

APPROXIMATION OF THE DIFFERENCE OF TWO POISSON-LIKE COUNTS BY SKELLAM

H. L. GAN,* *Northwestern University*

ERIC D. KOLACZYK,** *Boston University*

Abstract

Poisson-like behavior for event count data is ubiquitous in nature. At the same time, differencing of such counts arises in the course of data processing in a variety of areas of application. As a result, the Skellam distribution – defined as the distribution of the difference of two independent Poisson random variables – is a natural candidate for approximating the difference of Poisson-like event counts. However, in many contexts strict independence, whether between counts or among events within counts, is not a tenable assumption. Here we characterize the accuracy in approximating the difference of Poisson-like counts by a Skellam random variable. Our results fully generalize existing, more limited, results in this direction and, at the same time, our derivations are significantly more concise and elegant. We illustrate the potential impact of these results in the context of problems from network analysis and image processing, where various forms of weak dependence can be expected.

Keywords: Skellam approximation; Stein’s method; Poisson approximation

2010 Mathematics Subject Classification: Primary 62E17

Secondary 60F05; 60J27

1. Introduction

Given two independent Poisson random variables X and Y with means λ_1 and λ_2 , the Skellam distribution, originally attributed to [24], is defined as the distribution of the difference of X and Y . Formally, a random variable W defined on the integers is said to have Skellam distribution with parameters $\lambda_1, \lambda_2 > 0$, which we will denote by $\text{Sk}(\lambda_1, \lambda_2)$, if for all $k \in \mathbb{Z}$,

$$\mathbb{P}(W = k) = e^{-(\lambda_1 + \lambda_2)} \left(\sqrt{\frac{\lambda_1}{\lambda_2}} \right)^k I_k(2\sqrt{\lambda_1 \lambda_2}),$$

where $I_k(2\sqrt{\lambda_1 \lambda_2})$ denotes the modified Bessel function of the first kind with index k and argument $2\sqrt{\lambda_1 \lambda_2}$.

In light of the ubiquity of Poisson-like behavior in nature and the ease with which differencing can arise in data processing, it is perhaps no surprise that the Skellam distribution has seen use in a variety of areas of application. These include application to neural decoding in computational neuroscience [23], denoising [14] and edge detection [16] in image processing, conservation laws in particle physics (see, for example, [12] and [21]), x-ray fluoroscopy in radiology [13],

Received 17 August 2017; revision received 20 March 2018.

* Postal address: Mathematics Department, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208, USA.
Email address: ganhl@math.northwestern.edu

** Postal address: Department of Mathematics and Statistics, Boston University, 111 Cummington Mall, Boston, MA 02215, USA. Email address: kolaczyk@bu.edu

and the identification of genetic variants in bioinformatics [2]. Most recently, the Skellam distribution has been found to have a role in network analysis [4].

In each of these contexts, there are two categories of events being counted and the resulting sums (i.e. denoted X and Y above) are then differenced. The counting, of course, motivates the use of the Poisson distribution in modeling. The events being counted might be the spiking of neurons in two areas of the brain, the arrival of particles in two adjacent detectors in an array, the genetic variants in two nearby regions of the genome, or the presence/absence of a given subgraph across subsets of nodes in a network. Ideally, indicators of these events are independent, both within each type of event category and across the two categories. Independence within is ideal for arguing a Poisson approximation to the counts in each of the two event categories (i.e. in arguing Poisson approximations to the distributions of each of X and Y). At the same time, strictly speaking, independence across the two categories would seem to be necessary, as it is inherent to the definition of the Skellam distribution (i.e. the distribution of $X - Y$).

However, just as it is known that a Poisson approximation to event counts can be accurate under various forms of weak dependence, it is natural to expect that the *difference* of Poisson-like counts might be similarly well-approximated under some form of weak dependence. If the events are *dependent within* each category but *independent between* categories, then formal results of this nature follow from a trivial extension of existing results for Poisson approximation. On the other hand, if events are *dependent between* categories then such results are not immediate.

Motivated by the problem of subgraph counting in noisy networks, where it was noted that such complex dependencies can arise easily, Balachandran *et al.* [4] initiated work on such a general Skellam approximation using Stein's method. However, the results provided in [4] are limited, in that the bounds for the Stein factors therein were derived using a purely analytic approach for the Kolmogorov metric and were restricted to the $\lambda_1 = \lambda_2$ case. In pursuing the same problem of general Skellam approximation here, also using Stein's method, our approach in this paper will use the so-called probabilistic method by exploiting properties of generators of Markov processes, in contrast to the direct analytic approach used in [4]. The main advantages of our approach here are that we can derive bounds for the more general $\lambda_1 \neq \lambda_2$ case, and that the proofs via this approach are significantly easier to derive.

The importance of our work is fundamental in nature, yet it has the potential to be wide-ranging in its practical impact. In each of the application domains described above, there is the very real possibility of general weak dependence among event counts (i.e. both within and between categories). For example, dependencies arise naturally when counting subgraphs in noisy networks, either through dependency in the measurements underlying the construction of the network in the first place or through overlap of vertex subsets while counting [4]. Alternatively, dependency can be expected in particle counts obtained by the types of charge-coupled device (CCD) imaging instruments commonly used in astrophysics, due to so-called spillover effects; see, for example, [3] and [25]. We will expand more on both of these examples in a later section.

There is by now, of course, a large and rich literature on the use of Stein's method to characterize the accuracy of Poisson approximation to event counts, see [8] for a monograph on the topic. However, the focus of this paper is on approximating the *difference* of two Poisson-like counts, which to the best of the authors' knowledge is yet to be studied in depth other than the work of [4]. In [4], the focus was on approximating the distribution of what were termed 'noisy' subgraph counts, i.e. subgraph counts in graphs wherein our knowledge of the

presence/absence status of edges among vertex pairs is uncertain. There the focus was on a centered version of such counts, which was found upon manipulation to yield a difference of two Poisson-like sums and, hence, motivated approximation by Skellam. We use a simple version of the same type of problem as one of two illustrations of our results later in this paper. Nevertheless, as also pointed out by Balachandran *et al.* [4], the use of Stein's method for noisy graphs is different from that used traditionally for random graphs. Stein's method was first introduced to approximation theory for random graphs by Barbour [5], wherein both Poisson and normal approximation results for isolated trees in random graphs were derived. The results for the normal case were expanded in [9] to a variety of applications such as subgraph counts and the number of isolated vertices. For summaries of Stein's method results for random graphs, see [8] and [18]; particularly the former for Poisson approximation results that are more relevant to the work in this paper for obvious reasons.

This paper is organized as follows. In Section 2 we construct our framework for Stein's method for the Skellam distribution, and derive bounds for the relevant Stein factors. In Section 3 we utilise each of these bounds in two example applications: counting subgraphs in noisy networks and counting particles in imaging. Both examples are relatively simple but can be easily generalised. The paper concludes with a discussion of our results and some open questions in Section 4.

2. Stein's method for the Skellam distribution

Our results are derived using an adaptation of the multivariate Poisson approximation. While the Skellam distribution is univariate, the objects we are typically interested in approximating with the Skellam distribution are differences of two random variables. Our approach reflects this by initially considering the bivariate Poisson approximation and then choosing test functions that project down to the univariate case appropriately.

We begin by noting the bivariate Poisson–Stein identity. Note that (X, Y) are said to be bivariate Poisson with parameters (λ_1, λ_2) if X and Y have marginal distributions $\text{Poisson}(\lambda_1)$ and $\text{Poisson}(\lambda_2)$ and are independent.

Lemma 1. (Bivariate Poisson–Stein identity.) *We say that (X, Y) is a bivariate Poisson-distributed random vector with parameters (λ_1, λ_2) if and only if for all functions h in a family of suitable functions, $\mathbb{E}(\mathcal{A}h(X, Y)) = 0$, where*

$$\begin{aligned} \mathcal{A}h(x, y) = & \lambda_1[h(x+1, y) - h(x, y)] + x[h(x-1, y) - h(x, y)] \\ & + \lambda_2[h(x, y+1) - h(x, y)] + y[h(x, y-1) - h(x, y)]. \end{aligned} \quad (1)$$

Details about multivariate Poisson approximation via Stein's method can be found in [6] and [7]. For the Skellam distribution, we seek to modify the bivariate Poisson approximation by considering test functions that depend only upon the difference between X and Y . Noting that we will be abusing notation slightly by often writing bivariate functions that depend only upon the difference as a univariate function, for example $f(x, y) = f(x - y)$, for any function f we define the Stein equation where we set $h_f(x, y) =: h(x, y)$ by

$$\mathcal{A}h_f(x, y) = f(x - y) - \text{Sk}(\lambda_1, \lambda_2)\{f\}, \quad (2)$$

where $\text{Sk}(\lambda_1, \lambda_2)\{f\} := \mathbb{E}f(Z)$ and $Z \stackrel{D}{=} \text{Sk}(\lambda_1, \lambda_2)$. Hence, by taking expectations it is sufficient to find a uniform bound for $\mathbb{E}\mathcal{A}h_f(X, Y)$ to bound $\mathbb{E}f(X - Y) - \text{Sk}(\lambda_1, \lambda_2)\{f\}$ for any f . We will consider all f from the family of test functions corresponding to indicator

functions on the difference of the two coordinates, which encapsulates the total variation (TV) distance. That is, $\mathcal{F}_{TV} = \{f : f(x, y) = \mathbf{1}_A(x - y), A \subset \mathbb{Z}\}$, where $\mathbf{1}_A$ is the indicator function of the event A .

Let $\Delta_i h(x, y) = h((x, y) + \mathbf{e}^{(i)}) - h(x, y)$, where $\mathbf{e}^{(i)}$ denotes a unit vector in coordinate i for $i \in \{1, 2\}$. Also, let $\Delta_{ij}^2 h(x, y) = \Delta_i(\Delta_j h(x, y))$, where $j \in \{1, 2\}$ also. To apply Stein's method successfully, bounds of the correct order are required for the Stein factors

$$\begin{aligned} \|\Delta_i h\| &= \sup_{f \in \mathcal{F}_{TV}} \sup_{x, y} |\Delta_i h(x, y)|, \\ \|\Delta_{ij}^2 h\| &:= \sup_{f \in \mathcal{F}_{TV}} \sup_{x, y} |\Delta_{ij}^2 h(x, y)| \\ &= \sup_{f \in \mathcal{F}_{TV}} \sup_{x, y} |h((x, y) + \mathbf{e}^{(i)} + \mathbf{e}^{(j)}) - h((x, y) + \mathbf{e}^{(i)}) - h((x, y) + \mathbf{e}^{(j)}) \\ &\quad + h(x, y)|. \end{aligned}$$

Theorem 1. For $i, j \in \{1, 2\}$,

$$\|\Delta_i h\| \leq \min \left\{ 1, \sqrt{\frac{2}{e \max\{\lambda_1, \lambda_2\}}} \right\}, \tag{3}$$

$$\|\Delta_{ij}^2 h\| \leq \min \left\{ 1, \frac{1}{2 \max\{\lambda_1, \lambda_2\}^2} + \frac{\sqrt{2} \log^+(\sqrt{2} \max\{\lambda_1, \lambda_2\})}{\max\{\lambda_1, \lambda_2\}} \right\}, \tag{4}$$

where $\log^+(x) = \max\{\log(x), 0\}$. Furthermore,

$$\begin{aligned} \|\Delta_i h\| &\leq \int_0^\infty e^{-t} \max\{1, e^{-(\lambda_1 + \lambda_2)(1 - e^{-t})} I_0((\lambda_1 + \lambda_2)(1 - e^{-t}))\} dt \\ &\sim \sqrt{\frac{2}{\pi(\lambda_1 + \lambda_2)}}, \end{aligned} \tag{5}$$

where the asymptotic equivalence is for when both λ_1 and λ_2 are large.

Proof. Our proof will follow similar ideas and techniques used in the univariate Poisson approximation, for example Lemma 10.2.5 of [8]. Note that we will prove the bounds in the $i = j = 1$ case, and the other cases follow essentially the same proof and, hence, are not included.

It can be shown that for any bounded function f , the (well defined) solution to the Stein equation (2) is

$$h_f(x, y) = - \int_0^\infty [\mathbb{E}f(Z_{x,y}(t)) - \text{Sk}(\lambda_1, \lambda_2)\{f\}] dt,$$

where $Z_{x,y}(t)$ is a Markov process starting at (x, y) and following generator (1). Hence,

$$\Delta_1 h(x, y) = - \int_0^\infty \mathbb{E}[f(Z_{x+1,y}(t)) - f(Z_{x,y}(t))] dt, \tag{6}$$

$$\Delta_{11}^2 h(x, y) = - \int_0^\infty \mathbb{E}[f(Z_{x+2,y}(t)) - f(Z_{x+1,y}(t)) - f(Z_{x+1,y}(t)) + f(Z_{x,y}(t))] dt.$$

We will construct couplings by defining the following *independent* processes:

$D_1(t)$, a pure death process with rate 1 and $D_1(0) = 1$;

$D_2(t)$, a pure death process with rate 1 and $D_2(0) = 1$;

$D_x(t)$, a pure death process with unit-per-capita death rate and $D_x(0) = x$;

$D_y(t)$, a pure death process with unit-per-capita death rate and $D_y(0) = y$;

$Z_0^{\lambda_1}(t)$, an immigration–death process with immigration rate λ_1 , unit-per-capita death rate, and $Z_0^{\lambda_1}(0) = 0$;

$Z_0^{\lambda_2}(t)$, an immigration–death process with immigration rate λ_2 , unit-per-capita death rate, and $Z_0^{\lambda_2}(0) = 0$.

We can define a coupling (see Theorem 2.1 of [7] for more details) such that

$$\begin{aligned} Z_{x+1,y}(t) &:= Z_{x,y}(t) + (D_1(t), 0), \\ Z_{x,y}(t) &:= (Z_0^{\lambda_1}(t), 0) + (0, Z_0^{\lambda_2}(t)) + (D_x(t), 0) + (0, D_y(t)). \end{aligned}$$

Using this coupling, (6) now becomes

$$\begin{aligned} \Delta_1 h(x, y) = & - \int_0^\infty e^{-t} \mathbb{E}[f(Z_0^{\lambda_1}(t) + D_x(t) - Z_0^{\lambda_2}(t) - D_y(t) + D_1(t)) \\ & - f(Z_0^{\lambda_1}(t) + D_x(t) - Z_0^{\lambda_2}(t) - D_y(t)) \mid \mathbf{1}_{\{D_1(t)=1\}}] dt. \end{aligned}$$

Note that if $D_1(t) = 0$ then the two terms in the expectation cancel out. Given $f \in \mathcal{F}_{TV}$, then as f is either 0 or 1, the constant bound is immediate. For the (λ_1, λ_2) dependent bound, the term in the expectation can be evaluated as

$$\begin{aligned} & \sum_{k=0}^\infty \{ \mathbb{E}[f(k + D_x(t) - Z_0^{\lambda_2}(t) - D_y(t) + 1) \mid Z_0^{\lambda_1}(t) = k] \mathbb{P}(Z_0^{\lambda_1}(t) = k) \\ & - \mathbb{E}[f(k + 1 + D_x(t) - Z_0^{\lambda_2}(t) - D_y(t)) \mid Z_0^{\lambda_1}(t) = k + 1] \mathbb{P}(Z_0^{\lambda_1}(t) = k + 1) \} \\ & - f(D_x(t) - Z_0^{\lambda_2}(t) - D_y(t)) \mathbb{P}(Z_0^{\lambda_1}(t) = 0) \\ & = \sum_{k=0}^\infty \mathbb{E}[f(k + 1 + D_x(t) - Z_0^{\lambda_2}(t) - D_y(t))] (\mathbb{P}(Z_0^{\lambda_1}(t) = k) - \mathbb{P}(Z_0^{\lambda_1}(t) = k + 1)) \\ & - f(D_x(t) - Z_0^{\lambda_2}(t) - D_y(t)) \mathbb{P}(Z_0^{\lambda_1}(t) = 0). \end{aligned} \tag{7}$$

Noting that $Z_0^{\lambda_1}(t) \stackrel{D}{=} \text{Poisson}(\lambda_1(1 - e^{-t}))$ (see [15, p. 101]), the above can be bounded by

$$\begin{aligned} \sum_{k=0}^\infty |\mathbb{P}(Z_0^{\lambda_1}(t) = k + 1) - \mathbb{P}(Z_0^{\lambda_1}(t) = k)| + \mathbb{P}(Z_0^{\lambda_1}(t) = 0) &= 2 \max_{x \geq 0} \mathbb{P}(Z_0^{\lambda_1}(t) = x) \\ &\leq 2 \frac{1}{\sqrt{2e\lambda_1(1 - e^{-t})}}, \end{aligned} \tag{8}$$

where the final bound on the Poisson probabilities can be found from Equation (A.2.7) of [8]. Recall that the functions f under consideration are indicator functions on the real line. Now, given that each of the first differences of the Poisson probabilities is multiplied by f in (7), then the worst case for the function f would be to include either all the positive or negative differences from $\mathbb{P}(Z_0^{\lambda_1}(t) = k) - \mathbb{P}(Z_0^{\lambda_1}(t) = k + 1)$. As the bound in (8) contains both the

positive and negative differences, we can drop a factor of 2 in our final bound.

$$\begin{aligned}
 |\Delta_1 h(x, y)| &\leq \int_0^\infty e^{-t} \min \left\{ 1, \frac{1}{\sqrt{2e\lambda_1(1 - e^{-t})}} \right\} dt \\
 &= \int_0^{-\log(1-1/2e\lambda_1)} e^{-t} dt + \int_{-\log(1-1/2e\lambda_1)}^\infty \frac{e^{-t}}{\sqrt{2e\lambda_1(1 - e^{-t})}} dt \\
 &= \sqrt{\frac{2}{e\lambda_1}} - \frac{1}{2e\lambda_1}.
 \end{aligned}$$

The final result in (3) is achieved by noting that instead of conditioning upon $Z_0^{\lambda_1}(t)$ we could equally have conditioned upon $Z_0^{\lambda_2}(t)$ with the same corresponding final result.

For the second bound (5), instead of conditioning upon only $Z_0^{\lambda_1}(t)$, we will condition on both $Z_0^{\lambda_1}(t)$ and $Z_0^{\lambda_2}(t)$. Therefore, similarly to earlier, we need to bound

$$\begin{aligned}
 \sum_{k=0}^\infty \mathbb{E}[f(k + 1 + D_x(t) - D_y(t))][\mathbb{P}(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) = k) - \mathbb{P}(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) = k + 1)] \\
 - f(D_x(t) - D_y(t))\mathbb{P}(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) = 0)
 \end{aligned} \tag{9}$$

and, hence, we need a suitable bound for $\max_k \{\mathbb{P}(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) = k)\}$. Recalling the distributions of $Z_0^{\lambda_1}(t)$ and $Z_0^{\lambda_2}(t)$, this essentially amounts to finding a uniform bound for the maximum of a Skellam distribution. Using the characteristic function inversion formula,

$$\begin{aligned}
 \text{Sk}(\lambda_1, \lambda_2)\{k\} &= \frac{1}{2\pi} \int_{-\pi}^\pi e^{-itk} \exp(\lambda_1(e^{it} - 1)) \exp(\lambda_2(e^{-it} - 1)) dt \\
 &\leq \frac{1}{2\pi} \int_{-\pi}^\pi e^{(\lambda_1 + \lambda_2)(\cos t - 1)} dt \\
 &= e^{-(\lambda_1 + \lambda_2)} I_0(\lambda_1 + \lambda_2),
 \end{aligned}$$

where the last equality follows from Equation (9.6.19) of [1]. The final bound in the theorem is now clear by starting with (9), following the same argument as for the bound which only depends upon λ_1 , and then where a bound is required for $\max_{x \geq 0} \mathbb{P}(Z_0^{\lambda_1}(t) = x)$ in the earlier argument in (8), use the above Skellam bound. The asymptotic result can be derived from the fact that $I_0(z) \sim e^z / \sqrt{2\pi z}$ from Equation (9.7.1) of [1].

The bounds for the second difference are derived in a similar manner, i.e.

$$\begin{aligned}
 \Delta_{11}^2 h(x, y) &= - \int_0^\infty e^{-2t} \mathbb{E}[f(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t) + D_1(t) + D_2(t)) \\
 &\quad - f(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t) + D_1(t)) \\
 &\quad - f(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t) + D_2(t)) \\
 &\quad + f(Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t)) | \mathbf{1}_{\{D_1(t)=D_2(t)=1\}}] dt. \tag{10}
 \end{aligned}$$

Similarly to earlier, we have conditioned upon $D_1(t) = D_2(t) = 1$ in the above equation. Note that, as $f \in \mathcal{F}_{TV}$, we can bound the expectation in the integral by 2. This immediately gives the first of the two bounds in the theorem.

We now work on a (λ_1, λ_2) -dependent bound in a similar fashion as for the first difference. Without loss of generality, assume that $\lambda_1 \geq \lambda_2$. The term in the expectation can be evaluated as follows:

$$\begin{aligned} & \sum_{k=-2}^{\infty} \{ \mathbb{E}[f(k - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t) + 2) \mid Z_0^{\lambda_1}(t) = k] \mathbb{P}(Z_0^{\lambda_1}(t) = k) \\ & - 2\mathbb{E}[f(k + 1 - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t) + 1) \mid Z_0^{\lambda_1}(t) = k + 1] \mathbb{P}(Z_0^{\lambda_1}(t) = k + 1) \\ & + \mathbb{E}[f(k + 2 - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t)) \mid Z_0^{\lambda_1}(t) = k + 2] \mathbb{P}(Z_0^{\lambda_1}(t) = k + 2) \} \\ & = \sum_{k=-2}^{\infty} \{ \mathbb{E}[f(k + 2 - Z_0^{\lambda_2}(t) + D_x(t) - D_y(t))] \\ & \quad \times (\mathbb{P}(Z_0^{\lambda_1}(t) = k) - 2\mathbb{P}(Z_0^{\lambda_1}(t) = k + 1) + \mathbb{P}(Z_0^{\lambda_1}(t) = k + 2)) \}. \end{aligned} \tag{11}$$

Note that in order to bound $\Delta_{12}^2 h(x, y)$, we modify this approach by conditioning on $Z_0^{\lambda_1}(t)$ being equal to $k, k - 1, k + 1, k$, respectively, for the four terms in (10). The other cases follow by symmetry. Given $|f(x)| \leq 1$, the absolute value of the above is bounded by

$$\sum_{k=-2}^{\infty} | \mathbb{P}(Z_0^{\lambda_1}(t) = k) - 2\mathbb{P}(Z_0^{\lambda_1}(t) = k + 1) + \mathbb{P}(Z_0^{\lambda_1}(t) = k + 2) |,$$

which has a natural bound of 2. Recalling $Z_0^{\lambda_1}(t) \stackrel{D}{=} \text{Poisson}(\lambda_1(1 - e^{-t}))$, the above becomes a sum of second differences of the Poisson probabilities. For $X \stackrel{D}{=} \text{Poisson}(\lambda)$,

$$\begin{aligned} \sum_{k=0}^{\infty} |p_k - 2p_{k-1} + p_{k-2}| &= \frac{1}{\lambda^2} \sum_{k=0}^{\infty} p_k |\lambda^2 - 2k\lambda + k(k - 1)| \\ &= \frac{1}{\lambda^2} \mathbb{E}|\lambda^2 - 2X\lambda + X(X - 1)| \\ &\leq \frac{1}{\lambda^2} \sqrt{\mathbb{E}[(\lambda^2 - 2X\lambda + X(X - 1))]^2} \\ &= \frac{\sqrt{2}}{\lambda}, \end{aligned}$$

where the inequality is from Hölder’s inequality. If $\max\{\lambda_1, \lambda_2\} < 1/\sqrt{2}$ then we achieve the constant bound in (4), so assuming $\max\{\lambda_1, \lambda_2\} \geq 1/\sqrt{2}$, this yields

$$\begin{aligned} \|\Delta_{11}^2 h\| &\leq \int_0^\infty e^{-2t} \min \left\{ 2, \frac{\sqrt{2}}{\max\{\lambda_1, \lambda_2\}(1 - e^{-t})} \right\} dt \\ &= \frac{1}{2 \max\{\lambda_1, \lambda_2\}^2} + \frac{\sqrt{2} \log(\sqrt{2} \max\{\lambda_1, \lambda_2\})}{\max\{\lambda_1, \lambda_2\}}. \end{aligned} \quad \square$$

Remark 1. Noting that $\frac{1}{2}(\lambda_1 + \lambda_2) \leq \max\{\lambda_1, \lambda_2\} \leq \lambda_1 + \lambda_2$, we can replace the maximum terms in (3) and (4) with the following more aesthetically pleasing but less sharp bounds:

$$\begin{aligned} \|\Delta_i h\| &\leq \min \left\{ 1, \sqrt{\frac{4}{e(\lambda_1 + \lambda_2)}} \right\}, \\ \|\Delta_{ij}^2 h\| &\leq \min \left\{ 1, \frac{2}{(\lambda_1 + \lambda_2)^2} + \frac{2\sqrt{2} \log^+(\sqrt{2}(\lambda_1 + \lambda_2))}{\lambda_1 + \lambda_2} \right\}. \end{aligned}$$

3. Applications

We illustrate the use of our results on approximation by the Skellam distribution through two applications. Each is a simple caricature of a more complicated application in which such an approximation has been explored in the context of a specific real application. The first pertains to the problem of subgraph counts in noisy networks, as introduced in [4], while the second relates to photon counting devices in image processing.

3.1. Measurement errors in Erdős–Rényi graph edge counts

The analysis of network data is widespread across the scientific disciplines; see, for example, [17], [19], and [22]. In applied network analysis, a common *modus operandi* is to first gather basic measurements relevant to the interactions among elements in a system of interest, then construct a graph-based representation of that system, with nodes serving as elements and links indicating interactions between pairs of elements, and finally summarize the structure of the resulting graph using a variety of numerical and visual tools. See [19, Chapters 3 and 4] for background and several case studies illustrating this process. Key here is the point that the process of network analysis usually rests upon some collection of measurements of a more basic nature and there are usually errors inherent in those measurements. Unfortunately, the uncertainty in approximating some true graph $G = (V, E)$ by some estimated graph $\hat{G} = (V, \hat{E})$, which manifests as errors in our knowledge of the presence/absence of edges between vertex pairs, must necessarily propagate to any estimates of network summaries $\eta(G)$ we seek. Yet currently there is little in the literature by way of formal and principled statistical methodology for dealing with this propagation of error. A natural first step in this direction is a distributional analysis.

This problem was first formalized by Balachandran *et al.* [4], where the focus was on the distribution of subgraph count statistics in noisy networks. And, since it is standard in the applied network analysis literature to cite observed subgraph counts, the quantity studied in [4] was the discrepancy between observed and true subgraph counts. Particular emphasis was placed on the simplest case where the subgraph of interest is an edge, and the corresponding subgraph count, the total number of edges. The statistic of interest therefore was the discrepancy $D = |E| - |\hat{E}|$. Accordingly, we consider the same statistic here, but in the specific case where the true underlying graph G is a classical random graph.

Formally, suppose that G is an Erdős–Rényi random graph with n possible edges (i.e. for notational simplicity, n refers to the number of vertex pairs rather than the number of vertices). This graph is not necessarily a complete graph, but rather each vertex pair has an edge independently with probability p_i . We will denote by U_i , $i \in \{1, \dots, n\}$, the indicator random variable such that $U_i = 1$ if an edge exists between the i th vertex pair.

Motivated by the discussion above, suppose that instead of observing the true graph G , we instead observe a version \hat{G} with errors. Let V_i , $i \in \{1, \dots, n\}$, be the associated edge indicator variable for the observed graph and furthermore set the conditionally independent error probabilities to be

$$\mathbb{P}(V_i = 0 \mid U_i = 1) = r_i, \quad \mathbb{P}(V_i = 1 \mid U_i = 0) = s_i.$$

In this setup, let $U = \sum_{i=1}^n U_i$ and $V = \sum_{i=1}^n V_i$ with V_i independent of U_j , $j \neq i$. In this case, $U - V$ would therefore represent the difference in the number of edges of each graph. That is, $U - V = |E| - |\hat{E}|$. We will aim to explicitly quantify the accuracy of a Skellam approximation for $U - V$.

The details of our problem statement differ slightly from that of [4], in that the true underlying graph G is random, but the spirit remains the same, in that the discrepancy D is the difference of two random variables U and V that are certainly not independent. Furthermore, and a significant departure from [4], we do not require that $\mathbb{E}[U - V] = 0$. Leveraging the main result of this paper, we have the following result.

Theorem 2. *In the above setup, if we set $\lambda_1 = \sum_{i=1}^n r_i p_i$ and $\lambda_2 = \sum_{i=1}^n s_i(1 - p_i)$, then*

$$d_{TV}(\mathcal{L}(U - V), \text{Sk}(\lambda_1, \lambda_2)) \leq \sum_{i=1}^n (p_i r_i + (1 - p_i) s_i)^2 \left[\frac{2}{[\sum_{i=1}^n (p_i r_i + (1 - p_i) s_i)]^2} + \frac{2\sqrt{2} \log(\sqrt{2} \sum_{i=1}^n (p_i r_i + (1 - p_i) s_i))}{\sum_{i=1}^n (p_i r_i + (1 - p_i) s_i)} \right].$$

Proof. The first thing to note that is while we are trying to estimate the difference of U and V , we do not need to consider edges that exist in both random graphs. Let \hat{U} denote the number of edges that are in the true graph but not the observed graph, and similarly let \hat{V} be the number of edges that are not in the true graph but are in the observed graph. In this fashion, $U - V = \hat{U} - \hat{V}$. (As an aside, we note that in [4] the problem was necessarily formulated directly in terms of what we refer to as $\hat{U} - \hat{V}$, since there the true graph G was assumed to be nonrandom.) We similarly define \hat{U}_i and \hat{V}_i as indicators for individual edges, noting that $\mathbb{P}(\hat{U}_i = 1) = p_i r_i$ and $\mathbb{P}(\hat{V}_i = 1) = (1 - p_i) s_i$. We are required to bound

$$\begin{aligned} \mathbb{E} \mathcal{A}h(\hat{U}, \hat{V}) &= \mathbb{E} \left[\sum_{i=1}^n [p_i r_i (h(\hat{U} + 1, \hat{V}) - h(\hat{U}, \hat{V}))] + \hat{U} (h(\hat{U} - 1, \hat{V}) - h(\hat{U}, \hat{V})) \right] \\ &\quad + \mathbb{E} \left[\sum_{i=1}^n [(1 - p_i) s_i (h(\hat{U}, \hat{V} + 1) - h(\hat{U}, \hat{V}))] \right. \\ &\quad \left. + \hat{V} (h(\hat{U}, \hat{V} - 1) - h(\hat{U}, \hat{V})) \right]. \end{aligned} \tag{12}$$

We begin with

$$\begin{aligned} &\mathbb{E}[\hat{U} (h(\hat{U} - 1, \hat{V}) - h(\hat{U}, \hat{V}))] \\ &= \mathbb{E} \sum_{i=1}^n \hat{U}_i (h(\hat{U} - 1, \hat{V}) - h(\hat{U}, \hat{V})) \\ &= \sum_{i=1}^n \mathbb{E}[\hat{U}_i (h(\hat{U} - 1, \hat{V}) - h(\hat{U}, \hat{V})) \mid \hat{U}_i = 1, \hat{V}_i = 0] \mathbb{P}(\hat{U}_i = 1, \hat{V}_i = 0) \\ &= \sum_{i=1}^n p_i r_i \mathbb{E}[h(\hat{U}^{(i)}, \hat{V}^{(i)}) - h(\hat{U}^{(i)} + 1, \hat{V}^{(i)})], \end{aligned}$$

where $\hat{U}^{(i)} = \hat{U} - \hat{U}_i$ and $\hat{V}^{(i)} = \hat{V} - \hat{V}_i$. Hence, the first half of (12) becomes

$$\sum_{i=1}^n p_i r_i \mathbb{E}[(h(\hat{U} + 1, \hat{V}) - h(\hat{U}, \hat{V})) + (h(\hat{U}^{(i)}, \hat{V}^{(i)}) - h(\hat{U}^{(i)} + 1, \hat{V}^{(i)}))]. \tag{13}$$

We now consider three cases:

- (i) $U_i = 0$ and $V_i = 0$;
- (ii) $U_i = 1$ and $V_i = 0$;
- (iii) $U_i = 0$ and $V_i = 1$.

Case (ii) can be termed a false negative, and (iii) a false positive. Note that it is impossible for an edge to be a false positive and false negative at the same time. In case (i), the terms in (13) will cancel out to 0, and in (ii) and (iii) we have exactly a second difference of the function h , and these two cases take probability $p_i r_i$ and $(1 - p_i) s_i$, respectively. Therefore, (13) can be bounded by

$$\|\Delta_{ij}^2 h\| \sum_{i=1}^n p_i r_i (p_i r_i + (1 - p_i) s_i). \tag{14}$$

An analogous argument follows for the second half of (12) and, therefore, the entirety of (12) can be bounded by

$$\|\Delta_{ij}^2 h\| \sum_{i=1}^n (p_i r_i + (1 - p_i) s_i)^2,$$

and the final bound follows from Theorem 1 and Remark 1. □

As a simplification to aid with interpretation of the bound, if we set $p_i = p$, $r_i = r$ and $s_i = s$, the bound becomes

$$\frac{2}{n} + (pr + (1 - p)s) 2\sqrt{2} \log(\sqrt{2n}(pr + (1 - p)s)).$$

The assumption that the error probabilities r_i and s_i are constant across the graph is referred to as a homogeneity assumption in [4]. While likely not strictly true in practice, it is a useful assumption for better illustrating how the relevant aspects of the problem combine to influence the accuracy of approximation by Skellam. If we further assume that λ_1 and λ_2 are equal to some common value, say λ , our setup is then roughly equivalent to that of [4]. This assumption can be viewed as imposing a type of centering on the noise at the level of individual edges, since it dictates that in expectation we have $|E|$ equal to $|\hat{E}|$. In this case, since $rp = s(1 - p) = \lambda/n$, the bound becomes

$$\frac{1}{n} [2 + 4\sqrt{2}\lambda \log(2\sqrt{2}\lambda)].$$

When it is not unreasonable to expect that λ vary with n , we then find that the accuracy of approximation by Skellam in this problem – for this special case – varies like $O(\lambda_n \log(\lambda_n)/n)$.

The method of proof of this bound is unsurprisingly similar to the Poisson approximation of the sum of independent but not necessarily identical Bernoulli trials. In our case, there are essentially three components of the error terms that we would expect to appear: two of them will result from the individual Poisson approximations of \hat{U} and \hat{V} and then there should be a third term which deals with the fact that \hat{U} and \hat{V} are not independent. For the reader familiar with Poisson approximation, it should be clear where the ‘third’ component of the error appears in (14). The difference arises because the conditioning we make upon U_i has ramifications on V_i as they are not independent. One would expect a single univariate Poisson approximation would only have a sum of the $p_i^2 r_i^2$ in (14), but we require the second term in our scenario. However, in some sense, this extra term disappears in the final bound because our Stein factor has both λ_1 and λ_2 in the denominator.

3.2. Haar wavelet coefficients under photon imaging with spillover effects

Current state-of-the-art high-quality imaging applications, such as those encountered in medicine and scientific research, make heavy use of what is known as a charge-coupled device (CCD). A CCD converts electrical charges to digital values. In the context of imaging, these electrical charges in turn derive from the conversion of photons – essentially particles of light – into an electrical signal. Therefore, CCDs (and a variety of other related devices) are central to modern image acquisition and digital image processing, in that by assembling arrays of CCDs and orienting them towards an object of interest it is possible to represent that object through a matrix of photon counts over the individual CCDs in the array.

Ideally, the count in each CCD would be independent of the others and relevant only to a certain corresponding portion of the imaged object. However, for technical reasons, there can be various types of degradation. For example, it is typically the case that photons that should be counted in a given CCD can actually be counted in others. This effect is sometimes referred to as ‘spillover’ and can be thought of as inducing a type of blurring in the image. Standard practice is to calibrate imaging instruments before use, yielding a (usually) probabilistic mapping function that characterizes the blurring. Depending on the extent of such degradation and the application at hand, this may be used, in turn, for deblurring in the image processing stage. See, for example, [12] and [21] for a detailed description of this paradigm in the context of X-ray imaging in astrophysics.

Here we set up a simple caricature of the type of image degradation problem just described, in which a weak dependence among photon counts results. Without loss of generality, we consider a one-dimensional signal rather than a two-dimensional image. In practice, the indexing in this dimension is typically photon energy, rather than photon source location. But the same type of degradation issues can be present. For our signal processing, we consider the use of wavelets, a work-horse in signal and image processing for over 20 years; see [20]. Specifically, both for simplicity and to match most closely the focus of this paper, we consider the use of the Haar wavelet. The result of applying a Haar wavelet transform to a one-dimensional signal is to produce a collection of Haar coefficients which, as the inner product of the wavelet and the signal, are proportional to the difference of the sums of the signal values over two adjacent windows.

Suppose we have n bins (for example, corresponding to CCDs), and note in the following that all defined vectors will be of length n . Let the vector \mathbf{X} be the true signal and suppose that $\mathbf{X} \stackrel{D}{=} \text{Poisson}(\mathbf{f})$, so the $X_i \stackrel{D}{=} \text{Poisson}(f_i)$ and are also independent from each other. Hirakawa and Wolfe [14] showed that both the wavelet and scaling coefficients for the Haar wavelet are distributed as (proportional to) Skellam random variables with parameters comprised of sums and differences of the elements of \mathbf{f} . Set \mathbf{P} where $P_i \in \{0, 1\}$ denote the positive inclusions for a given Haar wavelet coefficient, similarly \mathbf{N} with $N_i \in \{0, 1\}$ for the negative inclusions and $\mathbf{P} + \mathbf{N} = \{0, 1\}^n$, i.e. there is no overlap of the 1’s. Then the Haar wavelet coefficient can be represented as $U - V$, where $U = \mathbf{P} \cdot \mathbf{X}$, $V = \mathbf{N} \cdot \mathbf{X}$ and ‘ \cdot ’ denotes the dot product. Furthermore, $U - V \stackrel{D}{=} \text{Sk}(\mathbf{P} \cdot \mathbf{f}, \mathbf{N} \cdot \mathbf{f})$. In the following we will investigate how measurement errors would impact the distribution of these coefficients.

A simple variant of the type of spillover referred to above, in the context of a one-dimensional signal, is when a particle may actually end up being observed at a lower energy level than its true energy. In our model we will assume that each particle that arrives is independent and there is a probability p that the particle will be observed in exactly one level lower than its true energy. Let Y_i denote the number of particles in bin i that are observed correctly, and Y_i^* denote

the number of particles in bin i that are the result of errors in measurement. That is, Y_i^* is the number of particles of energy level $i + 1$ but are measured at level i .

Due to the thinning property of Poisson random variables, Y and Y^* are independent. Set $U' = P \cdot Y + P \cdot Y^*$ and $V' = N \cdot Y + N \cdot Y^*$. The observed Haar wavelet coefficient satisfies

$$U' - V' \stackrel{D}{=} \text{Sk}((1 - p)P \cdot f + pP \cdot f^{(-1)}, (1 - p)N \cdot f + pN \cdot f^{(-1)}),$$

where $f_i^{(-1)} = f_{i+1}$. Note that we can set $P_{-1} = N_{-1} = 0$ and $f_{n+1} = 0$ to avoid boundary issues. So our question is, what is the difference between these two different Skellam distributions, i.e. between the distributions of the true and observed Haar wavelet coefficients.

Theorem 3. *In the above set up,*

$$d_{\text{TV}}(\mathcal{L}(U' - V'), \mathcal{L}(U - V)) \leq \sqrt{\frac{2p^2}{e \max(P \cdot f, N \cdot f)}} [|P \cdot f - P \cdot f^{(-1)}| + |N \cdot f - N \cdot f^{(-1)}|].$$

Proof. To bound this difference in total variation, we use a simple adaptation of Theorem 1.C(i) of [8]. Using the true distribution of $U - V$ as our ‘reference’ measure, we need to bound $|\mathbb{E}\mathcal{A}(U', V')|$ from (1), where $\lambda_1 = P \cdot f$ and $\lambda_2 = N \cdot f$. Note that using the usual Poisson–Stein identity,

$$\begin{aligned} & \mathbb{E}[U'[h(U' - 1, V') - h(U', V')] \mid V'] \\ &= -((1 - p)P \cdot f + pP \cdot f^{(-1)})\mathbb{E}[h(U' + 1, V') - h(U', V') \mid V'], \end{aligned}$$

therefore to bound the first half of (1),

$$\begin{aligned} & |\mathbb{E}[P \cdot f[h(U' + 1, V') - h(U', V')] + U'[h(U' - 1, V') - h(U', V')]]| \\ &= |((P - (1 - p)P) \cdot f - pP \cdot f^{(-1)})\mathbb{E}[\mathbb{E}[h(U' + 1, V') - h(U', V') \mid V']]| \\ &\leq \|\Delta_i h\|_p |P \cdot f - P \cdot f^{(-1)}|. \end{aligned}$$

An analogous bound can be derived for the second half of (1) and this yields the final result. The proof is complete. □

Note that the bound in the above theorem is larger when, relative to the larger of total signal intensity in the positive or negative window (i.e. the larger of $P \cdot f$ or $N \cdot f$), the discrepancy in those totals resulting from a shift of the windows by 1 is large. That is, when the windows are near a spike or jump in the underlying signal f . Therefore, in particular, the effects of spillover are minimal in regions of the signal that are smooth.

Remark 2. If we want to generalise this result to allow the error probability to be random, for example the error rate for bin i could depend upon X_i , this should in theory be possible by adapting Theorem 1.C(ii) of [8].

4. Discussion

There is one notable drawback in the approach used in this paper. Given that our approach is to project from two dimensions to one using appropriate test functions, this will only be applicable when approximating the difference of two random variables. If one wishes to approximate a single univariate random variable with the Skellam distribution directly, then this approach will not be useful. It remains an open question as to whether a direct one-dimensional approach is possible.

Poisson approximation via the generator method involves characterising the Stein identity as the generator of an immigration–death Markov process where the immigration rate is constant λ and the death rate is unit-per-capita. Such a generator characterises the Poisson distribution as it is the unique stationary distribution of such a process. Intuitively, for the Skellam distribution one would aim to construct a generator defined on the integers such that λ_1 would denote the rate of increase of ‘positive particles’, λ_2 the rate of increase of ‘negative particles’, and then an offsetting death-type rate that would remove particles appropriately, thus ensuring the process does not explode in either direction so that the associated stationary distribution is Skellam. The problem with attempting such a construction, from a one-dimensional viewpoint, is that if we only know the difference between the two counts of positive and negative particles, this is not enough information to properly define the transition rates of the process. For example, if we know that the difference of the two counts is 0, there are infinitely many possibilities for the number of positive and negative particles, and to properly define the process we need to know how many positive and negative particles there are. The problem described above with constructing an appropriate one-dimensional generator for the process is what leads us to believe that a one-dimensional approach is not possible using the generator method, however we concede that it is possible that there may exist a generator representation that would be amenable to analysis.

An interesting question is whether there exists a nice clean bound for the first difference of h of the order $1/\sqrt{\lambda_1 + \lambda_2}$ as opposed to our two bounds in (3) and (5). Our bound (5) was derived via the inversion formula for characteristic functions. The ‘usual’ method that is used in Poisson approximation does not seem viable in the Skellam scenario, primarily because it involves finding a uniform bound for the maximum of the Poisson mass function in terms of λ . For the Skellam distribution, one might suspect an analogous approach, however given we have one quantity to bound but two parameters to work with, this method seems unfruitful. We expect that it should be possible to find such a bound, and this remains an interesting open problem.

Similarly, for the second difference, our bound involving the maximum of λ_1 and λ_2 should be able to have all the maximum terms replaced with the sum of the two parameters without the penalty invoked in Remark 1. The correct way to derive such a bound would be to condition upon the difference $Z_0^{\lambda_1}(t) - Z_0^{\lambda_2}(t)$ in (11) rather than just one of the two processes. This would ultimately require a bound upon the sum of the absolute second differences of Skellam probabilities. For $p_k = \mathbb{P}(Z = k)$, where $Z \stackrel{D}{=} \text{Sk}(\lambda_1, \lambda_2)$, numerical results indicate that $\sum_k |p_k - 2p_{k-1} + p_k| \leq 1/(\lambda_1 + \lambda_2)$, which intuitively makes sense, given the Poisson bound, as both λ_1 and λ_2 will ‘flatten’ out the mass function as they increase. However, we were unable to prove such a result, as the Bessel functions proved to be not very tractable.

It is worth comparing our bound for the second difference to the Stein solution (4) to the corresponding bounds derived in Theorem 4 of [4], where it was shown that $\|\Delta_{11}^2 h\| \leq 160/2\lambda$, but limited to the case where $\lambda = \lambda_1 = \lambda_2$ and for the Kolmogorov metric. However, given the test functions for the total variation distance and Kolmogorov distance are not completely dissimilar, a comparison is still worthwhile. Our bound in this paper is of order $\log(\lambda_1 + \lambda_2)/(\lambda_1 + \lambda_2)$, so for very large λ_1, λ_2 this will fare worse. However, our constant is much better so this will only be worse on very large λ_1 and λ_2 . And, obviously, our bounds have the significant added flexibility of not requiring $\lambda_1 = \lambda_2$.

In light of the bounds of order $1/(\lambda_1 + \lambda_2)$ in [4], an interesting question is whether our bounds in this paper of $\log(\lambda_1 + \lambda_2)/(\lambda_1 + \lambda_2)$ are of the right order. Given that Barbour [7] has shown that for the multivariate Poisson approximation, the Stein factors are of strict

order $\log(\lambda_1 + \lambda_2)/(\lambda_1 + \lambda_2)$, and since our approach involves adapting the bivariate Poisson approximation on specific test functions, we believe that our order may be the best possible using our approach.

Using the generator approach, the standard bound for the Poisson–Stein factor (see, for example, Corollary 2.12 of [11]) involves a coupling based upon hitting times of an immigration–death process. However, this coupling is difficult to use in the multivariate case as hitting times become significantly more complicated when there are multiple dimensions. Logarithmic terms are quite common in Poisson related approximation theory, such as multivariate Poisson as discussed above, and also for process approximation where it has been shown that logarithmic terms are strictly necessary if we wish to use uniform bounds for the Stein factors [10]. Our approach in this paper has both aspects of multivariate and univariate analysis, multivariate in the sense that we are essentially considering a special case of bivariate Poisson approximation, but the ultimate target is Skellam which is univariate. As a result, it is not clear what the correct order should be. We would lean towards the correct order not including a logarithmic term, but such a bound is likely beyond the methods used in this paper. Whether a direct analytic or alternative approach would yield a better result is unknown.

Acknowledgements

We would like to thank two anonymous referees for their careful reading of this manuscript and their helpful comments and suggestions. This work was supported in part by the Air Force Office of Scientific Research (grant number 12RSL042). We would also like to thank Nathan Ross for pointing out a slightly sharper bound for the second differences of the Poisson probabilities used in the proof for Theorem 1.

References

- [1] ABRAMOWITZ, M. AND STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Nat. Bureau Standards Appl. Math. Ser. **55**). U.S. Government Printing Office, Washington, D.C.
- [2] ALTMANN, A. *et al.* (2011). VipR: variant identification in pooled DNA using R. *Bioinformatics* **27**, 77–84.
- [3] ATHIRAY, P. S., SREEKUMAR, P., NARENDRANATH, S. AND GOW, J. P. D. (2015). Simulating charge transport to understand the spectral response of swept charge devices. *Astronomy Astrophys.* **583**, A97.
- [4] BALACHANDRAN, P., KOLACZYK, E. D. AND VILES, W. D. (2017). On the propagation of low-rate measurement error to subgraph counts in large networks. *J. Mach. Learn. Res.* **18**, 61.
- [5] BARBOUR, A. D. (1982). Poisson convergence and random graphs. *Math. Proc. Camb. Philos. Soc.* **92**, 349–359.
- [6] BARBOUR, A. D. (1988). Stein’s method and Poisson process convergence. In *A Celebration of Applied Probability* (J. Appl. Prob. Spec. Vol. **25A**), Applied Probability Trust, Sheffield, pp. 175–184.
- [7] BARBOUR, A. D. (2005). Multivariate Poisson-binomial approximation using Stein’s method. In *Stein’s Method and Applications* (Lect. Notes Ser. Inst. Math. Sci. Nat. Univ. Singap. **5**). Singapore University Press, pp. 131–142.
- [8] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation* (Oxford Studies Prob. **2**). Oxford University Press.
- [9] BARBOUR, A. D., KAROŃSKI, M. AND RUCIŃSKI, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory B* **47**, 125–145.
- [10] BROWN, T. C. AND XIA, A.-H. (1995). On Stein-Chen factors for Poisson approximation. *Statist. Prob. Lett.* **23**, 327–332.
- [11] BROWN, T. C. AND XIA, A. (2001). Stein’s method and birth-death processes. *Ann. Prob.* **29**, 1373–1403.
- [12] BZDAK, A., KOCH, V. AND SKOKOV, V. (2013). Baryon number conservation and the cumulants of the net proton distribution. *Phys. Rev. C* **87**, 014901.
- [13] CESARELLI, M. *et al.* (2013). X-ray fluoroscopy noise modeling for filter design. *Internat. J. Comput. Assisted Radiology Surgery* **8**, 269–278.
- [14] HIRAKAWA, K. AND WOLFE, P. J. (2012). Skellam shrinkage: wavelet-based intensity estimation for inhomogeneous Poisson data. *IEEE Trans. Inf. Theory* **58**, 1080–1093.

- [15] HOEL, P. G., PORT, S. C. AND STONE, C. J. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin, Boston, MA.
- [16] HWANG, Y., KIM, J.-S. AND KWEON, I.-S. (2007). Sensor noise modeling using the Skellam distribution: application to the color edge detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007*, IEEE.
- [17] JACKSON, M. O. (2008). *Social and Economic Networks*. Princeton University Press.
- [18] JANSON, S., ŁUCZAK, T. AND RUCINSKI, A. (2000). *Random Graphs*. John Wiley, New York.
- [19] KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer, New York.
- [20] MALLAT, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA.
- [21] MORITA, K., FRIMAN, B., REDLICH, K. AND SKOKOV, V. (2013). Net quark number probability distribution near the chiral crossover transition. *Phys. Rev. C* **88**, 034903.
- [22] NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
- [23] SHIN, H.-C. *et al.* (2010). Neural decoding of finger movements using Skellam-based maximum-likelihood decoding. *IEEE Trans. Biomed. Eng.* **57**, 754–760.
- [24] SKELLAM, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. R. Statist. Soc. (N.S.)* **109**, 296.
- [25] TOWNSLEY, L. K., BROOS, P. S., GARMIRE, G. P. AND NOUSEK, J. A. (2000). Mitigating charge transfer inefficiency in the *Chandra x-ray observatory* advanced CCD imaging spectrometer. *Astrophys. J. Lett.* **534**, L139.