

STEIN'S METHOD AND STOCHASTIC ORDERINGS

FRASER DALY,* *Universität Zürich*

CLAUDE LEFÈVRE,** *Université Libre de Bruxelles*

SERGEY UTEV,*** *University of Nottingham*

Abstract

A stochastic ordering approach is applied with Stein's method for approximation by the equilibrium distribution of a birth–death process. The usual stochastic order and the more general s -convex orders are discussed. Attention is focused on Poisson and translated Poisson approximations of a sum of dependent Bernoulli random variables, for example, k -runs in independent and identically distributed Bernoulli trials. Other applications include approximation by polynomial birth–death distributions.

Keywords: Stein's method; birth–death process; stochastic ordering; total variation distance; (in)dependent indicators; (translated) Poisson approximation; total negative and positive dependence; (approximate) local dependence; polynomial birth–death approximation; k -runs

2010 Mathematics Subject Classification: Primary 62E17

Secondary 60F05; 60J80

1. Introduction

Stein's method has proved to be an effective tool in probability approximation, and has the advantage of being applicable in the presence of dependence. See, for example, Stein (1986), Barbour and Chen (2005), and Chen *et al.* (2011) for more recent developments. It is well known that error bounds obtained via Stein's method may be simplified under some assumptions on the dependence present. For example, in the presence of a negative or positive relation, Stein's method gives simple error bounds in the Poisson approximation of a sum of indicator random variables. This is exploited throughout the work of Barbour *et al.* (1992), and will be returned to in our Section 4.

In this work we consider the more general situation of approximation by the equilibrium distribution of a birth–death process, and examine the situations in which Stein's method leads to simple, easily calculable error bounds. These error bounds will typically be differences of moments of our random variables. As we will see, the assumptions under which we can obtain such error bounds are naturally phrased in terms of stochastic orderings.

Consider a birth–death process on (some subset of) $\mathbb{Z}^+ = \{0, 1, \dots\}$ with birth rates α_j and death rates β_j for $j \geq 0$. Suppose that $\beta_0 = 0$. Let π be the stationary distribution of such a process, with $\pi_j = P(\pi = j)$, $j \geq 0$. In this work we combine Stein's method with a

Received 11 November 2009; revision received 21 October 2011.

* Current address: Heilbronn Institute for Mathematical Research, Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK. Email address: fraser.daly@bristol.ac.uk

** Postal address: Département de Mathématique, Université Libre de Bruxelles, Campus de la Plaine, CP 210, B-1050 Bruxelles, Belgium. Email address: clefevre@ulb.ac.be

*** Postal address: School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK. Email address: sergey.utev@nottingham.ac.uk

stochastic ordering construction to consider the approximation by π of some random variable W on \mathbb{Z}^+ .

Our random variable π satisfies the identity $E[Ag(\pi)] = 0$ for any bounded function $g: \mathbb{Z}^+ \mapsto \mathbb{R}$, where A is the linear operator defined by

$$Ag(j) = \alpha_j g(j + 1) - \beta_j g(j), \quad j \geq 0; \tag{1}$$

A is a characterising operator for π , in the sense that a random variable $Z \stackrel{D}{=} \pi$ if and only if $E[Ag(Z)] = 0$ for all bounded g . The construction of such a characterising operator is the basis of Stein’s method for probability approximation. See the books by Stein (1986), Barbour *et al.* (1992), Barbour and Chen (2005), Chen *et al.* (2011), and the references therein. For Stein’s method applied to birth–death processes, see Brown and Xia (2001), Holmes (2004), and Eichelsbacher and Reinert (2008).

Given some bounded test function h , the so-called Stein equation is defined by

$$h(j) - E[h(\pi)] = Af(j), \quad j \geq 0. \tag{2}$$

Its solution is denoted by $f = f_h = Sh$. We call S the Stein operator. Bounds on S are an essential ingredient of Stein’s method.

Note that the solution f of the Stein equation depends on the chosen test function h . However, for notational convenience, in much of the work that follows we will write f rather than f_h or Sh . We will often choose $h(j) = \mathbf{1}_{\{j \in B\}}$ for some $B \subseteq \mathbb{Z}^+$, in which case the solution f will depend on the chosen set B .

Some bounds on the solution of the Stein equation in this general framework are given in the work of Brown and Xia (2001) and Eichelsbacher and Reinert (2008). For example, Theorem 2.10 of Brown and Xia (2001) gives the following result.

Proposition 1. (Brown and Xia (2001).) *If $\beta_0 = 0$ and $\alpha_k - \alpha_{k-1} \leq \beta_k - \beta_{k-1}$ for $k = 1, 2, \dots$, then, for all $h: \mathbb{Z}^+ \mapsto [0, 1]$,*

$$|f_h(i + 1) - f_h(i)| \leq \min \left\{ \frac{1}{\alpha_i}, \frac{1}{\beta_i} \right\}.$$

Section 2 of Eichelsbacher and Reinert (2008) also gives several bounds on $|f_h(i)|$ and $|f_h(i + 1) - f_h(i)|$, some of which rely on conditions on the birth and death rates analogous to those in Proposition 1.

There are several common distributions π covered by this framework. For many of these examples, the bounds mentioned above may be applied.

- If $\alpha_j = \lambda$ and $\beta_j = j$, then $\pi \sim \text{Po}(\lambda)$, the Poisson distribution with mean λ . See Barbour *et al.* (1992) and the references therein.
- If $\alpha_j = q(r + j)$ and $\beta_j = j$, then $\pi \sim \text{NB}(r, 1 - q)$ has a negative binomial distribution. See Brown and Phillips (1999).
- If $\alpha_j = (n - j)p$ and $\beta_j = (1 - p)j$, then $\pi \sim \text{Bin}(n, p)$. See Ehm (1991).
- In the geometric case we may, of course, use the negative binomial operator above. Alternatively, we may choose $\alpha_j = q$ and $\beta_j = \mathbf{1}_{\{j \geq 1\}}$, so that $\pi \sim \text{Geom}(1 - q)$. See Peköz (1996).

The present work is organised as follows. In Section 2 we derive abstract error bounds using Stein's method combined with some stochastic ordering assumptions in the setting of approximation by the equilibrium distribution of a birth–death process. In Section 3, a simple sufficient condition under which these stochastic ordering assumptions hold is considered, and some applications are given. In Section 4 we discuss Poisson approximations for a sum of dependent indicators. We will see how concepts of negative and positive relation relate to our stochastic ordering assumptions, and present generalisations of error bounds derived in Barbour *et al.* (1992). Based on this work, in Section 5 we consider translated Poisson approximations. Applications here will include the approximation of the number of k -runs in independent and identically distributed (i.i.d.) Bernoulli trials. Finally, in Section 6 we give another abstract approximation theorem and consider its application to a sum of independent indicator random variables.

2. An abstract approximation theorem

Consider Stein's method for approximating the equilibrium distribution of a birth–death process. Our aim in this section is to derive abstract error bounds under some stochastic ordering assumptions.

2.1. A first-order bound

Suppose that W is a random variable supported on (some subset of) \mathbb{Z}^+ with $\mu_j = P(W = j)$, $j \geq 0$. Set $\mu_{-1} = 0$. We are interested in the approximation of such a variable W by π , specifically by estimating the difference $|Eh(W) - Eh(\pi)|$, i.e. $|E[Af(W)]|$. To this end, a simple representation of this difference will be applied with some stochastic ordering assumptions to yield bounds using Stein's method. We may then bound, for example, the total variation distance between $\mathcal{L}(W)$ and $\mathcal{L}(\pi)$, defined by

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) = \sup_{B \subseteq \mathbb{Z}^+} |P(W \in B) - P(\pi \in B)|.$$

Although we are primarily interested in the approximation in the total variation distance, the results we derive may also be used with other probability metrics.

Let Δ be the forward difference operator. Since, with operator (1), the choice of $f(0)$ is arbitrary, we follow Brown and Xia (2001) and choose $f(0) = 0$. Writing $f(j) = \Delta f(0) + \dots + \Delta f(j - 1)$, we thus obtain the representation

$$Eh(W) - Eh(\pi) = \sum_{k=0}^{\infty} \Delta f(k) \sum_{j=k+1}^{\infty} (\alpha_{j-1} \mu_{j-1} - \beta_j \mu_j). \tag{3}$$

In the next subsection, we will extend (3) to include the l th forward differences of $f(\cdot)$ for all $l \geq 1$.

We now consider how this representation may be applied in conjunction with the usual stochastic ordering, denoted by ' \succeq_{st} '. Define two random variables W_α and W_β by

$$P(W_\alpha = j) = \frac{\alpha_{j-1} \mu_{j-1}}{E \alpha_W} \quad \text{and} \quad P(W_\beta = j) = \frac{\beta_j \mu_j}{E \beta_W}, \quad j \geq 1. \tag{4}$$

If $W_\alpha \succeq_{st} W_\beta$ and $E \alpha_W \geq E \beta_W$, we have $\sum_{j=i}^{\infty} \alpha_{j-1} \mu_{j-1} \geq \sum_{j=i}^{\infty} \beta_j \mu_j$ for all $i \geq 1$. In this case, (3) may be bounded to obtain

$$|Eh(W) - Eh(\pi)| \leq \|\Delta f\|_\infty E[\alpha_W(W + 1) - \beta_W W].$$

A similar argument holds if we instead assume that $W_\beta \succeq_{st} W_\alpha$ and $E \beta_W \geq E \alpha_W$. We thus obtain the following result.

Proposition 2. *Assume that one of the following two conditions holds:*

$$(i) W_\alpha \succeq_{st} W_\beta \text{ with } E \alpha_W \geq E \beta_W \text{ or } (ii) W_\beta \succeq_{st} W_\alpha \text{ with } E \beta_W \geq E \alpha_W. \tag{5}$$

Then, for bounded $h: \mathbb{Z}^+ \mapsto \mathbb{R}$,

$$|E h(W) - E h(\pi)| \leq \|\Delta S h\|_\infty |E[\alpha_W(W + 1) - \beta_W W]|. \tag{6}$$

2.2. An s -order bound

We will now establish our main abstract result. For that, we will have recourse to the concept of discrete s -convex stochastic ordering, denoted by ' \succeq_{s-cx} ', for any integer $s \geq 1$. See, for example, Lefèvre and Utev (1996) for this notion. Briefly, given any two nonnegative integer-valued random variables X and Y , we say that $Y \succeq_{s-cx} X$ when

$$E[f(X)] \leq E[f(Y)] \text{ for all } s\text{-convex functions } f,$$

that is, for all functions f satisfying $\Delta^s f(j) \geq 0, j \geq 0$. Note that this ordering implies that $E[X^l] \leq E[Y^l]$ for $l = 1, \dots, s - 1$.

This is a generalisation of the usual stochastic ordering (corresponding to $s = 1$). As we will see, this is a natural generalisation to apply when considering representations analogous to (3) with higher powers of Δ . Functions of W_α and W_β appearing in such representations that we employ can be seen to be s -convex for some s .

We begin by introducing a Bernoulli random variable v_p with

$$P(v_p = 1) = p = 1 - P(v_p = 0),$$

independently of all other entries. We write $\alpha = E \alpha_W$ and $\beta = E \beta_W$, and, in an analogous way to (4), we define the random variables W_α and W_β by

$$P(W_\alpha \in B) = \alpha^{-1} E[\alpha_W \mathbf{1}_{\{W+1 \in B\}}] \text{ and } P(W_\beta \in B) = \beta^{-1} E[\beta_W \mathbf{1}_{\{W \in B\}}] \tag{7}$$

for any Borel set B . For notational convenience, we choose to write $C_n^k = \binom{n}{k}$, and define $C_n^k = 0$ for $k > n$.

We now give the key proposition and an immediate corollary, followed by the proof of the proposition.

Proposition 3. *Assume that there exists a random variable Y on \mathbb{Z}^+ such that $W_\beta - Y \geq 0$ almost surely (a.s.) and that*

$$W_\alpha \succeq_{s-cx} v_p(W_\beta - Y). \tag{8}$$

Then, for bounded $h: \mathbb{Z}^+ \mapsto \mathbb{R}$,

$$\begin{aligned} |E h(W) - E h(\pi)| \leq & \sum_{t=0}^{s-1} |\Delta^t S h(0)| |E[\alpha_W C_{W+1}^t] - E[\beta_W C_W^t]| \\ & + \|\Delta^s S h\|_\infty (\alpha E[C_{W_\alpha}^s] - 2\alpha p E[C_{W_\beta - Y}^s] + (\alpha p + |\alpha p - \beta|) E[C_{W_\beta}^s]). \end{aligned} \tag{9}$$

Corollary 1. Assume that $\alpha = \beta$ and that one of the following two conditions holds:

$$(i) W_\alpha \succeq_{s-cx} W_\beta \quad \text{or} \quad (ii) W_\beta \succeq_{s-cx} W_\alpha. \tag{10}$$

Then, for bounded $h: \mathbb{Z}^+ \mapsto \mathbb{R}$,

$$|Eh(W) - Eh(\pi)| \leq \|\Delta^s Sh\|_\infty |E[\alpha_W C_{W+1}^s] - E[\beta_W C_W^s]|. \tag{11}$$

Corollary 1 follows from Proposition 3 by considering the special case of (8) when $p = 1$ and $Y = 0$ a.s. When $\alpha = \beta$ and condition (10) holds, we have

$$E[\alpha_W (W + 1)^t] = E[\beta_W W^t], \quad t = 0, \dots, s - 1,$$

so that inequality (9) reduces to (11).

We note that Proposition 2 does not follow as a special case of Corollary 1, since this latter result requires the condition that $\alpha = \beta$, which is not needed in Proposition 2.

Proof of Proposition 3. In the first step we derive a representation of $E[Af(W)]$ that generalises representation (3). Observe that (1) and (7) give

$$E[Af(W)] = E[\alpha_W f(W + 1)] - E[\beta_W f(W)] = \alpha E[f(W_\alpha)] - \beta E[f(W_\beta)].$$

Expanding the function f by the discrete Taylor formula, we obtain, for any $s = 1, 2, \dots$,

$$f(x) = f(0) + \sum_{k=0}^{\infty} \Delta f(k) \mathbf{1}_{\{x > k\}} = \sum_{t=0}^{s-1} \Delta^t f(0) C_x^t + \sum_{k=0}^{\infty} \Delta^s f(k) C_{x-k-1}^{s-1}. \tag{12}$$

To see this, note that the $s = 1$ case is clear. Then proceed by induction, using (12) with $s = 1$ applied to the function $\Delta^{s-1} f$ to write

$$\Delta^{s-1} f(k) = \Delta^{s-1} f(0) + \sum_{j=0}^{\infty} \Delta^s f(j) C_{k-j-1}^0.$$

See also Lefèvre and Utev (1996, Section 2.2).

Now, using (12), we find that

$$\begin{aligned} E[Af(W)] &= \sum_{t=0}^{s-1} \Delta^t f(0) E[AC_W^t] + \sum_{k=0}^{\infty} \Delta^s f(k) E[AC_{W-k-1}^{s-1}] \\ &= \sum_{t=0}^{s-1} \Delta^t f(0) (\alpha E[C_{W_\alpha}^t] - \beta E[C_{W_\beta}^t]) \\ &\quad + \sum_{k=0}^{\infty} \Delta^s f(k) (\alpha E[C_{W_\alpha-k-1}^{s-1}] - \beta E[C_{W_\beta-k-1}^{s-1}]). \end{aligned} \tag{13}$$

Our next step is to derive an abstract metrics-ordering relationship result, which is stated below as a separate lemma. Using bound (15) in representation (13) then leads to the announced bound (11).

Lemma 1. Let $X, Y,$ and Z be random variables on \mathbb{Z}^+ such that

$$Z - Y \geq 0 \text{ a.s. and } X \succeq_{s\text{-cx}} v_p(Z - Y). \tag{14}$$

Then, for all $a, b \in \mathbb{R}^+,$

$$\sum_{k=0}^{\infty} |a E[C_{X-k-1}^{s-1}] - b E[C_{Z-k-1}^{s-1}]| \leq a E[C_X^s] - 2ap E[C_{Z-Y}^s] + (ap + |ap - b|) E[C_Z^s]. \tag{15}$$

Proof. Letting

$$w_k^{(s)}(x) = w_k(x) = C_{x-k-1}^{s-1},$$

we obtain

$$\begin{aligned} \sum_{k=0}^{\infty} |a E[C_{X-k-1}^{s-1}] - b E[C_{Z-k-1}^{s-1}]| &= \sum_{k=0}^{\infty} |a E[w_k(X)] - b E[w_k(Z)]| \\ &\leq a \sum_{k=0}^{\infty} |E[w_k(X)] - E[w_k(v_p(Z - Y))]| \\ &\quad + a \sum_{k=0}^{\infty} |E[w_k(v_p Z)] - E[w_k(v_p(Z - Y))]| \\ &\quad + \sum_{k=0}^{\infty} |a E[w_k(v_p Z)] - b E[w_k(Z)]| \\ &= S_1 + S_2 + S_3. \end{aligned} \tag{16}$$

Let us examine the three sums in (16). First, we easily check that

$$\sum_{k=0}^{\infty} E[w_k(Z)] = E[C_Z^s]. \tag{17}$$

Using (17), we successively find that

$$S_3 = |ap - b| \sum_{k=0}^{\infty} E[w_k(Z)] = |ap - b| E[C_Z^s];$$

since $Z - Y \geq 0$ and $Z \succeq_{st} Z - Y,$

$$S_2 = ap \sum_{k=0}^{\infty} (E[w_k(Z)] - E[w_k(Z - Y)]) = ap(E[C_Z^s] - E[C_{Z-Y}^s]);$$

finally, by assumption (14) and a standard property of the order ‘ $\succeq_{s\text{-cx}}$ ’,

$$S_1 = a \sum_{k=0}^{\infty} [E w_k(X) - p E w_k(Z - Y)] = a(E[C_X^s] - p E[C_{Z-Y}^s]).$$

Inserting these three terms in (16) yields bound (15).

Remark 1. For $s = p = 1$ and $a = b = 1$, Lemma 1 states that if $X \succeq_{st} Z - Y \geq 0$ then an upper bound for the Wasserstein distance between $\mathcal{L}(X)$ and $\mathcal{L}(Z)$ is

$$d_W(\mathcal{L}(X), \mathcal{L}(Z)) = \sum_{k=0}^{\infty} |\mathbb{P}(X > k) - \mathbb{P}(Z > k)| \leq 2EY + EX - EZ. \tag{18}$$

This bound is of interest in the stochastic ordering context investigated in Kamae *et al.* (1977), with random variables on \mathbb{Z}^+ here. Note that, by choosing the optimal coupling X, Z , and $Y = (Z - X)_+$, (18) gives the exact bound since

$$d_W(\mathcal{L}(X), \mathcal{L}(Z)) \leq 2E(Z - X)_+ + EX - EZ = E|Z - X| = d_W(\mathcal{L}(X), \mathcal{L}(Z)).$$

It is worth noting that an analogous argument allows us to show that the same bound (18) holds under the single condition $X + Y \succeq_{st} Z$. A priori, this result seems to be preferable, since the extra condition $Z - Y \geq 0$ is not required. One can see, however, that $X \succeq_{st} Z - Y$ does not imply that $X + Y \succeq_{st} Z$ in general. As an example, choose $X = U, Y = U$, and $Z = n$ a.s., where n is any fixed positive integer and U is discrete uniform on the set $\{0, 1, \dots, n\}$. Then, $X = U \stackrel{D}{=} n - U = Z - Y$ so that $X \succeq_{st} Z - Y$, but $X + Y \not\prec_{st} Z = n$.

3. A simple sufficient condition and examples

In practice, it may be difficult to check directly such conditions as stochastic ordering between W_α and W_β , as required in (5) and (10). It is thus useful to have available a simple sufficient condition which we may then apply.

Throughout this subsection, we assume that $\alpha = \beta$ and W_α and W_β have equal moments of order $t = 1, \dots, s - 1$. That is, we assume that

$$(A) \ E[\alpha_W(W + 1)^t] = E[\beta_W W^t], \ t = 0, \dots, s - 1.$$

A well-known Karlin–Novikoff sufficient condition to guarantee the s -convex ordering in (10) under (A) is that our sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j\mu_j\}$ has at most s changes of sign. See, for example, Denuit and Lefèvre (1997, p. 206).

Note that, with this Karlin–Novikoff cut criterion, if the sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j\mu_j\}$ is nonnegative after the final sign change then $W_\alpha \succeq_{s-cx} W_\beta$. Conversely, if this sequence is nonpositive after its final sign change, the ordering is reversed.

Proposition 4. *Suppose that condition (A) is satisfied and that the sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j\mu_j\}$ has at most s changes of sign. Then (11) holds.*

As a consequence of Proposition 4, we obtain the following corollary, which extends Proposition A.1 of Barbour and Pugliese (2000) to birth–death processes.

Corollary 2. *Suppose that $E\alpha_W = E\beta_W$. If the sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j\mu_j\}$ is monotone then W_α and W_β are stochastically ordered, so inequality (6) may be applied.*

We illustrate these results with the following examples.

Example 1. Our first example is motivated by Phillips and Weinberg (2000). Let W have a Bose–Einstein occupancy distribution. That is, given $m, d \geq 1$,

$$\mu_j = \mathbb{P}(W = j) = \binom{d + m - j - 2}{m - j} \binom{d + m - 1}{m}^{-1}, \quad 0 \leq j \leq m.$$

We wish to approximate W by $\pi \sim \text{Geom}(p)$, a geometric distribution on $\mathbb{Z}^+ = \{0, 1, \dots\}$, where $p = (d - 1)/(d + m - 1)$. Let $q = 1 - p$. To obtain our geometric law, we choose $\alpha_j = q$ and $\beta_j = \mathbf{1}_{\{j>0\}}$, $j \geq 0$, as the birth and death rates.

We can easily check that in this case $E\alpha_W = E\beta_W$ and the sequence $\{q\mu_{j-1} - \mu_j\}$ is nondecreasing, so $W_\alpha \succeq_{\text{st}} W_\beta$. Using Corollary 2, bound (6) then becomes

$$|Eh(W) - Eh(\pi)| \leq p\|\Delta Sh\|_\infty |EW - E\pi| \tag{19}$$

for bounded $h: \mathbb{Z}^+ \mapsto \mathbb{R}$. Moreover, it is known (see Peköz (1996, Section 2)) that the Stein operator S admits here the representation

$$Sh(j) = - \sum_{i=j}^\infty [h(i) - Eh(\pi)]q^{i-j}.$$

From this, we find that $\Delta Sh(k) = -\sum_{i=k}^\infty \Delta h(i)q^{i-k}$, which leads to the bound

$$\|\Delta Sh\|_\infty \leq p^{-1}\|\Delta h\|_\infty.$$

Inserting this bound into (19) yields the following result.

Corollary 3. *With W and π as above and $h: \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded,*

$$|Eh(W) - Eh(\pi)| \leq \|\Delta h\|_\infty \frac{m}{d(d-1)}.$$

In particular, $d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq m/(d\{d-1\})$.

Examples 2 and 3 centre around approximations by so-called polynomial birth–death distributions, defined in Brown and Xia (2001) as the equilibrium distribution of a birth–death process with birth and death rates α_j and β_j which are polynomial in j . With such choices, we will write $\pi \sim \text{PBD}(\alpha_j, \beta_j)$.

Example 2. Suppose that W satisfies $\mu_j = (a + bj^{-1})\mu_{j-1}$ for some $a, b \in \mathbb{R}$. That is, W belongs to the Katz (or Panjer) family of distributions (see Johnson *et al.* (1992, Section 2.3.1)). It is well known that in this case W must have either a binomial, Poisson, or negative binomial distribution.

We fix some $l \geq 1$ and consider the approximation of W by the polynomial birth–death distribution $\pi \sim \text{PBD}(\alpha, jQ_{l-1}(j))$. Here we have chosen a constant birth rate α and a death rate $\beta_j = jQ_{l-1}(j)$, where $Q_{l-1}(j)$ is a monic polynomial in j of degree $l - 1$. This gives us l parameters needed to specify the distribution of π . We choose these parameters in such a way that $E[\alpha_W(W + 1)^t] = E[\beta_W W^t]$ for $t = 0, \dots, l - 1$ is satisfied. We assume that we can do so in such a way that all the coefficients of $Q_{l-1}(j)$ are nonnegative. This is a stronger condition than we need for the results below to be applicable, but serves to illustrate the type of bounds we may obtain. In particular, this assumption guarantees that the death rate β_j is nonnegative and nondecreasing.

With our choice of birth and death rates we have

$$\alpha\mu_{j-1} - \beta_j\mu_j = \alpha\mu_{j-1} - jQ_{l-1}(j)(a + bj^{-1})\mu_{j-1} = \mu_{j-1}[\alpha - ajQ_{l-1}(j) - bQ_{l-1}(j)].$$

Noting that $\alpha - ajQ_{l-1}(j) - bQ_{l-1}(j)$ is a polynomial of degree l in j , and, therefore, has at most l real roots, the sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j\mu_j\}$ has at most l changes of sign, so that either $W_\alpha \succeq_{l\text{-cx}} W_\beta$ or $W_\beta \succeq_{l\text{-cx}} W_\alpha$.

With the assumption that β_j is nondecreasing in j , Theorem 2.10 of Brown and Xia (2001) gives

$$\sup\{\|\Delta Sh\|_\infty : h(j) = \mathbf{1}_{\{j \in B\}}, B \subseteq \mathbb{Z}^+\} \leq \alpha^{-1}.$$

Hence, with $h(j) = \mathbf{1}_{\{j \in B\}}$ for some $B \subseteq \mathbb{Z}^+$,

$$\|\Delta^l Sh\|_\infty \leq 2^{l-1} \|\Delta Sh\|_\infty \leq 2^{l-1} \alpha^{-1}.$$

From Corollary 1 we thus obtain Corollary 4.

Corollary 4. *With W and π as above, if the coefficients of $Q_{l-1}(j)$ are nonnegative,*

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq 2^{l-1} \alpha^{-1} \left| \mathbb{E} \left[\alpha \binom{W+1}{l} - W Q_{l-1}(W) \binom{W}{l} \right] \right|. \tag{20}$$

For example, consider the case where $W \sim \text{Bin}(n, p)$ and $\pi \sim \text{PBD}(\alpha, \gamma j + j(j-1))$, so that $l = 2$. Choosing our constants α and γ according to the prescription above, straightforward calculations give

$$\alpha = n(n-1)p(1-p) \quad \text{and} \quad \gamma = (n-1)(1-2p).$$

Furthermore,

$$\begin{aligned} \mathbb{E}[W(W+1)] &= np(np+2-p), \\ \mathbb{E}[W^2(W-1)] &= n(n-1)p^2(np+2-2p), \\ \text{and } \mathbb{E}[W^2(W-1)^2] &= n(n-1)p^2(n^2p^2+4np-5np^2-8p+6p^2+2). \end{aligned}$$

Evaluating bound (20) then gives the following result.

Corollary 5. *Assume that $W \sim \text{Bin}(n, p)$ and $\pi \sim \text{PBD}(\alpha, \gamma j + j(j-1))$. Then, if $\gamma \geq 1$,*

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq 2p^2. \tag{21}$$

We first note that the condition $\gamma \geq 1$ is stronger than we need for the death rate $\beta_j = \gamma j + j(j-1)$ to be nonnegative and nondecreasing, and that bound (21) also applies under this weaker condition.

We further note that (20) does not necessarily give a bound of the optimal order. In the case covered by (21), Theorem 3.1 of Brown and Xia (2001) gives a bound on the total variation distance of order $O(p^2/\sqrt{\lambda})$, where $\lambda = \mathbb{E}[W] = np$. This disparity is due to our rather crude use of the supremum norm in obtaining bounds such as (20). In Sections 5 and 6 we will consider more refined ways to bound the terms of our Stein equation in some particular cases when we have two parameters to choose in our approximating distribution π . Despite this disadvantage, we nevertheless note that (20) gives an explicit bound which may be applied in many contexts.

Example 3. Our final example of this section focuses on mixture distributions of the polynomial birth–death type. Suppose that $\pi \sim \text{PBD}(\alpha, \beta_j)$ and $W \sim \text{PBD}(\xi, \beta_j)$ for some constant birth rate α , polynomial death rate β_j , and random variable ξ on \mathbb{R}^+ . In this case we have

$$\mu_j = \frac{\mathbb{E}[\mu_0(\xi)\xi^j]}{\prod_{k=1}^j \beta_k}, \quad j \geq 0. \tag{22}$$

We choose α such that $\alpha = E \beta_W$, that is,

$$\alpha = E \sum_{j=0}^{\infty} \beta_j \mu_j = E \sum_{j=0}^{\infty} \xi \mu_{j-1} = E \xi.$$

Using (22), we obtain

$$\begin{aligned} \alpha \mu_j - \beta_{j+1} \mu_{j+1} &= E \left[\frac{\mu_0(\xi) \alpha^{j+1}}{\prod_{k=1}^j \beta_k} \left\{ \left(\frac{\xi}{\alpha} \right)^j - \left(\frac{\xi}{\alpha} \right)^{j+1} \right\} \right] \\ &= E \left[\frac{\alpha \mu_0(\xi)}{\mu_0(\alpha)} \left(1 - \frac{\xi}{\alpha} \right) \pi_j \left(\frac{\xi}{\alpha} \right)^j \right]. \end{aligned}$$

From this, we can see that the sequence $\{\alpha \mu_j - \beta_{j+1} \mu_{j+1}\}$ is monotone. Hence, Corollary 2 gives us the following result.

Corollary 6. *With W and π as above and $h: \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded,*

$$|E h(W) - E h(\pi)| \leq \|\Delta S h\|_{\infty} |E[\alpha(W + 1) - \beta_W W]|. \tag{23}$$

For example, if $\beta_j = j$ then $W \sim \text{Po}(\xi)$ and we take $\pi \sim \text{Po}(\lambda)$, where $\lambda = E \xi$. Using the well-known bound on the Stein operator S in this case, namely,

$$\|\Delta S h\|_{\infty} \leq \lambda^{-1} (1 - e^{-\lambda}) \|h\|_{\infty}, \tag{24}$$

evaluating (23) gives, after some straightforward calculations,

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \lambda^{-1} (1 - e^{-\lambda}) \text{var}(\xi),$$

a bound that has also been obtained in Barbour *et al.* (1992, Theorem 1.C).

4. Poisson approximation for a sum of indicators

Throughout this section, the random variable W of interest is a sum of indicators, i.e.

$$W = X_1 + \dots + X_n,$$

where the X_i are Bernoulli variables, possibly dependent, with

$$p_i = P(X_i = 1) = 1 - P(X_i = 0), \quad 1 \leq i \leq n.$$

Using Propositions 2 and 3, we are going to investigate the approximation of the sum W by a Poisson random variable $\pi \sim \text{Po}(\lambda)$.

Recall that our Poisson variable is derived from (1) when $\alpha_j = \lambda$ and $\beta_j = j$, so that, by (7),

$$W_{\alpha} = W + 1 \quad \text{and} \quad P(W_{\beta} \in B) = \frac{E[W \mathbf{1}_{\{W \in B\}}]}{E W} \tag{25}$$

for any Borel set B . Note that W_{β} has the W -size-biased distribution. See, for example, Goldstein and Rinott (1996). This is crucial in the work that follows.

In the analysis, an important role will be played by the variables

$$W_i = W - X_i, \quad 1 \leq i \leq n.$$

4.1. Total dependence

First, we consider the case where the indicators X_i are totally negatively dependent (TND) in the sense of Papadatos and Papathanasiou (2002). Let us recall that n random variables X_i , $1 \leq i \leq n$, are TND if

$$\text{cov}[g_1(X_i), g_2(W_i)] \leq 0, \quad 1 \leq i \leq n, \quad (26)$$

for all nondecreasing functions g_1 and g_2 such that the covariance exists.

Papadatos and Papathanasiou (2002, Theorem 3.1) showed that the class of TND indicators includes the standard class of negatively related indicators. Stein's method for the Poisson approximation of a sum of negatively related indicators is discussed in, for example, Barbour *et al.* (1992) and Erhardsson (2005). Recall that indicator random variables X_1, \dots, X_n are said to be negatively related if

$$\begin{aligned} & \mathbb{E}[g(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \mid X_i = 1] \\ & \leq \mathbb{E}[g(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)], \quad 1 \leq i \leq n, \end{aligned} \quad (27)$$

for all nondecreasing functions $g: \{0, 1\}^{n-1} \mapsto \{0, 1\}$.

We wish to bound the total variation distance between $\mathcal{L}(W)$ and $\text{Po}(\lambda)$. For that, we will apply Proposition 2. By (25), we have, for any function $g: \mathbb{Z}^+ \mapsto \mathbb{R}$,

$$\mathbb{E} g(W_\alpha) = \mathbb{E} g(W + 1) \quad \text{and} \quad \mathbb{E} g(W_\beta) = \frac{\mathbb{E}[Wg(W)]}{\mathbb{E} W}.$$

Thus, to show that $W_\alpha \succeq_{\text{st}} W_\beta$, we must prove that if g is nondecreasing then $\mathbb{E} W \mathbb{E} g(W + 1) \geq \mathbb{E}[Wg(W)]$. In fact, this was established by Papadatos and Papathanasiou (2002, Lemma 3.1).

Using bound (24) on the Stein operator in the Poisson case, (5) and (6) yield the following result.

Theorem 1. *If the indicators $\{X_i: 1 \leq i \leq n\}$ are TND then $W_\alpha \succeq_{\text{st}} W_\beta$. If, in addition, $\mathbb{E} W \geq \lambda$ then*

$$d_{\text{TV}}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} ([\lambda + 1] \mathbb{E} W - \mathbb{E}[W^2]).$$

Further results on, and examples of, TND indicator random variables can be found in Papadatos and Papathanasiou (2002).

Let us now consider the case where the indicators X_i are positively dependent in a certain sense. We adapt definition (26) and say that n random variables X_1, \dots, X_n are totally positively dependent (TPD) if

$$\text{cov}[g_1(X_i), g_2(W_i)] \geq 0, \quad 1 \leq i \leq n,$$

for all nondecreasing functions g_1 and g_2 such that the covariance exists.

Association or a positive relation is sufficient for TPD. This can be established analogously to the proof of Theorem 3.1 of Papadatos and Papathanasiou (2002). Recall that our indicator random variables are said to be positively related if (27) holds with the inequality reversed for all nondecreasing functions $g: \{0, 1\}^{n-1} \mapsto \{0, 1\}$. This standard property is used with Stein's method in, for example, Barbour *et al.* (1992) and Erhardsson (2005).

In the sequel, it is assumed that $\mathbb{E} W = \lambda$. To obtain a bound for the total variation distance, we will apply Proposition 3, using the lemma stated below. To begin with, we introduce a

random variable X_V , a mixing of our n indicators, in which the index V is a random variable of law

$$P(V = i) = \frac{E X_i}{\lambda}, \quad 1 \leq i \leq n. \tag{28}$$

Lemma 2. *If $E W = \lambda$ and the indicators $\{X_i : 1 \leq i \leq n\}$ are TPD, then*

$$W_\beta \succeq_{st} W_\alpha - X_V, \tag{29}$$

where $W_\alpha - X_V \geq 0$ a.s.

Proof. As seen in (25), $W_\alpha = W + 1$ and, thus, $W_\alpha - X_V \geq 0$ a.s. Moreover, W_β has the so-called W -size-biased distribution; see, for example, Goldstein and Rinott (1996). With W being a sum of indicators, it is then known that W_β admits the representation

$$W_\beta = \sum_{i \neq V} \hat{X}_i + 1, \tag{30}$$

where V is a random variable of law (28), and, if $V = v$,

$$\hat{X}_i \stackrel{D}{=} (X_i \mid X_v = 1), \quad i \neq v.$$

Thus, by (30), ordering (29) is equivalent to $\sum_{i \neq V} \hat{X}_i \succeq_{st} W - X_V$. To establish this, it is enough to prove that

$$\sum_{i \neq v} \hat{X}_i \succeq_{st} W - X_v, \quad 1 \leq v \leq n;$$

see Shaked and Shanthikumar (2007, Theorem 1.A.3(d)). Now, by (30) and the TPD assumption, we obtain, for any real $a \geq 0$,

$$P\left(\sum_{i \neq v} \hat{X}_i > a\right) = P\left(\sum_{i \neq v} X_i > a \mid X_v = 1\right) \geq P\left(\sum_{i \neq v} X_i > a\right) = P(W - X_v > a),$$

which is the desired result.

Thanks to Lemma 2, we may apply Proposition 3 with $s = p = 1$. Noting that, by (28),

$$E X_V = \sum_{i=1}^n p_i P(V = i) = \frac{1}{\lambda} \sum_{i=1}^n p_i^2,$$

we then obtain the following result.

Theorem 2. *If $E W = \lambda$ and the indicators $\{X_i : 1 \leq i \leq n\}$ are TPD, then*

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \left\{ E[W^2] + 2 \sum_{i=1}^n p_i^2 - \lambda(\lambda + 1) \right\}.$$

This bound is obtained (and applied) in Barbour *et al.* (1992, Corollary 2.C.4) under the condition of positive relation. See also Erhardsson (2005).

4.2. Approximate local dependence

Approximate local dependence is becoming a rather popular topic in probability. For works related to this idea, see, for example, Chen (1975), Barbour *et al.* (1992), and Chatterjee *et al.* (2005). We now derive an abstract Poisson approximation theorem by combining stochastic ordering with such an approach.

We say that the n indicators X_1, \dots, X_n are approximately locally negatively dependent (ALND) if there exist n nonnegative reals $\delta_1, \dots, \delta_n$ (of sum $\delta > 0$) and n random variables Y_1, \dots, Y_n on \mathbb{Z}^+ such that

$$E[X_i g(W_i - Y_i)] \leq \delta_i E[g(W_i - Y_i)], \quad 1 \leq i \leq n, \tag{31}$$

for all nonnegative, nondecreasing functions g . Similarly, X_1, \dots, X_n are said to be approximately locally positively dependent (ALPD) if

$$E[X_i g(W_i - Y_i)] \geq \delta_i E[g(W_i - Y_i)], \quad 1 \leq i \leq n, \tag{32}$$

for all nonnegative, nondecreasing functions g .

Define

$$\varepsilon = \sum_{i=1}^n E[X_i Y_i] \quad \text{and} \quad \varepsilon_* = \varepsilon + \sum_{i=1}^n \delta_i E[X_i + Y_i],$$

and let

$$c_\lambda = \frac{(\lambda + 1)(1 - e^{-\lambda})}{\lambda} + 2d_\lambda \quad \text{with} \quad d_\lambda = 1 \wedge \sqrt{\frac{2}{e\lambda}}.$$

Theorem 3. *If $E W = \lambda$ and the indicators $\{X_i : 1 \leq i \leq n\}$ are ALND, then*

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} (|\text{var}(W) - \lambda| + 2\varepsilon) + c_\lambda |\delta - \lambda|, \tag{33}$$

while if the indicators $\{X_i : 1 \leq i \leq n\}$ are ALPD then

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} (|\text{var}(W) - \lambda| + 2\varepsilon_*) + c_\lambda |\delta - \lambda|.$$

Before proving Theorem 3, we give an example of its application.

Example 4. We examine a variation of the classical birthday problem; see also Barbour *et al.* (1992). Suppose that we independently colour $N \geq 2$ points with one of m colours, each colour being chosen equiprobably. Let Γ be the set of all subsets $i \subseteq \{1, \dots, N\}$ of size 2. For $i \in \Gamma$, let Z_i be the indicator that the points indexed by i have the same colour. Moreover, suppose that we uniformly choose r of the $|\Gamma| = \binom{N}{2}$ pairs of points, independently of the colourings chosen. For $i \in \Gamma$, we let $\xi_i = 0$ if the pair of points indexed by i is chosen, and set $\xi_i = 1$ otherwise.

Set $W = \sum_{i \in \Gamma} Z_i \xi_i$. This counts the number of pairs of points with the same colour, excluding those r pairs of points we have chosen. In the case where $r = 0$, this corresponds to the classical birthday problem. A bound in the Poisson approximation of W in this case is given in Arratia *et al.* (1989, Example 2).

We observe that, for all $i, j \in \Gamma$, $E[Z_i] = m^{-1}$ and $E[Z_i Z_j] = m^{-2}$. Furthermore,

$$E[\xi_i] = \frac{\binom{N}{2} - r}{\binom{N}{2}} \quad \text{and} \quad E[\xi_i \xi_j] = \frac{\binom{N}{2} - r}{\binom{N}{2}} \left(\frac{\binom{N}{2} - r - 1}{\binom{N}{2} - 1} \right), \quad i \neq j.$$

Straightforward calculations then give

$$\lambda = E[W] = \frac{\binom{N}{2} - r}{m} \quad \text{and} \quad \lambda - \text{var}(W) = \frac{\binom{N}{2} - r}{m^2}.$$

Now, we write $W_i = W - Z_i \xi_i$ and choose

$$Y_i = \sum_{j \neq i} Z_j \xi_j \mathbf{1}_{\{i \cap j \neq \emptyset\}} \quad \text{and} \quad \delta_i = E[Z_i \xi_i].$$

Condition (31) holds true with these choices. Indeed, $W_i - Y_i$ is independent of Z_i and the ξ_i are negatively related by construction. Thus, for all nondecreasing functions g , we have

$$E[Z_i \xi_i g(W_i - Y_i)] = E[Z_i \xi_i] E[g(W_i - Y_i) \mid \xi_i = 1] \leq E[Z_i \xi_i] E[g(W_i - Y_i)],$$

as required. We further see that

$$\begin{aligned} \varepsilon &= \sum_{i \in \Gamma} E[Z_i \xi_i Y_i] \\ &= \sum_{i \in \Gamma} \sum_{j \neq i} E[Z_i Z_j] E[\xi_i \xi_j] \mathbf{1}_{\{i \cap j \neq \emptyset\}} \\ &= \frac{2(N-1) \{ \binom{N}{2} - r \} \{ \binom{N}{2} - r - 1 \}}{m^2 \{ \binom{N}{2} - 1 \}}. \end{aligned}$$

Evaluating (33) then gives the bound in Corollary 7 below.

Corollary 7. *With W as above,*

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{m} \left\{ 1 + 4(N-1) \left(\frac{\binom{N}{2} - r - 1}{\binom{N}{2} - 1} \right) \right\}.$$

In the $r = 0$ case, a bound of the same order was established in Arratia *et al.* (1989, Example 2).

4.3. Proof of Theorem 3

The ALND case. The method of proof combines the ideas given in Barbour *et al.* (1992), Goldstein and Rinott (1996), Papadatos and Papathanasiou (2002), and Reinert (2005).

We suppose first that f is any nonnegative, nondecreasing function. We thus have

$$\begin{aligned} E[Wf(W)] &= \sum_{i=1}^n E[X_i f(W)] \\ &= \sum_{i=1}^n E[X_i f(W_i + 1)] \\ &= \sum_{i=1}^n E[X_i f(W_i - Y_i + 1)] + \sum_{i=1}^n E[X_i \{f(W_i + 1) - f(W_i - Y_i + 1)\}], \end{aligned}$$

which we denote by $T_1 + T_2$. We bound the sum T_2 by noting that

$$|f(x) - f(y)| \leq \|\Delta f\|_\infty |x - y|,$$

which yields

$$T_2 \leq \|\Delta f\|_\infty \sum_{i=1}^n E(X_i Y_i) = \|\Delta f\|_\infty \varepsilon.$$

For the sum T_1 , by (31) and since f is nondecreasing, we obtain

$$T_1 \leq \sum_{i=1}^n \delta_i E[f(W_i - Y_i + 1)] \leq \sum_{i=1}^n \delta_i E[f(W + 1)] = \delta E[f(W + 1)].$$

Using these two bounds, we find that

$$\begin{aligned} E[Af(W)] &= \lambda E[f(W + 1)] - E[Wf(W)] \\ &\geq -(\delta - \lambda) E[f(W + 1)] - \|\Delta f\|_\infty \varepsilon. \end{aligned} \tag{34}$$

To obtain an upper bound, we define a function \tilde{f} on $\{0, 1, \dots, n - 1\}$ by

$$\tilde{f}(x) = \|f\|_\infty + \|\Delta f\|_\infty x - f(x). \tag{35}$$

Note that \tilde{f} is, as f , a nonnegative, nondecreasing function. By assumption, $E W = \lambda$ so that $E[A1] = 0$; observe also that $E[AW] = \lambda E[W + 1] - E[W^2] = -[\text{var}(W) - \lambda]$. Thus,

$$\begin{aligned} E[A\tilde{f}(W)] &= \|g\|_\infty E[A1] + \|\Delta f\|_\infty E[AW] - E[Af(W)] \\ &= -\|\Delta f\|_\infty [\text{var}(W) - \lambda] - E[Af(W)]. \end{aligned}$$

On the other hand, (34) is applicable to the function \tilde{f} , so that

$$E[A\tilde{f}(W)] \geq -(\delta - \lambda) E[\tilde{f}(W + 1)] - \|\Delta \tilde{f}\|_\infty \varepsilon.$$

From these two formulae, we deduce that

$$E[Af(W)] \leq \|\Delta \tilde{f}\|_\infty \varepsilon + (\delta - \lambda) E[\tilde{f}(W + 1)] + \|\Delta f\|_\infty |\text{var}(W) - \lambda|. \tag{36}$$

Now, let f be an arbitrary function. We start with the standard decomposition $f = f_+ - f_-$, where f_+ and f_- are nonnegative, nondecreasing functions with, of course,

$$\|\Delta^j f_+\|_\infty \leq \|\Delta^j f\|_\infty \quad \text{and} \quad \|\Delta^j f_-\|_\infty \leq \|\Delta^j f\|_\infty, \quad j = 0, 1. \tag{37}$$

By (34) and (36), we obtain the upper bound

$$\begin{aligned} E[Af(W)] &= E[Af_+(W)] - E[Af_-(W)] \\ &\leq \|\Delta \tilde{f}_+\|_\infty \varepsilon + (\delta - \lambda) E[\tilde{f}_+(W + 1)] + \|\Delta f_+\|_\infty |\text{var}(W) - \lambda| \\ &\quad + (\delta - \lambda) E[f_-(W + 1)] + \|\Delta f_-\|_\infty \varepsilon \\ &= \|\Delta f_+\|_\infty |\text{var}(W) - \lambda| + (\|\Delta \tilde{f}_+\|_\infty + \|\Delta f_-\|_\infty) \varepsilon \\ &\quad + (\delta - \lambda) \{\|f_+\|_\infty + \|\Delta f_+\|_\infty (\lambda + 1) - E[f(W + 1)]\}, \end{aligned}$$

where we have used (35) and the fact that $E W = \lambda$ for the last equality. By a similar method, we obtain the lower bound

$$\begin{aligned} E[Af(W)] &\geq -(\delta - \lambda) E[f_+(W + 1)] - \|\Delta f_+\|_\infty \varepsilon \\ &\quad - \|\Delta \tilde{f}_-\|_\infty \varepsilon - (\delta - \lambda) E[\tilde{f}_-(W + 1)] - \|\Delta f_-\|_\infty |\text{var}(W) - \lambda| \\ &= -\|\Delta f_-\|_\infty |\text{var}(W) - \lambda| - (\|\Delta f_+\|_\infty + \|\Delta \tilde{f}_-\|_\infty) \varepsilon \\ &\quad - (\delta - \lambda) \{\|f_-\|_\infty + \|\Delta f_-\|_\infty (\lambda + 1) + E[f(W + 1)]\}. \end{aligned}$$

By (37) and since $\|\Delta \tilde{f}\|_\infty \leq \|\Delta f\|_\infty$, combining the two previous bounds then yields

$$|E[Af(W)]| \leq \|\Delta f\|_\infty(|\text{var}(W) - \lambda| + 2\varepsilon) + |\delta - \lambda|[2\|f\|_\infty + \|\Delta f\|_\infty(\lambda + 1)]. \tag{38}$$

With $f = Sh$, it now suffices to apply the standard bounds in (38), i.e.

$$\|\Delta Sh\|_\infty \leq \lambda^{-1}(1 - e^{-\lambda})\|h\|_\infty \quad \text{and} \quad \|Sh\|_\infty \leq d_\lambda \|h\|_\infty,$$

which gives (33).

The ALPD case. This case is treated analogously to the ALND case. For f nonnegative and nondecreasing, we first write

$$\begin{aligned} E[Wf(W)] &= \sum_{i=1}^n E[X_i f(W_i - Y_i + 1)] + \sum_{i=1}^n E[X_i \{f(W_i + 1) - f(W_i - Y_i + 1)\}] \\ &\geq \sum_{i=1}^n E[X_i f(W_i - Y_i + 1)] - \|\Delta f\|_\infty \varepsilon. \end{aligned}$$

By (32), we then obtain

$$\begin{aligned} E[Wf(W)] &\geq \sum_{i=1}^n \delta_i E[f(W_i - Y_i + 1)] - \|\Delta f\|_\infty \varepsilon \\ &= \delta E[f(W + 1)] - \sum_{i=1}^n \delta_i E[f(W + 1) - f(W_i - Y_i + 1)] - \|\Delta f\|_\infty \varepsilon \\ &\geq \delta E[f(W + 1)] - \|\Delta f\|_\infty \sum_{i=1}^n \delta_i E(X_i + Y_i) - \|\Delta f\|_\infty \varepsilon \\ &= \delta E[f(W + 1)] - \|\Delta f\|_\infty \varepsilon_*. \end{aligned}$$

Overall, we find that

$$E[Af(W)] = \lambda E[f(W + 1)] - E[Wf(W)] \geq -(\delta - \lambda) E[f(W + 1)] + \|\Delta f\|_\infty \varepsilon_*.$$

The rest of the proof follows as in the ALND case.

5. Translated Poisson approximation

We assume, as in Section 4, that $W = X_1 + \dots + X_n$ is a sum of (possibly dependent) indicator random variables, with $p_i = P(X_i = 1)$. Define

$$\lambda_k = \sum_{i=1}^n p_i^k, \quad \lambda = \lambda_1 = E[W], \quad \text{and} \quad \sigma^2 = \text{var}(W).$$

We are going to discuss the approximation of W by a translated Poisson distribution.

5.1. Main results

A random variable Z has a translated Poisson distribution $TP(\lambda, \sigma^2)$ if Z is distributed as $Z' + \rho$, where $Z' \sim \text{Po}(\sigma^2 + \gamma)$ with

$$\rho = \lambda - \sigma^2 - \gamma \quad \text{and} \quad \gamma = \langle \lambda - \sigma^2 \rangle \in [0, 1),$$

where $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x .

We note that $E[Z] = \lambda$ and $\sigma^2 \leq \text{var}(Z) = \sigma^2 + \gamma < \sigma^2 + 1$, so that our approximating translated Poisson distribution has a mean equal to, and variance close to, that of W . We would thus expect a closer approximation than could be obtained by simply using the one-parameter Poisson distribution. The variances of W and Z cannot necessarily be made to match exactly, as we must shift our Poisson distribution by an integer. However, the error term arising from this mismatch does not adversely affect the order of the bounds we obtain, as we will see below.

The following results give us bounds in the translated Poisson approximation for W under some stochastic ordering assumptions. We defer the proof of Theorem 4 until Section 5.3, giving first some examples of its application in Section 5.2.

Our bounds demonstrate convergence to a translated Poisson distribution if $\sigma \rightarrow \infty$ as $n \rightarrow \infty$. Bounds on the total variation distance between $\mathcal{L}(W)$ and a translated Poisson random variable may still be found if this is not the case, but require a different analysis of the error terms. For example, in proving Theorem 4, we write $P(W - \rho < 0) \leq \sigma^{-2}$. This error term may be reduced, or even omitted altogether, depending on the problem at hand, with a more careful analysis. This could give us good bounds in the cases where $\sigma \rightarrow \sigma_\infty < \infty$ as $n \rightarrow \infty$.

In the sequel, we let W^s be a random variable having the W -size-biased distribution, and v_q be an indicator random variable, independent of all else, with $P(v_q = 1) = q$. As before, we write $W_i = W - X_i$, $1 \leq i \leq n$, and, for any random index V , we let $W_V = W - X_V$.

Theorem 4. *Suppose that X_1, \dots, X_n are positively related, and that there exist $q \in [0, 1]$ and $l \in \mathbb{Z}^+$ such that*

$$(W + 1 \mid X_k = 0) \geq_{\text{st}} (W - l - v_q \mid X_k = 1), \quad 1 \leq k \leq n. \tag{39}$$

Then

$$d_{\text{TV}}(\mathcal{L}(W), \text{TP}(\lambda, \sigma^2)) \leq \frac{2}{\sigma^2} + \frac{\lambda_2 + (l + q)(\lambda - \lambda_2)}{\lambda\sigma} + \frac{(l + 1)(l + 2q)(\lambda - \lambda_2)}{\sigma^2} d_{\text{TV}}(\mathcal{L}(W^s), \mathcal{L}(W^s + 1)). \tag{40}$$

Suppose instead that X_1, \dots, X_n are negatively related, and that there exist $q \in [0, 1]$ and $l \in \mathbb{Z}^+$ such that

$$(W + 1 \mid X_k = 0) \leq_{\text{st}} (W + l + v_q \mid X_k = 1), \quad 1 \leq k \leq n. \tag{41}$$

Then

$$d_{\text{TV}}(\mathcal{L}(W), \text{TP}(\lambda, \sigma^2)) \leq \frac{2}{\sigma^2} + \frac{\lambda_2 + (l + q)(\lambda - \lambda_2)}{\lambda\sigma} + \frac{l(l + 2q - 1)(\lambda - \lambda_2)}{\sigma^2} d_{\text{TV}}(\mathcal{L}(W^s), \mathcal{L}(W^s + 1)). \tag{42}$$

Consider the stochastic ordering assumptions (39) and (41). We note that the choice of the pair l, q is not unique, in that choosing $l = m$ and $q = 1$ gives the same assumption as choosing $l = m + 1$ and $q = 0$. It is easily checked, however, that each of these choices gives rise to the same bounds in (40) and (42). In the examples below, we will verify the validity of such stochastic orderings by using an appropriate coupling argument.

5.2. Applications

Example 5. Suppose that X_1, \dots, X_n are independent. Thus, they are also negatively related. Moreover, condition (41) is true for $q = l = 0$. Therefore, (42) is applicable and yields the following result.

Corollary 8. *With W as above,*

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \leq \frac{\lambda_2}{\lambda\sigma} + \frac{2}{\sigma^2}.$$

This bound is of the order we would expect, see also Čekanavičius and Vaitkus (2001).

Example 6. Suppose that m balls are placed into N urns in such a way that no urn contains more than one ball and all arrangements are equally likely. Let W be the number of balls in the first n urns. Thus, W has a hypergeometric distribution with

$$\lambda = \frac{mn}{N} \quad \text{and} \quad \sigma^2 = \frac{mn(N - m)(N - n)}{(N - 1)N^2}.$$

We set X_i to be the indicator that the i th urn contains a ball, so that $W = X_1 + \dots + X_n$. By construction, these indicators are negatively related. Condition (41) holds for $q = 1$ and $l = 0$. To see this, we construct $(W + 1 \mid X_k = 0)$ by considering the N urns and excluding the k th urn. Distribute the m balls in these $N - 1$ urns, such that all arrangements are equally likely, and count the number of the first n urns that are occupied. Adding one to this count gives us our random variable. We then choose (uniformly and independently of what has gone before) one of the occupied urns. Take the ball from this urn and place it in urn k . This gives $(W + 1 \mid X_k = 1)$. If the ball chosen is from one of the first n urns, the number of occupied urns is the same as before. Otherwise, we have increased the number of occupied urns within the first n . Evaluating bound (42) then gives Corollary 9 below.

Corollary 9. *For W having our hypergeometric distribution,*

$$\begin{aligned} d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) &\leq \frac{1}{\sigma} + \frac{2}{\sigma^2} \\ &= \sqrt{\frac{N^2(N - 1)}{mn(N - m)(N - n)}} + \frac{2N^2(N - 1)}{mn(N - m)(N - n)}. \end{aligned}$$

Röllin (2007, Section 4.1) considered the translated Poisson approximation for the hypergeometric distribution, and showed that if $m = O(n)$ and $N = O(n)$ then one obtains a bound in the total variation distance of order $O(1/\sqrt{n})$. This order is also reflected in our result.

Example 7. Suppose that ξ_1, \dots, ξ_n are i.i.d. Bernoulli random variables with

$$p = P(\xi_i = 1) = 1 - P(\xi_i = 0), \quad 1 \leq i \leq n.$$

Fix an integer $k \geq 2$, and define

$$X_i = \xi_i \xi_{i+1} \dots \xi_{i+k-1} \quad \text{and} \quad W = \sum_{i=1}^n X_i,$$

in which, to avoid edge effects, all indices are treated modulo n . Thus, W counts the number

of k -runs in our Bernoulli trials. Observe that

$$\lambda = np^k, \quad \lambda_2 = np^{2k}, \quad \text{and} \quad \sigma^2 = \frac{np^k}{1-p}(1+p-p^k[2+(2k-1)(1-p)]).$$

The translated Poisson approximation for k -runs was treated in Röllin (2005, Section 3.2), who gave a bound in the total variation distance of the form K/\sqrt{n} for some constant $K = K(k, p)$ independent of n . Barbour and Xia (1999, Section 5) also gave a bound of this order for 2-runs. We will use our Theorem 4 to give an explicit bound with this same order.

It is easily seen that the variables X_1, \dots, X_n are positively related. Condition (39) holds by choosing $q = 1$ and $l = 2k - 3$. To see this, consider the following construction. Given the Bernoulli random variables ξ_1, \dots, ξ_n , fix some $m \leq n$ and set $\xi_m = \xi_{m+1} = \dots = \xi_{m+k-1} = 1$, while the others remain independent Bernoulli random variables with parameter p . Counting the number of k -runs in these n Bernoulli trials gives $(W \mid X_m = 1)$. Suppose now that we resample the random variables $\xi_m, \dots, \xi_{m+k-1}$, conditional on at least one of these being 0. Counting the number of k -runs now gives $(W \mid X_m = 0)$. In this resampling procedure, we can remove at most $2k - 1$ of the k -runs that were originally present. Thus, our construction implies that

$$(W \mid X_m = 0) + 2k - 1 \geq (W \mid X_m = 1),$$

or, equivalently,

$$(W + 1 \mid X_m = 0) \geq (W - 2k + 2 \mid X_m = 1),$$

and, hence, the announced values of q and l .

Following the work of Section 4, to construct W^s , we choose an index V uniformly from $\{1, \dots, n\}$, and set $\xi_V = \xi_{V+1} = \dots = \xi_{V+k-1} = 1$, while the other ξ_i remain independent Bernoulli random variables with parameter p . Lemma 2.1 of Wang and Xia (2008) thus gives

$$d_{TV}(\mathcal{L}(W^s), \mathcal{L}(W^s + 1)) \leq 1 \wedge \frac{2.3}{\sqrt{(n-k-1)p^k(1-p)^3}}.$$

Using this, Theorem 4 yields the following result.

Corollary 10. *Let W count the number of k -runs in n independent Bernoulli trials, each with success probability p . Then,*

$$\begin{aligned} & d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \\ & \leq \frac{2}{\sigma^2} + \frac{p^k + (2k-2)(1-p^k)}{\sigma} \\ & \quad + \frac{(2k-2)(2k-1)np^k(1-p^k)}{\sigma^2} \left(1 \wedge \frac{2.3}{\sqrt{(n-k-1)p^k(1-p)^3}} \right). \end{aligned} \tag{43}$$

Our bound (43) has the same order as that of Röllin (2005, Theorem 5) and Barbour and Xia (1999, Theorem 5.2) (this latter result applying only to the 2-runs case). Numerical comparison of the bounds shows that ours generally performs well compared to these other bounds, often giving a better result. In Table 1 we give some illustrations, with values for comparison taken from Röllin (2005).

TABLE 1: Numerical comparisons for 2-runs. Upper bounds on the total variation distance from (a) our result (43), (b) Röllin (2005), and (c) Barbour and Xia (1999). Missing values are due to restrictions on the choice of parameters.

n	Bound	P				
		0.10	0.25	0.50	0.75	0.90
10^6	(a)	0.1553	0.0675	0.0500	0.0814	0.2512
	(b)	0.4463	0.2334	0.1747	0.5528	>1
	(c)	0.0304	—	0.1251	0.6014	—
10^8	(a)	0.0155	0.0067	0.0050	0.0081	0.0251
	(b)	0.0445	0.0233	0.0175	0.0553	0.2554
	(c)	0.0030	—	0.0125	0.0601	—
10^{10}	(a)	0.0016	0.0007	0.0005	0.0008	0.0025
	(b)	0.0045	0.0023	0.0017	0.0055	0.0255
	(c)	0.0003	—	0.0013	0.0060	—

5.3. Proof of Theorem 4

Our proof is based on that of Propositions 2 and 3, using the characterising operator for the Poisson distribution. We find representations of our Stein equation in conjunction with which our dependence and stochastic ordering assumptions may be applied.

Throughout this section, we let $f = Sh$ be the solution to the Stein equation (2) with the choices $\alpha_j = \sigma^2 + \gamma$ and $\beta_j = j$, corresponding to the Poisson distribution with mean $\sigma^2 + \gamma$. We suppose that the test function h has the form $h(j) = \mathbf{1}_{\{j \in B\}}$ for some $B \subseteq \mathbb{Z}^+$. We write $g_B(j) = f(j - \rho)$. We note that g_B depends on the choice of the set B , though for notational convenience we will often simply write g for g_B . We note further that bounds on the supremum norm of f also apply to g , so that in particular $\|\Delta g_B\|_\infty \leq \sigma^{-2}$ for each $B \subseteq \mathbb{Z}^+$.

Following Röllin (2007, Section 3), from the Stein equation we obtain

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \leq \sup_{B \subseteq \mathbb{Z}^+} |E[(\sigma^2 + \gamma)g_B(W + 1) - (W - \rho)g_B(W)]| + P(W - \rho < 0). \tag{44}$$

Using Chebyshev’s inequality, we can write

$$P(W - \rho < 0) = P(W - \lambda < -\sigma^2 - \gamma) \leq P(|W - \lambda| > \sigma^2 + \gamma) \leq P(|W - \lambda| > \sigma^2) \leq \sigma^{-2}.$$

So, we now concentrate on the first term on the right-hand side of (44). Throughout our proof, we will make use of the following equalities in distribution:

$$(W \mid X_V = 1) \stackrel{D}{=} W^s \quad \text{and} \quad (W_V \mid X_V = 0) \stackrel{D}{=} (W \mid X_V = 0). \tag{45}$$

Step 1. For this part of the proof, we will establish the following bound, which may be of some independent interest.

Proposition 5. *With the above notation,*

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \leq (\lambda - \lambda_2) \sup_{B \subseteq \mathbb{Z}^+} \{\Lambda_B\} + \frac{\lambda_2}{\lambda\sigma} + \frac{2}{\sigma^2}, \tag{46}$$

where

$$\Lambda_B = \mathbb{E} \sum_{j=0}^{\infty} |\Delta g_B(j) - \Delta g_B(W)| \times |\mathbb{P}(W_V + 1 > j \mid X_V = 0) - \mathbb{P}(W_V + 1 > j \mid X_V = 1)|. \quad (47)$$

To prove this result, we separately consider the cases where $\sigma^2 \leq \lambda$ and $\sigma^2 \geq \lambda$. We begin by assuming that $\sigma^2 \leq \lambda$, so that $\rho \geq 0$. Recall that

$$\mathbb{E}[Wg(W)] = \lambda \mathbb{E}[g(W^s)]. \quad (48)$$

Using (48), we can then write

$$\mathbb{E}[(\sigma^2 + \gamma)g(W + 1) - (W - \rho)g(W)] = \lambda \mathbb{E}[g(\tilde{W}) - g(W^s)], \quad (49)$$

where

$$\mathbb{P}(\tilde{W} = j) = \lambda^{-1}[(\sigma^2 + \gamma)\mathbb{P}(W + 1 = j) + \rho \mathbb{P}(W = j)], \quad j \geq 0.$$

That is, $\tilde{W} = W + v_r$, where v_r is a Bernoulli variable with success probability $r = \lambda^{-1}(\sigma^2 + \gamma)$. Note that $r \leq 1$ by assumption. We rewrite (49) as

$$\lambda \mathbb{E}[g(\tilde{W}) - g(W^s)] = \lambda \mathbb{E}[g(\tilde{W}) - g(\bar{W})] + \lambda \mathbb{E}[g(\bar{W}) - g(W^s)], \quad (50)$$

by defining $\bar{W} = W_V + 1$, where V is a random index chosen according to (28). For the first term in (50), we note that, by conditioning on v_r ,

$$\lambda \mathbb{E}g(\tilde{W}) = \lambda \mathbb{E}g(W + v_r) = (\sigma^2 + \gamma) \mathbb{E} \Delta g(W) + \lambda \mathbb{E}g(W). \quad (51)$$

Furthermore, by conditioning on X_V and using the equalities in (45),

$$\lambda \mathbb{E}g(\bar{W}) = \lambda \mathbb{E}g(W_V + 1) = \lambda_2 \mathbb{E}g(W^s) + (\lambda - \lambda_2) \mathbb{E}[g(W) \mid X_V = 0], \quad (52)$$

since $\mathbb{P}(X_V = 1) = \lambda^{-1}\lambda_2$. Again, by considering conditioning on X_V and using (45), we have

$$(\lambda - \lambda_2) \mathbb{E}[g(W) \mid X_V = 0] = \lambda \mathbb{E}g(W + 1) - \lambda_2 \mathbb{E}g(W^s + 1). \quad (53)$$

Combining (51), (52), and (53), we obtain

$$\begin{aligned} \lambda \mathbb{E}[g(\tilde{W}) - g(\bar{W})] &= (\sigma^2 + \gamma - \lambda) \mathbb{E} \Delta g(W) + \lambda_2 \mathbb{E} \Delta g(W^s) \\ &= \lambda_2 \mathbb{E}[\Delta g(W^s) - \Delta g(W)] + \gamma \mathbb{E} \Delta g(W) \\ &\quad + (\sigma^2 - \lambda + \lambda_2) \mathbb{E} \Delta g(W). \end{aligned} \quad (54)$$

Now consider the second term in (50). Let us combine it with the final term in (54). Since

$$\mathbb{E}[\bar{W} - W^s] = -\lambda^{-1}(\sigma^2 - \lambda + \lambda_2),$$

and proceeding as we did in deriving (3), we obtain

$$\begin{aligned} \lambda \mathbb{E}[g(\tilde{W}) - g(W^s)] + (\sigma^2 - \lambda + \lambda_2) \mathbb{E} \Delta g(W) \\ = \lambda \mathbb{E} \sum_{j=0}^{\infty} (\Delta g(j) - \Delta g(W)) [\mathbb{P}(\bar{W} > j) - \mathbb{P}(W^s > j)]. \end{aligned} \quad (55)$$

Using the definition of \bar{W} , conditioning on X_V , and employing (45), we have

$$\begin{aligned} & \lambda[\mathbb{P}(\bar{W} > j) - \mathbb{P}(W^s > j)] \\ &= (\lambda - \lambda_2)[\mathbb{P}(W_V + 1 > j \mid X_V = 0) - \mathbb{P}(W_V + 1 > j \mid X_V = 1)]. \end{aligned} \tag{56}$$

Hence, the right-hand side of (55) becomes

$$\begin{aligned} & (\lambda - \lambda_2) \mathbb{E} \sum_{j=0}^{\infty} (\Delta g(j) - \Delta g(W)) [\mathbb{P}(W_V + 1 > j \mid X_V = 0) \\ & \qquad \qquad \qquad - \mathbb{P}(W_V + 1 > j \mid X_V = 1)]. \end{aligned} \tag{57}$$

Substituting representations (54) and (57) into (49), then substituting in turn in (44) yields

$$\begin{aligned} d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) & \leq (\lambda - \lambda_2) \sup_{B \subseteq \mathbb{Z}^+} \{\Lambda_B\} + \lambda_2 \sup_{B \subseteq \mathbb{Z}^+} |\mathbb{E}[\Delta g_B(W^s) - \Delta g_B(W)]| \\ & \qquad \qquad \qquad + \gamma \sup_{B \subseteq \mathbb{Z}^+} |\mathbb{E} \Delta g_B(W)| + \mathbb{P}(W - \rho < 0), \end{aligned}$$

where Λ_B is given by (47).

Recalling that $\mathbb{P}(W - \rho < 0) \leq \sigma^{-2}$, $\gamma \leq 1$, and $\|\Delta g_B\|_{\infty} \leq \sigma^{-2}$, we have

$$\gamma \sup_{B \subseteq \mathbb{Z}^+} |\mathbb{E} \Delta g_B(W)| + \mathbb{P}(W - \rho < 0) \leq 2\sigma^{-2}.$$

Furthermore, the random variable W^s having the W -size-biased distribution satisfies

$$\mathbb{P}(W^s = j) = \lambda^{-1} j \mathbb{P}(W = j), \quad 0 \leq j \leq n,$$

and so

$$2d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s)) = \sum_{j=0}^{\infty} |\mathbb{P}(W = j) - \mathbb{P}(W^s = j)| = \mathbb{E}|1 - \lambda^{-1}W| \leq \lambda^{-1}\sigma. \tag{58}$$

We thus have

$$\lambda_2 |\mathbb{E}[\Delta g_B(W^s) - \Delta g_B(W)]| \leq 2\lambda_2 \|\Delta g_B\|_{\infty} d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s)) \leq \frac{\lambda_2}{\lambda\sigma}.$$

Combining the above bounds, we obtain the desired result.

In the second step of the proof, we consider how Λ_B can be bounded. Before doing this, we show that if $\sigma^2 \geq \lambda$ then bound (46) continues to hold.

Consider now the case where $\sigma^2 \geq \lambda$, so that $\rho \leq 0$. We will use an analogous argument to show that bound (46) continues to hold. In place of (50), we write

$$\begin{aligned} & \mathbb{E}[(\sigma^2 + \gamma)g(W + 1) - (W - \rho)g(W)] \\ &= (\sigma^2 + \gamma) \mathbb{E}[g(W + 1) - g(\hat{W})] + (\sigma^2 + \gamma) \mathbb{E}[g(\hat{W}) - g(W^*)], \end{aligned} \tag{59}$$

where $\hat{W} = W + v_t(1 - X_V)$, $W^* = v_t W^s + (1 - v_t)W$, and $t = \lambda(\sigma^2 + \gamma)^{-1}$. Consider the first term on the right-hand side of (59). For this term, we argue as we did to derive (54). Conditioning on v_t and X_V , and employing the equalities in (45), we find, as for (54), that

$$\begin{aligned} & (\sigma^2 + \gamma) \mathbb{E}[g(W + 1) - g(\hat{W})] \\ &= \lambda_2 \mathbb{E}[\Delta g(W^s) - \Delta g(W)] + \gamma \mathbb{E} \Delta g(W) + (\sigma^2 - \lambda + \lambda_2) \mathbb{E} \Delta g(W). \end{aligned}$$

As we have

$$E[\hat{W} - W^*] = -(\sigma^2 + \gamma)^{-1}(\sigma^2 - \lambda + \lambda_2),$$

we then write

$$\begin{aligned} & (\sigma^2 + \gamma) E[g(\hat{W}) - g(W^*)] + (\sigma^2 - \lambda + \lambda_2) E \Delta g(W) \\ &= (\sigma^2 + \gamma) E \sum_{j=0}^{\infty} (\Delta g(j) - \Delta g(W)) [P(\hat{W} > j) - P(W^* > j)]. \end{aligned} \tag{60}$$

Using the definitions of \hat{W} and W^* , and conditioning on v_t , we find that

$$P(\hat{W} > j) - P(W^* > j) = t[P(\bar{W} > j) - P(W^s > j)].$$

Comparing this with (55), recalling the definition of t , and using (56), we find that (57) also gives us a representation of (60). Continuing the argument as before, bound (46) holds too in the present case.

Step 2. In this part of the proof we bound Λ_B , defined in (47), and, thus, obtain the bounds of our theorem. In doing so, we will use our stochastic ordering and dependence assumptions. The cases where X_1, \dots, X_n are positively and negatively related will be discussed separately. In the positively related case, the argument of Lemma 2 shows that

$$P(W_V + 1 > j \mid X_V = 0) - P(W_V + 1 > j \mid X_V = 1) \leq 0, \quad j \geq 0.$$

Noting that $(W_V + 1 \mid X_V = 1) \stackrel{D}{=} W^s$, we fix some $l \in \mathbb{Z}^+$ and write

$$\begin{aligned} & P(W_V + 1 > j \mid X_V = 1) - P(W_V + 1 > j \mid X_V = 0) \\ &= P(W_V + 1 > j + l \mid X_V = 1) - P(W_V + 1 > j \mid X_V = 0) \\ &+ \sum_{i=1}^l P(W^s = j + i). \end{aligned} \tag{61}$$

Suppose now that there exists some $q \in [0, 1]$ such that, for each $j \geq 0$,

$$\begin{aligned} & P(W_V + 1 > j + l \mid X_V = 1) - P(W_V + 1 > j \mid X_V = 0) \\ &\leq q P(W_V = j + l \mid X_V = 1) \end{aligned} \tag{62}$$

$$= q P(W^s = j + l + 1). \tag{63}$$

We will show presently that this is implied by the stochastic ordering assumption (39). Using (61) and (63), we find that

$$\begin{aligned} \Lambda_B &\leq q E |\Delta g_B(W^s - l - 1) - \Delta g_B(W)| + \sum_{i=1}^l E |\Delta g_B(W^s - i) - \Delta g_B(W)| \\ &\leq 2q \|\Delta g_B\|_{\infty} d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s - l - 1)) \\ &+ 2\|\Delta g_B\|_{\infty} \sum_{i=1}^l d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s - i)). \end{aligned} \tag{64}$$

Using our bound on $\|\Delta g_B\|_\infty$ and the triangle inequality for the total variation distance, the first term in (64) is bounded by

$$2q\sigma^{-2}\{d_{TV}(\mathcal{L}(W), \mathcal{L}(W^S)) + (l + 1)d_{TV}(\mathcal{L}(W^S), \mathcal{L}(W^S + 1))\} \leq 2q\sigma^{-2}\left\{\frac{\sigma}{2\lambda} + (l + 1)d_{TV}(\mathcal{L}(W^S), \mathcal{L}(W^S + 1))\right\}, \tag{65}$$

where this last inequality uses (58). Similarly, the second term in (64) can be bounded by

$$2\sigma^{-2}\sum_{i=1}^l\{d_{TV}(\mathcal{L}(W), \mathcal{L}(W^S)) + id_{TV}(\mathcal{L}(W^S), \mathcal{L}(W^S + 1))\} \leq \sigma^{-2}\left\{\frac{l\sigma}{\lambda} + l(l + 1)d_{TV}(\mathcal{L}(W^S), \mathcal{L}(W^S + 1))\right\}. \tag{66}$$

Combining (64), (65), and (66) with bound (46) yields the desired inequality (40).

So, the proof of the first part of Theorem 4 is completed upon showing that the stochastic ordering condition (39) implies inequality (62). Writing

$$P(W_V = j + l \mid X_V = 1) = P(W_V + 1 > j + l \mid X_V = 1) - P(W_V > j + l \mid X_V = 1)$$

for $0 \leq j \leq n$, it can be seen that (62) is equivalent to

$$P(W_V + 1 > j \mid X_V = 0) \geq (1 - q)P(W_V + 1 - l > j \mid X_V = 1) + qP(W_V - l > j \mid X_V = 1)$$

for $j \geq 0$. This, in turn, is equivalent to the stochastic ordering

$$(W + 1 \mid X_V = 0) \succeq_{st} (1 - v_q)(W - l \mid X_V = 1) + v_q(W - l - 1 \mid X_V = 1), \tag{67}$$

which can be seen using (45). Some rearranging shows that the stochastic ordering assumption (39) implies the stochastic ordering (67); hence, we obtain the result of Theorem 4 in the positively related case.

We now turn our attention to the case of a negative relation, and complete the proof of that portion of Theorem 4. When X_1, \dots, X_n are negatively related, we can use a similar argument to the above. We have

$$P(W_V + 1 > j \mid X_V = 0) - P(W_V + 1 > j \mid X_V = 1) \geq 0, \quad 0 \leq j \leq n.$$

Analogously to the positively related case, we write, for some fixed $l \in \mathbb{Z}^+$,

$$P(W_V + 1 > j \mid X_V = 0) - P(W_V + 1 > j \mid X_V = 1) = P(W_V + 1 > j \mid X_V = 0) - P(W_V + 1 > j - l \mid X_V = 1) + \sum_{i=0}^{l-1} P(W^S = j - i).$$

This time, we suppose that there exists $q \in [0, 1]$ such that

$$P(W_V + 1 > j \mid X_V = 0) - P(W_V + 1 > j \mid X_V = 1) \leq qP(W_V + 1 + l = j \mid X_V = 1). \tag{68}$$

Following a similar argument to that used in the case of a positive relation, we find that

$$\Lambda_B \leq \frac{l + q}{\lambda\sigma} + \frac{l(l + 2q - 1)}{\sigma^2}d_{TV}(\mathcal{L}(W^S), \mathcal{L}(W^S + 1)).$$

Combining this with (46) gives us the desired inequality (42). It remains to show that the stochastic ordering assumption (41) implies inequality (68), which can be achieved as above.

6. Another abstract approximation theorem

Our aim hereafter is to consider an alternative approximation theorem which can be found within the present framework. For concreteness, we suppose that the birth rates α_j and death rates β_j are such that the random variable π has two parameters. This will be the case in the application presented later.

Let us return to the basic representation (13). To choose the two parameters of π , it seems natural, in our context, to consider $s = 2$ and introduce the two conditions $\alpha = \beta$ and $E W_\alpha = E W_\beta$ (i.e. $E[\alpha_W(W + 1)] = E[\beta_W W]$). With these choices, representation (13) becomes

$$E h(W) - E h(\pi) = \alpha \sum_{i=0}^{\infty} \Delta^2 f(i) E[(W_\alpha - i - 1)_+ - (W_\beta - i - 1)_+]. \tag{69}$$

Moreover, suppose that we can construct W_α and W_β on the same probability space in such a way that $W_\beta = W_\alpha + Y$ for some random variable Y which takes values in the set $\{-1, 0, 1\}$. Under this assumption, $E[W_\alpha] = E[W_\beta] = E[W_\alpha + Y]$, which implies that $E[Y] = 0$. It is easily seen that representation (69) can be rewritten as

$$\begin{aligned} E h(W) - E h(\pi) &= -\alpha \sum_{i=0}^{\infty} \Delta^2 f(i) E[Y \mathbf{1}_{\{W_\alpha - 1 \geq i + 1\}} + Y_+ \mathbf{1}_{\{W_\alpha - 1 = i\}}] \\ &= -\alpha E[\mathbf{1}_{\{Y=1\}} \Delta^2 f(W_\alpha - 1) + Y \Delta f(W_\alpha - 1)]. \end{aligned} \tag{70}$$

Noting that

$$|E[\mathbf{1}_{\{Y=1\}} \Delta^2 f(W_\alpha - 1)]| \leq 2 \|\Delta f\|_\infty d_{TV}(\mathcal{L}(W_\alpha), \mathcal{L}(W_\alpha + 1)) \sup_w \{P(Y = 1 \mid W_\alpha = w)\}$$

and

$$|E[Y \Delta f(W_\alpha - 1)]| \leq \|\Delta f\|_\infty E|E[Y \mid W_\alpha]| \leq \|\Delta f\|_\infty \sqrt{\text{var}(E[Y \mid W_\alpha])},$$

we can immediately bound the right-hand side of (70) to obtain the following result.

Proposition 6. *Suppose that $\alpha = \beta$ and $E W_\alpha = E W_\beta$. If W_α and W_β can be constructed on the same probability space such that*

$$W_\beta = W_\alpha + Y \text{ for some random variable } Y \text{ valued in } \{-1, 0, 1\}, \tag{71}$$

then, for bounded $h: \mathbb{Z}^+ \mapsto \mathbb{R}$,

$$\begin{aligned} |E h(W) - E h(\pi)| &\leq 2\alpha \|\Delta S h\|_\infty d_{TV}(\mathcal{L}(W_\alpha), \mathcal{L}(W_\alpha + 1)) \sup_w \{P(Y = 1 \mid W_\alpha = w)\} \\ &\quad + \alpha \|\Delta S h\|_\infty \sqrt{\text{var}(E[Y \mid W_\alpha])}. \end{aligned} \tag{72}$$

Clearly, if such a random variable Y takes values on a bounded set other than $\{-1, 0, 1\}$, a representation analogous to (70) may still be found, and a result analogous to Proposition 6 is available. We now apply Proposition 6 to approximate a sum of independent indicator random variables.

Example 8. Suppose that $W = X_1 + \dots + X_n$ is the sum of independent Bernoulli random variables with success probabilities $p_i, 1 \leq i \leq n$. Brown and Xia (2001, Section 3) showed that in this case we can improve on a Poisson or binomial approximation for W by using a

so-called polynomial birth–death distribution, with the choices $\alpha_j = \alpha$ and $\beta_j = \gamma j + j(j - 1)$ for some constants α and γ .

We will follow that approach and choose here α and γ such that $\alpha = \beta$ and $E[\alpha_W(W + 1)] = E[\beta_W W]$. Straightforward computations then give us expressions for these parameters:

$$\gamma = \lambda^2 \lambda_2^{-1} - 1 - 2\lambda + 2\lambda_3 \lambda_2^{-1} \quad \text{and} \quad \alpha = \gamma \lambda + \lambda^2 - \lambda_2. \tag{73}$$

Here $\lambda_k = \sum_{i=1}^n p_i^k$ and $\lambda = \lambda_1 = E[W]$ (as in Section 5). Note that the parameter choices (73) are the same as those employed in Brown and Xia (2001), who based their selection on minimising the error bound obtained in their result.

To begin with, let us prove that condition (71) is satisfied. Since the birth rate is constant (as in the Poisson case), we again have $W_\alpha = W + 1$. Let us turn our attention to W_β . We let $W_i = W - X_i$ and $W_{i,j} = W - X_i - X_j$, $0 \leq i, j \leq n$, and observe that $W(W - 1) = \sum_{1 \leq i \neq j \leq n} X_i X_j$. By the definition of W_β , we obtain

$$\begin{aligned} P(W_\beta = k) &= \alpha^{-1} E[(\gamma W + W(W - 1)) \mathbf{1}_{\{W=k\}}] \\ &= \alpha^{-1} \left[\gamma \sum_{i=1}^n p_i P(W_i + 1 = k) + \sum_{1 \leq i \neq j \leq n} p_i p_j P(W_{i,j} + 2 = k) \right] \end{aligned}$$

for $1 \leq k \leq n$. In the spirit of the size-biasing construction of Section 4, we now define two random indices $T, U \in \{1, \dots, n\}$ chosen according to the distribution

$$P(T = i, U = j) = \frac{p_i p_j}{\lambda^2 - \lambda_2}, \quad i \neq j, \quad P(T = U = i) = 0.$$

Recall also the random index V given in (28). Combining these definitions with the above, we can write

$$P(W_\beta = k) = \alpha^{-1} \gamma \lambda P(W + 1 - X_V = k) + \alpha^{-1} (\lambda^2 - \lambda_2) P(W + 2 - X_T - X_U = k)$$

for $1 \leq k \leq n$. Let $q = \alpha^{-1} \gamma \lambda$; note from (73) that $0 \leq q \leq 1$ whenever $\gamma \geq 0$. In the sequel we will assume that this is indeed the case. Introduce a Bernoulli random variable v_q with success probability q , independent of all other entries. We can then write

$$W_\beta = v_q(W + 1 - X_V) + (1 - v_q)(W + 2 - X_T - X_U) = W + 1 + Y = W_\alpha + Y,$$

where

$$Y = (1 - v_q)(1 - X_T - X_U) - v_q X_V, \tag{74}$$

Y being valued in $\{-1, 0, 1\}$ with $E[Y] = 0$, as desired.

Now let us evaluate bound (72). First, we need a bound on the solution f of the Stein equation in this situation. By Theorem 2.10 of Brown and Xia (2001), we know that

$$\sup\{\|\Delta S h\|_\infty : h(j) = \mathbf{1}_{\{j \in B\}}, B \subseteq \mathbb{Z}^+\} \leq \alpha^{-1}. \tag{75}$$

Furthermore, W being a sum of independent indicators, we have (from Barbour and Jensen (1989, Lemma 1))

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(W + 1)) \leq \frac{1}{2\sqrt{\sum_{i=1}^n p_i(1 - p_i)}}. \tag{76}$$

Finally, consider the two conditional terms in (72). Note from (74) that $Y = 1$ if and only if $v_q = X_T = X_U = 0$, so that

$$\begin{aligned} P(Y = 1 \mid W) &= (1 - q) P(X_T = X_U = 0 \mid W) \\ &= (1 - q) E[(1 - X_T)(1 - X_U) \mid W] \\ &= \alpha^{-1} \sum_{1 \leq i \neq j \leq n} p_i p_j E[(1 - X_i)(1 - X_j) \mid W]. \end{aligned}$$

This probability takes its greatest value when $W = 0$, with $E[(1 - X_i)(1 - X_j) \mid W = 0] = 1$ for all i and j . Hence,

$$\sup_w \{P(Y = 1 \mid W = w)\} = \alpha^{-1} \sum_{1 \leq i \neq j \leq n} p_i p_j = \alpha^{-1}(\lambda^2 - \lambda_2). \tag{77}$$

Now, let $\|Z\| = (E[Z^2])^{1/2}$ be the L_2 norm for any random variable Z . Since $T \stackrel{D}{=} U$ and $E[Y] = 0$, we write

$$E[Y \mid W] = -q(E[X_V \mid W] - E[X_V]) - 2(1 - q)(E[X_T \mid W] - E[X_T]),$$

and, thus,

$$\begin{aligned} \sqrt{\text{var}(E[Y \mid W])} &= \|E[Y \mid W]\| \\ &\leq q \sum_{j=1}^n \|E[X_j \mid W] - E[X_j]\| P(V = j) \\ &\quad + 2(1 - q) \sum_{j=1}^n \|E[X_j \mid W] - E[X_j]\| P(T = j) \\ &\leq (q + 2(1 - q)) \max_{1 \leq j \leq n} \sqrt{\text{var}(E[X_j \mid W])}. \end{aligned}$$

When $p_j = p$ for $j = 1, \dots, n$, $E[X_j \mid W] = W/n$ and so the bound becomes the equality

$$\sqrt{\text{var}(E[Y \mid W])} = (2 - q) \sqrt{\text{var}\left(\frac{W}{n}\right)}. \tag{78}$$

Inserting (75), (76), (77), and (78) into (72) then yields the bound

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq \frac{p}{(1 - p)\sigma} + \frac{(2 - q)\sigma}{n} = O\left(\frac{p}{\sqrt{\lambda}}\right),$$

where $\sigma^2 = \text{var}(W) = np(1 - p)$.

By exploring the explicit structure of the auxiliary variable Y , it is possible to derive better bounds. Throughout this part, we let $\bar{a} = 1 - a$ for any $a \in \mathbb{R}$ and $\sigma_k = \sqrt{\sum_{i=k+1}^n \rho_i}$, where ρ_i is the i th largest number of $p_1(1 - p_1), \dots, p_n(1 - p_n)$. From Barbour and Jensen (1989, Lemma 1) we have, for all $i, j = 1, \dots, n$ and $i \neq j$,

$$2d_{\text{TV}}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1)) \leq \sigma_1^{-1} \quad \text{and} \quad 2d_{\text{TV}}(\mathcal{L}(W_{i,j}), \mathcal{L}(W_{i,j} + 1)) \leq \sigma_2^{-1}.$$

Note that, from representation (74),

$$\mathbf{1}_{\{Y=1\}} = \bar{v}_q \bar{X}_T \bar{X}_U, \quad \mathbf{1}_{\{Y=-1\}} = v_q X_V + \bar{v}_q X_T X_U. \tag{79}$$

The derivations below are based on the conditional independences of X_T and W_T , given T , X_U and W_U , given U , and X_V and W_V , given V . By substituting (79) into (70), integrating with respect to v_q , separating linear and quadratic terms, and noting that $T \stackrel{D}{=} U$, we derive, after some simple calculations,

$$\begin{aligned} I &= E h(W) - E h(\pi) \\ &= -\alpha E[\bar{v}_q \bar{X}_T \bar{X}_U \Delta f(W + 1)] + \alpha E[(v_q X_V + \bar{v}_q X_T X_U) \Delta f(W)] \\ &= -\alpha \bar{q} E[X_T X_U \Delta^2 f(W)] + 2\alpha \bar{q} E[X_T \Delta^2 f(W)] \\ &\quad - \alpha (\bar{q} E[\Delta f(W + 1)] - E[(2\bar{q} X_T + q X_V) \Delta f(W)]) \\ &= I_1 + I_2 + I_3. \end{aligned}$$

Using the conditional independence of $W_{T,U}$ and X_T, X_U given T and U , the first term I_1 is bounded by

$$\begin{aligned} |I_1| &= \alpha \bar{q} |E E[X_T X_U \mid T, U] E[\Delta^2 f(W_{T,U} + 2)]| \\ &\leq 2\alpha \bar{q} \|\Delta f\|_\infty E[X_T X_U] \max_{i \neq j} \{d_{TV}(\mathcal{L}(W_{i,j}), \mathcal{L}(W_{i,j} + 1))\} \\ &\leq \frac{\lambda_2^2 - \lambda_4}{\alpha \sigma_2}. \end{aligned}$$

By conditioning on T ,

$$\begin{aligned} |I_2| &= 2\alpha \bar{q} |E E[X_T \mid T] E[\Delta^2 f(W_T + 1)]| \\ &\leq 4\alpha \bar{q} \|\Delta f\|_\infty E[X_T] \max_i \{d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1))\} \\ &\leq \frac{2(\lambda \lambda_2 - \lambda_3)}{\alpha \sigma_1}. \end{aligned}$$

To bound I_3 , we first note that, since $E[Y] = 0$,

$$\bar{q} = 2\bar{q} E[X_T] + q E[X_V].$$

Thus,

$$\begin{aligned} |I_3| &= |2\alpha \bar{q} E[X_T (E[\Delta f(W_T + 1) \mid T] - E[\Delta f(W_T + X_T + 1)])] \\ &\quad + \alpha q E[X_V (E[\Delta f(W_V + 1) \mid V] - E[\Delta f(W_V + X_V + 1)])]| \\ &\leq 2\alpha \{2\bar{q} E[\{E[X_T \mid T]\}^2] + q E[\{E[X_V \mid V]\}^2]\} \\ &\quad \times \|\Delta f\|_\infty \max_i \{d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1))\} \\ &\leq \frac{2(\lambda \lambda_3 - \lambda_4)}{\alpha \sigma_1} + \frac{\gamma \lambda_3}{\alpha \sigma_1}. \end{aligned}$$

By combining the bounds on I_1, I_2 , and I_3 we derive the following result.

Proposition 7. *With W and π as above,*

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq \frac{\lambda_2^2 - \lambda_4}{\alpha \sigma_2} + \frac{2(\lambda \lambda_2 - \lambda_3)}{\alpha \sigma_1} + \frac{2(\lambda \lambda_3 - \lambda_4)}{\alpha \sigma_1} + \frac{\gamma \lambda_3}{\alpha \sigma_1}. \tag{80}$$

Let us conclude by comparing our result with that of Brown and Xia (2001, Theorem 3.1), who obtained

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq \frac{\gamma\lambda_3}{\alpha\sigma_1} + \frac{2\lambda\lambda_2}{\alpha\sigma_2}. \quad (81)$$

When $p_i = p \rightarrow 0$ for each i and $\lambda \rightarrow \infty$, both the bounds (80) and (81) are asymptotically equivalent to $3p^2/\sqrt{\lambda}$.

Acknowledgements

Fraser Daly gratefully acknowledges the financial support of the Schweizerische Nationalfonds, and thanks the Belgian Fonds National de la Recherche Scientifique for support during a visit to the Université Libre de Bruxelles. Thanks are also due to Andrew Barbour for several useful discussions and to an anonymous referee who made several suggestions which greatly improved the presentation of this work.

References

- ARRATIA, R., GOLDSTEIN, L. AND GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17**, 9–25.
- BARBOUR, A. D. AND CHEN, L. H. Y. (eds) (2005). *An Introduction to Stein's Method* (Lecture Notes Ser., Inst. Math. Sci. **4**), Singapore University Press.
- BARBOUR, A. D. AND JENSEN, J. L. (1989). Local and tail approximations near the Poisson limit. *Scand. J. Statist.* **16**, 75–87.
- BARBOUR, A. D. AND PUGLIESE, A. (2000). On the variance-to-mean ratio in models of parasite distributions. *Adv. Appl. Prob.* **32**, 701–719.
- BARBOUR, A. D. AND XIA, A. (1999). Poisson perturbations. *ESAIM Prob. Statist.* **3**, 131–150.
- BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation*. Oxford University Press.
- BROWN, T. C. AND PHILLIPS, M. J. (1999). Negative binomial approximation with Stein's method. *Methodology Comput. Appl. Prob.* **1**, 407–421.
- BROWN, T. C. AND XIA, A. (2001). Stein's method and birth–death processes. *Ann. Prob.* **29**, 1373–1403.
- ČEKANAČIUS, V. AND VAITKUS, P. (2001). Centered Poisson approximation via Stein's method. *Lithuanian Math. J.* **41**, 319–329.
- CHATTERJEE, S., DIACONIS, P. AND MECKES, E. (2005). Exchangeable pairs and Poisson approximation. *Prob. Surveys* **2**, 64–106.
- CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Prob.* **3**, 534–545.
- CHEN, L. H. Y., GOLDSTEIN, L. AND SHAO Q.-M. (2011). *Normal Approximation by Stein's Method*. Springer, Heidelberg.
- DENUIT, M. AND LEFÈVRE, C. (1997). Some new classes of stochastic order relations among arithmetic random variables, with applications in actuarial sciences. *Insurance Math. Econom.* **20**, 197–213.
- EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Prob. Lett.* **11**, 7–16.
- EICHELSBACHER, P. AND REINERT, G. (2008). Stein's method for discrete Gibbs measures. *Ann. Appl. Prob.* **18**, 1588–1618.
- ERHARDSSON, T. (2005). Stein's method for Poisson and compound Poisson approximation. In *An Introduction to Stein's Method* (Lecture Notes Ser., Inst. Math. Sci. **4**), eds A. D. Barbour and L. H. Y. Chen, Singapore University Press, pp. 61–113.
- GOLDSTEIN, L. AND RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.
- HOLMES, S. (2004). Stein's method for birth and death chains. In *Stein's Method: Expository Lectures and Applications* (IMS Lecture Notes Monogr. Ser. **46**), eds P. Diaconis and S. Holmes, Beachwood, OH, pp. 45–67.
- JOHNSON, N. L., KOTZ, S. AND KEMP, A. W. (1992). *Univariate Discrete Distributions*. John Wiley, New York.
- KAMAE, T., KRENGEL, U. AND O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Prob.* **5**, 899–912.
- LEFÈVRE, C. AND UTEV, S. (1996). Comparing sums of exchangeable Bernoulli random variables. *J. Appl. Prob.* **33**, 285–310.
- PAPADATOS, N. AND PAPATHANASIOU, V. (2002). Poisson approximation for a sum of dependent indicators: an alternative approach. *Adv. Appl. Prob.* **34**, 609–625.
- PEKÖZ, E. A. (1996). Stein's method for geometric approximation. *J. Appl. Prob.* **33**, 707–713.

- PHILLIPS, M. J. AND WEINBERG, G. V. (2000). Non-uniform bounds for geometric approximation. *Statist. Prob. Lett.* **49**, 305–311.
- REINERT, G. (2005). Three general approaches to Stein's method. In *An Introduction to Stein's Method* (Lecture Notes Ser., Inst. Math. Sci. **4**), eds A. D. Barbour and L. H. Y. Chen, Singapore University Press, pp. 183–221.
- RÖLLIN, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.
- RÖLLIN, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Prob.* **17**, 1596–1614.
- SHAKED, M. AND SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. Springer, New York.
- STEIN, C. (1986). *Approximate Computation of Expectations* (IMS Lecture Notes Monogr. Ser. **7**). Institute of Mathematical Statistics, Hayward, CA.
- WANG, X. AND XIA, A. (2008). On negative binomial approximation to k -runs. *J. Appl. Prob.* **45**, 456–471.