Bernoulli **19**(4), 2013, 1243–1267
DOI: 10.3150/12-BEJSP18

Approximating dependent rare events

LOUIS H. Y. CHEN[1] and ADRIAN RÖLLIN[2]

[1]*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076. E-mail: matchyl@nus.edu.sg*
[2]*Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546. E-mail: adrian.roellin@nus.edu.sg*

In this paper we give a historical account of the development of Poisson approximation using Stein's method and present some of the main results. We give two recent applications, one on maximal arithmetic progressions and the other on bootstrap percolation. We also discuss generalisations to compound Poisson approximation, Poisson process approximation and multivariate Poisson approximation, and state a few open problems.

Keywords: Bernoulli random variables; bootstrap percolation; compound Poisson approximation; local dependence; maximal arithmetic progressions; monotone coupling; multivariate Poisson approximation; Poisson approximation; Poisson process approximation; rare events; size-bias coupling; Stein's method

1. Introduction

The Poisson limit theorem as commonly found in textbooks of probability states that the number of successes in n independent trials converges in distribution to a Poisson distribution with mean $\lambda > 0$ if the maximum of the success probabilities tends to 0 and their sum converges to λ. The case where the trials have equal success probabilities was implicitly proved by Abraham de Moivre (1712) in his solution to the problem of finding the number of trials that gives an even chance of getting k successes. However, it was Siméon-Denis Poisson (1837) who first gave an explicit form of the Poisson distribution and proved the limit theorem for independent trials with equal success probabilities, that is, for the binomial distribution. The Poisson distribution was not much used before Ladislaus von Bortkiewicz (1898) expounded its mathematical properties and statistical usefulness.

In his book *Ars Conjectandi*, published posthumously in 1713, Jacob Bernoulli (1654–1705) considered games of chance and urn models with two possible outcomes and proved what is now known as the weak law of large numbers. He stressed that the probability of winning a game or of drawing a ball of a particular color from an urn (with replacement) remains the same when the game or the drawing of a ball is repeated. This has led to the use of the term Bernoulli trials to represent independent trials with the same probability of success. Representing success by 1 and failure by 0, a random variable taking values 0 and 1 is called a Bernoulli random variable. However, in this article, a set or a sequence of Bernoulli random variables need not be independent nor take the value 1 with equal probabilities. Also, if the success probability of a Bernoulli random variable is small, the event corresponding to success is called *rare*.

The Poisson limit theorem suggests that the distribution of a sum of independent Bernoulli random variables with small success probabilities can be approximated by the Poisson distribution

1350-7265

with the same mean if the success probabilities are small and the number of random variables is large. A measure of the accuracy of the approximation is the total variation distance. For two distributions P and Q over $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, the total variation distance between them is defined by

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

which is also equal to

$$\frac{1}{2} \sup_{|h|=1} \left| \int h \, dP - \int h \, dQ \right| = \frac{1}{2} \sum_{i \in \mathbb{Z}} \bigl|P\{i\} - Q\{i\}\bigr|.$$

For the binomial distribution $\mathrm{Bi}(n, p)$, Prohorov (1953) proved that

$$d_{\mathrm{TV}}\bigl(\mathrm{Bi}(n, p), \mathrm{Po}(np)\bigr) \le p\left[\frac{1}{\sqrt{2\pi e}} + \mathrm{O}\left(1 \wedge \frac{1}{\sqrt{np}} + p\right)\right],$$

where $\mathrm{Po}(np)$ denotes the Poisson distribution with mean np. Here, following Barbour, Holst and Janson (1992), the formulation corrects a minor error in the original paper. This result is remarkable in that the approximation is good so long as p is small, regardless of how large np is.

The result of Prohorov was generalised by Le Cam (1960) to sums of independent Bernoulli random variables $X_1, \dots, X_n$ with success probabilities $p_1, \dots, p_n$ that are not necessarily equal. Let $W = \sum X_i$ and $\lambda = \sum p_i$. Using the method of convolution operators, Le Cam (1960) obtained the error bounds

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{Po}(\lambda)\bigr) \le \sum_{i=1}^{n} p_i^2, \tag{1.1}$$

and

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{Po}(\lambda)\bigr) \le \frac{8}{\lambda} \sum_{i=1}^{n} p_i^2 \qquad \text{if } \max_{1 \le i \le n} p_i \le \frac{1}{4}. \tag{1.2}$$

In terms of order, the bound in (1.1) is better than that in (1.2) if $\lambda < 1$ and vice versa if $\lambda \ge 1$. Combining (1.1) and (1.2), one obtains a bound of the order $(1 \wedge \lambda^{-1}) \sum p_i^2$, which is small so long as $\max p_i$ is small, regardless of how large λ is. This form of the error bound has become the characteristic of Poisson approximation in subsequent developments of the subject.

In this article we will discuss the use of Stein's ideas in the Poisson approximation to the distributions of sums of dependent Bernoulli random variables, its historical development, applications, and some generalisations and open problems. The article is not intended to be a survey paper but an exposition with a focus on explaining Stein's ideas and presenting some results and recent applications. The references are not exhaustive but contain only those papers that are relevant to the objective of this article.

This paper is organised as follows. Section 2 is a brief introduction to Stein's method. Section 3 gives a brief overview of two approaches to Poisson approximation using Stein's method, and

Sections 4 and 5 discuss the developments of these two approaches. Section 6 is devoted to two recent applications of Poisson approximation and Section 7 discusses three generalisations of Poisson approximation.

2. Stein's method

In his seminal 1972 paper published in the Sixth Berkeley Symposium, Charles Stein introduced a new method of normal approximation. The method did not involve Fourier analysis but hinged on the solution of a differential equation. Although the method was developed for normal approximation, Stein's ideas were very general and the method was modified by Chen (1975) for Poisson approximation. Since then the method has been constantly developed and applied to many approximations beyond normal and Poisson and in finite as well as infinite dimensional spaces. It has been applied in many areas including computational biology, computer science, combinatorial probability, random matrices, reliability and many more. The method, together with its applications, continues to grow and remains a very active research area. See, for example, Stein (1986), Arratia, Goldstein and Gordon (1990), Barbour, Holst and Janson (1992), Diaconis and Holmes (2004), Barbour and Chen (2005a, 2005b), Chatterjee, Diaconis and Meckes (2005), Chen, Goldstein and Shao (2011), Ross (2011), Shih (2011), Nourdin and Peccati (2012).

In a nutshell, Stein's method can be described as follows. Let W and Z be random elements taking values in a space $\mathcal{S}$ and let $\mathcal{X}$ and $\mathcal{Y}$ be some classes of real-valued functions defined on $\mathcal{S}$. In approximating the distribution $\mathcal{L}(W)$ of W by the distribution $\mathcal{L}(Z)$ of Z, we write $\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}Lf_h(W)$ for a test function $h \in \mathcal{Y}$, where L is a linear operator (Stein operator) from $\mathcal{X}$ into $\mathcal{Y}$ and $f_h \in \mathcal{X}$ a solution of the equation

$$Lf = h - \mathbb{E}h(Z) \qquad \text{(Stein equation).}$$

The error $\mathbb{E}Lf_h(W)$ can then be bounded by studying the solution f_h and exploiting the probabilistic properties of W. The operator L characterises $\mathcal{L}(Z)$ in the sense that $\mathcal{L}(W) = \mathcal{L}(Z)$ if and only if for a sufficiently large class of functions f we have

$$\mathbb{E}Lf(W) = 0 \qquad \text{(Stein identity).}$$

In normal approximation, where $\mathcal{L}(Z)$ is the standard normal distribution, the operator used by Stein (1972) is given by $Lf(w) = f'(w) - wf(w)$ for $w \in \mathbb{R}$, and in Poisson approximation, where $\mathcal{L}(Z)$ is the Poisson distribution with mean $\lambda > 0$, the operator L used by Chen (1975) is given by $Lf(w) = \lambda f(w+1) - wf(w)$ for $w \in \mathbb{Z}_+$. However the operator L is not unique even for the same approximating distribution but depends on the problem at hand. For example, for normal approximation L can also be taken to be the generator of the Ornstein–Uhlenbeck process, that is, $Lf(w) = f''(w) - wf'(w)$, and for Poisson approximation, L taken to be the generator of an immigration-death process, that is, $Lf(w) = \lambda[f(w+1) - f(w)] + w[f(w-1) - f(w)]$. This generator approach, which is due to Barbour (1988), allows extensions to multivariate and process settings. Indeed, for multivariate normal approximation, $Lf(w) = \Delta f(w) - w \cdot \nabla f(w)$, where f is defined on the Euclidean space; see Barbour (1990) and Götze (1991), and also Reinert and Röllin (2009) and Meckes (2009).

3. Poisson approximation

In Poisson approximation, the main focus has been on bounding the total variation distance between the distribution of a sum of dependent Bernoulli random variables and the Poisson distribution with the same mean. One of the main objectives has been to obtain a bound which is the "correct" generalisation of the bound obtained by Le Cam (1960), specifically, one with the multiplicative factor $1 \wedge \lambda^{-1}$.

Broadly speaking, there are two main approaches to Poisson approximation using Stein's method, the local approach and the size-bias coupling approach. The local approach was first studied by Chen (1975) and developed further by Arratia, Goldstein and Gordon (1989, 1990), who presented Chen's results in a form which is easy to use and applied them to a wide range of problems including problems in extreme values, random graphs and molecular biology. The size-bias coupling approach dates back to Barbour (1982) in his work on Poisson approximation for random graphs. Barbour, Holst and Janson (1992) presented a systematic development of monotone couplings, and applied their results to random graphs and many combinatorial problems. A recent review of Poisson approximation by Chatterjee, Diaconis and Meckes (2005) used Stein's method of exchangeable pairs to study classical problems in combinatorial probability. They also reviewed a size-bias coupling of Stein [(1986), p. 93] for any set of dependent Bernoulli random variables.

4. The local approach

The operator L given by $Lf(w) = f'(w) - wf(w)$ for $w \in \mathbb{R}$, which was used by Stein (1972) for normal approximation, is constructed by showing that $\mathbb{E}\{f'(Z) - Zf(Z)\} = 0$ for all bounded absolutely continuous functions f if $Z \sim \mathrm{N}(0, 1)$. This identity is proved by integration by parts. As a discrete counterpart, the operator L given by $Lf(w) = \lambda f(w+1) - wf(w)$ for $w \in \mathbb{Z}_+$, which was used by Chen (1975) for Poisson approximation, is constructed by showing that $\mathbb{E}\{\lambda f(Z) - Zf(Z)\} = 0$ for all bounded real-valued functions f if $Z \sim \mathrm{Po}(\lambda)$, using summation by parts.

Using the Stein equation

$$\lambda f(w+1) - wf(w) = h(w) - \mathbb{E}h(Z), \tag{4.1}$$

where $|h| = 1$ and Z has the Poisson distribution with mean $\lambda > 0$, Chen (1975) developed Stein's method for Poisson approximation for sums of φ-mixing sequences of Bernoulli random variables $X_1, \dots, X_n$ with success probabilities $p_1, \dots, p_n$. When specialised to independent Bernoulli random variables, his results yield

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{Po}(\lambda)\bigr) \le 3\biggl(1 \wedge \frac{1}{\sqrt{\lambda}}\biggr) \sum_{i=1}^{n} p_i^2$$

and

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{Po}(\lambda)\bigr) \le \frac{5}{\lambda} \sum_{i=1}^{n} p_i^2,$$

where $W = \sum X_i$. These results improve slightly those of Le Cam (1960).

Chen's proofs depend crucially on the bounds he obtained on the solution of (4.1) and its smoothness. These bounds were improved by Barbour and Eagleson (1983), who proved that for $h = I_A$, $A \subset \mathbb{Z}_+$,

$$\|f_h\|_\infty \le 1 \wedge \frac{1.4}{\sqrt{\lambda}} \tag{4.2}$$

and

$$\|\Delta f_h\|_\infty \le \frac{1 - e^{-\lambda}}{\lambda} \le 1 \wedge \frac{1}{\lambda}, \tag{4.3}$$

where $\|\cdot\|_\infty$ denotes the supremum norm and $\Delta f(w) = f(w+1) - f(w)$.

It is perhaps instructive to see how easily Le Cam's results, with substantially smaller constants, can be proved by Stein's method using (4.3).

Let W be the sum of independent Bernoulli random variables $X_1, \dots, X_n$ with success probabilities $p_1, \dots, p_n$, and let $W^{(i)} = W - X_i$ for $i = 1, \dots, n$. For any bounded real-valued function f,

$$\begin{aligned}
\mathbb{E}\{\lambda f(W+1) - W f(W)\} &= \sum_{i=1}^n \mathbb{E}\{p_i f(W+1) - X_i f(W)\} \\
&= \sum_{i=1}^n p_i \mathbb{E}\{f(W+1) - f(W^{(i)}+1)\} \\
&= \sum_{i=1}^n p_i \mathbb{E}\{X_i \Delta f(W^{(i)}+1)\} \\
&= \sum_{i=1}^n p_i^2 \mathbb{E}\Delta f(W^{(i)}+1).
\end{aligned}$$

By choosing $f = f_h$, a bounded solution of (4.1), where $h = I_A$ and $A \subset \mathbb{Z}_+$, we obtain

$$\begin{aligned}
d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{Po}(\lambda)) &= \sup_{A \subset \mathbb{Z}_+} |\mathbb{P}[W \in A] - \mathbb{P}[Z \in A]| \\
&\le \|\Delta f_h\|_\infty \sum_{i=1}^n p_i^2 \le \left(1 \wedge \frac{1}{\lambda}\right) \sum_{i=1}^n p_i^2.
\end{aligned} \tag{4.4}$$

We wish to remark that the solution f_h is unique except at $w = 0$, but the value of f_h at $w = 0$ is never used in the calculation. So it has been conveniently set to be 0.

The above proof of (4.4) is given in Barbour and Hall (1984), who also proved that

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{Po}(\lambda)) \ge \frac{1}{32}\left(1 \wedge \frac{1}{\lambda}\right) \sum_{i=1}^n p_i^2.$$

This shows that $(1 \wedge \lambda^{-1}) \sum p_i^2$ is of the best possible order for the Poisson approximation. Indeed, it has been proved by Deheuvels and Pfeifer (1986), using a semigroup approach, and also by Chen and Choi (1992) and Barbour, Chen and Choi (1995), using Stein's method, that $d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda))$ is asymptotic to $\sum p_i^2$ (respectively $(2\pi e)^{-1/2}\lambda^{-1} \sum p_i^2$) as $\max p_i \to 0$ and $\lambda \to 0$ (respectively $\lambda \to \infty$).

We end this section by stating a theorem of Arratia, Goldstein and Gordon [(1989, 1990), Theorem 1], which was proved using (4.2) and (4.3).

Theorem 4.1. *Let $\{X_\alpha : \alpha \in J\}$ be Bernoulli random variables with success probabilities p_α, $\alpha \in J$. Let $W = \sum_{\alpha \in J} X_\alpha$ and $\lambda = \mathbb{E}W = \sum_{\alpha \in J} p_\alpha$. Then, for any collection of sets $B_\alpha \subset J$, $\alpha \in J$,*

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{Po}(\lambda)\bigr) \le \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) + \left(1 \wedge \frac{1.4}{\sqrt{\lambda}}\right) b_3$$

and

$$\bigl|\mathbb{P}[W = 0] - e^{-\lambda}\bigr| \le \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2 + b_3),$$

where

$$b_1 = \sum_{\alpha \in J} \sum_{\beta \in B_\alpha} p_\alpha p_\beta, \qquad b_2 = \sum_{\alpha \in J} \sum_{\beta \in B_\alpha \setminus \{\alpha\}} \mathbb{E}(X_\alpha X_\beta),$$
$$b_3 = \sum_{\alpha \in J} \bigl|\mathbb{E}(X_\alpha | X_\beta, \beta \notin B_\alpha) - p_\alpha\bigr|.$$

If for each $\alpha \in J$, X_α is independent of $\{X_\beta : \beta \notin B_\alpha\}$, then $b_3 = 0$, and we call $\{X_\alpha : \alpha \in J\}$ locally dependent with dependence neighbourhoods $\{B_\alpha : \alpha \in J\}$. An m-dependent sequence of random variables, which is a special case of a φ-mixing sequence, is locally dependent.

The wide applicability of Theorem 4.1 is illustrated through many examples in Arratia, Goldstein and Gordon (1989, 1990). Many problems to which Theorem 4.1 is applied are concerned with locally dependent random variables.

5. The size-bias coupling approach

In his monograph, Stein [(1986), pp. 89–93] considered the following general problem of Poisson approximation. Let $X_1, \ldots, X_n$ be dependent Bernoulli random variables with success probabilities $p_i = \mathbb{P}[X_i = 1]$ for $i = 1, \ldots, n$. Let $W = \sum X_i$ and let $\lambda = \mathbb{E}W$ with $\lambda > 0$. Assume I to be uniformly distributed over $\{1, \ldots, n\}$ and independent of $X_1, \ldots, X_n$. Then for any bounded real-valued function f defined on $\{0, 1, \ldots, n\}$,

$$\mathbb{E}\bigl\{W f(W)\bigr\} = \lambda \mathbb{E}\bigl(f(W) \mid X_I = 1\bigr). \tag{5.1}$$

If W^* and W are defined on the same probability space such that the distribution of W^* equals the conditional distribution of W given $X_I = 1$, then (5.1) becomes

$$\mathbb{E}\{Wf(W)\} = \lambda \mathbb{E} f(W^*),$$

from which one obtains

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{Po}(\lambda)\big) \le \big(1 - e^{-\lambda}\big)\mathbb{E}\big|W + 1 - W^*\big|. \tag{5.2}$$

From (5.2), one can see that if the distribution of $W + 1$ is close to that of W^*, then the distribution of W is approximately Poisson with mean λ, and (5.2) gives an upper bound on the total variation distance.

This approach to Poisson approximation was reviewed in Chatterjee, Diaconis and Meckes (2005), who also applied (5.2) to a variety of problems, such as the matching problem, the coupon-collector's problem and the birthday problem.

In their monograph, Barbour, Holst and Janson (1992) studied Poisson approximation for Bernoulli random variables satisfying monotone coupling assumptions. We state their main theorem in this context as follows.

Theorem 5.1. *Let $\{X_\alpha : \alpha \in J\}$ be Bernoulli random variables with success probabilities p_α, $\alpha \in J$. Suppose for each $\alpha \in J$, there exists $\{Y_{\beta,\alpha} : \beta \in J\}$ defined on the same probability space as $\{X_\alpha : \alpha \in J\}$ such that*

$$\mathcal{L}\big(\{Y_{\beta,\alpha} : \beta \in J\}\big) = \mathcal{L}\big(\{X_\alpha : \alpha \in J \mid X_\alpha = 1\}\big).$$

Let $W = \sum X_\alpha$, $\lambda = \mathbb{E}W = \sum p_\alpha$, and $Z \sim \mathrm{Po}(\lambda)$.

1. *If*

$$Y_{\beta,\alpha} \le X_\beta \qquad \text{for all } \beta \in J \text{ (negatively related)}, \tag{5.3}$$

then

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{Po}(\lambda)\big) \le (1 \wedge \lambda)\left(1 - \frac{\mathrm{Var}(W)}{\lambda}\right). \tag{5.4}$$

2. *If*

$$Y_{\beta,\alpha} \ge X_\beta \qquad \text{for all } \beta \in J \text{ (positively related)}, \tag{5.5}$$

then

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{Po}(\lambda)\big) \le (1 \wedge \lambda)\left(\frac{\mathrm{Var}(W)}{\lambda} - 1 + \frac{2}{\lambda}\sum_{\alpha \in J} p_\alpha^2\right). \tag{5.6}$$

From (5.4) and (5.6), one can see that $\mathcal{L}(W)$ is approximately $\mathrm{Po}(\lambda)$ if $\mathrm{Var}(W)/\lambda$ is close to 1.

The proof of Theorem 5.1 is pretty similar to that for (5.2). Let $V_\alpha = \sum_{\beta\neq\alpha} Y_{\beta,\alpha}$ and $W^{(\alpha)} = W - X_\alpha$ for $\alpha \in J$. Then for any bounded real-valued function f defined on $\{0, 1, \dots, |J|\}$,

$$\begin{aligned} \mathbb{E}\{Wf(W)\} &= \sum_{\alpha\in J} p_\alpha \mathbb{E}\bigl(f\bigl(W^{(\alpha)}+1\bigr) \mid X_\alpha = 1\bigr) \\ &= \sum_{\alpha\in J} p_\alpha \mathbb{E} f(V_\alpha + 1) = \lambda \mathbb{E} f(V_I + 1), \end{aligned} \tag{5.7}$$

where I is independent of all the X_α and V_α, and $\mathbb{P}[I = \alpha] = p_\alpha/\lambda$, $\alpha \in J$.

Using the monotone properties (5.3) and (5.5), one gets

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{Po}(\lambda)\bigr) \le \bigl(1 - e^{-\lambda}\bigr)\mathbb{E}\bigl[(W+1) - (V_I + 1)\bigr]$$

for the negatively related case, and

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{Po}(\lambda)\bigr) \le \bigl(1 - e^{-\lambda}\bigr)\bigl(\mathbb{E}X_I + \mathbb{E}\bigl[(V_I+1) - \bigl(W^{(I)}+1\bigr)\bigr]\bigr)$$

for the positively related case. Straightforward calculations then yield (5.4) and (5.6).

Barbour, Holst and Janson (1992) also established conditions for existence of monotone couplings and applied Theorem 5.1 to large number of problems in random permutations, random graphs, occupancy and urn models, spacings, and exceedances and extremes.

The coupling approach of Stein (1986) and of Barbour, Holst and Janson (1992) can actually be formulated under the general framework of size-bias coupling. Here is the definition of size-biased distribution; see Goldstein and Rinott (1996).

Definition 5.1. *Let W be a non-negative random variable with mean $\lambda > 0$. We say that W^s has the W-*size biased distribution *if*

$$\mathbb{E}\{Wf(W)\} = \lambda \mathbb{E} f\bigl(W^s\bigr)$$

for all real-valued functions f such that the expectations exist.

If W is a non-negative integer-valued random variable, then $\mathbb{P}[W^s = k] = k\mathbb{P}[W = k]/\lambda$ for $k \ge 1$. The following theorem follows immediately.

Theorem 5.2. *Let W be a non-negative integer valued random variable with $\mathbb{E}W = \lambda > 0$. Assume that W^s and W are defined on the same probability space, that is, assume that there is size-bias coupling. Then we have*

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{Po}(\lambda)\bigr) \le \bigl(1 - e^{-\lambda}\bigr)\mathbb{E}\bigl|W + 1 - W^s\bigr|.$$

Note that in the case where W is a sum of Bernoulli random variables, W^s can be taken to be W^* in (5.2) or $V_I + 1$ in (5.7). Furthermore, it is clear from Theorem 5.2 that the Poisson distribution is the only distribution such that its size-biased distribution is the original distribution shifted by one.

We conclude by saying that a large portion of the literature on the coupling approach to Poisson approximation falls under the general framework of size-bias coupling. Indeed, (5.7) provides a general way for constructing size-bias coupling for sums of Bernoulli random variables. Couplings involving the size-biased distribution, however, have found applications beyond Poisson approximation; see for example Peköz and Röllin (2011) and Peköz, Röllin and Ross (2013).

6. Applications

A remarkable feature of Theorem 4.1 and Theorem 5.1 is that the error bounds depend only on the first two moments of the random variables. It also happens that many interesting scientific problems can be formulated as occurrences of dependent rare events. For example, one is often interested in the maximum of a set of random variables $\xi_1, \dots, \xi_n$. For a threshold t, define $X_i = I[\xi_i > t]$ for $i = 1, \dots, n$, and let $W = \sum X_i$. Then

$$P[\max \xi_i \le t] = \mathbb{P}[W = 0]. \tag{6.1}$$

Often t is large, so that $\{\xi_1 > t\}, \dots, \{\xi_n > t\}$ are rare events. If the $X_1, \dots, X_n$ satisfy the conditions of Theorem 4.1 or Theorem 5.1 and the error bound is small, then

$$P[\max \xi_i \le t] \approx e^{-\lambda_t} \qquad \text{where } \lambda_t = \sum_{i=1}^{n} P[\xi_i > t].$$

Since the appearance of Theorems 4.1 and 5.1, Poisson approximation has been applied to a large number of problems in many different fields, which include computational biology, random graphs and large-scale networks, computer science, statistical physics, epidemiology, reliability theory, game theory, and financial mathematics. In computational biology, Poisson approximation is typically used to calculate p-values in sequence comparison, while in random graphs, it is used to count the copies of a small graph in a large graph. Here is a sample of publications on problems in different fields, in which Poisson approximation is applied: Dembo, Karlin and Zeitouni (1994), Neuhauser (1994), Waterman and Vingron (1994), Waterman (1995), Embrechts, Klüppelberg and Mikosch (1997), Karlin and Chen (2000), Barbour and Reinert (2001), Lange (2002), Lippert, Huang and Waterman (2002), Grimmett and Janson (2003), Franceschetti and Meester (2006), Hart, Rinott and Weiss (2008), Draief and Massoulié (2010), Falk, Hüsler and Reiss (2011).

In what follows, we will present two recent applications of Poisson approximation, one by Benjamini, Yadin and Zeitouni (2007) on maximal arithmetic progressions, and the other by Bollobás *et al.* (2013) on bootstrap percolation.

6.1. Maximal arithmetic progressions

The occurrences of arithmetic progressions in subsets of the set of positive integers are of interest in number theory. Tao (2007) gave a historic account of the topic, in particular, Szemerédi's theorem, which states that any "dense" subset of positive integers must contain arbitrarily long arithmetic progressions.

Benjamini, Yadin and Zeitouni (2007, 2012) analyse the following probabilistic variant of arithmetic progressions. Let $\xi_1, \dots, \xi_n$ be i.i.d. Bernoulli random variables with success probability $0 < p < 1$. We say that there is an arithmetic progression of length at least t, starting at $a + s$ with a common difference s, if $\xi_a = 0$ and $\xi_{a+s} = \xi_{a+2s} = \cdots = \xi_{a+ts} = 1$ as long as $a + ts \le n$. Let U_n denote the length of the maximal arithmetic progression among $\xi_1, \dots, \xi_n$. We have the following result.

Theorem 6.1 [Benjamini, Yadin and Zeitouni (2007)]. *Let $x \in \mathbb{R}$ be fixed and let $0 \le \delta_n < 1$ be such that $x - \frac{2\log n}{\log p} + \frac{\log\log n}{\log p} - \delta_n$ is integer valued. Then*

$$\mathbb{P}\left[U_n + \frac{2\log n}{\log p} - \frac{\log\log n}{\log p} < x - \delta_n\right] \sim \exp\left(\frac{(1-p)p^{x-\delta_n}\log p}{4}\right) \tag{6.2}$$

as $n \to \infty$.

Note that the distribution of U_n is of Gumbel-type. However, the rounding effect of δ_n does not vanish, since U_n is integer-valued and, as one can show, the variance of U_n is of order 1. Therefore, limiting distributions only exist along subsequences $n_1, n_2, \dots$ for which $\lim_{m\to\infty} \delta_{n_m}$ exists, in which case the limiting distribution is a discretised Gumbel distribution.

Idea of proof. Denote by $\mathcal{I}_{n,t}$ the set of pairs (a, s) of positive integers that satisfy $a + ts \le n$, and for each such $(a, s) \in \mathcal{I}_{n,t}$ define

$$X_{a,s} = \mathrm{I}[\xi_a = 0, \xi_{a+s} = \xi_{a+2s} = \cdots = \xi_{a+ts} = 1].$$

Let $W_{n,t} = \sum_{(a,s)} X_{a,s}$, were the sum ranges over all pairs $(a, s) \in \mathcal{I}_{n,t}$. Then $W_{n,t}$ counts the arithmetic progressions of length at least t in $\{1, 2, \dots, n\}$. Following (6.1), we have

$$\mathbb{P}[U_n < t] = \mathbb{P}[W_{n,t} = 0].$$

We claim that

$$\mathbb{P}[W_{n,t} = 0] \approx e^{-\lambda_{n,t}}, \tag{6.3}$$

where

$$\lambda_{n,t} = |\mathcal{I}_{n,t}| q p^t$$

with $q = 1 - p$. It is not difficult to see that

$$|\mathcal{I}_{n,t}| = \sum_{s=1}^{\lfloor \frac{n-1}{t} \rfloor} (n - ts) \sim \frac{n^2}{2t}$$

if $n, t \to \infty$ as long as $t = \mathrm{o}(n)$. We let

$$t = x - \frac{2\log n}{\log p} + \frac{\log\log n}{\log p} - \delta_n, \tag{6.4}$$

which is integer-valued by definition of δ_n. Since with this choice of t we have

$$\lambda_{n,t} \sim \frac{-qp^{x-\delta_n}\log p}{4}$$

as $n \to \infty$, we have established (6.2).

It remains to justify (6.3) for t defined as in (6.4), which we will accomplish via Theorem 4.1. To this end, let $A_{a,s} = \{a, a+s, \dots, a+ts\}$ for each $(a,s) \in \mathcal{I}_{n,t}$. Note that $X_{a,s}$ and $X_{a',s'}$ are independent whenever the sets $A_{a,s}$ and $A_{a',s'}$ are disjoint. Denote by $D_{a,s}(k)$ be number of pairs $(a', s') \in \mathcal{I}_{n,t}$ with $s' \neq s$, such that $|A_{a,s} \cap A_{a',s'}| = k$. From Benjamini, Yadin and Zeitouni [(2007), Proposition 4] we have the estimate

$$D_{a,s}(k) \leq \begin{cases} (t+1)^2 n, & \text{if } k = 1, \\ (t+1)^2 t^2, & \text{if } 2 \leq k \leq t/2+1, \\ 0, & \text{if } k > t/2+1. \end{cases}$$

We can now apply Theorem 4.1. Let $N_{a,s} \subset \mathcal{I}_{n,t}$ be the set of pairs (a', s') such that $A_{a,s} \cap A_{a',s'}$ is non-empty. It is clear then that $b_3 = 0$. Now,

$$b_1 \leq \sum_{a,s}\left(1 + \sum_{k=1}^{t} D_{a,s}(k)\right)p^{2(t+1)} = \mathrm{O}\left(\frac{n^2}{t} \cdot \left(1 + t^2 n + t^5\right) \cdot p^{2t}\right) = \mathrm{o}(1),$$

where we used that $p^{2t} = \mathrm{O}(\log(n)/n^4)$. Since $\mathbb{E}(X_{a,s}X_{a',s'}) = p^{2(t+1)-k}$ if $|A_{a,s} \cap A_{a',s'}| = k$, we also have

$$b_2 \leq \sum_{a,s}\sum_{k=1}^{t} D_{a,s}(k)p^{2(t+1)-k} = \mathrm{O}\left(\frac{n^2}{t} \cdot \left(t^2 n + t^4 p^{-t/2}\right) \cdot p^{2t}\right) = \mathrm{o}(1).$$

Hence, by Theorem 4.1, $|\mathbb{P}[W_{n,t} = 0] - e^{-\lambda_{n,t}}| \to 0$ as $n \to \infty$, justifying (6.3). $\square$

We refer to Benjamini, Yadin and Zeitouni (2007, 2012) and Zhao and Zhang (2012) for further details and refinements.

6.2. Bootstrap percolation

Consider the d-dimensional torus lattice $\mathbb{T}_n^d = \mathbb{Z}^d/n\mathbb{Z}^d$, along with the canonical ℓ_1 distance, that is, the smallest number of edges connecting to points. Two sites are connected if their ℓ_1 distance is 1. Bollobás *et al.* (2013) considered *d-neighbour bootstrap percolation* on $\mathbb{T}_n^d$, a special type of a *cellular automaton*. A vertex can be either *infected* or *uninfected*. At each time step, an uninfected vertex becomes infected if d or more of its neighbours are infected (at each time step, this rule is applied *simultaneously* for all vertices). Once a vertex is infected, it stays infected.

The rules of cellular automata are usually deterministic, and in the model considered by Bollobás *et al.* (2013), randomness is added only at the beginning: at time 0, each vertex is infected with probability p and remains uninfected with probability $q = 1 - p$, independently of all other vertices. With $A_t \subset \mathbb{T}_n^d$ denoting the set of all infected sites at time t, we shall be interested in the first time

$$T_n = \inf\{t \geq 0 : A_t = \mathbb{T}_n^d\}$$

when all the sites are infected.

The following result says that, if we let p converge to 1 at the right speed as $n \to \infty$, T_n is essentially concentrated on one or two points. In order to formulate the result, we define the combinatorial quantity

$$m_t = \sum_{r=0}^{t} \sum_{j=0}^{r} \binom{d}{j}.$$

Theorem 6.2 [Bollobás *et al.* (2013), Theorem 3]. *Fix a positive integer t. If, for some function $\omega(n) \to \infty$,*

$$\left(\frac{\omega(n)}{n^d}\right)^{\frac{1}{m_{t-1}}} \leq q_n \leq \left(\frac{1}{\omega(n)n^d}\right)^{\frac{1}{m_t}} \tag{6.5}$$

then $\mathbb{P}_{p_n}[T_n = t] \to 1$. If instead, for some slowly varying function $\omega(n)$,

$$\left(\frac{1}{\omega(n)n^d}\right)^{\frac{1}{m_t}} \leq q_n \leq \left(\frac{\omega(n)}{n^d}\right)^{\frac{1}{m_t}} \tag{6.6}$$

then $\mathbb{P}_{p_n}[T_n \in \{t, t+1\}] \to 1$.

Idea of proof. As in the previous application, we reformulate the problem as an extremal problem. Although we assume that t is fixed throughout, the arguments can be extended to $t = \mathrm{o}(\log n/\log\log n)$. For each $i \in \mathbb{T}_n^d$, let Y_i be the time when vertex i becomes infected, that is

$$Y_i = \inf\{t \geq 0 : i \in A_t\}.$$

Now clearly $T_n = \max_{i \in \mathbb{T}_n^d} Y_i$. For each $i \in \mathbb{T}_n^d$, let $X_{t,i} = \mathrm{I}[Y_i > t]$ be the indicator that vertex i is uninfected at time t. Note that, although Y_i and Y_j are not independent for any i and j, the indicators $X_{t,i}$ and $X_{t,j}$ are independent whenever the ℓ_1-distance between i and j is larger than $2t + 1$, since infections can only propagate an ℓ_1-distance 1 per time step. With $W_{n,t} = \sum_{i \in \mathbb{T}_n^d} X_{t,i}$, we have

$$\mathbb{P}_p[T_n \leq t] = \mathbb{P}_p[W_{n,t} = 0]. \tag{6.7}$$

We claim that

$$\mathbb{P}_p[W_{n,t} = 0] \approx e^{-\lambda_{n,t}}, \tag{6.8}$$

where

$$\lambda_{n,t} = \sum_{i \in \mathbb{T}_n^d} \mathbb{P}_p[i \text{ is uninfected at time } t] = n^d \rho_{n,t}(p)$$

with $\rho_{n,t}(p) = \mathbb{P}_p[0 \text{ is uninfected at time } t]$.

Bollobás *et al.* [(2013), Theorem 17] gave the following results about the behaviour of $\rho_{n,t}(p)$. If there exists $C = C(t,d) > 0$ such that

$$q_n^{m_t} \le \frac{C}{n^d}, \tag{6.9}$$

for all n, then

$$\rho_{n,t}(p_n) \sim d^3 2^{d-1} q_n^{m_t} \tag{6.10}$$

as $n \to \infty$. Hence, if

$$q_n^{m_t} \le \frac{1}{n^d \omega(n)} \tag{6.11}$$

for some function $\omega(n) \to \infty$, we have that $\lambda_{n,t}(p_n) = n^d \rho_{n,t}(p_n) \to 0$, so that, under (6.11),

$$\mathbb{P}_{p_n}[T_n \le t] \to 1. \tag{6.12}$$

If, in contrast, we have

$$q_n^{m_t} \ge \frac{\omega(n)}{n^d} \tag{6.13}$$

for some function $\omega(n) \to \infty$, we can argue as follows. A simple coupling argument yields that the system is monotone, that is, if $\tilde{p} \le p$, we have $\rho_{n,t}(\tilde{p}) \ge \rho_{n,t}(p)$, and hence $\lambda_{n,t}(\tilde{p}) \ge \lambda_{n,t}(p)$. Since by (6.10) we have

$$\lambda_{n,t}(p_n) \sim C d^3 2^{d-1}$$

for arbitrarily large C, we must have

$$\lambda_{n,t}(p_n) \to \infty$$

if (6.13) is true, thus yielding

$$\mathbb{P}_{p_n}[T_n \le t] \to 0. \tag{6.14}$$

Since the first inequality in (6.5) is just (6.13) with t replaced by $t-1$, we have from (6.14) that $\mathbb{P}_{p_n}[T_n \le t-1] \to 0$. On the other hand, the second inequality of (6.5) is just (6.11), hence

(6.12) implies $\mathbb{P}_{p_n}[T_n \le t] \to 1$. Thus, $\mathbb{P}_{p_n}[T_n = t] \to 1$. The proof of the second statement is analogous by observing that (6.6) implies

$$\left(\frac{\tilde{\omega}(n)}{n^d}\right)^{\frac{1}{m_{t-1}}} \le q_n \le \left(\frac{1}{\tilde{\omega}(n)n^d}\right)^{\frac{1}{m_{t+1}}},$$

where, with $\alpha = m_{t-1}/m_t < 1$,

$$\tilde{\omega}(n) = \frac{n^{d(1-\alpha)}}{\omega(n)^\alpha} \to \infty.$$

It remains to justify (6.8). Again, by monotonicity it is enough to consider (6.9), since $\tilde{p} \le p$ implies $\mathbb{P}_{\tilde{p}}[T_n \le t] \le \mathbb{P}_p[T_n \le t]$. Let

$$\tilde{\rho}_{n,t}(p) = \max_{j:d(0,j)\le 2t} \mathbb{P}_p[0 \text{ and } j \text{ are uninfected at time } t].$$

Bollobás *et al.* [(2013), Lemma 19] showed that, if (6.9) holds, then

$$\tilde{\rho}_{n,t}(p_n) = \mathrm{o}\big(\rho_{n,t}(p_n)\big).$$

Let now $N_i = \{j \in \mathbb{T}_n^d : d(i,j) \le 2t\}$. It is clear that $X_{t,i}$ is independent of $(X_{t,j})_{j\notin N_i}$, hence $b_3 = 0$. With the crude bound $|N_i| \le t^d$, we have

$$\frac{b_1}{\lambda_{n,t}(p_n)} \le \frac{n^d t^d \rho_{n,t}(p_n)^2}{\lambda_{n,t}(p_n)} = t^d \rho_{n,t}(p_n) = \mathrm{o}(1)$$

and

$$\frac{b_2}{\lambda_{n,t}(p_n)} \le \frac{n^d t^d \tilde{\rho}_{n,t}(p_n)}{\lambda_{n,t}(p_n)} = t^d \frac{\tilde{\rho}_{n,t}(p_n)}{\rho_{n,t}(p_n)} = \mathrm{o}(1),$$

justifying (6.8). □

7. Generalisations and open problems

In this section we will discuss three generalisations of Poisson approximation and touch briefly on two other generalisations, the three generalisations being compound Poisson approximation, Poisson process approximation and multivariate Poisson approximation. Compound Poisson distributions on the real line, the distributions of Poisson point processes, and multivariate Poisson distributions are all compound Poisson distributions if viewed in an appropriate way, but the three approximations have been studied separately because of the different contexts in which they arise and the different problems to which they are applied.

7.1. Compound Poisson approximation

In many probability models (see Aldous (1989)), events occur in clumps at widely scattered localities or at long irregular intervals in time. In such situations, the Poisson approximation for the number of events occurring either fails or performs poorly. If the number of clumps is approximately Poisson, the clumps are roughly independent and their sizes close to identically distributed, then the number of events occurring can be approximated by a compound Poisson distribution. A typical example of events occurring in clumps is earthquakes exceeding certain magnitude. Often such an earthquake is followed by a quick succession of several earthquakes before normalcy is resumed.

We illustrate further the notion of clumps by presenting the example of the longest head run discussed in Arratia, Goldstein and Gordon (1989, 1990). Note that this example is a special case of the maximal arithmetic progressions in Section 6.1. Suppose a coin is tossed repeatedly where the probability of falling heads is p $(0 < p < 1)$. Let R_n be the length of the longest run of heads starting from within the first n tosses. What is the asymptotic distribution of R_n as $n \to \infty$?

Let $Z_1, Z_2, \ldots$ be independent Bernoulli random variables with success probability p $(0 < p < 1)$, where $\{Z_i = 1\}$ represents the event that the coin falls heads at the ith toss. Let $J = \{1, 2, \ldots, n\}$ and let $t \geq 1$. Define $Y_i = Z_i Z_{i+1} \cdots Z_{i+t-1}$ for $i = 1, 2, \ldots, n$, and define

$$X_i = \begin{cases} Y_1, & \text{if } i = 1, \\ (1 - Z_{i-1})Y_i, & \text{if } 2 \leq i \leq n. \end{cases}$$

Let $W = \sum X_i$ and let $\lambda = \mathbb{E}W$. Then $\{R_n < t\} = \{W = 0\}$.

Define $B_i = \{j \in J : |i - j| \leq t\}$, $i = 1, 2, \ldots, n$. Then $\{X_i : i \in J\}$ is locally dependent with dependence neighbourhoods $\{B_i : i \in J\}$. Applying Theorem 4.1, we obtain $b_3 = b_2 = 0$, and $b_1 < \lambda^2(2t + 1)/n + \lambda p^t$.

Hence

$$\left|\mathbb{P}[R_n < t] - e^{-\lambda}\right| \leq \left(1 \wedge \frac{1}{\lambda}\right)\left(\lambda^2(2t + 1)/n + \lambda p^t\right). \tag{7.1}$$

Requiring that λ remains bounded away from 0 and from ∞ and that the error bound tends to 0 as $n \to \infty$ leads to the following conclusion: for a fixed integer c, $\mathbb{P}[R_n - \lfloor \log_{1/p}(n(1 - p)) \rfloor < c] \to \exp\{-p^{c-r}\}$ along a subsequence of n if and only if $\log_{1/p}(n(1 - p)) - \lfloor \log_{1/p}(n(1 - p)) \rfloor \to r \in [0, 1]$ along the same subsequence.

Now let $V = \sum Y_i$ and let $\mu = \mathbb{E}V$. Then we also have $\{R_n < t\} = \{V = 0\}$. The difference between the X_i and the Y_i is that while X_i indicates a run of at least t heads starting from the ith toss preceded by a tail, Y_i indicates a run of at least t heads starting from the ith toss regardless of what precedes it. For a run of more than t heads starting from the ith toss, say, $Z_{i-1} = 0$, $Z_i = \cdots = Z_{i+m-1} = 1$, $Z_{i+m} = 0$, where $m > t$, $X_i = 1$, $X_{i+1} = \cdots = X_{i+m-t} = 0$, whereas $Y_i = Y_{i+1} = \cdots = Y_{i+m-t} = 1$. Thus while W counts the clumps, which consist of runs of at least t heads, V counts the clumps and their sizes. The way the X_i are defined so that W counts only the clumps is called *declumping*.

If we apply Theorem 4.1 to V, we will obtain a bound on $|\mathbb{P}[R_n < t] - e^{-\mu}|$. Since $\{Y_i : i \in J\}$ is locally dependent, $b_3 = 0$. But b_2 does not tend to 0 if we require μ to be bounded away from 0 and from ∞. Thus Poisson approximation fails. However, Arratia, Goldstein and Gordon (1990) showed that the distribution of V is approximately compound Poisson through an extension of Poisson approximation to Poisson process approximation.

We pause for a moment to remark that there are two equivalent representations of the compound Poisson distribution on $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$. Let $\xi_1, \xi_2, \ldots$ be i.i.d. positive integer-valued random variables with $\mathbb{P}[\xi_1 = k] = \gamma_k$ for $k = 1, 2, \ldots$, and let N be a Poisson random variable with mean $\nu > 0$, independent of the ξ_i. The distribution of $\xi_1 + \xi_2 + \cdots + \xi_N$, which is compound Poisson, is the same as that of $\sum i Z_i$, where the Z_i are independent Poisson random variables with means $\nu\gamma_i$ respectively. Let γ be the common distribution of the ξ_i. Then $\gamma = \sum \gamma_i \delta_i$, where δ_i is the Dirac measure at i. We denote this compound Poisson distribution by $\mathrm{CP}(\nu\gamma) = \mathrm{CP}(\sum \nu\gamma_i\delta_i)$ and call $\nu\gamma$ the generating measure.

Arratia, Goldstein and Gordon (1990) showed that by representing $\{Y_i : i \in J\}$ as a Bernoulli process indexed by $J \times \{1, 2, \ldots\}$ where J denotes the location of clumps and $\{1, 2, \ldots\}$ the clump sizes, $\{Y_i : i \in J\}$ can be approximated in total variation by a Poisson process, which is a collection of independent Poisson random variables indexed by $J \times \{1, 2, \ldots\}$. By taking an appropriate projection and using the above alternative representation of the compound Poisson distribution, Arratia, Goldstein and Gordon (1990) obtained a bound on the total variation distance between the distribution of V and a compound Poisson distribution. This in turn provides an error bound for $|\mathbb{P}[R_n < t] - e^{-\nu}|$, where ν is the mean of the Poisson number of terms in the compound Poisson distribution and is less than $\mu = \mathbb{E}V$. This error bound is of the same order as that in (7.1), but without the factor $1 \wedge \lambda^{-1}$. However, it leads to the same asymptotic distribution for R_n as $n \to \infty$ because λ is bounded away from 0 and from ∞.

The factor $1 \wedge \lambda^{-1}$ is lost because Poisson process approximation for the Bernoulli process representing $\{Y_i : i \in J\}$ requires too much information extraneous to the compound Poisson approximation for V. A direct approach using Stein's method, which partially recovers the factor $1 \wedge \lambda^{-1}$, was developed by Barbour, Chen and Loh (1992). Let $\lambda_i \geq 0$, $i = 1, 2, \ldots$ such that $\sum \lambda_i < \infty$. Barbour, Chen and Loh (1992) used the Stein equation

$$\sum i\lambda_i f(w + i) - wf(w) = \mathrm{I}(w \in A) - \mathbb{P}[Z \in A] \qquad \text{for } w \in \mathbb{Z}_+, \tag{7.2}$$

where A is a subset of $\mathbb{Z}_+$, Z_i, $i \geq 1$, are independent $\mathrm{Po}(\lambda_i)$, and $Z = \sum i Z_i$.

By solving (7.2) analytically as well as writing $f(w) = g(w) - g(w - 1)$ and using the generator approach to solve (7.2), they obtained the following bounds on the solution f_A. For $A \subset \mathbb{Z}_+ = \{0, 1, 2, \ldots\}$,

$$\|f_A\|_\infty \leq \bigl(1 \wedge \lambda_1^{-1}\bigr)e^{\nu}, \qquad \|\Delta f_A\|_\infty \leq \bigl(1 \wedge \lambda_1^{-1}\bigr)e^{\nu}, \tag{7.3}$$

where $\nu = \sum \lambda_i$, and if $i\lambda_i \downarrow 0$, then

$$\|f_A\|_\infty \leq \begin{cases} 1, & \text{if } \lambda_1 - 2\lambda_2 \leq 1, \\ \dfrac{2}{(\lambda_1 - 2\lambda_2)^{1/2}} - \dfrac{1}{\lambda_1 - 2\lambda_2}, & \text{if } \lambda_1 - 2\lambda_2 > 1, \end{cases}$$

and

$$\|\Delta f_A\|_\infty \leq 1 \wedge \left(\frac{1}{4(\lambda - 2\lambda_2)^2} + \frac{\log^+(2(\lambda_1 - 2\lambda_2))}{\lambda_1 - 2\lambda_2}\right).$$

As in the case of Poisson approximation, the solution f_A is unique except at $w = 0$. Since its value at $w = 0$ is never used in the calculation, it has been conveniently set to be 0. Using the bounds on $\|\Delta f_A\|_\infty$, Barbour, Chen and Loh (1992) proved the following theorem for locally dependent Bernoulli random variables.

Theorem 7.1. *Suppose* $\{X_\alpha : \alpha \in J\}$ *are locally dependent Bernoulli random variables with success probabilities* p_α *and dependence neighbourhoods* $B_\alpha \subset C_\alpha$, $\alpha \in J$, *such that for each* $\alpha \in J$, X_α *is independent of* $\{X_\beta : \beta \in B_\alpha^c\}$ *and* $\{X_\beta : \beta \in B_\alpha\}$ *is independent of* $\{X_\beta : \beta \in C_\alpha^c\}$. *Let* $W = \sum X_\alpha$ *and let* $Y_\alpha = \sum_{\beta \in B_\alpha} X_\beta$. *Define* $\lambda_i = i^{-1} \sum \mathbb{E} X_\alpha \mathrm{I}[Y_\alpha = i]$ *for* $i = 1, 2, \ldots$, *let* $\nu = \sum \mathbb{E} X_\alpha Y_\alpha^{-1} = \sum \lambda_i$, *and let* $\gamma = \sum (\lambda_i/\nu)\delta_i$.

1. *We have*

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{CP}(\nu\gamma)\big) \le \big(1 \wedge \lambda_1^{-1}\big) e^{\nu} \sum_{\alpha \in J} \sum_{\beta \in C_\alpha} p_\alpha p_\beta.$$

2. *If* $i\lambda_i \downarrow 0$ *as* $i \to \infty$, *then we have*

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{CP}(\nu\gamma)\big) \\ &\quad \le 2\left[1 \wedge \left(\frac{1}{4(\lambda_1 - 2\lambda_2)^2} + \frac{\log^+(2(\lambda_1 - 2\lambda_2))}{\lambda_1 - 2\lambda_2}\right)\right] \sum_{\alpha \in J} \sum_{\beta \in C_\alpha} p_\alpha p_\beta. \end{aligned}$$

If $\lambda_i = 0$ for $i \ge 3$ and $\lambda_1 < 2\lambda_2$, it can be shown that both $\|f_A\|_\infty$ and $\|\Delta f_A\|_\infty$ grow exponentially fast with ν (see Barbour and Utev, Barbour and Utev (1998, 1999)). This shows that the bounds in (7.3) cannot be much improved. To circumvent this difficulty Barbour and Utev (1998, 1999) considered bounds on

$$H_0^a(\nu\gamma) := \sup_{A \subset \mathbb{Z}_+} \sup_{w > a} \big|f_A(w)\big|,$$

$$H_1^a(\nu\gamma) := \sup_{A \subset \mathbb{Z}_+} \sup_{w > a} \big|f_A(w+1) - f_A(w)\big|.$$

Assuming that the generating function of $\gamma = \sum (\lambda_i/\nu)\delta_i$ has a radius of convergence $R > 1$ and assuming some other conditions, Barbour and Utev (1999) proved that there exist constants C_0, C_1 and C_2 depending on γ such that for any $a > C_2 \nu m_1 + 1$, where m_1 is the mean of γ,

$$H_0^a(\nu\gamma) \le C_0 \nu^{-1/2},$$

and

$$H_1^a(\nu\gamma) \le C_1 \nu^{-1}.$$

The expressions for C_0, C_1 and C_2 are complicated but explicit. Sufficient conditions can be found under which these constants are uniformly bounded. Using the bound on $H_1^a(\nu\gamma)$, Barbour and Månsson (2000) proved the following theorem.

Theorem 7.2. *For $n \geq 1$, let $\lambda_{in} > 0$ for $i = 1, 2, \ldots$. Let W_n, $n = 1, 2, \ldots$, be a sequence of non-negative, integer-valued random variables such that for each $n \geq 1$ and each bounded $f : \mathbb{Z}_+ \to \mathbb{R}$,*

$$\left| \mathbb{E}\left(\sum_{i \geq 1} i \lambda_{in} f(W_n + i) - W_n f(W_n) \right) \right| \leq \|\Delta f\|_\infty \epsilon_n.$$

Let $\nu_n = \sum_{i \geq 1} \lambda_{in} < \infty$ and $\gamma_{in} = \lambda_{in}/\nu_n$. Assume that

$$\text{(i)} \quad \lim_{n \to \infty} \gamma_{in} = \gamma_i \text{ for each } i \geq 1, \qquad \text{(ii)} \quad \inf_{n \geq 1} \gamma_{1n} > 0,$$

$$\text{(iii)} \quad \sup_{n \geq 1} \sum_{i \geq 1} \gamma_{in} R^i < \infty \text{ for some } R > 1, \qquad \text{(iv)} \quad \inf_{n \geq 1} \nu_n > 2.$$

Then there exist positive constants $K < \infty$ and $c < 1$ such that for any x satisfying $c < x < 1$ and any n for which $\mathbb{E}W_n \geq (x - c)^{-1}$,

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), \mathrm{CP}(\Gamma_n)\big) \leq K(1 - x)^{-1}\big(\nu_n^{-1}\epsilon_n + \mathbb{P}\big(W_n \leq (1 + x)\mathbb{E}W_n/2\big)\big),$$

where the generating measure $\Gamma_n = \sum \lambda_{in}\delta_i = \nu_n \sum \gamma_{in}\delta_i$.

In their efforts to obtain bounds on the solution of the Stein equation (7.2) so that the bounds resemble or "correctly" generalise those in the Poisson approximation, Barbour and Xia [(1999), Theorem 2.5] obtained the following theorem by treating compound Poisson approximation as a perturbation of Poisson approximation.

Theorem 7.3. *Let $\lambda_i \geq 0$, $i \geq 1$, satisfy*

$$\theta := \frac{1}{\lambda} \sum i(i - 1)\lambda_i < \frac{1}{2} \qquad \text{where } \lambda = \sum i\lambda_i < \infty.$$

Then for any subset $A \subset \mathbb{Z}_+$, the Stein equation (7.2) has a bounded solution $f = f_A$ satisfying

$$\|f_A\|_\infty \leq \frac{1}{(1 - 2\theta)\lambda^{1/2}}, \qquad \|\Delta f_A\|_\infty \leq \frac{1}{(1 - 2\theta)\lambda}.$$

Using the bound on $\|\Delta f_A\|_\infty$ for the locally dependent Bernoulli random variables defined in Theorem 7.1, we obtain the following theorem.

Theorem 7.4. *Let $\{X_\alpha : \alpha \in J\}$ be locally dependent Bernoulli random variables as defined in Theorem 7.1. Let $W = \sum X_\alpha$ and let $Y_\alpha = \sum_{\beta \in B_\alpha} X_\beta$. Define $\lambda_i = i^{-1} \sum \mathbb{E}X_\alpha \mathrm{I}[Y_\alpha = i]$ for $i = 1, 2, \ldots$. If $\theta := \lambda^{-1} \sum i(i - 1)\lambda_i < \frac{1}{2}$, where $\lambda = \sum i\lambda_i < \infty$, and $\gamma = \sum(\lambda_i/\nu)\delta_i$, then*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{CP}(\nu\gamma)\big) \leq 2\left(1 \wedge \frac{1}{(1 - 2\theta)\lambda}\right) \sum_{\alpha \in J} \sum_{\beta \in C_\alpha} p_\alpha p_\beta.$$

Much progress has been made on bounding the solution of the Stein equation (7.2) in compound Poisson approximation. The results presented in this section are quite satisfactory although many conditions on the λ_i or the generating measure are required. It still remains a tantalising question as to what general results one could obtain by using a different Stein equation or by using a non-uniform bound on its solution, and to what extent one could do away with those conditions on the λ_i. Roos (2003) used the generating function approach of Kerstan to study compound Poisson approximation for sums of independent random variables without imposing any condition on the λ_i, but the method works only under the condition of independence. Even for sums of independent random variables it is unclear if the results of Roos (2003) can be proved using Stein's method. For further reading on compound Poisson approximation, see Barbour and Chryssaphinou (2001), and Erhardsson (2005).

7.2. Poisson process approximation

In Poisson process approximation, both the number of rare events that occur and the respective locations at which they occur are approximated by a Poisson point process on a metric space. In the longest head run example discussed in Arratia, Goldstein and Gordon (1990), the information on the locations where the events occur is used in the calculation of the compound Poisson approximation. In Leung *et al.* (2005), a Poisson process approximation for palindromes in a DNA is used to provide a mathematical basis for modelling the palindromes as i.i.d. uniform random variables on an interval. The total variation distance is used for the Poisson process approximation in the longest head run example, but in general such a distance is not appropriate. For example, the total variation distance between a Bernoulli process indexed by $\{i/n : i = 1, 2, \ldots, n\}$ with success probability λ/n and a Poisson process on $[0, 1]$ with rate λ is always 1, although the Bernoulli process converges weakly to the Poisson process as $n \to \infty$. A distance which is commonly used in process approximations is the Wasserstein distance.

By writing $f(w) = g(w) - g(w - 1)$, Barbour (1988) converted the Stein equation (4.1) to a second order difference equation and introduced the generator approach to extend Poisson approximation to higher dimensions and to Poisson process approximation. Following the generator approach, Barbour and Brown (1992) established a general framework for Poisson process approximation. In this framework, a compact metric space Γ endowed with a metric $d_0 \leq 1$ is the carrier space, Ξ is a point process on Γ with finite intensity measure $\boldsymbol{\lambda}$ of total mass λ, where $\boldsymbol{\lambda}(A) = \mathbb{E}\Xi(A)$ for every Borel set in Γ, and Z is a Poisson point process on Γ with the same intensity measure $\boldsymbol{\lambda}$. Let $\mathcal{X}$ be the configuration space $\{\sum_{1\leq i\leq k} \delta_{\alpha_i} : \alpha_i \in \Gamma, k \geq 0\}$. Define a metric $d_1 \leq 1$ on $\mathcal{X}$ by

$$d_1\Bigg(\sum_{1\leq i\leq m} \delta_{x_i}, \sum_{1\leq i\leq n} \delta_{y_i}\Bigg) = \begin{cases} 1, & \text{if } m \neq n, \\ n^{-1} \min \sum d_0(x_i, y_{\pi(i)}), & \text{if } m = n, \end{cases}$$

where the minimum is taken over all permutations π of $\{1, 2, \ldots n\}$.

Define $\mathcal{H} = \{h : \mathcal{X} \to \mathbb{R} : |h(\xi_1) - h(\xi_2)| \leq d_1(\xi_1, \xi_2)\}$. The Wasserstein distance with respect to d_1 between the distributions of two point processes Ξ_1 and Ξ_2 on Γ with finite intensity measures is defined by

$$d_2\big(\mathcal{L}(\Xi_1), \mathcal{L}(\Xi_2)\big) = \sup_{h\in\mathcal{H}}\big|\mathbb{E}h(\Xi_1) - \mathbb{E}h(\Xi_2)\big|.$$

Note that d_2 is a metric bounded by 1. The Stein equation for approximating the distribution of the point process Ξ by that of the Poisson point process Z is

$$\begin{aligned}\mathcal{A}g(\xi) &:= \int_\Gamma \big[g(\xi+\delta_x) - g(\xi)\big]\boldsymbol{\lambda}(dx) + \int \big[g(\xi-\delta_x) - g(\xi)\big]\xi(dx)\\ &= h(\xi) - \mathbb{E}h(Z),\end{aligned} \tag{7.4}$$

where $h \in \mathcal{H}$ and $\mathcal{A}$ is the generator of a measure-valued immigration-death process $Y_\xi(t)$ with immigration intensity $\boldsymbol{\lambda}$, per capita unit death rate, $Y_\xi(0) = \xi$, and stationary distribution $\mathcal{L}(Z)$.

The Stein equation (7.4) has a solution $g = g_h$ given by

$$g_h(\xi) = -\int_0^\infty \big[Eh\big(Y_\xi(t)\big) - Eh(Z)\big]\,dt.$$

Using coupling, Barbour and Brown (1992) obtained the following bounds on g_h:

$$\big|\Delta_\alpha g_h(\xi)\big| := \big|g_h(\xi+\delta_\alpha) - g_h(\xi)\big| \le 1 \wedge 1.65\lambda^{-1/2}, \tag{7.5}$$

$$\begin{aligned}\big|\Delta^2_{\alpha\beta} g_h(\xi)\big| &:= \big|g_h(\xi+\delta_\alpha+\delta_\beta) - g_h(\xi+\delta_\alpha) - g_h(\xi+\delta_\beta) + g_h(\xi)\big|\\ &\le 1 \wedge \frac{5(1+2\log^+(2\lambda/5)}{2\lambda},\end{aligned} \tag{7.6}$$

where λ is the total mass of $\boldsymbol{\lambda}$.

In applications, the logarithmic term in (7.5) carries over to the error bounds in the approximation. Attempts were made to remove the logarithmic terms. Xia (1997, 2000) succeeded in some special cases. A general result in the form of a non-uniform bound on $|\Delta^2_{\alpha\beta} g_h(\xi)|$ was obtained by Brown, Weinberg and Xia (2000) and later improved by Xia (2005), which is given as

$$\big|\Delta^2_{\alpha\beta} g_h(\xi)\big| \le 1 \wedge \left(\frac{3.5}{\lambda} + \frac{2.5}{|\xi|+1}\right), \tag{7.7}$$

where $|\xi|$ is the number of points in ξ, that is, the total measure of ξ.

Using (7.7), the error bound on the Wasserstein distance for Poisson process approximation for Bernoulli processes has the same factor as that on the total variation distance for the Poisson approximation for sums of independent Bernoulli random variables, namely, $1 \wedge \lambda^{-1}$.

Chen and Xia (2004) studied Stein's method for Poisson process approximation from the point of view of Palm theory. For a point process Ξ on Γ with finite intensity measure, the Palm process Ξ_α associated with Ξ at $\alpha \in \Gamma$ has the same distribution as the conditional distribution of Ξ given that a point has occurred at α. A point process Ξ on Γ with finite intensity measure $\boldsymbol{\lambda}$ is locally dependent with neighbourhoods $\{A_\alpha : \alpha \in \Gamma\}$ if $\mathcal{L}(\Xi_\alpha^{(\alpha)}) = \mathcal{L}(\Xi^{(\alpha)})$ $\boldsymbol{\lambda}$-a.s., where $\Xi_\alpha^{(\alpha)}$ and $\Xi^{(\alpha)}$ are respectively the restrictions of Ξ_α and Ξ to A_α^c for each $\alpha \in \Gamma$.

The following theorem, which uses (7.7), is Corollary 3.6 in Chen and Xia (2004).

Theorem 7.5. *Let Ξ be a locally dependent point process on the compact metric space Γ with finite intensity measure $\boldsymbol{\lambda}$ and with neighbourhoods $\{A_\alpha : \alpha \in \Gamma\}$, and let Z be a Poisson point process on Γ with the same intensity measure $\boldsymbol{\lambda}$. Let λ be the total measure of $\boldsymbol{\lambda}$. Then*

$$\begin{aligned} d_2\big(\mathcal{L}(\Xi), \mathcal{L}(Z)\big) \le \mathbb{E}\int_{\alpha\in\Gamma}\left(\frac{5}{\lambda} + \frac{3}{|\Xi^{(\alpha)}|+1}\right)\big(\Xi(A_\alpha)-1\big)\Xi(d\alpha) \\ + \int_{\alpha\in\Gamma}\int_{\beta\in A_\alpha}\left(\frac{5}{\lambda} + \mathbb{E}\frac{3}{|\Xi^{(\alpha\beta)}|+1}\right)\boldsymbol{\lambda}(d\alpha)\boldsymbol{\lambda}(d\beta), \end{aligned}$$

where $|\xi|$ is the total measure of ξ and $\xi^{(\alpha\beta)}$ is the restriction of ξ to $A_\alpha^c \cap A_\beta^c$.

This theorem gives the factor $1 \wedge \lambda^{-1}$ in the Wasserstein distance error bound for the Poisson approximation for Bernoulli Processes. It has also been applied to Poisson process approximation for palindromes in a DNA in Leung *et al.* (2005), and to Poisson point process approximation for the Matérn hard-core process in Chen and Xia (2004).

For further reading on Poisson process approximation, see Xia (2005).

7.3. Multivariate Poisson approximation

For the multivariate analogue of Poisson approximation, we consider independent Bernoulli random d-vectors, $X_1, \ldots, X_n$ with

$$\mathbb{P}\big[X_j = e^{(i)}\big] = p_{j,i}, \qquad \mathbb{P}[X_j = 0] = 1 - p_j, \qquad 1 \le i \le d, 1 \le j \le n,$$

where $e^{(i)}$ denotes the ith coordinate vector in $\mathbb{R}^d$ and $p_j = \sum_{1\le i\le d} p_{j,i}$.

Let $W = \sum X_j$, $\lambda = \sum p_j$, $\mu_i = \lambda^{-1}\sum_{1\le j\le n} p_{j,i}$, and let $Z = (Z_1, \ldots, Z_d)$, where $Z_1, \ldots, Z_d$ are independent Poisson random variables with means $\lambda\mu_1, \ldots, \lambda\mu_d$. Using the Stein equation

$$\begin{aligned} \mathcal{A}g(j) &= \sum \lambda\mu_i\{g(j+e^{(i)}) - g(j)\} + \sum j^{(i)}\{g(j-e^{(i)}) - g(j)\} \\ &= \mathrm{I}[j \in A] - \mathbb{P}[Z \in A], \end{aligned} \tag{7.8}$$

where A is a subset of $\mathbb{Z}_+^d$ and $\mathcal{A}$ the generator of a multivariate immigration-death process whose stationary distribution is $\mathcal{L}(Z)$, Barbour (1988) proved that

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathcal{L}(Z)\big) \le \sum_{1\le j\le n} p_j^2 \wedge \left(\frac{c_\lambda}{\lambda}\sum_{1\le i\le d}\frac{p_{j,i}^2}{\mu_i}\right), \tag{7.9}$$

where $c_\lambda = \frac{1}{2} + \log^+(2\lambda)$.

The error bound in (7.9) looks like the "correct" generalisation of $(1 \wedge \lambda^{-1})\sum_{1\le j\le n} p_j^2$ in the univariate case except for the factor c_λ, which grows logarithmically with λ.

Using the multivariate adaption of Kerstan's generating function method, Roos (1999) proved that

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathcal{L}(Z)\bigr) \le 8.8 \sum_{1\le j\le n} p_j^2 \wedge \left(\frac{1}{\lambda} \sum_{1\le i\le d} \frac{p_{j,i}^2}{\mu_i}\right), \tag{7.10}$$

which improves over (7.9) in removing c_λ from the error bound although the absolute constant is increased to 8.8.

The error bound in (7.9) was obtained by bounding $\Delta_{ik} g_A$ in the error term in the approximation where g_A is the solution of the Stein equation (7.8), $\Delta_i g(k) = g(k + e^{(i)}) - g(k)$ and $\Delta_{ik} = \Delta_i(\Delta_k)$. By studying the behaviour of $\Delta_{ik} g_A$, Barbour (2005) showed that the order of the bound in (7.9) is best possible for $d \ge 2$ if it is proved by bounding $\Delta_{ik} g_A$. By an indirect approach to bounding the error term Barbour (2005) obtained two error bounds, one of which comes very close to (7.10) and the other better than an earlier bound of Roos (1998).

There does not seem to be much progress on multivariate Poisson approximation using Stein's method since 2005. It still remains a question if one could prove (7.10) using Stein's method, but by another approach, perhaps by a non-uniform bound on $\Delta_{ik} g_A$ or by a different Stein equation.

7.4. Other generalisations

There are two interesting generalisations of Poisson approximation which we will not discuss in this paper but will mention in passing. First, Brown and Xia (2001) developed probabilistic methods for approximating general distributions on non-negative integers with a new family of distributions called polynomial birth-death distributions. These distributions include as special cases the Poisson, negative binomial, binomial and hyper-geometric distributions. Second, Peccati (2011) combined Stein's method with the Malliavin calculus of variations to study Poisson approximation for functionals of general Poisson random measures. This is a follow-up to his very successful work (see Nourdin and Peccati (2012)) in normal approximation for Gaussian functionals using Stein's method and the Malliavin calculus. Both the work of Brown and Xia (2001) and of Peccati (2011) open up new domains for Poisson-related approximations and applications of Stein's method.

Acknowledgments

We would like to thank Aihua Xia for helpful discussions on generalisations and open problems and Andrew Barbour for his very helpful comments.

References

Aldous, D. (1989). *Probability Approximations Via the Poisson Clumping Heuristic*. *Applied Mathematical Sciences* **77**. New York: Springer. MR0969362

Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5** 403–434. With comments and a rejoinder by the authors. MR1092983

Barbour, A.D. (1982). Poisson convergence and random graphs. *Math. Proc. Cambridge Philos. Soc.* **92** 349–359. MR0671189

Barbour, A.D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A** 175–184. MR0974580

Barbour, A.D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. MR1035659

Barbour, A.D. (2005). Multivariate Poisson-binomial approximation using Stein's method. In *Stein's Method and Applications. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 131–142. Singapore: Singapore Univ. Press. MR2205332

Barbour, A.D. and Brown, T.C. (1992). Stein's method and point process approximation. *Stochastic Process. Appl.* **43** 9–31. MR1190904

Barbour, A.D., Chen, L.H.Y. and Loh, W.L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20** 1843–1866. MR1188044

Barbour, A.D., Chen, L.H.Y. and Choi, K.P. (1995). Poisson approximation for unbounded functions. I. Independent summands. *Statist. Sinica* **5** 749–766. MR1347617

Barbour, A.D. and Chen, L.H.Y. (2005a). *An Introduction to Stein's Method. Lecture Notes Series* **4**. Singapore: Singapore Univ. Press and World Scientific.

Barbour, A.D. and Chen, L.H.Y. (2005b). *Stein's Method and Applications. Lecture Notes Series* **5**. Singapore: Singapore Univ. Press and World Scientific.

Barbour, A.D. and Chryssaphinou, O. (2001). Compound Poisson approximation: A user's guide. *Ann. Appl. Probab.* **11** 964–1002. MR1865030

Barbour, A.D. and Eagleson, G.K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.* **15** 585–600. MR0706618

Barbour, A.D. and Hall, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480. MR0755837

Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. New York: Oxford Univ. Press. MR1163825

Barbour, A.D. and Månsson, M. (2000). Compound Poisson approximation and the clustering of random points. *Adv. in Appl. Probab.* **32** 19–38. MR1765173

Barbour, A.D. and Reinert, G. (2001). Small worlds. *Random Structures Algorithms* **19** 54–74. MR1848027

Barbour, A.D. and Utev, S. (1998). Solving the Stein equation in compound Poisson approximation. *Adv. in Appl. Probab.* **30** 449–475. MR1642848

Barbour, A.D. and Utev, S. (1999). Compound Poisson approximation in total variation. *Stochastic Process. Appl.* **82** 89–125. MR1695071

Barbour, A.D. and Xia, A. (1999). Poisson perturbations. *ESAIM Probab. Stat.* **3** 131–150. MR1716120

Benjamini, I., Yadin, A. and Zeitouni, O. (2007). Maximal arithmetic progressions in random subsets. *Electron. Commun. Probab.* **12** 365–376. MR2350574

Benjamini, I., Yadin, A. and Zeitouni, O. (2012). Erratum: Maximal arithmetic progressions in random subsets. *Electron. Commun. Probab.* **17** 1.

Bernoulli, J. (1713). *Ars Conjectandi*. Basel: Thurnisiorum.

Bollobás, B., Holmgren, C., Smith, P. and Uzzell, A.J. (2013). The speed of bootstrap percolation. *Ann. Probab.* To appear.

Brown, T.C., Weinberg, G.V. and Xia, A. (2000). Removing logarithms from Poisson process error bounds. *Stochastic Process. Appl.* **87** 149–165. MR1751169

Brown, T.C. and Xia, A. (2001). Stein's method and birth-death processes. *Ann. Probab.* **29** 1373–1403. MR1872746

Chatterjee, S., Diaconis, P. and Meckes, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2** 64–106. MR2121796

Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. MR0428387

Chen, L.H.Y. and Choi, K.P. (1992). Some asymptotic and large deviation results in Poisson approximation. *Ann. Probab.* **20** 1867–1876. MR1188045

Chen, L.H.Y., Goldstein, L. and Shao, Q.M. (2011). *Normal Approximation by Stein's Method*. *Probability and Its Applications* (*New York*). Heidelberg: Springer. MR2732624

Chen, L.H.Y. and Xia, A. (2004). Stein's method, Palm theory and Poisson process approximation. *Ann. Probab.* **32** 2545–2569. MR2078550

de Moivre, A. (1712). De mensura sortis, seu, de probabilitate eventuum in ludis a casu fortuito pendentibus. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **27** 213–264.

Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* **14** 663–676. MR0832029

Dembo, A., Karlin, S. and Zeitouni, O. (1994). Limit distribution of maximal non-aligned two-sequence segmental score. *Ann. Probab.* **22** 2022–2039. MR1331214

Diaconis, P. and Holmes, S., eds. (2004). *Stein's Method*: *Expository Lectures and Applications*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **46**. Beachwood, OH: IMS. Papers from the Workshop on Stein's Method held at Stanford University, Stanford, CA, 1998. MR2118599

Draief, M. and Massoulié, L. (2010). *Epidemics and Rumours in Complex Networks*. *London Mathematical Society Lecture Note Series* **369**. Cambridge: Cambridge Univ. Press. MR2582458

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*: *For Insurance and Finance*. *Applications of Mathematics* (*New York*) **33**. Berlin: Springer. MR1458613

Erhardsson, T. (2005). Stein's method for Poisson and compound Poisson approximation. In *An Introduction to Stein's Method*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 61–113. Singapore: Singapore Univ. Press. MR2235449

Falk, M., Hüsler, J. and Reiss, R.D. (2011). *Laws of Small Numbers*: *Extremes and Rare Events*, extended ed. Basel: Birkhäuser. MR2732365

Franceschetti, M. and Meester, R. (2006). Critical node lifetimes in random networks via the Chen–Stein method. *IEEE Trans. Inform. Theory* **52** 2831–2837. MR2238568

Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949

Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

Grimmett, G. and Janson, S. (2003). On smallest triangles. *Random Structures Algorithms* **23** 206–223. MR1995691

Hart, S., Rinott, Y. and Weiss, B. (2008). Evolutionarily stable strategies of random games, and the vertices of random polygons. *Ann. Appl. Probab.* **18** 259–287. MR2380899

Karlin, S. and Chen, C. (2000). r-scan statistics of a marker array in multiple sequences derived from a common progenitor. *Ann. Appl. Probab.* **10** 709–725. MR1789977

Lange, K. (2002). *Mathematical and Statistical Methods for Genetic Analysis*, 2nd ed. *Statistics for Biology and Health*. New York: Springer. MR1892279

Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197. MR0142174

Leung, M.Y., Choi, K.P., Xia, A. and Chen, L.H.Y. (2005). Nonrandom clusters of palindromes in herpesvirus genomes. *J. Comput. Biol.* **12** 331–354. MR2906377

Lippert, R.A., Huang, H. and Waterman, M.S. (2002). Distributional regimes for the number of k-word matches between two random sequences. *Proc. Natl. Acad. Sci. USA* **99** 13980–13989 (electronic). MR1944413

Meckes, E. (2009). On Stein's method for multivariate normal approximation. In *High Dimensional Probability V*: *The Luminy Volume*. *Inst*. *Math*. *Stat*. *Collect*. **5** 153–178. Beachwood, OH: IMS. MR2797946

Neuhauser, C. (1994). A Poisson approximation for sequence comparisons with insertions and deletions. *Ann*. *Statist*. **22** 1603–1629. MR1311992

Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus*: *From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. MR2962301

Peccati, G. (2011). The Chen–Stein method for Poisson functionals. Preprint.

Peköz, E.A. and Röllin, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann*. *Probab*. **39** 587–608. MR2789507

Peköz, E., Röllin, A. and Ross, N. (2013). Degree asymptotics with rates for preferential attachment random graphs. *Ann*. *Appl*. *Probab*. **23** 1188–1218.

Poisson, S.D. (1837). *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile*, *Prétcedées des Règles Générales du Calcul des Probabilités*. Paris: Bachelier.

Prohorov, Y.V. (1953). Asymptotic behavior of the binomial distribution. *Uspehi Matem*. *Nauk* (*N*.*S*.) **8** 135–142. MR0056861

Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann*. *Probab*. **37** 2150–2173. MR2573554

Roos, B. (1998). Metric multivariate Poisson approximation of the generalized multinomial distribution. *Teor*. *Veroyatn*. *Primen*. **43** 404–413. MR1679016

Roos, B. (1999). On the rate of multivariate Poisson convergence. *J*. *Multivariate Anal*. **69** 120–134. MR1701409

Roos, B. (2003). Kerstan's method for compound Poisson approximation. *Ann*. *Probab*. **31** 1754–1771. MR2016599

Ross, N. (2011). Fundamentals of Stein's method. *Probab*. *Surv*. **8** 210–293. MR2861132

Shih, H.H. (2011). On Stein's method for infinite-dimensional Gaussian approximation in abstract Wiener spaces. *J*. *Funct*. *Anal*. **261** 1236–1283. MR2807099

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ*. *California*, *Berkeley*, *Calif*., 1970/1971), *Vol*. *II*: *Probability Theory* 583–602. Berkeley, CA: Univ. California Press. MR0402873

Stein, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. Hayward, CA: IMS. MR0882007

Tao, T. (2007). What is good mathematics? *Bull*. *Amer*. *Math*. *Soc*. (*N*.*S*.) **44** 623–634. MR2338369

von Bortkiewicz, L. (1898). *Das Gesetz der Kleinen Zahlen*. Leipzig: Teubner.

Waterman, M.S. (1995). *Introduction to Computational Biology*: *Maps*, *Sequences and Genomes*. Boca Raton, FL: Chapman & Hall/CRC.

Waterman, M.S. and Vingron, M. (1994). Sequence comparison significance and Poisson approximation. *Statist*. *Sci*. **9** 367–381. MR1325433

Xia, A. (1997). On using the first difference in the Stein–Chen method. *Ann*. *Appl*. *Probab*. **7** 899–916. MR1484790

Xia, A. (2000). Poisson approximation, compensators and coupling. *Stoch*. *Anal*. *Appl*. **18** 159–177. MR1739289

Xia, A. (2005). Stein's method and Poisson process approximation. In *An Introduction to Stein's Method*. *Lect*. *Notes Ser*. *Inst*. *Math*. *Sci*. *Natl*. *Univ*. *Singap*. **4** 115–181. Singapore: Singapore Univ. Press. MR2235450

Zhao, M.Z. and Zhang, H.Z. (2012). On the longest length of arithmetic progressions. Available at arXiv:1204.1149.