frac{(n+1-\lambda)(n+2)}{\lambda(n+1)(n+2-\lambda)} \le \frac{1}{\lambda}. \tag{5.15}$$

The second term of (5.14) depends on the difference between ζ and $\xi+\delta_z-\delta_{U_z}$. For convenience, write z as z_{n+1} and realize $\mathbf{Z}_{\xi+\delta_z}$ from

$$\mathbf{Z}_{\xi+\delta_z}(t) = \mathbf{Z}_0(t) + \sum_{i=1}^{n+1}\delta_{z_i}\mathbf{1}_{S_i>t},$$

where $S_1, \dots, S_{n+1}$ are independent exp(1) random variables and independent of $\mathbf{Z}_0$. At time τ_{n+1}^-, $\sum_{i=1}^{n+1} 1_{S_i > \tau_{n+1}^-}$ of $z_1, \dots, z_{n+1}$ are still alive and at least one of them has died. Since the death can happen to $z_1, \dots, z_{n+1}$ with equal chance, we can construct a coupling such that

$$\xi + \delta_z - \sum_{i=1}^{n+1} \delta_{z_i} 1_{S_i > \tau_{n+1}^-} = \sum_{i=1}^{n+1} \delta_{z_i} 1_{S_i \le \tau_{n+1}^-} = \delta_{U_z} + \eta$$

with random element $\eta \in \mathscr{H}$. Under this coupling,

$$\xi + \delta_z - \delta_{U_z} = \eta + \sum_{i=1}^{n+1} \delta_{z_i} 1_{S_i > \tau_{n+1}^-},$$

giving

$$nd_1(\xi + \delta_z - \delta_{U_z}, \zeta) \le |\mathbf{Z}_0(\tau_{n+1}^-)|. \tag{5.16}$$

On the other hand,

$$\mathbf{E}|\mathbf{Z}_0(\tau_{n+1}^-)| = n - \mathbf{E} \sum_{i=1}^{n+1} 1_{S_i > \tau_{n+1}^-} = n - (n+1)\mathbf{P}(S_1 > \tau_{n+1}^-).$$

Noting that $S_1 > \tau_{n+1}^-$ is equivalent to

$$S_1 > \inf\left\{t : \ |\mathbf{Z}_0(t)| + \sum_{i=2}^{n+1} 1_{S_i > t} = n - 1\right\} := \tau_n^-$$

and S_1 is independent of τ_n^-, we have from (5.6) that

$$\mathbf{E}|\mathbf{Z}_0(\tau_{n+1}^-)| = n - (n+1)\mathbf{E}\exp(-\tau_n^-) = -1 + (n+1)\left(\frac{\bar{F}(n-1)}{\bar{F}(n)} - \frac{n}{\lambda}\right). \tag{5.17}$$

Now, we claim that

$$\frac{n+1-\lambda}{\lambda}\mathbf{E}|\mathbf{Z}_0(\tau_{n+1}^-)| \le 1. \tag{5.18}$$

In fact, by (5.17), (5.18) is equivalent to

$$(n+1-\lambda)[\lambda\bar{F}(n-1) - n\bar{F}(n)] - \lambda\bar{F}(n) \le 0.$$

Expanding the formula into a power series of λ, we have that (5.18) is the same as

$$\sum_{i=n+1}^{\infty} \frac{\lambda^i}{(i-1)!}[2n + 1 - i - (n+1)n/i] \le 0,$$

which is obviously true.

Thus, it follows from Lemma 5.3, (5.16) and (5.18) that the second term of (5.14) is bounded by

$$\begin{aligned}(5.19)\qquad &\frac{n+1-\lambda}{\lambda}|\mathbf{E}h(\zeta)-\mathbf{E}h(\xi+\delta_z-\delta_{U_z})| \\ &\quad\le \frac{n+1-\lambda}{\lambda}\mathbf{E}|\mathbf{Z}_0(\tau_{n+1}^-)|\left(\frac{1}{\lambda}+\frac{1}{n+1}\right)\le\frac{1}{\lambda}+\frac{1}{n+1}.\end{aligned}$$

Combining (5.15) and (5.19) yields (5.13).

CASE II: $n+1\le\lambda$. Likewise, let $\tau_n^+=\inf\{t:|\mathbf{Z}_{\xi+\delta_z-\delta_{U_z}}(t)|=n+1\}$, then we get from strong Markov property that

$$h(\xi+\delta_z-\delta_{U_z})=-\mathbf{E}\int_0^{\tau_n^+}[f(\mathbf{Z}_{\xi+\delta_z-\delta_{U_z}}(t)-\mathrm{Po}(\boldsymbol{\lambda})(f)]dt+\mathbf{E}h(\varsigma),$$

where $\varsigma=\mathbf{Z}_{\xi+\delta_z-\delta_{U_z}}(\tau_n^+)$. Hence

$$\begin{aligned}(5.20)\qquad &\left|\frac{n+1-\lambda}{\lambda}[h(\xi+\delta_z)-\mathbf{E}h(\xi+\delta_z-\delta_{U_z})]\right| \\ &\quad\le\frac{\lambda-(n+1)}{\lambda}\mathbf{E}\tau_n^+ +\frac{\lambda-(n+1)}{\lambda}|\mathbf{E}h(\varsigma)-\mathbf{E}h(\xi+\delta_z)|.\end{aligned}$$

Using (2.11) and Proposition A.2.3 of Barbour, Holst and Janson (1992) again, we have

$$(5.21)\qquad \frac{\lambda-(n+1)}{\lambda}\mathbf{E}\tau_n^+=\frac{(\lambda-(n+1))F(n)}{\lambda^2\mathrm{Po}(\lambda)\{n\}}\le\frac{(\lambda-(n+1))\lambda}{\lambda^2(\lambda-n)}\le\frac{1}{\lambda}.$$

To estimate the second term of (5.20), without loss of generality, we may assume $U_z=z_{n+1}$ and realize

$$\mathbf{Z}_\xi(t)=\mathbf{Z}_0(t)+\sum_{i=1}^n\delta_{z_i}1_{S_i>t},$$

where $S_1,\dots,S_n$ and $\mathbf{Z}_0$ are as in Case I. At time τ_n^+, there are $\sum_{i=1}^n 1_{S_i>\tau_n^+}$ of $z_1,\dots,z_n$ still alive, so

$$(5.22)\qquad (n+1)d_1(\xi+\delta_z,\varsigma)\le|\mathbf{Z}_0(\tau_n^+)|.$$

But

$$\mathbf{E}|\mathbf{Z}_0(\tau_n^+)|=n+1-\mathbf{E}\sum_{i=1}^n 1_{S_i>\tau_n^+}=n+1-n\mathbf{P}(S_1>\tau_n^+).$$

Since $S_1>\tau_n^+$ is equivalent to

$$S_1>\inf\left\{t:\ |\mathbf{Z}_0(t)|+\sum_{i=2}^n 1_{S_i>t}=n\right\}:=\tau_{n-1}^+$$

and S_1 is independent of τ_{n-1}^+, we get from (5.6) that

$$(5.23)\quad \mathbf{E}|\mathbf{Z}_0(\tau_n^+)|=n+1-n\mathbf{E}\exp(-\tau_{n-1}^+)=1+\frac{nF(n)-\lambda F(n-1)}{F(n)}.$$

We now show

$$\frac{\lambda-(n+1)}{\lambda}\mathbf{E}|\mathbf{Z}_0(\tau_n^+)|\le 1. \tag{5.24}$$

Using (5.23), (5.24) can be rewritten as

$$(\lambda-(n+1))[nF(n)-\lambda F(n-1)]\le (n+1)F(n),$$

which can be verified by rearranging the terms into a series of λ, as in Case I.

Applying Lemma 5.3, (5.22) and (5.24), it follows that

$$\begin{aligned}&\frac{\lambda-(n+1)}{\lambda}|\mathbf{E}h(\varsigma)-\mathbf{E}h(\xi+\delta_z)|\\ &\quad\le\frac{\lambda-(n+1)}{\lambda}\mathbf{E}|\mathbf{Z}_0(\tau_n^+)|\left(\frac{1}{\lambda}+\frac{1}{n+1}\right)\le\frac{1}{\lambda}+\frac{1}{n+1},\end{aligned}$$

which, together with (5.20) and (5.21), gives (5.13). □

The implication of the improved estimate (5.5) is that the calculations for bounding the errors of Poisson process approximation in terms of d_2-metric can be significantly simplified. To illustrate the degree of simplification, we present a theorem based on Palm processes.

Recall that for a point process Ξ with mean measure $\boldsymbol{\lambda}$ the point process Ξ_x is said to be a Palm process associated with Ξ at $x\in\Gamma$ if for any measurable function $f:\Gamma\times\mathscr{H}\mapsto[0,\infty)$,

$$\mathbf{E}\left(\int_\Gamma f(x,\Xi)\Xi(dx)\right)=\mathbf{E}\left(\int_\Gamma f(x,\Xi_x)\boldsymbol{\lambda}(dx)\right)$$

[see Kallenberg (1976)]. The process $\Xi_x-\delta_x$ is called the reduced Palm process.

The metric we shall use for bounding the approximation errors is d_1'' as introduced in Brown and Xia (1995a): for $\xi_1=\sum_{i=1}^n\delta_{x_i}$, $\xi_2=\sum_{i=1}^m\delta_{y_i}\in\mathscr{H}$ with $m\ge n$,

$$d_1''(\xi_1,\xi_2):=\min_{\mathscr{P}}\sum_{i=1}^n d_0(x_i,y_{\mathscr{P}(i)})+(m-n),$$

where $\mathscr{P}$ ranges over all permutations of $(1,\dots,m)$. Note that $d_1''(\xi_1,\xi_2)\le\|\xi_1-\xi_2\|$, the total variation norm of the signed measure $\xi_1-\xi_2$. Also, if $|\xi_1|=|\xi_2|=n$, then $d_1''(\xi_1,\xi_2)=nd_1(\xi_1,\xi_2)$. This results in

LEMMA 5.5. *For $\xi,\eta\in\mathscr{H}$ and $x\in\Gamma$,*

$$\begin{aligned}&|[h(\xi+\delta_x)-h(\xi)]-[h(\eta+\delta_x)-h(\eta)]|\\ &\quad\le\frac{2}{|\eta|\wedge|\xi|+1}[d_1''(\xi,\eta)-\big||\eta|-|\xi|\big|]+\left(\frac{5}{\lambda}+\frac{3}{|\eta|\wedge|\xi|+1}\right)\big||\eta|-|\xi|\big|\\ &\quad\le\left(\frac{5}{\lambda}+\frac{3}{|\eta|\wedge|\xi|+1}\right)d_1''(\xi,\eta).\end{aligned}$$

PROOF. Write $\xi = \sum_{i=1}^n \delta_{x_i}$ and $\eta = \sum_{i=1}^m \delta_{y_i}$. Without loss of generality, we assume $m \geq n$ and

$$d_1''(\xi, \eta) = \sum_{i=1}^n d_0(x_i, y_i) + (m - n).$$

Define $\eta' = \sum_{i=1}^n \delta_{y_i}$, then

$$\begin{aligned} &|[h(\xi + \delta_x) - h(\xi)] - [h(\eta + \delta_x) - h(\eta)]| \\ &\quad \leq |[h(\xi + \delta_x) - h(\xi)] - [h(\eta' + \delta_x) - h(\eta')]| \\ &\qquad + |[h(\eta' + \delta_x) - h(\eta')] - [h(\eta + \delta_x) - h(\eta)]|. \end{aligned}$$

The proof is then complete by applying Lemma 5.4 and Theorem 5.1. □

As an immediate application of Lemma 5.5, we now state a theorem for estimating the errors of Poisson process approximation in terms of Palm processes.

THEOREM 5.6. *Let Ξ be a finite point process on Γ with mean measure $\boldsymbol{\lambda}$. Suppose for each $x \in \Gamma$ that Ξ_x is the Palm process associated with Ξ at x. Then*

$$d_2(\mathscr{L}\Xi, Po(\boldsymbol{\lambda})) \leq \mathbf{E}\int_\Gamma \frac{2}{|\Xi| \wedge |\Xi_x - \delta_x| + 1}[d_1''(\Xi, \Xi_x - \delta_x) - \big||\Xi| - |\Xi_x - \delta_x|\big|]\boldsymbol{\lambda}(dx)$$

$$+\mathbf{E}\int_\Gamma \left(\frac{5}{\lambda} + \frac{3}{|\Xi| \wedge |\Xi_x - \delta_x| + 1}\right)\big||\Xi| - |\Xi_x - \delta_x|\big|\boldsymbol{\lambda}(dx) \tag{5.25}$$

$$\leq \mathbf{E}\int_\Gamma \left(\frac{5}{\lambda} + \frac{3}{|\Xi| \wedge |\Xi_x - \delta_x| + 1}\right) d_1''(\Xi, \Xi_x - \delta_x)\boldsymbol{\lambda}(dx). \tag{5.26}$$

REMARK 5.7. These simplified bounds, with only one integrand in (5.26), are to be compared with Theorem 3.1 of Brown, Weinberg and Xia (2000), which has three terms to compute. Thus, Theorem 5.6 is more applicable and the order of the estimated bounds will be the same as that obtained from Theorem 3.1 of Brown, Weinberg and Xia (2000).

Acknowledgments. Many thanks to Louis H. Y. Chen and Kwok Pui Choi for many useful conversations and splendid hospitality. The work was done when Aihua Xia was a member of the Department of Statistics, The University of New South Wales, Sydney 2052, Australia.

After acceptance subject to revision of this paper, the authors learned that a former Ph.D student of Brown, G Weinberg, had submitted and obtained acceptance of a paper “Stein factor bounds for random variables” in *Journal of Applied Probability.* The work that Weinberg reports extends the work in Xia (1999) and the results are less general than those here.

REFERENCES

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: 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** 403–434.

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

BARBOUR, A. D. and BROWN, T. C. (1992). Stein's method and point process approximation. *Stochastic Processes Appl.* **43** 9–31.

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

BARBOUR, A. D. and EAGLESON, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. Appl. Prob.* **15** 585–600.

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 JENSEN, J. L. (1989). Local and tail approximations near the Poisson limit. *Scandinavian J. Statist.* **16** 75–87.

BARBOUR, A. D. and XIA, A. (1999). Poisson perturbation. *ESAIM: Probab. Statist.* **3** 131–150.

BROWN, T. C. and PHILLIPS, M. J. (1999). Negative binomial approximation with Stein's method. *Method. Comput. Appl. Probab.* **1** 407–421.

BROWN, T. C., WEINBERG, G. V. and XIA, A. (2000). Removing logarithms from Poisson process error bounds. *Stochastic Processes Appl.* **87** 149–165.

BROWN, T. C. and XIA, A. (1995a). On metrics in point process approximation. *Stochastics Stochastics Rep.* **52** 247–263.

BROWN, T. C. and XIA, A. (1995b). On Stein-Chen factors for Poisson approximation. *Statist. Probab. Lett.* **23** 327–332.

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

EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16.

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

JOHNSON, N. L., KOTZ, S. and KEMP, A. (1992). *Univariate Discrete Distributions.* Wiley, New York.

KALLENBERG, O. (1976). *Random Measures.* Academic Press, New York.

KEILSON, J. (1979). *Markov Chain Models–Rarity and Exponentiality.* Springer, New York.

LOH, W. (1992). Stein's method and multinomial approximation. *Ann. Appl. Probab.* **2** 536–554.

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

STEIN, C. (1971). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **3** 583–602. Univ. California Press, Berkeley.

WANG, Z. and YANG, X. (1992). *Birth and Death Processes and Markov Chains.* Springer, New York.

XIA, A. (1999). A probabilistic proof of Stein's factors. *J. Appl. Probab.* **36** 287–290.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MELBOURNE
PARKVILLE, VIC 3010
AUSTRALIA
E-MAIL: xia@ms.unimelb.edu.au