ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 122 (2012) 3033–3065

stochastic processes and their applications

www.elsevier.com/locate/spa

On the asymptotics of locally dependent point processes

Aihua Xia[a,*], Fuxi Zhang[b]

[a] *Department of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia*
[b] *School of Mathematical Sciences, Peking University, Beijing 100871, China*

Received 6 August 2011; received in revised form 25 April 2012; accepted 21 May 2012
Available online 27 May 2012

Abstract

We investigate a family of approximating processes that can capture the asymptotic behaviour of locally dependent point processes. We prove two theorems presented to accommodate respectively the positively and negatively related dependent structures. Three examples are given to illustrate that our approximating processes can circumvent the technical difficulties encountered in compound Poisson process approximation (see Barbour and Månsson (2002) [10]) and our approximation error bound decreases when the mean number of the random events increases, in contrast to the increasing of bounds for compound Poisson process approximation.

MSC: primary 60G55; secondary 60E15

Keywords: Polynomial birth–death point process; Poisson process; Compound Poisson process; The Wasserstein metric; Stein's factors

1. Introduction

Random events in space and time often exhibit a locally dependent structure. When the events are very rare and the dependent structure is not too complicated, a natural approach is to declump the events into clusters and then approximate the positions of the clusters by a suitable Poisson process and the sizes of the clusters by independent and identically distributed random elements, as is well documented in [1]. Consequently, compound Poisson and marked Poisson processes are often widely accepted as the 'best approximate models' for clustered rare events.

* Corresponding author. Tel.: +61 3 8344 4247; fax: +61 3 8344 4599.
E-mail addresses: aihuaxia@unimelb.edu.au, xia@ms.unimelb.edu.au (A. Xia), zhangfxi@math.pku.edu.cn (F. Zhang).

doi:10.1016/j.spa.2012.05.011

The first attempt to estimate the errors of Poisson process approximation seems to go back to [15] with errors measured in the total variation distance, while the errors in the Lévy–Prohorov distance were not studied until [22,25] (see also [32]). All of these studies are based on the stochastic calculus approach with a filtration, a compensator and coupling techniques as the tools used to quantify the distances. Barbour and Brown [4], clearly inspired by the success of Stein's method in multivariate Poisson approximation [3], laid down a general framework for using Stein's method to estimate the Poisson process approximation errors. Their framework can be well adjusted for errors expressed in terms of Janossy densities, Palm distributions and compensators (see [5,34]). In terms of compound Poisson process approximation, there seems to have been no major advance until Arratia et al. [2] replaced the original point process with a new one carrying the information of locations and cluster sizes separately so that the Stein–Chen method for Poisson approximation could be employed to obtain useful error bounds. There are enormous advantages for this approach if one can successfully declump the point process, but the procedure of declumping is far from obvious in applications. By contrast, Barbour and Månsson [10] avoided declumping totally by setting up a framework of Stein's method such that the quality of approximation can be studied directly, and the authors summarized that the direct approach 'has conceptual advantages, but entails technical difficulties' on p. 1492. One of the main difficulties is that Stein's factors, like their counterparts for compound Poisson random variable approximation (see [6,11,12]), are generally too crude to use unless more conditions are imposed such as that the compound Poisson process is very close to a Poisson process. An immediate consequence is that the error bounds obtained often deteriorate when the mean of the point process increases, i.e., more information is available. On the other hand, using the improved estimates for Stein's factors for Poisson process approximation in [34] (cf [16]), Chen and Xia [20] managed to produce error estimates for Poisson process approximation to short range dependent rare events and the estimates will remain small (but not improve either) when the average number of events increases.

It is well-known that the central limit theorem often exhibits the *large sample property*, i.e. the larger the sample size, the better the approximation, as evidenced by the Berry–Esseen bound (see [19]). If we are interested in the total counts of rare and weakly dependent events, the Poisson law of small numbers is the cornerstone of the area. However, the Poisson approximation error does not enjoy the large sample property when more rare events are counted [7]. The shortcoming is due to the fact that a Poisson distribution has only one parameter to fiddle with while a normal distribution has two parameters. When more parameters are introduced, this property can be recovered (see [26,24,18,13,17,28]). In fact, Brown and Xia [17] discovered a large family of distributions that can achieve the same purpose.

The success of compound Poisson process approximation essentially hinges on the fact that the events are very rare. It is tempting to ask whether the approximation theory is still valid when the events are less rare, more heavily dependent and the mean number of events increases. One way to tackle this problem is to keep the approximating process as a Poisson process but weaken the metric for quantifying the difference between point processes [30]. The weaker metric will naturally limit its applicability. The second approach is to introduce more parameters into the approximating point process models. To put the idea into practice, Xia and Zhang [35] introduced a family of point process counterparts of approximating distributions suggested in [17], and named them the polynomial birth–death point processes, or *PBDP* in short. In particular, Xia and Zhang [35] bounded the distance between the Bernoulli process with a constant success probability and a suitable PBDP in terms of the Wasserstein distance (defined in Section 2 below; see also [4]). The assumption of the constant success probability plays the crucial role there because the symmetric structure enables the authors to construct a suitable coupling for directly comparing the two distributions. The pilot study shows that, for the Bernoulli process with the same success probability, it is possible to recover the large sample property for PBDP approximation. The purpose of this paper is to demonstrate that the large sample property prevails among a large group of point processes when these PBDP are used as approximating models. To this end, we set up the Stein equation of PBDP approximation and establish its Stein factors so that one can directly estimate the difference between the distribution of a general point process and that of a PBDP.

Our paper is arranged as follows. In Section 2, we briefly review the polynomial birth–death point processes introduced in [35], lay down a foundation of Stein's method for their approximation and conclude the section with estimates of Stein's factors in terms of the Wasserstein metric. To make our paper reader-friendly, we postpone the technical proofs of Stein's factors to Section 5. Section 3 is devoted to point processes with locally dependent structures which are analogous to those in [19]. We state two theorems for error estimates of PBDP approximations, separately for positively and negatively related dependence. The proofs of these theorems are rather complicated so we leave them to the last two sections (Sections 6 and 7) of the paper. Examples are provided in Section 4 to illustrate the key steps of applying the main theorems.

2. Stein's method for polynomial birth–death point processes

The family of approximating distributions in [17] was introduced through the invariant distributions of birth–death processes. For ease of use, they focused on the birth and death rates as the polynomial functions of the states of the process, and consequently called the invariant distribution the *polynomial birth–death distribution*. More precisely, let

$$\alpha_k = a + bk, \quad \forall\, k \ge 0; \qquad \beta_k = k + \beta k(k-1), \quad \forall\, k \ge 0, \tag{2.1}$$

where $a > 0$, $0 \le b < 1$, $\beta \ge 0$. A birth–death process with birth rates $\{\alpha_k\}$ and death rates $\{\beta_k\}$ must be ergodic. As in [17], we let $Z_n(\cdot) := \{Z_n(t) : t \ge 0\}$ be such a process with initial value n and use $\pi_{a,b;\beta}$, or simply π when there is no confusion, to stand for the invariant distribution.

Let Γ be a compact metric space with metric d_0 bounded by 1 and Borel σ-algebra $\mathscr{B}(\Gamma)$ generated by d_0. Set $U, U_1, U_2, \ldots$ as independent and identically distributed Γ-valued random elements with distribution μ. In this paper, the expression $\sum_{i=1}^{X} \delta_{U_i}$ always implies that the nonnegative integer random variable X is independent of $\{U_i : i \ge 1\}$. We call $\mathbf{Z}$ a *polynomial birth–death point process* (see [35]) if it can be expressed as

$$\mathbf{Z} = \sum_{i=1}^{Z} \delta_{U_i}$$

for $Z \sim \pi_{a,b;\beta}$, and denote $\mathscr{L}(\mathbf{Z})$ by $\boldsymbol{\pi}_{a,b;\beta;\mu}$ or simply $\boldsymbol{\pi}$ when there is no confusion. One of the major advantages for using $\boldsymbol{\pi}_{a,b;\beta;\mu}$ as the approximating distribution is that if the approximation is good enough, it is easy to write down its likelihood function, so many classical statistical tools can be easily implemented for finding the estimates of parameters and making statistical inferences. We now give a few examples to illustrate that the definition is a natural extension of the polynomial birth–death distribution.

Example 1. Suppose Z follows Binomial(n, p); then $\mathbf{Z}$ reduces to a binomial process.

Example 2. If Z is a Poisson random variable with mean a, then $\mathbf{Z}$ becomes a Poisson process on Γ with mean measure $a\mu$.

Example 3. When Z has a negative binomial distribution, we call $\mathbf{Z}$ a *negative binomial process*.

Remark 2.1. There are two possible ways to define a negative binomial process. The one that we defined here does not have the property of independent increments, while if we define it as a compound Poisson process with clusters following a logarithmic distribution, then it does have the property of independent increments. Nevertheless, the two distributions converge when the intensity of the Poisson component becomes large (see Remark 4.8 below).

Now we construct a Markov process with invariant distribution $\boldsymbol{\pi} = \boldsymbol{\pi}_{a,b;\beta;\mu}$. Allowing repeats of points, each finite integer-valued measure on Γ can be written as $\xi = \sum_{i=1}^{n} \delta_{x_i}$. Since the points $x_1, \ldots, x_n$ are not necessarily distinct, we introduce the notation $\wr x_1, \ldots, x_n \int$ to stand for the collection of the n points. In this paper, we do not distinguish $\sum_{i=1}^{n} \delta_{x_i}$ from the collection $\wr x_1, \ldots, x_n \int$, or a configuration with n particles respectively located at $x_1, \ldots, x_n$. For example, when we say a site/point x or a particle at x in ξ, it means that $\xi(\{x\}) \geq 1$.

For each measure ξ on Γ, we denote its total mass by $|\xi|$. Let $\mathscr{H}$ be the class of all possible finite integer-valued measures (also known as the configurations of point processes) on Γ and let $\mathscr{B}(\mathscr{H})$ be the smallest σ-algebra in $\mathscr{H}$ making the mappings $\xi \mapsto \xi(C)$ measurable for all relatively compact Borel sets $C \subset \Gamma$. For each suitable measurable function h on $\mathscr{H}$, we define

$$\begin{aligned}
\mathscr{A}h(\xi) &:= \big(a + b|\xi|\big) \int_{\Gamma} \big(h(\xi + \delta_x) - h(\xi)\big)\mu(dx) \\
&\quad + \big(1 + \beta(|\xi| - 1)\big) \int_{\Gamma} \big(h(\xi - \delta_x) - h(\xi)\big)\xi(dx) \\
&= \big(a + b|\xi|\big) \left(\mathbb{E}h(\xi + \delta_U) - h(\xi)\right) \\
&\quad + \big(1 + \beta(|\xi| - 1)\big)|\xi| \left(\mathbb{E}h(\xi - \delta_{V(\xi)}) - h(\xi)\right),
\end{aligned} \tag{2.2}$$

where, for $\xi = \sum_{i=1}^{n} \delta_{x_i}$, $V(\xi)$ is a uniformly distributed random element on the collection $\wr x_1, \ldots, x_n \int$. In other words, $V(\xi)$ is equally likely to be one of $x_1, \ldots, x_n$. A particle system $\mathbf{Z}_\xi(\cdot) := \{\mathbf{Z}_\xi(t) : t \geq 0\}$ with the generator $\mathscr{A}$ and the initial configuration ξ evolves as follows:

- with rate a a new particle immigrates to Γ and settles at a site according to μ;
- with rate b an existing particle gives birth, and the new born particle is also located at a site chosen according to μ;
- with rate 1, an existing particle commits suicide;
- with rate β, an existing particle kills another existing particle.

We call such a Markov process a *birth–death system*. It is not difficult to check that the birth–death system has the unique invariant distribution $\boldsymbol{\pi}_{a,b;\beta;\mu}$. Noting that for any $\xi \in \mathscr{H}$, $\{|\mathbf{Z}_\xi(t)| : t \geq 0\}$ is a birth–death process with rates (2.1), we have $\mathscr{L}(|\mathbf{Z}_\xi(\cdot)|) = \mathscr{L}(Z_{|\xi|}(\cdot))$. Therefore, $\mathscr{L}\left(\mathbf{Z}_\xi(t)\right) = \mathscr{L}\left(\sum_{i=1}^{Z_n(t)} \delta_{U_i}\right)$ if $\mathscr{L}(\xi) = \mathscr{L}\left(\sum_{i=1}^{n} \delta_{U_i}\right)$. In particular, we have $\mathscr{L}\left(\mathbf{Z}_\emptyset(t)\right) = \mathscr{L}\left(\sum_{i=1}^{Z_0(t)} \delta_{U_i}\right)$.

Bearing in mind the Stein equation suggested by [4], the natural choice of the Stein equation for the generator $\mathscr{A}$ is

$$\mathscr{A}h(\xi) = f(\xi) - \boldsymbol{\pi}(f) \tag{2.3}$$

for suitable functions f on $\mathscr{H}$, where $\boldsymbol{\pi}(f) := \int f(\xi)\boldsymbol{\pi}(d\xi)$. We now consider the question of the existence of an h that solves the Eq. (2.3).

Proposition 2.2. *For any bounded function f on $\mathscr{H}$,*

$$h_f(\xi) := -\int_0^\infty \big(\mathbb{E}f(\mathbf{Z}_\xi(t)) - \boldsymbol{\pi}(f)\big)dt$$

is well defined, and is a solution of (2.3).

Proof. Let $\{U_i\}$ be independent μ-distributed random elements which are independent of $\{\mathbf{Z}_\xi(t) : t \geq 0\}$. For the pair $\{U_i,\ 1 \leq i \leq |\xi|\}$ with the points in ξ, define $\xi' = \sum_{i=1}^{|\xi|} \delta_{U_i}$, and construct $\{\mathbf{Z}_{\xi'}(t) : t \geq 0\}$ from $\{\mathbf{Z}_\xi(t) : t \geq 0\}$ by replacing the points in ξ with the paired counterparts in ξ'. Let $\tilde{\tau}$ be the last death time of all of the points in ξ. We have

$$\int_0^\infty |\mathbb{E}f(\mathbf{Z}_\xi(t)) - \mathbb{E}f(\mathbf{Z}_{\xi'}(t))|dt \leq \int_0^\infty \mathbb{E}\big(2\|f\|1_{\{\tilde{\tau}>t\}}\big)dt = 2\|f\|\mathbb{E}\tilde{\tau} < \infty,$$

since $\tilde{\tau}$ is stochastically smaller than the maximum of $|\xi|$ independent and identically distributed exp(1) random variables.

Next, define $\bar{f}(n) = \mathbb{E}f(\sum_{i=1}^n \delta_{U_i})$ for all $n \geq 0$; then

$$\int_0^\infty \left|\mathbb{E}f(\mathbf{Z}_{\xi'}(t)) - \boldsymbol{\pi}(f)\right| dt \leq \int_0^\infty \left|\mathbb{E}\bar{f}(Z_{|\xi|}(t)) - \pi(\bar{f})\right| dt < \infty$$

due to the positive recurrence of the Markov chain $\{Z_{|\xi|}(t),\ t \geq 0\}$. Hence,

$$\begin{aligned}
&\int_0^\infty |\mathbb{E}f(\mathbf{Z}_\xi(t)) - \boldsymbol{\pi}(f)|dt \\
&\quad \leq \int_0^\infty |\mathbb{E}f(\mathbf{Z}_\xi(t)) - \mathbb{E}f(\mathbf{Z}_{\xi'}(t))|dt + \int_0^\infty \left|\mathbb{E}f(\mathbf{Z}_{\xi'}(t)) - \boldsymbol{\pi}(f)\right| dt < \infty,
\end{aligned}$$

which implies that h_f is well-defined.

To establish (2.3), let $\boldsymbol{\tau}_\xi = \inf\{t : \mathbf{Z}_\xi(t) \neq \xi\}$, which has an exponential distribution with parameter $\alpha_{|\xi|} + \beta_{|\xi|}$. Then

$$\begin{aligned}
h_f(\xi) &= -\int_0^\infty \big(\mathbb{E}f(\mathbf{Z}_\xi(t)) - \boldsymbol{\pi}(f)\big)dt \\
&= -\big(f(\xi) - \boldsymbol{\pi}(f)\big)\mathbb{E}\boldsymbol{\tau}_\xi - \mathbb{E}\int_{\boldsymbol{\tau}_\xi}^\infty \big(\mathbb{E}f(\mathbf{Z}_\xi(t)) - \boldsymbol{\pi}(f)\big)dt \\
&= -\frac{f(\xi) - \boldsymbol{\pi}(f)}{\alpha_{|\xi|} + \beta_{|\xi|}} + \mathbb{E}h(\mathbf{Z}_\xi(\boldsymbol{\tau}_\xi)) \\
&= -\frac{f(\xi) - \boldsymbol{\pi}(f)}{\alpha_{|\xi|} + \beta_{|\xi|}} \\
&\quad + \frac{\alpha_{|\xi|}\int_\Gamma h(\xi + \delta_x)\mu(dx) + \big(1 + \beta(|\xi| - 1)\big)\int_\Gamma h(\xi - \delta_x)\xi(dx)}{\alpha_{|\xi|} + \beta_{|\xi|}},
\end{aligned}$$

and (2.3) follows by rearranging the above equation. □

The metric used for quantifying the differences of two point processes is defined as follows (see [4]). Let $\mathscr{K}$ be the class of d_0-Lipschitz functions u on Γ such that $|u(x)-u(y)| \le d_0(x, y)$ for all $x, y \in \Gamma$. For any two measures ρ_1 and ρ_2 on Γ, define

$$d_1(\rho_1, \rho_2) = \begin{cases} 0, & \text{if } |\rho_1| = |\rho_2| = 0, \\ \dfrac{1}{|\rho_1|} \sup_{u \in \mathscr{K}} \left| \int_\Gamma u d\rho_1 - \int_\Gamma u d\rho_2 \right|, & \text{if } |\rho_1| = |\rho_2| \ne 0, \\ 1, & \text{if } |\rho_1| \ne |\rho_2|. \end{cases}$$

For any configurations $\xi = \sum_{i=1}^n \delta_{x_i}$ and $\eta = \sum_{i=1}^n \delta_{y_i} \in \mathscr{H}$ with $n \ge 1$, $d_1(\xi, \eta)$ can be represented as

$$d_1(\xi, \eta) = \min_\sigma \frac{1}{n} \sum_{i=1}^n d_0(x_i, y_{\sigma(i)}),$$

where the minimum is taken over all permutations σ of $(1, \ldots, n)$. The Wasserstein metric d_2 between point process distributions is defined as

$$d_2(\mathbf{P}, \mathbf{Q}) := \sup_f |\mathbf{P}(f) - \mathbf{Q}(f)| = \inf_{\xi \sim \mathbf{P}, \eta \sim \mathbf{Q}} \mathbb{E} d_1(\xi, \eta),$$

where the supremum is taken over all functions in

$$\mathscr{F} := \{ f : |f(\xi) - f(\eta)| \le d_1(\xi, \eta),\ \forall\, \xi, \eta \in \mathscr{H} \},$$

and the last equation is due to the duality theorem (see [27, p. 168]). The metric d_2 is a particular kind of metric from the well-known family of Wasserstein metrics. It is worth pointing out that, since $d_1 \le 1$, all functions in $\mathscr{F}$ are bounded and Proposition 2.2 ensures the existence of solutions of Stein's equation (2.3) for these functions. Historically, the Wasserstein metrics were motivated by the classical Monge transportation problem. In our context, we will handle the 'transportation problem' in two steps, i.e. forming 'sandpiles' by assembling local points to designated centres and then transporting the 'sandpiles' of the point process being approximated to the corresponding 'sandpiles' of the PBDP.

The following lemma is often useful for comparing two different approximating polynomial birth–death point processes.

Lemma 2.3. *We have*

$$d_2\left(\boldsymbol{\pi}_{a_1,b_1;\beta_1;\mu_1}, \boldsymbol{\pi}_{a_2,b_2;\beta_2;\mu_2}\right) \le d_{tv}(\pi_{a_1,b_1;\beta_1}, \pi_{a_2,b_2;\beta_2}) + d_1(\mu_1, \mu_2),$$

where for two probability measures Q_1 and Q_2 on $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$,

$$d_{tv}(Q_1, Q_2) := \sup_{A \subset \mathbb{Z}_+} |Q_1(A) - Q_2(A)|.$$

Proof. Using the Kantorovich–Rubinstein duality theorem [27, Theorem 8.1.1, p. 168], we can couple together $Z_1 \sim \pi_{a_1,b_1;\beta_1}$, $Z_2 \sim \pi_{a_2,b_2;\beta_2}$, and two sequences of Γ-valued random elements $\tau_{1i} \sim \mu_1$ and $\tau_{2i} \sim \mu_2$, $i \ge 1$, such that

$$d_{tv}\left(\pi_{a_1,b_1;\beta_1}, \pi_{a_2,b_2;\beta_2}\right) = \mathbb{P}(Z_1 \ne Z_2),$$
$$\mathbb{E} d_0(\tau_{1i}, \tau_{2i}) = d_1(\mu_1, \mu_2) \quad \text{for all } i \ge 1,$$

and $\{(\tau_{1i}, \tau_{2i}),\ i \geq 1\}$ are independent and independent of (Z_1, Z_2). Then

$$\begin{aligned} d_2\left(\boldsymbol{\pi}_{a_1,b_1;\beta_1;\mu_1}, \boldsymbol{\pi}_{a_2,b_2;\beta_2;\mu_2}\right) &\leq \mathbb{E} d_1\left(\sum_{i=1}^{Z_1}\delta_{\tau_{1i}}, \sum_{i=1}^{Z_2}\delta_{\tau_{2i}}\right) \\ &\leq \mathbb{P}(Z_1 \neq Z_2) + \mathbb{E}\left\{d_1\left(\sum_{i=1}^{Z_1}\delta_{\tau_{1i}}, \sum_{i=1}^{Z_2}\delta_{\tau_{2i}}\right)\middle| Z_1 = Z_2\right\}\mathbb{P}(Z_1 = Z_2) \\ &\leq d_{tv}\left(\pi_{a_1,b_1;\beta_1}, \pi_{a_2,b_2;\beta_2}\right) + \mathbb{E}\left\{\frac{1}{Z_1}\sum_{i=1}^{Z_1} d_0\left(\tau_{1i}, \tau_{2i}\right)\middle| Z_1 = Z_2\right\}\mathbb{P}(Z_1 = Z_2) \\ &\leq d_{tv}(\pi_{a_1,b_1;\beta_1}, \pi_{a_2,b_2;\beta_2}) + d_1(\mu_1, \mu_2), \end{aligned}$$

completing the proof. □

In applications of Stein's equation, one will encounter the following quantities:

$$C_n := \sup\{|h_f(\xi + \delta_x) - h_f(\xi + \delta_y)| : f \in \mathscr{F}, \xi \in \mathscr{H}, |\xi| = n\}, \tag{2.4}$$

with $C_{-1} := 0$,

$$\Delta_2 h(\xi; x, y) := h(\xi + \delta_x + \delta_y) - h(\xi + \delta_x) - h(\xi + \delta_y) + h(\xi), \quad \xi \in \mathscr{H},\ x, y \in \Gamma,$$

and

$$\Delta_2 h(\xi) := \sup\{|\Delta_2 h(\xi; x, y)| :\ x, y \in \Gamma\}.$$

The following estimates, often known as Stein's factors, are usually needed in applying Stein's method. In fact, the success of Stein's method is centred around the quality of these estimates.

Theorem 2.4. (i) *For $n \geq 0$,*

$$C_n \leq \min\left\{1, \frac{1}{2(n+1)} + \frac{1}{a}, \frac{1}{(a \wedge b)(n+1)}\right\}. \tag{2.5}$$

(ii) *For any $f \in \mathscr{F}, \xi \in \mathscr{H}$,*

$$\Delta_2 h_f(\xi) \leq \frac{2}{|\xi| + 1} + \frac{5}{a}. \tag{2.6}$$

Remark 2.5. The estimates in Theorem 2.4 are of the correct order. In fact, if we take $\beta = b = 0$, the PBDP becomes a Poisson process and the estimates for the Poisson process are known to be of the correct order (see [34]).

3. Locally dependent point processes

A *point process* Ξ on Γ is defined as a measurable mapping of some fixed probability space into $(\mathscr{H}, \mathscr{B}(\mathscr{H}))$ and $\lambda(dx) = \mathbb{E}\Xi(dx)$ is said to be the *intensity* or *mean measure* of Ξ [23, pp. 13–14]. A point process is said to be *simple* if it has at most one point at each location. For a point process Ξ on Γ with finite mean measure λ, the family of point processes $\{\Xi_x : x \in \Gamma\}$ are said to be *reduced Palm processes* associated with Ξ (at $x \in \Gamma$) if for any measurable function $f : \Gamma \times \mathscr{H} \to \mathbb{R}_+ := [0, \infty)$,

$$\mathbb{E}\left(\int_\Gamma f(x, \Xi - \delta_x)\Xi(dx)\right) = \int_\Gamma \mathbb{E} f(x, \Xi_x)\lambda(dx) \tag{3.1}$$

(Kallenberg [23, Chapter 10]). Intuitively, the reduced Palm distribution $\mathscr{L}(\Xi_x)$ is defined through the Radon–Nikodym derivative as follows:

$$\mathbb{P}(\Xi_x \in B) = \frac{\mathbb{E}[\Xi(dx)1_{\{\Xi-\delta_x\in B\}}]}{\mathbb{E}\Xi(dx)}, \quad \text{for all } B \in \mathscr{B}(\mathscr{H}).$$

When Ξ is a simple point process, it can be interpreted as the distribution of Ξ save one point at x conditional on there being one point at x.

In this paper, we also need the *second-order reduced Palm processes* Ξ_{xy} of the point process Ξ at $x,\ y \in \Gamma$ defined as the processes satisfying

$$\mathbb{E}\left(\iint_{\Gamma^2} f(x, y; \Xi - \delta_x - \delta_y)\Xi(dx)(\Xi - \delta_x)(dy)\right) = \iint_{\Gamma^2} \mathbb{E}f(x, y; \Xi_{xy})\lambda^{[2]}(dx, dy) \tag{3.2}$$

for any measurable function $f : \Gamma^2 \times \mathscr{H} \to \mathbb{R}_+$, where $\lambda^{[2]}(dx, dy) = \mathbb{E}\Xi(dx)(\Xi - \delta_x)(dy)$ is called the *second-order factorial moment measure* of Ξ [23, Section 12.3]. The second-order reduced Palm distribution $\mathscr{L}(\Xi_{xy})$ can also be viewed as the Radon–Nikodym derivative

$$\mathbb{P}(\Xi_{xy} \in B) = \frac{\mathbb{E}[\Xi(dx)(\Xi - \delta_x)(dy)1_{\{\Xi-\delta_x-\delta_y\in B\}}]}{\mathbb{E}\Xi(dx)(\Xi - \delta_x)(dy)}, \quad \text{for all } B \in \mathscr{B}(\mathscr{H}).$$

For $\xi \in \mathscr{H}$ and a Borel set $B \subset \Gamma$, we denote as $\xi|_B$ the restriction of ξ to B, i.e. $\xi|_B(C) = \xi(B \cap C)$ for all Borel sets $C \subset \Gamma$. We call $\{A_x : x \in \Gamma\}$ *a type-I neighbourhood* if $x \in A_x \in \mathscr{B}(\Gamma)$ for all $x \in \Gamma$ and the mapping

$$\Gamma \times \mathscr{H} \to \mathscr{H} : (x, \xi) \mapsto \xi|_{A_x^c}$$

is product measurable (see an equivalent statement and further discussions in [20, pp. 2547–2548]). We say that $\{A_{xy} : x, y \in \Gamma\}$ is *a type-II neighbourhood* if $\{x, y\} \subset A_{xy} \in \mathscr{B}(\Gamma)$ for all $x, y \in \Gamma$ and the mapping

$$\Gamma^2 \times \mathscr{H} \to \mathscr{H} : ((x, y), \xi) \mapsto \xi|_{A_{xy}^c}$$

is product measurable. We now define the locally dependent structures studied in this paper.

Definition 3.1. A point process Ξ is said to satisfy the *type-I local dependence* if there exist two type-I neighbourhoods $\{A_x : x \in \Gamma\}$ and $\{B_x : x \in \Gamma\}$ such that $A_x \subset B_x, \mathscr{L}\left(\Xi_x|_{A_x^c}\right) = \mathscr{L}\left(\Xi|_{A_x^c}\right), \Xi|_{B_x^c}$ is independent of $\Xi|_{A_x}$, and $\Xi_x|_{B_x^c}$ is independent of $\Xi_x|_{A_x}$ for all $x \in \Gamma$. A point process Ξ is said to satisfy the *type-II local dependence* if there exist two type-II neighbourhoods $\{A_{xy} : x, y \in \Gamma\}$ and $\{B_{xy} : x, y \in \Gamma\}$ such that $A_{xy} \subset B_{xy}, \mathscr{L}\left(\Xi_{xy}|_{A_{xy}^c}\right) = \mathscr{L}\left(\Xi|_{A_{xy}^c}\right), \Xi|_{B_{xy}^c}$ is independent of $\Xi|_{A_{xy}}$, and $\Xi_{xy}|_{B_{xy}^c}$ is independent of $\Xi_{xy}|_{A_{xy}}$ for all $x, y \in \Gamma$.

The locally dependent structures introduced here are parallel to, but a little stronger than, those in [19]. The condition $\mathscr{L}\left(\Xi_x|_{A_x^c}\right) = \mathscr{L}\left(\Xi|_{A_x^c}\right)$ can be loosely interpreted as $\Xi(dx)$ being independent of $\Xi|_{A_x^c}$. One may easily establish sufficient conditions for the locally dependent structures by imposing conditions on neighbourhoods containing balls (see the descriptive definitions in [14]).

To state the error estimates of the PBDP approximation to locally dependent point processes, we need to introduce the following notation. Let $\mathcal{G} = \{G_1, \dots, G_k\} \subset \mathscr{B}(\Gamma)$ be a partition of Γ, and choose $t_i \in \Gamma$ such that $\sup_{s\in G_i} d_0(s, t_i)$ is as small as possible, $i = 1, \dots, k$. Note that t_i, regarded as the 'designated centre' of the set G_i, is not necessarily in G_i. We define $\mathscr{M}_{\mathcal{G}} \circ \eta := \sum_{i=1}^{k} \eta(G_i)\delta_{t_i}$ for $\eta \in \mathscr{H}$. The mapping is to 'assemble' all the points of the configuration η in each G_i to its centre t_i. If we set $d_0(\mathcal{G})$ as

$$d_0(\mathcal{G}) = \max_{1\le i\le k} \sup_{s\in G_i} d_0(s, t_i),$$

then it is easy to check that

$$d_1(\eta, \mathscr{M}_{\mathcal{G}} \circ \eta) \le d_0(\mathcal{G}). \tag{3.3}$$

Let u be a positive constant to be chosen in applications, and we take $u = 2$ for our examples in Section 4. Let $\mathscr{F}_{TV}$ be the set of indicator functions of all sets in $\mathscr{B}(\mathscr{H})$. For a point process Ξ, we define

$$\begin{aligned} r_x(\Xi) &:= 4\mathbb{P}\left(\Xi(B_x^c) + 1 \le \frac{a}{u}\,\middle|\, \Xi|_{B_x}\right) \\ &\quad + \frac{4u+10}{a} \max_{1\le j\le k} \sup_{f\in\mathscr{F}_{TV}} \left|\mathbb{E}\left[f\left(\mathscr{M}_{\mathcal{G}} \circ (\Xi|_{B_x^c})\right)\right.\right. \\ &\quad - \left.\left. f\left(\mathscr{M}_{\mathcal{G}} \circ (\Xi|_{B_x^c}) + \delta_{t_j}\right)\middle|\, \Xi|_{B_x}\right]\right|. \end{aligned}$$

Similarly, $\bar{r}_x(\Xi)$ is defined by replacing all the conditional expectations/probability in the definition of $r_x(\Xi)$ with expectations/probability. It is worth pointing out that the type-I local dependence implies $\bar{r}_x(\Xi) = \bar{r}_x(\Xi_x)$. Let

$$\begin{aligned} \epsilon_{1,x}(\Xi) &= r_x(\Xi)\Xi(A_x)\Xi(B_x \setminus A_x) \\ &\quad + \bar{r}_x(\Xi)\big[\Xi(A_x) + 1\big]\Xi(A_x)/2 + \Xi(A_x)\mathbb{E}[r_x(\Xi)\,\Xi(B_x)], \\ \epsilon_{1,x}(\Xi_x) &= r_x(\Xi_x)\Xi_x(A_x)\Xi_x(B_x \setminus A_x) \\ &\quad + \bar{r}_x(\Xi_x)\big[\Xi_x(A_x) + 1\big]\Xi_x(A_x)/2 + \Xi_x(A_x)\mathbb{E}[r_x(\Xi)\,\Xi(B_x)], \\ \epsilon_{2,x}(\Xi) &= r_x(\Xi)\Xi(B_x \setminus A_x) + \bar{r}_x(\Xi) + \mathbb{E}[r_x(\Xi)\Xi(B_x)], \\ \epsilon_{2,x}(\Xi_x) &= r_x(\Xi_x)\Xi_x(B_x \setminus A_x) + \bar{r}_x(\Xi_x) + \mathbb{E}[r_x(\Xi)\Xi(B_x)]. \end{aligned}$$

In terms of the type-II local dependence, we define $r_{x,y}$ and $\bar{r}_{x,y}$ in the same way as r_x and $\bar{r}_x$ respectively, but with B_x replaced by B_{xy}. We then set

$$\begin{aligned} \epsilon_{1,x,y}(\Xi) &= r_{x,y}(\Xi)\Xi(A_{xy})\Xi(B_{xy} \setminus A_{xy}) + \bar{r}_{x,y}(\Xi)(\Xi(A_{xy}) + 1)\Xi(A_{xy})/2 \\ &\quad + \Xi(A_{xy})\mathbb{E}[r_{x,y}(\Xi)\,\Xi(B_{xy})], \\ \epsilon_{1,x,y}(\Xi_{x,y}) &= r_{x,y}(\Xi_{x,y})\Xi_{x,y}(A_{xy})\Xi_{x,y}(B_{xy} \setminus A_{xy}) \\ &\quad + \bar{r}_{x,y}(\Xi_{x,y})(\Xi_{x,y}(A_{xy}) + 1)\Xi_{x,y}(A_{xy})/2 \\ &\quad + \Xi_{x,y}(A_{xy})\mathbb{E}[r_{x,y}(\Xi)\,\Xi(B_{xy})], \\ \epsilon_{2,x,y}(\Xi) &= r_{x,y}(\Xi)\Xi(B_{xy}) + \bar{r}_{x,y}(\Xi) + \mathbb{E}\big[r_{x,y}(\Xi)\,\Xi(B_{xy})\big], \\ \epsilon_{2,x,y}(\Xi_{x,y}) &= r_{x,y}(\Xi_{x,y})\Xi_{x,y}(B_{xy}) + \bar{r}_{x,y}(\Xi_{x,y}) + \mathbb{E}\big[r_{x,y}(\Xi)\,\Xi(B_{xy})\big]. \end{aligned}$$

Theorem 3.2. *Assume that the point process Ξ on Γ with finite mean measure λ satisfies* $\mathrm{Var}(|\Xi|) \ge \mathbb{E}|\Xi|$ *and the type-*I *local dependence. Let $\nu(dx) = \lambda(dx)/|\lambda|, b = [$*Var$(|\Xi|) -$

$\mathbb{E}|\Xi|]/\mathrm{Var}(|\Xi|)$, $a = (1-b)|\lambda|$; *then*

$$d_2(\mathscr{L}(\Xi), \boldsymbol{\pi}_{a,b;0;\nu}) \le 2d_0(\mathcal{G}) + \int_\Gamma \mathbb{E}\big[(1+b)(\epsilon_{1,y}(\Xi_y) + \epsilon_{1,y}(\Xi)) + b\bar{r}_y(\Xi)\Xi_y(A_y) + b\epsilon_{2,y}(\Xi_y)\big]\lambda(dy).$$

Theorem 3.3. *Assume that the point process Ξ on Γ with finite mean measure λ satisfies* $\mathrm{Var}(|\Xi|) < \mathbb{E}|\Xi|$, *the type-*I *and type-*II *local dependence. Let*

$$\beta = \frac{|\lambda| - \mathrm{Var}(|\Xi|)}{|\lambda| - \mathrm{Var}(|\Xi|) + \mathbb{E}|\Xi|^3 - (|\lambda|+1)\mathbb{E}|\Xi|^2}, \qquad a = |\lambda| + \beta(\mathbb{E}|\Xi|^2 - |\lambda|), \tag{3.4}$$

and

$$\nu(dx) = \frac{1}{a}\left(\lambda(dx) + \beta\int_{y\in\Gamma}\lambda^{[2]}(dx, dy)\right). \tag{3.5}$$

If $\beta \ge 0$, then

$$d_2(\mathscr{L}(\Xi), \boldsymbol{\pi}_{a,0;\beta;\nu}) \le 2d_0(\mathcal{G}) + \int_\Gamma \mathbb{E}\left(\epsilon_{1,x}(\Xi_x) + \epsilon_{1,x}(\Xi)\right)\lambda(dx) + \beta\iint_{\Gamma^2}\mathbb{E}\left(\epsilon_{1,x,y}(\Xi_{xy}) + \epsilon_{1,x,y}(\Xi) + \epsilon_{2,x,y}(\Xi_{xy})\right)\lambda^{[2]}(dx, dy).$$

Remark 3.4. When one applies these theorems, it is advisable to leave the choice of $\mathcal{G}$ to the last stage so that an optimal bound with the best possible order can be achieved.

A less noticeable fact is that if one takes $d_0(x, y) = 0$ for every x and y, i.e. a pseudometric on Γ, and $\mathcal{G} = \{\Gamma\}$, then d_2 reduces to d_{tv} for the total counts of point processes, so our theorems also cover the PBDP approximation to the total counts of locally dependent point processes in the total variation distance.

The proofs of the two theorems will be given in Sections 6 and 7. In the next section, let us look at three examples to see how the theorems perform in applications.

4. Applications

4.1. Bernoulli process

Let $\Gamma = [0, 1]$, $d_0(x, y) = |x - y|$, and $I_1, \ldots, I_n$ be independent Bernoulli random variables with

$$\mathbb{P}(I_i = 1) = 1 - \mathbb{P}(I_i = 0) = p_i, \quad 1 \le i \le n.$$

Define $\Xi = \sum_{i=1}^n I_i\delta_{i/n}$. This simple point process is particularly useful for proving the Poisson process limit theorems for the extreme value theory [21, Chapter 5]. It was proved in [33, Proposition 3.6] (see also [29]) that the accuracy of Poisson process approximation to $\mathscr{L}(\Xi)$ is of order $\sum_{i=1}^n p_i^2/\sum_{i=1}^n p_i$ and the order cannot be improved when n becomes large. For when the p_i's are equal to p, Xia and Zhang [35], making use of the symmetric nature of the distribution $\mathscr{L}(\Xi)$, proved that an appropriate PBDP can approximate $\mathscr{L}(\Xi)$ with approximation error of order $\left(\frac{1}{n}+p\right)\wedge\frac{1}{\sqrt{np}}$. However, when the p_i's are not the same, the techniques employed in [35] will not work and we demonstrate below that our theorems can be applied to this case.

First of all, it is easy to verify that Ξ has mean measure $\lambda(dx)=\sum_{i=1}^n p_i\delta_{i/n}(dx)$ and its second-order factorial moment measure is $\lambda^{[2]}(dx,dy)=\sum_{1\le i\ne j\le n} p_ip_j\delta_{i/n}(dx)\delta_{j/n}(dy)$. Clearly, $\mathbb{E}|\Xi|>\mathrm{Var}(|\Xi|)$, so we can apply Theorem 3.3 to estimate the approximation error for $\mathscr{L}(\Xi)$.

To identify the approximating PBDP distribution, we let

$$\lambda_j=\sum_{i=1}^n p_i^j,\ j\ge 2,$$

$$\beta=\frac{\lambda_2}{|\lambda|^2-\lambda_2-2|\lambda|\lambda_2+2\lambda_3},$$

$$a=|\lambda|+\beta(|\lambda|^2-\lambda_2)$$

(cf [17, Theorem 3.1]) and

$$\begin{aligned}\nu(dx)&=\frac{1}{a}\left(\lambda(dx)+\beta\int_{y\in\Gamma}\lambda^{[2]}(dx,dy)\right)\\&=\frac{1}{a}\left(\lambda(dx)+\beta\sum_{i=1}^n(|\lambda|-p_i)p_i\delta_{i/n}(dx)\right).\end{aligned}$$

Next, we set up an appropriate partition $\mathcal{G}$ of $\Gamma=\{G_1,\ldots,G_k\}$. Let $1\le u_1,\ldots,u_k\le n$ be such that $u_1+\cdots+u_k=n$, $s_0=0$, $s_j=s_{j-1}+u_j$ for $1\le j\le k$. Set $G_1=[0,s_1/n]$ and $G_j=\left(\frac{s_{j-1}}{n},\frac{s_j}{n}\right]$ for $2\le j\le k$. We choose t_j as the middle point of the interval G_j, $1\le j\le k$, so that $d_0(\mathcal{G})=\max_{1\le j\le k}u_j/(2n)$. Define $W_j=\sum_{i=s_{j-1}+1}^{s_j}I_i$, $1\le j\le k$ and

$$\begin{aligned}\kappa&:=\max_{1\le j\le k}\max_{s_{j-1}+1\le l_1\ne l_2\le s_j}d_{tv}(\mathscr{L}(W_j-I_{l_1}-I_{l_2}),\mathscr{L}(W_j-I_{l_1}-I_{l_2}+1))\\&\le\max_{1\le j\le k}\max_{s_{j-1}+1\le l_1\ne l_2\le s_j}1\wedge\frac{1}{2\sqrt{\sum_{l=s_{j-1}+1}^{s_j}p_l(1-p_l)-p_{l_1}(1-p_{l_1})-p_{l_2}(1-p_{l_2})}},\end{aligned}$$

where the inequality is due to Lemma 1 of [9]. We take $A_x=B_x=\{x\}$, $A_{xy}=B_{xy}=\{x,y\}$, $u=2$; then $\Xi_x(A_x)=\Xi_x(B_x)=\Xi_{xy}(A_{xy})=\Xi_{xy}(B_{xy})=0$,

$$\mathbb{P}\left(|\Xi|-I_{l_1}-I_{l_2}\le\frac{a}{2}\right)\le O(|\lambda|^{-2}),$$

and hence all of $r_x,\bar r_x,r_{x,y}$ and $\bar r_{x,y}$ are bounded by $O(\kappa/a)$. Applying Theorem 3.3 gives the following estimate.

Theorem 4.1. *With the above set-up, if $\beta\ge 0$, then*

$$d_2(\mathscr{L}(\Xi),\boldsymbol{\pi}_{a,0;\beta;\nu})\le\max_{1\le j\le k}u_j/n+O(\kappa\lambda_2/|\lambda|).$$

Remark 4.2. The bound in Theorem 4.1 is sharp in the following sense. Let $n=m^2$, $0.2\le p_i\le 0.3$ for $1\le i\le m$, $p_i=0$ for $m+1\le i\le n$ and $u_j=m$ for $1\le j\le m$;

then Theorem 4.1 implies that $d_2(\mathscr{L}(\Xi), \boldsymbol{\pi}_{a,0;\beta;\nu}) = O(m^{-1/2})$. However, it was shown in Theorem 3.1 of [17] that $d_{tv}(\mathscr{L}(|\Xi|), \pi_{a,0;\beta}) = O(m^{-1/2})$ and one can easily verify that $d_2(\mathscr{L}(\Xi), \boldsymbol{\pi}_{a,0;\beta;\nu}) \geq d_{tv}(\mathscr{L}(|\Xi|), \pi_{a,0;\beta})$.

As a special case, we now assume that the p_i's are equal to p, and take $k = O\big((n(1-p)/p)^{1/3}\big)$, $u_j = O\left((pn^2/(1-p))^{1/3}\right)$, $j = 1, \ldots, k$, then

$$\kappa = O\left(1 \wedge \frac{1}{(np^2(1-p))^{1/3}}\right).$$

Hence, the following corollary is immediate.

Corollary 4.3. *For the Bernoulli point process $\Xi = \sum_{i=1}^n I_i \delta_{i/n}$, where the $\{I_i,\ 1 \leq i \leq n\}$ are independent and identically distributed Bernoulli random variables with $\mathbb{P}(I_1 = 1) = p$, let $\beta = \frac{1}{(n-1)(1-2p)}$, $a = np(1-p)/(1-2p)$, $\nu(dx) = \frac{1}{n}\sum_{i=1}^n \delta_{i/n}(dx)$; then*

$$d_2(\mathscr{L}(\Xi), \boldsymbol{\pi}_{a,0;\beta;\nu}) \leq O\left(\frac{p^{1/3}}{(n(1-p))^{1/3}}\right), \tag{4.1}$$

provided $p < 1/2$.

Remark 4.4. The bound (4.1) is not as good as the bound $O\left(\left(\frac{1}{n} + p\right) \wedge \frac{1}{\sqrt{np}}\right)$ derived in [35] when p is fixed and n becomes large. This is due to the fact that our method does not rely on the specific symmetric structure of the Bernoulli process Ξ and the bound is of the most general nature.

Remark 4.5. A Poisson process approximation to the Bernoulli process is justified when $p \to 0$ and $np \to \lambda$. However, in applications of extreme value theory, the value p is often fixed while n is large, so our theory provides a more practical alternative.

4.2. The compound Poisson process

Barbour and Månsson [10] considered compound Poisson process approximation in d_2 distance. The Stein factors for both compound Poisson random variable and process approximations are generally too crude to use unless they are sufficiently close to their Poisson counterparts or satisfy some other restrictive conditions. In this example, we will show that our PBDP, suitably chosen, will converge to the compound Poisson process when its cluster distribution is fixed, and has a finite third moment, and the mean of the Poisson process component becomes large, regardless of whether the compound Poisson process is sufficiently close to a Poisson process or not.

To begin with, let $\Xi = \sum_{i=1}^\infty iX_i$, where the $\{X_i\}$ are independent Poisson processes on Γ with mean measures $\{\mu_i\}$ respectively. For brevity, we write $\Xi \sim \mathrm{CP}(\mu_1, \mu_2, \ldots)$. It is easy to see that $\mathrm{Var}(|\Xi|) \geq \mathbb{E}|\Xi|$ with equality holds if and only if $\mu_j = 0$ for all $j \geq 2$.

Suppose that we have a partition $\mathcal{G} = \{G_1, \ldots, G_k\}$ of Γ.

Theorem 4.6. *Let $\lambda(dx) = \sum_{i=1}^\infty i\mu_i(dx)$, $\nu(dx) = \lambda(dx)/|\lambda|$, and*

$$b = \frac{\sum_{i=2}^\infty i(i-1)|\mu_i|}{\sum_{i=1}^\infty i^2|\mu_i|}, \qquad a = \frac{|\lambda|^2}{\sum_{i=1}^\infty i^2|\mu_i|}.$$

Then

$$d_2(\mathrm{CP}(\mu_1, \mu_2, \ldots), \boldsymbol{\pi}_{a,b;0;\nu}) \le O\left(\max_{1\le i\le k} 1 \wedge \frac{1}{\sqrt{\mu_1(G_i)}}\right) \frac{\sum_{i=1}^{\infty} i^3|\mu_i|}{a} + 2d_0(\mathcal{G}). \tag{4.2}$$

Remark 4.7. Suppose the cluster distribution is fixed everywhere and $\mu_1(G) \to \infty$ for every $G \in \mathscr{B}(\Gamma)$ such that $\mu_1(G) > 0$; then the upper bound given in (4.2) has the order $o(1)$. To this end, one can partition Γ into sets with small enough diameters; then for each set G_i with $\mu_1(G_i) > 0$, one can find $\mu_1(G_i)$ as large as one wishes. Furthermore, suppose that Γ is a simply connected domain in $\mathbb{R}^d$ with smooth boundary, $d_0(x, y) = |x - y| \wedge 1$, and μ_1 is proportional to the Lebesgue measure, i.e. points are homogeneous on Γ. Then, the upper bound given in (4.2) has the order $O\left(|\mu_1|^{-\frac{1}{d+2}}\right)$. As a matter of fact, one can partition Γ into boxes with the same diameter of order $O\left(|\mu_1|^{-\frac{1}{d+2}}\right)$, then combine the parts at the boundary of Γ to their adjacent boxes totally belonging to Γ, to obtain $\mathcal{G}$.

Remark 4.8. Alternatively to using the definition given in Example 3 we can define a negative binomial process as a compound Poisson process having a Poisson process of clusters and each cluster carrying a random number of points that follows a logarithmic distribution. Remark 4.7 ensures that if the logarithmic distribution for the clusters is fixed and the Poisson process is homogeneous, then the process will converge to our PBDP distribution when the mean measure of the Poisson process becomes large.

Proof of Theorem 4.6. A measure μ is called *diffuse* if for every point $x \in \Gamma$, $\mu(\{x\}) = 0$. If $\{\mu_i\}$ are not diffuse, we can enlarge the space Γ if necessary and take diffuse measures μ_i^n such that $|\mu_i^n| = |\mu_i|$ for $i \ge 1$ and $\max_{i\ge1} d_1(\mu_i^n, \mu_i) \to 0$ as $n \to \infty$. We then apply the Kantorovich–Rubinstein duality theorem [27, Theorem 8.1.1, p. 168] to couple two sequences of Γ-valued random elements $\tau_{ij} \sim \mu_i/|\mu_i|$ and $\tau_{ij}^n \sim \mu_i^n/|\mu_i^n|$, $i, j \ge 1$, such that

$$\mathbb{E}d_0(\tau_{ij}, \tau_{ij}^n) = d_1\left(\mu_i/|\mu_i|, \mu_i^n/|\mu_i^n|\right) = d_1\left(\mu_i, \mu_i^n\right),$$

and $\{(\tau_{ij}, \tau_{ij}^n),\ i, j \ge 1\}$ are independent and independent of $\{X_i,\ i \ge 1\}$. Let $\Xi^n \sim \mathrm{CP}(\mu_1^n, \mu_2^n, \ldots)$; then

$$\begin{aligned}
d_2(\mathscr{L}(\Xi), \mathscr{L}(\Xi^n)) &\le \mathbb{E}d_1\left(\sum_{i=1}^{\infty} i \sum_{j=1}^{|X_i|} \delta_{\tau_{ij}}, \sum_{i=1}^{\infty} i \sum_{j=1}^{|X_i|} \delta_{\tau_{ij}^n}\right) \\
&\le \mathbb{E}\left(\frac{\sum_{i=1}^{\infty} i \sum_{j=1}^{|X_i|} d_0\left(\tau_{ij}, \tau_{ij}^n\right)}{\sum_{i=1}^{\infty} i|X_i|}\right) \\
&\le \mathbb{E}\left(\frac{\sum_{i=1}^{\infty} i \sum_{j=1}^{|X_i|} d_1\left(\mu_i, \mu_i^n\right)}{\sum_{i=1}^{\infty} i|X_i|}\right) \\
&\le \max_{i\ge1} d_1(\mu_i^n, \mu_i).
\end{aligned}$$

This observation, together with Lemma 2.3, ensures that we can assume, without loss of generality, that the $\{\mu_i\}$ are all diffuse. Otherwise, we can approximate each Ξ^n with a suitable PBDP distribution and then take the limits.

Direct computation gives

$$|\lambda| = \sum_{i=1}^{\infty} i|\mu_i|, \qquad \mathrm{Var}(|\Xi|) = \sum_{i=1}^{\infty} i^2|\mu_i|,$$
$$b = \frac{\sum_{i=2}^{\infty} i(i-1)|\mu_i|}{\sum_{i=1}^{\infty} i^2|\mu_i|}, \qquad a = \frac{|\lambda|^2}{\sum_{i=1}^{\infty} i^2|\mu_i|}.$$

Because the compound Poisson process has independent increments, we let $A_x = B_x = \{x\}$; then

$$\begin{aligned} r_x(\Xi) &= \bar{r}_x(\Xi) = r_x(\Xi_x) = \bar{r}_x(\Xi_x) \\ &= 4\mathbb{P}\left(|\Xi| + 1 \le \frac{a}{u}\right) \\ &\quad + \frac{4u+10}{a} \max_{1\le j\le k} d_{TV}\left(\mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ \Xi), \mathscr{L}\left(\mathscr{M}_{\mathcal{G}} \circ \Xi + \delta_{t_j}\right)\right), \end{aligned}$$

where for any two point process distributions $\mathbf{P}$ and $\mathbf{Q}$ on $\mathscr{H}$, $d_{TV}(\mathbf{P}, \mathbf{Q}) := \inf_{\xi\sim\mathbf{P},\eta\sim\mathbf{Q}} \mathbb{P}(\xi \neq \eta)$. Noting that $\{\mu_i\}$ are all diffuse and consequently $\Xi(\{x\}) = 0$ a.s. for each $x \in \Gamma$, we have

$$\begin{aligned} &\epsilon_{1,x}(\Xi) = 0, \qquad \epsilon_{1,x}(\Xi_x) = \bar{r}_x(\Xi_x)(\Xi_x(\{x\}) + 1)\Xi_x(\{x\})/2, \\ &\epsilon_{2,x}(\Xi_x) = \bar{r}_x(\Xi_x). \end{aligned} \tag{4.3}$$

If $\{Y_i,\ i \ge 1\}$ are independent Poisson random variables, we can construct a maximal coupling of (Y_1', Y_1'') on the same probability space such that $\mathscr{L}(Y_1') = \mathscr{L}(Y_1)$, $\mathscr{L}(Y_1'') = \mathscr{L}(Y_1 + 1)$,

$$d_{tv}(\mathscr{L}(Y_1), \mathscr{L}(Y_1 + 1)) = \mathbb{P}(Y_1' \neq Y_1'')$$

(see [8, p. 254]). Enlarging the probability space if necessary, we can construct a copy $\{Y_i',\ i \ge 2\}$ of $\{Y_i,\ i \ge 2\}$ such that $\{Y_i',\ i \ge 2\}$ are independent of (Y_1', Y_1''); then

$$\begin{aligned} &d_{tv}\left(\mathscr{L}\left(\sum_{i=1}^{\infty} iY_i\right), \mathscr{L}\left(\sum_{i=1}^{\infty} iY_i + 1\right)\right) \\ &\quad \le \mathbb{P}\left(Y_1' + \sum_{i=2}^{\infty} iY_i' \neq Y_1'' + \sum_{i=2}^{\infty} iY_i'\right) = \mathbb{P}(Y_1' \neq Y_1'') = d_{tv}(\mathscr{L}(Y_1), \mathscr{L}(Y_1 + 1)) \\ &\quad = \max_{j\ge 0} \mathbb{P}(Y_1 = j) \le \frac{1}{\sqrt{2e\mathbb{E}(Y_1)}}, \end{aligned}$$

where the last inequality follows from Proposition A.2.7 of [8, p. 262]. Hence, we have

$$\begin{aligned} d_{TV}\big(\mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ \Xi), \mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ \Xi + \delta_{t_j})\big) &\le d_{tv}\big(\mathscr{L}(\Xi(G_j)), \mathscr{L}(\Xi(G_j) + 1)\big) \\ &\le \frac{1}{\sqrt{2e\mu_1(G_j)}} \le \frac{1}{\sqrt{\mu_1(G_j)}}. \end{aligned}$$

It is easy to see that we can write $|\Xi| = \sum_{i=1}^{V} \eta_i$, where all the random variables V and η_i's are independent, $V \sim \text{Poisson}(|\mu'|)$ with $\mu' = \sum_{i=1}^{\infty} \mu_i$, and the η_i's have the same distribution $\mathbb{P}(\eta_i = j) = |\mu_j|/|\mu'|,\ j \geq 1$. If we take $u = 2$, noting that $a \leq |\mu'|$, we have

$$\mathbb{P}\left(|\Xi| + 1 \leq \frac{a}{2}\right) \leq \mathbb{P}\left(V \leq \frac{|\mu'|}{2}\right) \leq O(|\mu'|^{-2}) \leq O(a^{-2}).$$

Hence,

$$\bar{r}_x(\Xi_x) = O\left(a^{-1} \max_{1 \leq i \leq k} 1 \wedge \frac{1}{\sqrt{\mu_1(G_i)}}\right). \tag{4.4}$$

Using the independent increments again, we get

$$\begin{aligned}
\text{Var}(|\Xi|) &= \mathbb{E}\int_\Gamma (|\Xi| - |\lambda|)\Xi(dx) = \int_\Gamma \mathbb{E}(|\Xi_x| + 1 - |\lambda|)\lambda(dx) \\
&= \int_\Gamma \mathbb{E}(\Xi_x(\{x\}) + 1)\lambda(dx), \\
\mathbb{E}(|\Xi| - 1)|\Xi|^2 &= \mathbb{E}\int_\Gamma |\Xi|(|\Xi - \delta_x|)\Xi(dx) = \int_\Gamma \mathbb{E}(|\Xi_x| + 1)|\Xi_x|\lambda(dx) \\
&= |\lambda|\mathbb{E}|\Xi|^2 + 2|\lambda|\int_\Gamma \mathbb{E}\Xi_x(\{x\})\lambda(dx) + |\lambda|^2 \\
&\quad + \int_\Gamma \mathbb{E}(\Xi_x(\{x\}) + 1)\Xi_x(\{x\})\lambda(dx),
\end{aligned}$$

which in turn imply

$$\int_\Gamma \mathbb{E}\Xi_x(\{x\})\lambda(dx) = \text{Var}(|\Xi|) - |\lambda|, \tag{4.5}$$

$$\int_\Gamma \mathbb{E}(\Xi_x(\{x\}) + 1)\Xi_x(\{x\})\lambda(dx) = \mathbb{E}(|\Xi| - |\lambda|)^3 - \text{Var}(|\Xi|). \tag{4.6}$$

Applying Theorem 3.2, and (4.3)–(4.6), together with $0 \leq b < 1$, gives

$$\begin{aligned}
&d_2(\text{CP}(\mu_1, \mu_2, \ldots), \boldsymbol{\pi}_{a,b;0;\nu}) \\
&\quad \leq 2d_0(\mathcal{G}) + \int_\Gamma \mathbb{E}\left[\bar{r}_x(\Xi_x)\left(\left(\Xi_x(\{x\}) + 1\right)\Xi_x(\{x\}) + \Xi_x(\{x\}) + 1\right)\right]\lambda(dx) \\
&\quad \leq 2d_0(\mathcal{G}) + O\left(a^{-1} \max_{1 \leq i \leq k} 1 \wedge \frac{1}{\sqrt{\mu_1(G_i)}}\right) \\
&\qquad \times \int_\Gamma \mathbb{E}\left[\left(\Xi_x(\{x\}) + 1\right)\Xi_x(\{x\}) + \Xi_x(\{x\}) + 1\right]\lambda(dx) \\
&\quad = 2d_0(\mathcal{G}) + O\left(a^{-1} \max_{1 \leq i \leq k} 1 \wedge \frac{1}{\sqrt{\mu_1(G_i)}}\right)\mathbb{E}(|\Xi| - |\lambda|)^3.
\end{aligned} \tag{4.7}$$

Finally, one can verify directly that

$$\mathbb{E}(|\Xi| - |\lambda|)^3 = \mathbb{E}\left(\sum_{i=1}^{\infty} i(|X_i| - |\mu_i|)\right)^3 = \mathbb{E}\sum_{i=1}^{\infty} i^3(|X_i| - |\mu_i|)^3 = \sum_{i=1}^{\infty} i^3|\mu_i|,$$

which, together with (4.7), implies (4.2). □

4.3. Runs

In the final example, we consider the point process of k-runs of 1's in a sequence of independent and identically distributed Bernoulli random variables (cf Example 5.2 of [10, p. 1527]). It is easy to see from our derivation that, at the cost of more notational complexity, one can lift the assumption of identical distribution.

To begin with, let $I_1, \ldots, I_n$ be independent Bernoulli random variables with identical distribution

$$\mathbb{P}(I_i = 1) = 1 - \mathbb{P}(I_i = 0) = p, \quad 1 \le i \le n.$$

Let $X_i = \prod_{j=i}^{i+k-1} I_j$ with $k \ge 2$, where we take $I_j = I_{j-n}$ for $j > n$ to avoid the edge effect. We define the point process of runs as

$$\Xi = \sum_{i=1}^{n} X_i \delta_{i/n}$$

on $\Gamma = [0, 1]$, with 0 being identified as the same as 1 and the distance on the circle $d_0(x, y) = |x - y| \wedge (1 - |x - y|)$. A point of Ξ at location i/n indicates that there is a run of k 1's starting at index i and it is clear that the run may overlap with others around it. Wang and Xia [31] demonstrated that $\mathrm{Var}(|\Xi|) \ge \mathbb{E}|\Xi|$ if and only if $2 + (2k - 1)p^k - (2k + 1)p^{k-1} \ge 0$, and the latter is easily satisfied if $p < 2/3$. Hence we only consider negative binomial process approximation to the distribution of Ξ.

Theorem 4.9. *Let $k \ge 2$ be a fixed integer,*

$$a = \frac{(1 - p)np^k}{1 + p - (2k + 1)p^k + (2k - 1)p^{k+1}},$$

$$b = \frac{p\big[2 - (2k + 1)p^{k-1} + (2k - 1)p^k\big]}{1 + p - (2k + 1)p^k + (2k - 1)p^{k+1}},$$

and $\nu(dx) = \frac{1}{n} \sum_{i=1}^{n} \delta_{i/n}(dx)$. *Assume that $p < 2/3$; then*

$$d_2(\mathscr{L}(\Xi), \boldsymbol{\pi}_{a,b;0;\nu}) \le \begin{cases} O\left(\dfrac{p^{2/3}}{(np^k)^{1/3}}\right), & \text{if } np^k \ge 1, \\ O(p), & \text{if } np^k < 1. \end{cases}$$

Remark 4.10. The point process of runs in [10, Example 5.2], is defined on the carrier space $\Gamma' = [0, n]$ with 0 being identified as the same as n and metric $\tilde{d}_0(x, y) = \big(|x - y|p^k\big) \wedge 1$, where $|\cdot|$ is the distance on the circle. Although $\tilde{d}_0$ seems to be a natural choice in the context of compound Poisson process approximation, it depends on the mean of the process being approximated. An unexpected effect is that, when the parameters vary, it is impossible to judge from the error estimates whether the approximations become better or worse. Another defect of the approach in [10] is that a factor $\ln n$ appears inevitably in the approximation bound, which makes it useless when n becomes large. In practical applications, p is often fixed while n tends to be large, so approximate distributions are needed. Our approximating distribution uses fewer parameters but achieves an approximation bound that decreases when p becomes small and/or n becomes large.

Proof of Theorem 4.9. It is easy to verify that the mean measure of Ξ is $\lambda(dx) = p^k \sum_{i=1}^n \delta_{i/n}(dx)$, $\mathbb{E}|\Xi| = |\lambda| = np^k$ and $\mathrm{Var}(|\Xi|) = \frac{np^k}{1-p}\big(1 + p - (2k+1)p^k + (2k-1)p^{k+1}\big)$; hence we set

$$\nu = \frac{1}{n}\sum_{i=1}^n \delta_{i/n},$$
$$b = \frac{\mathrm{Var}(|\Xi|) - \mathbb{E}|\Xi|}{\mathrm{Var}(|\Xi|)} = \frac{p\big[2 - (2k+1)p^{k-1} + (2k-1)p^k\big]}{1 + p - (2k+1)p^k + (2k-1)p^{k+1}},$$
$$a = (1-b)np^k = \frac{(1-p)np^k}{1 + p - (2k+1)p^k + (2k-1)p^{k+1}}.$$

We assume that $|\lambda| \geq 1$ first. To tackle the dependence resulting from the overlapping runs, we introduce the neighbourhoods $A_{i/n} = \{j/n : \ i - k + 1 \leq j \leq i + k - 1\}$ and $B_{i/n} = \{j/n : \ i - 2k + 2 \leq j \leq i + 2k - 2\}$, where j is interpreted as $j + n$ if $j < 0$, and $j - n$ if $j > n$. Next, we choose $\mathcal{G} = \{G_j : \ 1 \leq j \leq l_n\}$ by taking $l_n = O\left(n^{1/3}p^{(k-2)/3}\right)$, $G_j = (s_{j-1}/n, s_j/n]$ for $j = 1, \ldots, l_n$, where $s_0 = 0, s_j = s_{j-1} + u_j$ for $j = 1, \ldots, l_n$ with $u_1, \ldots, u_{l_n} = O\left(n^{2/3}p^{(2-k)/3}\right)$ and $\sum_{j=1}^{l_n} u_j = n$.

To estimate r_x, we take $u = 2$, write $x = i/n$ and $Y_i = \sum_{|j-i|>4k-4} X_j$. Applying the Bienaymé–Chebyshev inequality gives

$$\mathbb{P}\left(\Xi\left(B_x^c\right) + 1 \leq \frac{a}{u}\bigg|\, \Xi|_{B_x}\right) \leq \mathbb{P}\left(Y_i + 1 \leq \frac{a}{u}\right) \leq \mathbb{P}\left(Y_i - \mathbb{E}Y_i \leq \frac{a}{u} - \mathbb{E}Y_i\right)$$
$$\leq \mathbb{P}\left(|Y_i - \mathbb{E}Y_i| \geq \left|\frac{a}{2} - (n - (8k-7))p^k\right|\right) \leq \frac{\mathbb{E}(Y_i - \mathbb{E}Y_i)^4}{((1+b)np^k/2 - (8k-7)p^k)^4}. \tag{4.8}$$

However,

$$\mathbb{E}(Y_i - \mathbb{E}Y_i)^4 = \sum_{|j_v - i| > 4k-4,\ v=1,2,3,4} \mathbb{E}\prod_{v=1}^4 \left(X_{j_v} - \mathbb{E}X_{j_v}\right)$$

and the summand reduces to 0 if one of the j_v's is not in the neighbourhoods of the others; hence

$$\mathbb{E}(Y_i - \mathbb{E}Y_i)^4 \leq 12|\lambda|^2\left(\frac{1-p^k}{1-p}\right)^2 + 9|\lambda|(12k-9) = O\left(|\lambda|^2\right).$$

This, together with (4.8), implies

$$\mathbb{P}\left(\Xi\left(B_x^c\right) + 1 \leq \frac{a}{u}\bigg|\, \Xi|_{B_x}\right) \leq O\left(|\lambda|^{-2}\right). \tag{4.9}$$

The same argument also leads to

$$\mathbb{P}\left(\Xi_x\left(B_x^c\right) + 1 \leq \frac{a}{u}\bigg|\, \Xi_x|_{B_x}\right) \leq O\left(|\lambda|^{-2}\right). \tag{4.10}$$

For $f \in \mathscr{F}_{TV}$, we will show that

$$\left|\mathbb{E}\left[f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right)\right) - f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right) + \delta_{t_j}\right)\right| \Xi|_{B_x}\right]\right| \leq O\left(n^{-1/3}p^{-(k+1)/3}\right), \tag{4.11}$$
$$\left|\mathbb{E}\left[f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right)\right) - f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right) + \delta_{t_j}\right)\right]\right| \leq O\left(n^{-1/3}p^{-(k+1)/3}\right). \tag{4.12}$$

In fact, if we write $t_j = (s_{j-1} + s_j)/(2n)$, $x = i/n$, then there are two cases to consider.

Case 1. $s_{j-1} < i \le s_j$. Because of the symmetry of our argument, we assume without loss of generality that $i \le \frac{s_{j-1}+s_j}{2}$. We write $\mathbf{I}_1 = (I_1, \ldots, I_{i+2k-2}, I_{s_j-k+1}, \ldots, I_n)$, $\mathbf{I}_2 = (I_{i+2k-1}, \ldots, I_{s_j-k})$, $\mathbf{v} = (v_1, \ldots, v_{i+2k-2}, v_{s_j-k+1}, \ldots, v_n)$. For any vector $\mathbf{v}$ with $v_l \in \{0, 1\}$, $\forall l$, due to [31, Lemma 2.1], the number $W(\mathbf{v}, \mathbf{I}_2)$ of k-runs of the sequence

$$v_{s_{j-1}+1}, \ldots, v_{i+2k-2}, I_{i+2k-1}, \ldots, I_{s_j-k}, v_{s_j-k+1}, \ldots, v_{s_j}$$

satisfies

$$d_{tv}(\mathscr{L}(W(\mathbf{v}), \mathbf{I}_2), \mathscr{L}(W(\mathbf{v}, \mathbf{I}_2) + 1)) \le O\left(n^{-1/3} p^{-(k+1)/3}\right). \tag{4.13}$$

For ease of notation, we use $\Xi(\mathbf{v}, \mathbf{I}_2)$ to stand for the point process of k-runs of the sequence

$$v_1, \ldots, v_{i+2k-2}, I_{i+2k-1}, \ldots, I_{s_j-k}, v_{s_j-k+1}, \ldots, v_n.$$

Then, for $f \in \mathscr{F}_{TV}$,

$$\begin{aligned}
&\left|\mathbb{E}\left[f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right)\right) - f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right) + \delta_{t_j}\right)\middle| \mathbf{I}_1 = \mathbf{v}\right]\right| \\
&\quad \le d_{TV}\left(\mathscr{L}\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi\left(\mathbf{v}, \mathbf{I}_2\right)|_{B_x^c}\right)\right), \mathscr{L}\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi\left(\mathbf{v}, \mathbf{I}_2\right)|_{B_x^c}\right) + \delta_{t_j}\right)\right) \\
&\quad = d_{tv}\left(\mathscr{L}(W(\mathbf{v}, \mathbf{I}_2)), \mathscr{L}\left(W(\mathbf{v}, \mathbf{I}_2) + 1\right)\right),
\end{aligned}$$

and this, together with (4.13), yields that

$$\left|\mathbb{E}\left[f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right)\right) - f\left(\mathscr{M}_{\mathcal{G}} \circ \left(\Xi|_{B_x^c}\right) + \delta_{t_j}\right)\middle| \Xi|_{B_x}\right]\right| \le O\left(n^{-1/3} p^{-(k+1)/3}\right).$$

Case 2. $i \notin \left(s_{j-1}, s_j\right]$. The proof is omitted since it is essentially the same as that of case 1 with some minor changes of notation only.

The proof of (4.12) is similar. Now, combining (4.9)–(4.12) yields $r_x(\Xi)$ and $\bar{r}_x(\Xi)$ are both bounded by $O\left(|\lambda|^{-1}\left(n^{-1/3} p^{-(k+1)/3}\right)\right)$. These, together with some crude estimates, e.g. $\mathbb{E}\Xi_x(A_x) \le (2k-2)p$, $\mathbb{E}\Xi_x(A_x)\Xi_x(B_x \setminus A_x) \le (2k-2)^2 p$ etc., imply that $\mathbb{E}\epsilon_{1,x}(\Xi_x)$, $\mathbb{E}\epsilon_{1,x}(\Xi)$ and $b\mathbb{E}\epsilon_{2,x}(\Xi_x)$ are all bounded by $O\left(|\lambda|^{-1}\left(n^{-1/3} p^{-(k+1)/3}\right)\right) p$. Therefore, if $|\lambda| \ge 1$, the proof is completed by substituting these estimates for the corresponding terms in Theorem 3.2.

Finally, if $|\lambda| < 1$, we take $l_n = O\left(p^{-1}\right)$, $u_1, \ldots, u_{l_n} = O(np)$. Then the right hand side of (4.9) and (4.10) can be replaced with 0, and the upper bounds for (4.11) and (4.12) become $O(1)$, which in turn imply that $r_x(\Xi)$ and $\bar{r}_x(\Xi)$ are both bounded by $O\left(|\lambda|^{-1}\right)$. Consequently, $\mathbb{E}\epsilon_{1,x}(\Xi_x)$, $\mathbb{E}\epsilon_{1,x}(\Xi)$ and $b\mathbb{E}\epsilon_{2,x}(\Xi_x)$ are all bounded by $O\left(|\lambda|^{-1}\right) p$. We then employ Theorem 3.2 to obtain the bound p, as claimed. □

5. Proof of Theorem 2.4

The proof of Theorem 2.4 relies on the coupling and analysis techniques. The main obstacle in coupling various birth–death systems together is the difficulty of identifying the individual particles from their locations. To circumvent the repeats of points, we need to lift the space to a higher-dimensional carrier space and tackle the problem in the lifted space. Such a technique has proved very effective in handling such situations [20,34].

5.1. Lifting the carrier space

In this subsection, we lift the carrier space from Γ to $\tilde{\Gamma} := \Gamma \times [0, 1]$ equipped with the product topology and its Borel σ-algebra $\mathscr{B}(\Gamma) \otimes \mathscr{B}([0, 1])$. We then define a pseudometric $\tilde{d}_0$

on $\tilde{\Gamma}$ as

$$\tilde{d}_0((x_1, t_1), (x_2, t_2)) = d_0(x_1, x_2).$$

It is worth mentioning that the topology generated by $\tilde{d}_0$ is coarser than the product topology. Let $\tilde{\mathscr{H}}$ be the class of all finite integer-valued measures on $\tilde{\Gamma}$ and $\tilde{d}_1$ be the pseudometric induced from $\tilde{d}_0$ in the same way as d_1 is from d_0. For $\tilde{\xi} \in \tilde{\mathscr{H}} = \sum_{i=1}^n \delta_{(x_i,t_i)}$, we define $\tilde{\xi}^{(\Gamma)} = \sum_{i=1}^n \delta_{x_i}$ and extend a function $f \in \mathscr{F}$ to a function on $\tilde{\mathscr{H}}$ via

$$\tilde{f}(\tilde{\xi}) = f(\tilde{\xi}^{(\Gamma)}).$$

It is not hard to check that for each $f \in \mathscr{F}$, $\tilde{f}$ is a $\tilde{d}_1$-Lipschitz function: $|\tilde{f}(\tilde{\xi}_1) - \tilde{f}(\tilde{\xi}_2)| \le \tilde{d}_1(\tilde{\xi}_1, \tilde{\xi}_2)$ for all $\tilde{\xi}_1, \tilde{\xi}_2 \in \tilde{\mathscr{H}}$.

Next, we define $\tilde{\mu}$ as the product measure of μ and the Lebesgue measure on [0, 1]. Regardless of whether μ is diffuse or not, the measure $\tilde{\mu}$ is always diffuse on $\tilde{\Gamma}$. Let

$$\begin{aligned}\tilde{\mathscr{A}}\tilde{h}(\tilde{\xi}) = (a + b|\tilde{\xi}|) \int_{\tilde{\Gamma}} (\tilde{h}(\tilde{\xi} + \delta_{\tilde{x}}) - \tilde{h}(\tilde{\xi}))\tilde{\mu}(d\tilde{x})\\ + (1 + \beta(|\tilde{\xi}| - 1)) \int_{\tilde{\Gamma}} (\tilde{h}(\tilde{\xi} - \delta_{\tilde{x}}) - \tilde{h}(\tilde{\xi}))\tilde{\xi}(d\tilde{x}).\end{aligned}$$

Birth–death systems on $\tilde{\Gamma}$ with the generator $\tilde{\mathscr{A}}$ evolve in the same way as birth–death systems on Γ with the generator $\mathscr{A}$.

To carry out the proof of Theorem 2.4, for a given birth–death system $\mathbf{Z}_\xi(\cdot)$ with $\xi = \sum_{i=1}^n \delta_{x_i}$, one can lift it to $\tilde{\mathbf{Z}}_{\tilde{\xi}}(\cdot)$ by setting up a $\tilde{\xi} \in \tilde{\mathscr{H}}$ consisting of distinct particles at (x_i, t_i), $1 \le i \le n$, where $t_1, \ldots, t_n$ are distinct elements of [0, 1], and throwing each new born particle at z, all with equal likelihood, onto $\{z\} \times [0, 1]$, independently of the others. Then,

$$\tilde{f}(\tilde{\mathbf{Z}}_{\tilde{\xi}}(t)) = f(\tilde{\mathbf{Z}}_{\tilde{\xi}}(t)^{(\Gamma)}) = f(\mathbf{Z}_\xi(t)), \quad \forall\, t \ge 0.$$

This procedure enables us to assume from now on that, without loss of generality, μ is diffuse and the particles at ξ, η, x and y are all distinct.

5.2. *Proof of (2.5)*

First of all, the proportion of the surviving initial particles at time t can be estimated as

$$\mathbb{E}\frac{|\eta \cap \mathbf{Z}_\eta(t)|}{|\mathbf{Z}_\eta(t)|} \le \min\left\{\left(1 + \frac{a}{2|\eta|}(e^t - 1)\right)^{-1}, e^{-(a\wedge b)t}\right\}. \tag{5.1}$$

To this end, we define $g(\zeta) := |\eta \cap \zeta|/|\zeta|$ for fixed $\eta \in \mathscr{H}$, where 0/0 is interpreted as 0. Recall that $V(\zeta)$ has a uniform distribution on the sites in ζ, and we have $\mathbb{E}g(\zeta - \delta_{V(\zeta)}) = g(\zeta)$. Hence

$$\mathscr{A}g(\zeta) = \big(a + b|\zeta|\big)\,(\mathbb{E}g(\zeta + \delta_U) - g(\zeta))$$

since the last term in (2.2) vanishes. Noticing that with probability 1, $U \notin \eta$, we have

$$g(\zeta + \delta_U) - g(\zeta) = |\eta \cap \zeta|\left(\frac{1}{|\zeta + \delta_U|} - \frac{1}{|\zeta|}\right) = -\frac{|\eta \cap \zeta|}{|\zeta|(|\zeta| + 1)} \quad \text{a.s.}$$

It follows that

$$\mathscr{A} g(\zeta) \le \min\left\{-\frac{a|\eta\cap\zeta|}{2|\zeta|^2}, -(a\wedge b)g(\zeta)\right\}.$$

Therefore, setting $\varphi(t) = \mathbb{E}g(\mathbf{Z}_\eta(t))$, we have

$$\varphi'(t) = \mathbb{E}\mathscr{A} g(\mathbf{Z}_\eta(t)) \le \min\left\{-\frac{a}{2}\mathbb{E}\frac{|\eta\cap\mathbf{Z}_\eta(t)|}{|\mathbf{Z}_\eta(t)|^2}, -(a\wedge b)\varphi(t)\right\}. \tag{5.2}$$

By the Cauchy inequality,

$$\left(\mathbb{E}\frac{|\eta\cap\mathbf{Z}_\eta(t)|}{|\mathbf{Z}_\eta(t)|}\right)^2 \le \mathbb{E}\frac{|\eta\cap\mathbf{Z}_\eta(t)|}{|\mathbf{Z}_\eta(t)|^2}\mathbb{E}|\eta\cap\mathbf{Z}_\eta(t)| \le |\eta|e^{-t}\mathbb{E}\frac{|\eta\cap\mathbf{Z}_\eta(t)|}{|\mathbf{Z}_\eta(t)|^2},$$

where the second inequality holds since each particle dies with rate at least 1. Therefore,

$$\mathbb{E}\frac{|\eta\cap\mathbf{Z}_\eta(t)|}{|\mathbf{Z}_\eta(t)|^2} \ge \frac{e^t}{|\eta|}\left(\mathbb{E}\frac{|\eta\cap\mathbf{Z}_\eta(t)|}{|\mathbf{Z}_\eta(t)|}\right)^2.$$

This, together with (5.2), yields

$$\varphi'(t) \le \min\left\{-\frac{ae^t}{2|\eta|}\varphi(t)^2, -(a\wedge b)\varphi(t)\right\}.$$

Therefore, (5.1) follows from the fact that $\varphi(0) = 1$.

Next, suppose that $\eta \in \mathscr{H}$, $|\eta| = n$ and the particles at x, y and η are all distinct. We start with $\mathbf{Z}_{\eta+\delta_x}(\cdot)$ and construct $\mathbf{Z}_{\eta+\delta_y}(\cdot)$ by replacing x with y. Let $\tau_z = \inf\{t : z \notin \mathbf{Z}_{\eta+\delta_x}(t)\}$ for $z \in \eta + \delta_x$. Then, $\mathbf{Z}_{\eta+\delta_x}(t) = \mathbf{Z}_{\eta+\delta_y}(t)$ for $t \ge \tau_x$. For $t < \tau_x$,

$$|f(\mathbf{Z}_{\eta+\delta_x}(t)) - f(\mathbf{Z}_{\eta+\delta_y}(t))| \le d_1(\mathbf{Z}_{\eta+\delta_x}(t), \mathbf{Z}_{\eta+\delta_y}(t)) \le \frac{1}{|\mathbf{Z}_{\eta+\delta_x}(t)|}.$$

Therefore,

$$|h(\eta+\delta_x) - h(\eta+\delta_y)| \le \int_0^\infty \mathbb{E}\frac{1_{\{\tau_x>t\}}}{|\mathbf{Z}_{\eta+\delta_x}(t)|}dt. \tag{5.3}$$

Notice that for all $z \in \eta + \delta_x$,

$$\mathbb{E}\frac{1_{\{\tau_x>t\}}}{|\mathbf{Z}_{\eta+\delta_x}(t)|} = \mathbb{E}\frac{1_{\{\tau_z>t\}}}{|\mathbf{Z}_{\eta+\delta_x}(t)|},$$

which implies that

$$\mathbb{E}\frac{1_{\{\tau_x>t\}}}{|\mathbf{Z}_{\eta+\delta_x}(t)|} = \frac{1}{n+1}\mathbb{E}\frac{\sum_{z\in\eta+\delta_x} 1_{\{\tau_z>t\}}}{|\mathbf{Z}_{\eta+\delta_x}(t)|} = \frac{1}{n+1}\mathbb{E}\frac{|(\eta+\delta_x)\cap\mathbf{Z}_{\eta+\delta_x}(t)|}{|\mathbf{Z}_{\eta+\delta_x}(t)|}.$$

This, together with (5.3) and (5.1), implies that

$$\begin{aligned} C_n &\le \frac{1}{n+1}\int_0^\infty \frac{1}{1+\frac{a}{2(n+1)}(e^t-1)}dt = \frac{\ln(n+1) - \ln\frac{a}{2}}{n+1-\frac{a}{2}} \\ &\le \frac{1}{2}\left(\frac{1}{n+1} + \frac{2}{a}\right) = \frac{1}{2(n+1)} + \frac{1}{a}, \end{aligned}$$

where the result also includes the case $a = 2(n+1)$, and

$$C_n \le \int_0^\infty \frac{1}{n+1} e^{-(a\wedge b)t} dt = \frac{1}{(a \wedge b)(n+1)}.$$

On the other hand,

$$\mathbb{E}\frac{1_{\{\tau_x > t\}}}{|\mathbf{Z}_{\eta+\delta_x}(t)|} \le \mathbb{P}(\tau_x > t) \le e^{-t},$$

and hence $C_n \le 1$. $\square$

5.3. Proof of (2.6)

Suppose that $|\xi| = n$ and particles at ξ, η, x and y are all distinct. Recall that $\mathscr{A}h(\xi + \delta_x) = f(\xi + \delta_x) - \boldsymbol{\pi}(f)$, i.e.

$$\alpha_{n+1}\mathbb{E}h(\xi + \delta_x + \delta_U) + \beta_{n+1}\mathbb{E}h(\xi + \delta_x - \delta_{V(\xi+\delta_x)}) - \big(\alpha_{n+1} + \beta_{n+1}\big)h(\xi + \delta_x)$$
$$= f(\xi + \delta_x) - \boldsymbol{\pi}(f). \tag{5.4}$$

It follows that

$$\mathbb{E}h(\xi + \delta_x + \delta_U) = \frac{f(\xi + \delta_x) - \boldsymbol{\pi}(f)}{\alpha_{n+1}} + \frac{\alpha_{n+1} + \beta_{n+1}}{\alpha_{n+1}} h(\xi + \delta_x) - \frac{\beta_{n+1}}{\alpha_{n+1}}\mathbb{E}h(\xi + \delta_x - \delta_{V(\xi+\delta_x)}).$$

Hence

$$\begin{aligned}\Delta_2 h(\xi; x, y) &= h(\xi + \delta_x + \delta_y) - \mathbb{E}h(\xi + \delta_x + \delta_U) + h(\xi + \delta_x) - h(\xi + \delta_y) \\ &\quad + \mathbb{E}h(\xi + \delta_x + \delta_U) - 2h(\xi + \delta_x) + h(\xi) \\ &= h(\xi + \delta_x + \delta_y) - \mathbb{E}h(\xi + \delta_x + \delta_U) + h(\xi + \delta_x) - h(\xi + \delta_y) \\ &\quad + \frac{f(\xi + \delta_x) - \boldsymbol{\pi}(f)}{\alpha_{n+1}} + \mathbb{E}\left(h(\xi) - h\left(\xi + \delta_x - \delta_{V(\xi+\delta_x)}\right)\right) \\ &\quad + \frac{\alpha_{n+1} - \beta_{n+1}}{\alpha_{n+1}}\mathbb{E}\left(h\left(\xi + \delta_x - \delta_{V(\xi+\delta_x)}\right) - h(\xi + \delta_x)\right).\end{aligned} \tag{5.5}$$

Swapping x and y, we get

$$\begin{aligned}\Delta_2 h(\xi; y, x) &= h(\xi + \delta_y + \delta_x) - \mathbb{E}h(\xi + \delta_y + \delta_U) + h(\xi + \delta_y) - h(\xi + \delta_x) \\ &\quad + \frac{f(\xi + \delta_y) - \boldsymbol{\pi}(f)}{\alpha_{n+1}} + \mathbb{E}\left(h(\xi) - h\left(\xi + \delta_y - \delta_{V(\xi+\delta_y)}\right)\right) \\ &\quad + \frac{\alpha_{n+1} - \beta_{n+1}}{\alpha_{n+1}}\mathbb{E}\left(h\left(\xi + \delta_y - \delta_{V(\xi+\delta_y)}\right) - h(\xi + \delta_y)\right).\end{aligned} \tag{5.6}$$

Since $\Delta_2 h(\xi; x, y) = \Delta_2 h(\xi; y, x)$ and $|f - \boldsymbol{\pi}(f)| \le 1$, we take the average of (5.5) and (5.6) to reach the bound

$$|\Delta_2 h(\xi; x, y)| \le \frac{1}{\alpha_{n+1}} + C_{n-1} + C_{n+1} + \left|\frac{\alpha_{n+1} - \beta_{n+1}}{\alpha_{n+1}}\right| \Delta_n, \tag{5.7}$$

where

$$\Delta_n = \sup\{|h(\eta+\delta_x) - h(\eta)| : |\eta| = n, x \in \Gamma\}.$$

On the other hand, we use (5.4) again to obtain

$$\mathbb{E}h(\xi+\delta_x-\delta_{V(\xi+\delta_x)}) = \frac{f(\xi+\delta_x)-\boldsymbol{\pi}(f)}{\beta_{n+1}} + \frac{\alpha_{n+1}+\beta_{n+1}}{\beta_{n+1}}h(\xi+\delta_x) - \frac{\alpha_{n+1}}{\beta_{n+1}}\mathbb{E}h(\xi+\delta_x+\delta_U),$$

and argue in the same way as for (5.7) to get

$$|\Delta_2 h(\xi;x,y)| \le \frac{1}{\beta_{n+1}} + C_{n-1} + C_{n+1} + \left|\frac{\beta_{n+1}-\alpha_{n+1}}{\beta_{n+1}}\right|\Delta_{n+1}. \tag{5.8}$$

In Section 5.4 below, we will prove that

$$\begin{cases} \dfrac{\alpha_{n+1}-\beta_{n+1}}{\alpha_{n+1}}\Delta_n \le \dfrac{1}{\alpha_{n+1}} + C_n, & \text{if } \alpha_{n+1} \ge \beta_{n+1}, \\ \dfrac{\beta_{n+1}-\alpha_{n+1}}{\beta_{n+1}}\Delta_{n+1} \le \dfrac{1}{\beta_{n+1}} + C_n, & \text{otherwise,} \end{cases} \tag{5.9}$$

and so it follows from (5.7) and (5.8) that

$$|\Delta_2 h(\xi)| \le C_{n-1} + C_n + C_{n+1} + 2\left(\frac{1}{\alpha_{n+1}} \wedge \frac{1}{\beta_{n+1}}\right). \tag{5.10}$$

For $n = 0$, (2.5) enables us to conclude that $C_k \le 1$, $C_{-1} = 0$, and it follows from (5.10) that

$$|\Delta_2 h(\xi)| \le 2 + \frac{2}{a} \le \frac{2}{n+1} + \frac{5}{a}.$$

For $n \ge 1$, using the estimate $C_k \le \frac{1}{2(k+1)} + \frac{1}{a}$ in (2.5), the fact that $2n \ge n+1$, and the bound given in (5.10), we have

$$|\Delta_2 h(\xi)| \le \frac{1}{2n} + \frac{1}{2(n+1)} + \frac{1}{2(n+2)} + \frac{5}{a} \le \frac{2}{n+1} + \frac{5}{a}.$$

This completes the proof of (2.6). □

5.4. *Proof of (5.9)*

Since $\{|\mathbf{Z}_\eta(t)|,\ t \ge 0\}$ is a birth–death process with birth rates $\{\alpha_k\}$, death rates $\{\beta_k\}$ and initial value $|\eta|$, we follow the convention in [17] to define $\tau_{|\eta|,k} = \inf\{t : |\mathbf{Z}_\eta(t)| = k\}$, $\tau_m^+ = \tau_{m,m+1}$ and $\tau_m^- = \tau_{m,m-1}$.

For any $\eta \in \mathscr{H}$ with $|\eta| = n$, by the strong Markov property of $\{\mathbf{Z}_\eta(t),\ t \ge 0\}$,

$$h(\eta) = -\mathbb{E}\int_0^{\tau_n^+}(f(\mathbf{Z}_\eta(t)) - \boldsymbol{\pi}(f))dt + \mathbb{E}h(\mathbf{Z}_\eta(\tau_n^+)),$$

which implies that

$$\left|(h\eta) - \mathbb{E}h\left(\mathbf{Z}_\eta(\tau_n^+)\right)\right| \le \mathbb{E}\tau_n^+. \tag{5.11}$$

Now we compare $\mathbb{E}h\left(\mathbf{Z}_\eta(\tau_n^+)\right)$ with $h(\eta+\delta_x)$. Let K_n^+ be the number of particles in η that have died before τ_n^+. Clearly, $0 \le K_n^+ \le n$. Given $K_n^+ = k$, there are at most $k+1$ pairs of mismatched points between $\mathbf{Z}_\eta(\tau_n^+)$ and $\eta+\delta_x$; consequently,

$$\left|\mathbb{E}\left(h\left(\mathbf{Z}_\eta(\tau_n^+)\right)|K_n^+ = k\right) - h(\eta+\delta_x)\right| \le C_n(k+1).$$

This in turn leads to

$$\left|\mathbb{E}h\left(\mathbf{Z}_\eta(\tau_n^+)\right) - h(\eta+\delta_x)\right| \le C_n(\mathbb{E}K_n^+ + 1). \tag{5.12}$$

Combining (5.11) and (5.12) gives

$$\Delta_n \le \mathbb{E}\tau_n^+ + C_n(\mathbb{E}K_n^+ + 1).$$

Likewise, for $\eta \in \mathscr{H}$ with $|\eta| = n+1$, it follows from the strong Markov property of $\{\mathbf{Z}_{\eta+\delta_x}(t),\ t \ge 0\}$ that

$$h(\eta+\delta_x) = -\mathbb{E}\int_0^{\tau_{n+2}^-}(f(\mathbf{Z}_\eta(t)) - \boldsymbol{\pi}(f))dt + \mathbb{E}h(\mathbf{Z}_{\eta+\delta_x}(\tau_{n+2}^-)),$$

giving

$$|h(\eta+\delta_x) - \mathbb{E}h(\mathbf{Z}_{\eta+\delta_x}(\tau_{n+2}^-))| \le \mathbb{E}\tau_{n+2}^-. \tag{5.13}$$

Let K_{n+2}^- be the number of particles in $\eta+\delta_x$ that have died before τ_{n+2}^-; then there are at most K_{n+2}^- mismatched pairs of points between $\mathbf{Z}_{\eta+\delta_x}(\tau_{n+2}^-)$ and η, leading to the bound

$$|\mathbb{E}h(\mathbf{Z}_{\eta+\delta_x}(\tau_{n+2}^-)) - h(\eta)| \le C_n\mathbb{E}K_{n+2}^-. \tag{5.14}$$

Collecting the estimates (5.13) and (5.14), we obtain

$$\Delta_{n+1} \le \mathbb{E}\tau_{n+2}^- + C_n\mathbb{E}K_{n+2}^-.$$

Put $F(k) = \sum_{i=0}^k \pi_i$ and $\bar{F}(k) = \sum_{i=k}^\infty \pi_i$. By Lemmas 2.2 and 2.4 in [17],

$$\mathbb{E}\tau_k^+ = \frac{F(k)}{\alpha_k\pi_k}, \qquad \mathbb{E}\tau_k^- = \frac{\bar{F}(k)}{\beta_k\pi_k},$$

$$\frac{F(k)}{F(k-1)} \ge \frac{\alpha_k}{\beta_k} \ge \frac{\bar{F}(k+1)}{\bar{F}(k)} \tag{5.15}$$

since $\alpha_k - \alpha_{k-1} \le \beta_k - \beta_{k-1}$ for all k. It follows from the first inequality of (5.15) that

$$\frac{(\alpha_{n+1}-\beta_{n+1})F(n)}{\alpha_n\pi_n} \le \frac{\beta_{n+1}F(n+1) - \beta_{n+1}F(n)}{\alpha_n\pi_n} = \frac{\beta_{n+1}\pi_{n+1}}{\alpha_n\pi_n} = 1,$$

which in turn yields

$$\mathbb{E}\tau_n^+ = \frac{F(n)}{\alpha_n\pi_n} \le \frac{1}{\alpha_{n+1}-\beta_{n+1}}, \quad \text{if } \beta_{n+1} < \alpha_{n+1}.$$

Likewise, using the second inequality of (5.15), we get

$$\mathbb{E}\tau_{n+2}^- = \frac{\bar{F}(n+2)}{\beta_{n+2}\pi_{n+2}} \le \frac{1}{\beta_{n+1}-\alpha_{n+1}}, \quad \text{if } \beta_{n+1} > \alpha_{n+1}.$$

To complete the proof of (5.9), it remains to show that

$$\frac{\alpha_{n+1}-\beta_{n+1}}{\alpha_{n+1}}(\mathbb{E}K_n^+ + 1) \le 1, \quad \text{if } \beta_{n+1} < \alpha_{n+1}, \tag{5.16}$$

$$\frac{\beta_{n+1}-\alpha_{n+1}}{\beta_{n+1}}\mathbb{E}K_{n+2}^- \le 1, \quad \text{if } \beta_{n+1} > \alpha_{n+1}. \tag{5.17}$$

To this end, we derive a recursive formula for $\mathbb{E}K_m^+$ and $\mathbb{E}K_m^-$, $m \ge 1$, in Lemma 5.1 later and give their estimates in the following lemma, Lemma 5.2. In particular, since $\alpha_k - \beta_k$ decreases in k and α_k increases in k, it follows from Lemma 5.2 that, if $\alpha_{n+1} > \beta_{n+1}$,

$$1 + \mathbb{E}K_n^+ \le \frac{\alpha_n}{\alpha_n - \beta_n} \le \frac{\alpha_{n+1}}{\alpha_{n+1} - \beta_{n+1}},$$

which is equivalent to (5.16). On the other hand, noting that

$$\beta_{n+2}/(\beta_{n+2} - \alpha_{n+2}) \le \beta_{n+1}/(\beta_{n+1} - \alpha_{n+1})$$

as $\beta_{n+1} - \alpha_{n+1} > 0$, applying Lemma 5.2 again, we obtain $\mathbb{E}K_{n+2}^- \le \beta_{n+1}/(\beta_{n+1} - \alpha_{n+1})$ and hence (5.17) follows. □

Lemma 5.1. *The following recursive formulae hold for $m \ge 1$:*

$$\mathbb{E}K_m^+ = \frac{m\beta_m\left(1 + \mathbb{E}K_{m-1}^+\right)}{m\alpha_m + \beta_m\left(1 + \mathbb{E}K_{m-1}^+\right)}, \qquad \mathbb{E}K_m^- = 1 + \frac{(m-1)\alpha_m \mathbb{E}K_{m+1}^-}{\alpha_m \mathbb{E}K_{m+1}^- + (m+1)\beta_m}.$$

Proof. Noting that all particles are equally likely to die, an initial particle in the initial configuration η with $|\eta| = m$ dies before τ_m^+ with probability $\frac{1}{m}\mathbb{E}K_m^+$, and if it survives, it dies before $\tau_{m,m+2}$ with probability $\frac{1}{m+1}\mathbb{E}K_{m+1}^+$. That is, the probability that an initial particle dies before $\tau_{m,m+2}$ is

$$\frac{1}{m}\mathbb{E}K_m^+ + \left(1 - \frac{1}{m}\mathbb{E}K_m^+\right)\frac{1}{m+1}\mathbb{E}K_{m+1}^+.$$

Therefore, there are on average $\mathbb{E}K_m^+ + \frac{m - \mathbb{E}K_m^+}{m+1}\mathbb{E}K_{m+1}^+$ initial particles that die before $\tau_{m,m+2}$.

On the other hand, $K_m^+ = 0$ means that the first change of the configuration of the birth–death system $\mathbf{Z}_\eta(\cdot)$ is a birth, so $K_m^+ = 0$ with probability $\frac{\alpha_m}{\alpha_m+\beta_m}$. However, if the first change is a death, which happens with probability $\frac{\beta_m}{\alpha_m+\beta_m}$, then one particle at some site x of η will die at $\boldsymbol{\tau}_\eta = \inf\{t : \mathbf{Z}_\eta(t) \ne \eta\}$. In the latter case, using the conclusion in the preceding paragraph, the mean number of particles in $\eta - \delta_x$ dying before the birth–death system $\mathbf{Z}_{\eta-\delta_x}(\cdot)$ reaches the size $m+1$ is $\mathbb{E}K_{m-1}^+ + \frac{m-1-\mathbb{E}K_{m-1}^+}{m}\mathbb{E}K_m^+$. In summary, we have established the relationship

$$\mathbb{E}K_m^+ = \frac{\beta_m}{\alpha_m + \beta_m}\left(1 + \mathbb{E}K_{m-1}^+ + \frac{m-1-\mathbb{E}K_{m-1}^+}{m}\mathbb{E}K_m^+\right),$$

which is equivalent to the first recursive formula.

The same argument can be adapted to prove the second recursive formula. In fact, assuming that $|\eta| = k \ge 2$, an initial particle in η dies before $\tau_{k,k-2}$ with probability

$$\frac{1}{k}\mathbb{E}K_k^- + \left(1 - \frac{1}{k}\mathbb{E}K_k^-\right)\frac{1}{k-1}\mathbb{E}K_{k-1}^-.$$

Now, let $|\eta| = m$. With probability $\frac{\beta_m}{\alpha_m+\beta_m}$, the first change of $\mathbf{Z}_\eta(\cdot)$ is a death, giving $K_m^- = 1$. Assume next that the first change is a birth; then, as shown above, each initial particle dies before the size reaches $m-1$ with probability $\frac{1}{m+1}\mathbb{E}K_{m+1}^- + \left(1 - \frac{1}{m+1}\mathbb{E}K_{m+1}^-\right)\frac{1}{m}\mathbb{E}K_m^-$. It then follows that

$$\mathbb{E}K_m^- = \frac{\beta_m}{\alpha_m+\beta_m} + \frac{\alpha_m}{\alpha_m+\beta_m}\frac{m}{m+1}\left(\mathbb{E}K_{m+1}^- + \frac{m+1-\mathbb{E}K_{m+1}^-}{m}\mathbb{E}K_m^-\right),$$

and reorganizing the equation yields the second recursive formula. □

Lemma 5.2. *If $\alpha_m > \beta_m$, then*

$$1 + \mathbb{E}K_m^+ \le \frac{\alpha_m}{\alpha_m - \beta_m}.$$

If $\beta_m > \alpha_m$, then,

$$\mathbb{E}K_m^- \le \frac{\beta_m}{\beta_m - \alpha_m}.$$

Proof. Suppose $\alpha_m > \beta_m$. By Lemma 5.1 and because $\mathbb{E}K_{m-1}^+ \ge 0$, we have

$$1 + \mathbb{E}K_m^+ \le 1 + \frac{\beta_m}{\alpha_m}\left(1 + \mathbb{E}K_{m-1}^+\right), \quad \forall\, m \ge 1. \tag{5.18}$$

Iterating (5.18) and noticing that β_k/α_k is increasing in k as well as that $\mathbb{E}K_0^+ = 0$, we conclude that

$$1 + \mathbb{E}K_m^+ \le \sum_{i=0}^{m-1}\left(\frac{\beta_m}{\alpha_m}\right)^l + \left(\frac{\beta_m}{\alpha_m}\right)^m\left(1 + \mathbb{E}K_0^+\right) \le \frac{1}{1 - \frac{\beta_m}{\alpha_m}} = \frac{\alpha_m}{\alpha_m - \beta_m}.$$

Assume that $\alpha_m < \beta_m$. Using Lemma 5.1 again together with the fact that $\mathbb{E}K_{m+1}^- \ge 1$, we have

$$\mathbb{E}K_m^- \le 1 + \frac{\alpha_m}{\beta_m}\mathbb{E}K_{m+1}^-. \tag{5.19}$$

Noticing that α_k/β_k is decreasing in k, we conclude that

$$\mathbb{E}K_m^- \le \sum_{i=0}^{l-1}\left(\frac{\alpha_m}{\beta_m}\right)^i + \left(\frac{\alpha_m}{\beta_m}\right)^l \mathbb{E}K_{m+l}^-$$

by iterating (5.19). Recalling that $\mathbb{E}K_{m+l}^- \le (m+l)$, we have, on letting $l \to \infty$, that

$$\mathbb{E}K_m^- \le \sum_{i=0}^{\infty}\left(\frac{\alpha_m}{\beta_m}\right)^i = \frac{1}{1 - \frac{\alpha_m}{\beta_m}} = \frac{\beta_m}{\beta_m - \alpha_m}. \quad \square$$

6. Proof of Theorem 3.2

Let X be a point process with distribution $\boldsymbol{\pi}_{a,b;0;\nu}$; then by the triangle inequality, we have

$$\begin{aligned} d_2(\mathscr{L}(\Xi), \boldsymbol{\pi}_{a,b;0;\nu}) \le\ & d_2(\mathscr{L}(\Xi), \mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ \Xi)) \\ & + d_2(\mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ \Xi), \mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ X)) + d_2(\mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ X), \boldsymbol{\pi}_{a,b;0;\nu}). \end{aligned}$$

It follows from (3.3) that both $d_2(\mathscr{L}(\Xi), \mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ \Xi))$ and $d_2(\mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ X), \boldsymbol{\pi}_{a,b;0;\nu})$ are bounded by $d_0(\mathcal{G})$, so it remains to estimate $d_2(\mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ \Xi), \mathscr{L}(\mathscr{M}_{\mathcal{G}} \circ X))$. Clearly, $\mathscr{M}_{\mathcal{G}} \circ X \sim \boldsymbol{\pi}_{a,b;0;\nu'}$, where

$$\nu'(dx) = \sum_{i=1}^{k} \nu(G_i)\delta_{t_i}(dx).$$

Using the Stein equation (2.3) with $\boldsymbol{\pi} = \boldsymbol{\pi}_{a,b;0;\nu'}$, it suffices to show that for each $f \in \mathscr{F}$,

$$\begin{aligned} &|\mathbb{E}\mathscr{A} h_f(\mathscr{M}_{\mathcal{G}} \circ \Xi)| = |\mathbb{E} f(\mathscr{M}_{\mathcal{G}} \circ \Xi) - \boldsymbol{\pi}_{a,b;0;\nu'}(f)| \\ &\quad \le \int_{\Gamma} \mathbb{E}\left[(1+b)(\epsilon_{1,y}(\Xi_y) + \epsilon_{1,y}(\Xi)) + b\bar{r}_y(\Xi)\Xi_y(A_y) + b\epsilon_{2,y}(\Xi_y)\right]\lambda(dy). \end{aligned} \tag{6.1}$$

To simplify the notation, we fix $f \in \mathscr{F}$; write $f'(\eta) = f(\mathscr{M}_{\mathcal{G}} \circ \eta)$, $h'(\eta) = h_f(\mathscr{M}_{\mathcal{G}} \circ \eta)$ and define

$$\Delta h'(\xi; x) = h'(\xi + \delta_x) - h'(\xi).$$

Noting that h' acts on the 'shuffled' configurations and so one can swap ν' for ν in $\mathscr{A} h'$, we apply (3.1) to expand $\mathbb{E}\mathscr{A} h'(\Xi)$ as

$$\begin{aligned} \mathbb{E}\mathscr{A} h'(\Xi) = b &\int_{\Gamma}\int_{\Gamma} \mathbb{E}[\Delta h'(\Xi_y + \delta_y; x) - \Delta h'(\Xi; x)]\lambda(dy)\nu(dx) \\ &+ \int_{\Gamma} \mathbb{E}[-\Delta h'(\Xi_x; x) + \Delta h'(\Xi; x)]\lambda(dx) \\ &+ \int_{\Gamma} \mathbb{E}\Delta h'(\Xi; x)[a\nu(dx) + b|\lambda|\nu(dx) - \lambda(dx)]. \end{aligned} \tag{6.2}$$

The last term vanishes since $(a + b|\lambda|)\nu = \lambda$, which is ensured by the facts that $|\nu| = 1$ and $a = (1-b)|\lambda|$.

To study the first term in (6.2), we take a coupling $(\Theta_y, \Upsilon_y, \Pi_y)$ of $\Xi|_{A_y^c}$ (notice that it has the same distribution as $\Xi_y|_{A_y^c}$), $\Xi|_{A_y}$, and $\Xi_y|_{A_y}$, such that $\mathscr{L}(\Theta_y + \Upsilon_y) = \mathscr{L}(\Xi)$ and $\mathscr{L}(\Theta_y + \Pi_y) = \mathscr{L}(\Xi_y)$. Dropping the subscript y from $(\Theta_y, \Upsilon_y, \Pi_y)$, we can write

$$\begin{aligned} &\mathbb{E}\{\Delta h'(\Xi_y + \delta_y; x) - \Delta h'(\Xi; x)\} \\ &\quad = \mathbb{E}\{\Delta h'(\Theta + \Pi + \delta_y; x) - \Delta h'(\Theta + \Upsilon; x)\} \\ &\quad = \mathbb{E}\left\{[\Delta h'(\Theta + \Pi + \delta_y; x) - \Delta h'(\Theta; x)] + [\Delta h'(\Theta; x) - \Delta h'(\Theta + \Upsilon; x)]\right\}. \end{aligned}$$

When expanded telescopically, it is the sum of $|\Pi|+1$ positive $\Delta_2 h'$-functions for the term in the first pair of square brackets, and $|\Upsilon|$ negative $\Delta_2 h'$-functions for the term in the second pair of square brackets. Similarly, the second term in (6.2) can be expressed as the sum of $|\Upsilon|$ positive $\Delta_2 h'$-functions and $|\Pi|$ negative $\Delta_2 h'$-functions. Therefore, when

$$b\int_{\Gamma} \mathbb{E}(\Xi_y(A_y) + 1 - \Xi(A_y))\lambda(dy) + \int_{\Gamma} \mathbb{E}(\Xi(A_y) - \Xi_y(A_y))\lambda(dy) = 0, \tag{6.3}$$

the expected numbers of positive and negative $\Delta_2 h'$-functions are then balanced. Noting that

$$\int_{\Gamma} \mathbb{E}(\Xi_y(A_y) - \Xi(A_y))\lambda(dy) = \mathrm{Var}(|\Xi|) - \mathbb{E}|\Xi|, \tag{6.4}$$

we obtain (6.3) by taking $b = \frac{\mathrm{Var}(|\Xi|)-\mathbb{E}|\Xi|}{\mathrm{Var}(|\Xi|)}$. Now, we define $\Pi = \sum_{j=1}^{|\Pi|} \delta_{X_j}$, $\Upsilon = \sum_{j=1}^{|\Upsilon|} \delta_{Y_j}$, and for $\eta = \sum_{i=1}^{n} \delta_{z_i}$, write $\langle\eta\rangle_0 = 0$, $\langle\eta\rangle_j = \sum_{i=1}^{j} \delta_{z_i}$ for $1 \le j \le n$. Taking $\hat{\Xi}$ as an independent copy of Ξ, we can expand $\mathbb{E}\mathscr{A}h'(\Xi)$ into

$$\mathbb{E}\mathscr{A}h'(\Xi) = e_1 + \cdots + e_5,$$

where, with $z \in \Gamma$ a fixed point,

$$e_1 = b \iint_{\Gamma^2} \mathbb{E}\sum_{j=1}^{|\Pi|}[\Delta_2 h'(\Theta + \langle\Pi\rangle_{j-1} + \delta_y; x, X_j) - \mathbb{E}\Delta_2 h'(\hat{\Xi}; z, z)]\lambda(dy)\nu(dx),$$

$$e_2 = b \iint_{\Gamma^2} \mathbb{E}[\Delta_2 h'(\Theta; x, y) - \mathbb{E}\Delta_2 h'(\hat{\Xi}; z, z)]\lambda(dy)\nu(dx),$$

$$e_3 = -b \iint_{\Gamma^2} \mathbb{E}\sum_{j=1}^{|\Upsilon|}[\Delta_2 h'(\Theta + \langle\Upsilon\rangle_{j-1}; x, Y_j) - \mathbb{E}\Delta_2 h'(\hat{\Xi}; z, z)]\lambda(dy)\nu(dx),$$

$$e_4 = -\int_{\Gamma} \mathbb{E}\sum_{j=1}^{|\Pi|}[\Delta_2 h'(\Theta + \langle\Pi\rangle_{j-1}; x, X_j) - \mathbb{E}\Delta_2 h'(\hat{\Xi}; z, z)]\lambda(dx),$$

$$e_5 = \int_{\Gamma} \mathbb{E}\sum_{j=1}^{|\Upsilon|}[\Delta_2 h'(\Theta + \langle\Upsilon\rangle_{j-1}; x, Y_j) - \mathbb{E}\Delta_2 h'(\hat{\Xi}; z, z)]\lambda(dx).$$

Now we concentrate on estimating e_1, since other cases are similar. Recalling that $\Xi_y|_{A_y^c}$ is not independent of $\Xi_y|_{A_y}$ while $\Xi_y|_{B_y^c}$ is, we can extract the part as $\Xi_y|_{B_y^c}$ from $\Theta \sim \mathscr{L}(\Xi_y|_{A_y^c})$, and denote it by Θ_1. Take a more detailed coupling $(\Theta_1, \Theta_2, \Upsilon, \Pi)$ such that (Θ_1, Θ_2) is a coupling of $\Xi|_{B_y^c}$ and $\Xi|_{B_y\setminus A_y}$ (as well as $\Xi_y|_{B_y^c}$ and $\Xi_y|_{B_y\setminus A_y}$), and Θ_1 is dependent of (Υ, Π). We then take $(\hat{\Theta}_2, \hat{\Upsilon})$ as a copy of (Θ_2, Υ) such that $(\hat{\Theta}_2, \hat{\Upsilon})$ is independent of Π and $\mathscr{L}(\Theta_1 + \hat{\Theta}_2 + \hat{\Upsilon}) = \mathscr{L}(\Xi)$. We insert $\Delta_2 h'(\Theta_1; x, X_j)$ and $\Delta_2 h'(\Theta_1; z, z)$ into the square brackets in e_1 to obtain

$$e_1 = b \iint_{\Gamma^2} (e_{11} + \cdots + e_{15})\lambda(dy)\nu(dx),$$

where

$$e_{11} = \mathbb{E}\sum_{j=1}^{|\Pi|}[\Delta_2 h'(\Theta_1 + \Theta_2 + \langle\Pi\rangle_{j-1} + \delta_y; x, X_j) - \Delta_2 h'(\Theta_1 + \langle\Pi\rangle_{j-1} + \delta_y; x, X_j)],$$

$$e_{12} = \mathbb{E}\sum_{j=1}^{|\Pi|}[\Delta_2 h'(\Theta_1 + \langle\Pi\rangle_{j-1} + \delta_y; x, X_j) - \Delta_2 h'(\Theta_1 + \delta_y; x, X_j)],$$

$$e_{13} = \mathbb{E}\sum_{j=1}^{|\Pi|}[\Delta_2 h'(\Theta_1 + \delta_y; x, X_j) - \Delta_2 h'(\Theta_1; x, X_j)],$$

$$e_{14} = \mathbb{E}\sum_{j=1}^{|\Pi|}[\Delta_2 h'(\Theta_1; x, X_j) - \Delta_2 h'(\Theta_1; z, z)],$$

$$e_{15} = \mathbb{E}|\Pi|\mathbb{E}[\Delta_2 h'(\Theta_1; z, z) - \Delta_2 h'(\Theta_1 + \hat{\Theta}_2 + \hat{\Upsilon}; z, z)].$$

Estimates of e_{11} *and* e_{15}. Notice that e_{11} can be further decomposed as

$$\mathbb{E}\sum_{j=1}^{|\Pi|}\sum_{i=1}^{|\Theta_2|}\big[\Delta_2 h'(\Theta_1+\delta_y+\langle\Theta_2,\Pi\rangle_{i,j-1};x,X_j)\\ -\Delta_2 h'(\Theta_1+\delta_y+\langle\Theta_2,\Pi\rangle_{i-1,j-1};x,X_j)\big],$$

where $\langle\Theta_2,\Pi\rangle_{i,j} = \langle\Theta_2\rangle_i + \langle\Pi\rangle_j$ are measurable functions of (Θ_2,Π). When we take the expectation conditional on $\Xi_y|_{B_y}$, or equivalently on (Θ_2,Π), it can be interchanged with the sums. Therefore, we concentrate on the conditional expectation

$$\mathbb{E}\left(\Delta_2 h'(\Theta_1+\delta_y+\langle\Theta_2,\Pi\rangle_{i,j-1};x,X_j)-\Delta_2 h'(\Theta_1+\delta_y\\ +\langle\Theta_2,\Pi\rangle_{i-1,j-1};x,X_j)\,\middle|\,\Theta_2,\Pi\right). \tag{6.5}$$

Since by Remark 2.5 there is in general no uniform bound of the form const.$/a$ for $\Delta_2 h'$, we write

$$\Delta_2 h' = h^{(1)} + h^{(2)},$$

where

$$h^{(1)} = \min\left\{\max\left(\Delta_2 h', -\frac{2u+5}{a}\right), \frac{2u+5}{a}\right\}, \qquad h^{(2)} = \Delta_2 h' - h^{(1)}.$$

Since

$$|\Delta_2 h'(\xi;x,y)| \le \frac{2u+5}{a} \text{ for } 1+|\xi| > \frac{a}{u},$$

we have

$$|h^{(1)}| \le \frac{2u+5}{a}, \qquad |h^{(2)}| \le 2, \quad \text{and} \quad h^{(2)}(\xi;x,y)=0 \quad \text{for } 1+|\xi| > \frac{a}{u}. \tag{6.6}$$

For the quantity given in (6.5), the differences based on $h^{(1)}$ and $h^{(2)}$ are respectively bounded by the second and the first terms of $r_y(\Xi_y)$, recalling that $\Xi_y|_{B_y}$ is equivalent to (Θ_2,Π). Hence,

$$|e_{11}| \le \mathbb{E}|\Pi|\cdot|\Theta_2| r_y(\Xi_y) = \mathbb{E} r_y(\Xi_y)\Xi_y(A_y)\Xi_y(B_y\setminus A_y). \tag{6.7}$$

Similarly, taking conditional expectation on $(\hat\Theta_2, , \hat\Upsilon)$, we get

$$|e_{15}| \le \mathbb{E}|\Pi|\mathbb{E}\big(|\hat\Theta_2+\hat\Upsilon| r_y(\Theta_1+\hat\Theta_2+\hat\Upsilon)\big) = \mathbb{E} r_y(\Xi)\Xi(B_y)\mathbb{E}\Xi_y(A_y). \tag{6.8}$$

Estimates of e_{12} *and* e_{13}. Notice that Θ_2 disappears now and Θ_1 is independent of Π. We use the conditional expectation on Π, and find that each conditional expectation, actually being the mean, is less than $\bar r_y(\Xi_y) = \bar r_y(\Xi)$. Hence

$$|e_{12}| \le \mathbb{E}\frac{|\Pi|(|\Pi|-1)}{2}\bar r_y(\Xi) = \bar r_y(\Xi)\mathbb{E}\frac{\Xi_y(A_y)(\Xi_y(A_y)-1)}{2}, \tag{6.9}$$

$$|e_{13}| \le \bar r_y(\Xi)\mathbb{E}|\Pi| = \bar r_y(\Xi)\mathbb{E}\Xi_y(A_y). \tag{6.10}$$

An estimate of e_{14}. In fact, e_{14} is another kind of difference that is very different from the other four since the two point processes have the same size. Let us state a result which tells us the cost of shuffling points x and y in $\Delta_2 h(\xi; x, y)$. Define

$$Dh'(\xi; x, y) = h'(\xi + \delta_x) - h'(\xi + \delta_y),$$
$$D_2 h'(\xi; x, y; z) = Dh'(\xi + \delta_z; x, y) - Dh'(\xi; x, y).$$

Then, one can directly verify the following equation:

$$\Delta_2 h'(\xi; x, y) - \Delta_2 h'(\xi; z, z) = D_2 h'(\xi; y, z; x) + D_2 h'(\xi; x, z; z). \tag{6.11}$$

Consequently, we can rewrite

$$e_{14} = \mathbb{E}\sum_{j=1}^{|\Pi|}[D_2 h'(\Theta_1; X_j, z; x) + D_2 h'(\Theta_1; x, z; z)],$$

bearing in mind $\Pi = \sum_{j=1}^{|\Pi|} \delta_{X_j}$. Now we estimate $D_2 h'$. Recalling $|Dh'| \le C_n$ defined in (2.4) and estimated in (2.5), we have

$$|Dh'(\xi; x, y)| \le 1 \wedge \left(\frac{1}{2(|\xi| + 1)} + \frac{1}{a}\right).$$

If we set

$$Dh' = h^{(3)} + h^{(4)},$$

where

$$h^{(3)} = \max\left\{\min\left(Dh', \frac{u + 2.5}{a}\right), -\frac{u + 2.5}{a}\right\} \quad \text{and} \quad h^{(4)} = Dh' - h^{(3)},$$

then

$$|h^{(3)}| \le \frac{u + 2.5}{a}, \qquad |h^{(4)}| \le 1 \quad \text{and} \quad h^{(4)}(\xi; x, y) = 0 \quad \text{for } 1 + |\xi| > \frac{a}{u}. \tag{6.12}$$

Comparing with (6.6), we conclude that $D_2 h'$, as the difference of Dh', has conditional expectation (that reduces to its expectation) less than half of $\bar{r}_y(\Xi)$. Therefore,

$$|e_{14}| \le \bar{r}_y(\Xi)\mathbb{E}|\Pi| = \bar{r}_y(\Xi)\mathbb{E}\Xi_y(A_y). \tag{6.13}$$

Collecting (6.7)–(6.10) and (6.13), we obtain

$$|e_1| \le b\int_\Gamma [\mathbb{E}r_y(\Xi_y)\Xi_y(A_y)\Xi_y(B_y \setminus A_y) + \bar{r}_y(\Xi)\mathbb{E}(\Xi_y(A_y) + 3)\Xi_y(A_y)/2 + \mathbb{E}r_y(\Xi)\Xi(B_y)\mathbb{E}\Xi_y(A_y)]\lambda(dy). \tag{6.14}$$

The same procedure can be applied to estimate e_2 to e_5 by first selecting the 'stepping stones' $\Xi_y|_{B_y^c}$ and $\hat{\Xi}|_{B_y^c}$ to 'bridge' $\Xi_y|_{A_y^c}$ and $\hat{\Xi} \sim \mathscr{L}(\Xi)$ for e_2 and e_4, and $\Xi|_{B_y^c}$ and $\hat{\Xi}|_{B_y^c}$ to 'bridge' $\Xi|_{A_y^c}$ and $\hat{\Xi} \sim \mathscr{L}(\Xi)$ in e_3 and e_5, then telescoping within the layer of dependence and using (6.11) and (6.12) to deal with the relocation of points. We omit the details here and the estimates are summarized below:

$$|e_2| \le b\int_\Gamma [\mathbb{E}r_y(\Xi_y)\Xi_y(B_y \setminus A_y) + \bar{r}_y(\Xi) + \mathbb{E}r_y(\Xi)\Xi(B_y)]\lambda(dy),$$

$$\begin{aligned}|e_3| &\le b\int_\Gamma [\mathbb{E}r_y(\Xi)\Xi(A_y)\Xi(B_y\setminus A_y)+\bar r_y(\Xi)\mathbb{E}(\Xi(A_y)+1)\Xi(A_y)/2\\&\quad+\mathbb{E}r_y(\Xi)\Xi(B_y)\mathbb{E}\Xi(A_y)]\lambda(dy),\\|e_4| &\le \int_\Gamma [\mathbb{E}r_x(\Xi_x)\Xi_x(A_x)\Xi_x(B_x\setminus A_x)+\bar r_x(\Xi)\mathbb{E}(\Xi_x(A_x)+1)\Xi_x(A_x)/2\\&\quad+\mathbb{E}r_x(\Xi)\Xi(B_x)\mathbb{E}\Xi_x(A_x)]\lambda(dx),\\|e_5| &\le \int_\Gamma [\mathbb{E}r_x(\Xi)\Xi(A_x)\Xi(B_x\setminus A_x)+\bar r_x(\Xi)\mathbb{E}(\Xi(A_x)+1)\Xi(A_x)/2\\&\quad+\mathbb{E}r_x(\Xi)\Xi(B_x)\mathbb{E}\Xi(A_x)]\lambda(dx).\end{aligned}$$

Now, the above four estimates, together with (6.14), yield (6.1), completing the proof of Theorem 3.2. □

7. Proof of Theorem 3.3

The proof is similar to that of Theorem 3.2 with some modification to suit the estimation involving the second-order reduced Palm processes. Let Y be a point process with distribution $\boldsymbol{\pi}_{a,0;\beta;\nu}$; it follows from the triangle inequality that

$$\begin{aligned}d_2(\mathscr{L}(\Xi),\boldsymbol{\pi}_{a,0;\beta;\nu}) &\le d_2(\mathscr{L}(\Xi),\mathscr{L}(\mathscr{M}_\mathcal{G}\circ\Xi))+d_2(\mathscr{L}(\mathscr{M}_\mathcal{G}\circ\Xi),\mathscr{L}(\mathscr{M}_\mathcal{G}\circ Y))\\&\quad+d_2(\mathscr{L}(\mathscr{M}_\mathcal{G}\circ Y),\boldsymbol{\pi}_{a,0;\beta;\nu}).\end{aligned}$$

Again, (3.3) implies that $d_2(\mathscr{L}(\Xi),\mathscr{L}(\mathscr{M}_\mathcal{G}\circ\Xi))$ and $d_2(\mathscr{L}(\mathscr{M}_\mathcal{G}\circ Y),\boldsymbol{\pi}_{a,0;\beta;\nu})$ are bounded by $d_0(\mathcal{G})$, so $d_2(\mathscr{L}(\mathscr{M}_\mathcal{G}\circ\Xi),\mathscr{L}(\mathscr{M}_\mathcal{G}\circ Y))$ is the only term to be estimated.

We replace $\boldsymbol{\pi}$ by $\boldsymbol{\pi}_{a,0;\beta;\nu'}$ in the Stein equation (2.3) with $\nu'(dx)=\sum_{i=1}^k\nu(G_i)\delta_{t_i}(dx)$. It is sufficient to prove

$$|\mathbb{E}\mathscr{A}h_f(\mathscr{M}_\mathcal{G}\circ\Xi)|\le\int_\Gamma\mathbb{E}\left(\epsilon_{1,x}(\Xi_x)+\epsilon_{1,x}(\Xi)\right)\lambda(dx)$$
$$+\beta\iint_{\Gamma^2}\mathbb{E}\left(\epsilon_{1,x,y}(\Xi_{xy})+\epsilon_{1,x,y}(\Xi)+\epsilon_{2,x,y}(\Xi_{xy})\right)\lambda^{[2]}(dx,dy)\tag{7.1}$$

for all $f\in\mathscr{F}$. For the fixed $f\in\mathscr{F}$, we set $f'(\eta)=f(\mathscr{M}_\mathcal{G}\circ\eta)$, $h'(\eta)=h_f(\mathscr{M}_\mathcal{G}\circ\eta)$ and then apply (3.1) and (3.2) to deduce the following expansion:

$$\begin{aligned}\mathbb{E}\mathscr{A}h'(\Xi)&=\int_\Gamma\mathbb{E}[-\Delta h'(\Xi_x;x)+\Delta h'(\Xi;x)]\lambda(dx)\\&\quad+\beta\iint_{\Gamma^2}\mathbb{E}[-\Delta h'(\Xi_{xy}+\delta_y;x)+\Delta h'(\Xi;x)]\lambda^{[2]}(dx,dy)\\&\quad+\int_\Gamma\Delta\mathbb{E}h'(\Xi;x)\left(a\nu(dx)-\lambda(dx)-\beta\int_{y\in\Gamma}\lambda^{[2]}(dx,dy)\right).\end{aligned}\tag{7.2}$$

The last term of (7.2) vanishes because of the definition of ν in (3.5), and $\nu(\Gamma)=1$ ensures that

$$a=|\lambda|+\beta\iint_{\Gamma^2}\lambda^{[2]}(dx,dy)=|\lambda|+\beta(\mathbb{E}|\Xi|^2-|\lambda|).$$

We take $\hat\Xi$ as an independent copy of Ξ which is also independent of all Ξ_x's and Ξ_{xy}'s. Denote the points in $\Xi|_{A_x}$, $\Xi_x|_{A_x}$, $\Xi|_{A_{xy}}$, $\Xi_{xy}|_{A_{xy}}$ respectively by X_j, Y_j, W_j, V_j. Then using the two types of local dependence, we have

$$\begin{aligned}
\mathbb{E}\mathscr{A}h'(\Xi) = & \int_\Gamma \{\mathbb{E}[-\Delta h'(\Xi_x;x) + \Delta h'(\Xi_x|_{A_x^c};x)] + \mathbb{E}[\Delta h'(\Xi;x) \\
& - \Delta h'(\Xi|_{A_x^c};x)]\}\lambda(dx) \\
& - \beta \iint_{\Gamma^2} [\mathbb{E}\Delta h'(\Xi_{xy};x) - \mathbb{E}\Delta h'(\Xi_{xy}|_{A_{xy}^c};x)]\lambda^{[2]}(dx,dy) \\
& - \beta \iint_{\Gamma^2} [\mathbb{E}\Delta h'(\Xi_{xy}+\delta_y;x) - \mathbb{E}\Delta h'(\Xi_{xy};x)]\lambda^{[2]}(dx,dy) \\
& + \beta \iint_{\Gamma^2} [\mathbb{E}\Delta h'(\Xi;x) - \mathbb{E}\Delta h'(\Xi|_{A_{xy}^c};x)]\lambda^{[2]}(dx,dy) \\
= & -\int_\Gamma \mathbb{E}\sum_{j=1}^{|\Xi_x(A_x)|} [\Delta_2 h'(\Xi_x|_{A_x^c} + \langle \Xi_x|_{A_x}\rangle_{j-1};x,Y_j) \\
& - \Delta_2 h'(\hat{\Xi};X_1,X_1)]\lambda(dx) \\
& + \int_\Gamma \mathbb{E}\sum_{j=1}^{|\Xi(A_x)|} [\Delta_2 h'(\Xi|_{A_x^c} + \langle \Xi|_{A_x}\rangle_{j-1};x,X_j) \\
& - \Delta_2 h'(\hat{\Xi};X_1,X_1)]\lambda(dx) \\
& - \beta \iint_{\Gamma^2} \mathbb{E}\sum_{j=1}^{|\Xi_{xy}(A_{xy})|} [\Delta_2 h'(\Xi_{xy}|_{A_{xy}^c} + \langle \Xi_{xy}|_{A_{xy}}\rangle_{j-1};x,V_j) \\
& - \Delta_2 h'(\hat{\Xi};X_1,X_1)]\lambda^{[2]}(dx,dy) \\
& - \beta \iint_{\Gamma^2} \mathbb{E}[\Delta_2 h'(\Xi_{xy};x,y) - \Delta_2 h'(\hat{\Xi};X_1,X_1)]\lambda^{[2]}(dx,dy) \\
& + \beta \iint_{\Gamma^2} \mathbb{E}\sum_{j=1}^{|\Xi(A_{xy})|} [\Delta_2 h'(\Xi|_{A_{xy}^c} + \langle \Xi|_{A_{xy}}\rangle_{j-1};x,W_j) \\
& - \Delta_2 h'(\hat{\Xi};X_1,X_1)]\lambda^{[2]}(dx,dy) \\
& - \mathbb{E}\Delta_2 h'(\hat{\Xi};X_1,X_1)\left[\int_\Gamma \mathbb{E}(\Xi_x(A_x) - \Xi(A_x))\lambda(dx)\right. \\
& \left. + \beta \iint_{\Gamma^2} \mathbb{E}(\Xi_{xy}(A_{xy}) + 1 - \Xi(A_{xy}))\lambda^{[2]}(dx,dy)\right] \\
=: & \ \phi_1 + \cdots + \phi_6. \qquad (7.3)
\end{aligned}$$

The term ϕ_6 becomes 0 if we set

$$\int_\Gamma \mathbb{E}(\Xi_x(A_x) - \Xi(A_x))\lambda(dx) + \beta \iint_{\Gamma^2} \mathbb{E}(\Xi_{xy}(A_{xy}) + 1 - \Xi(A_{xy}))\lambda^{[2]}(dx,dy) = 0,$$

and hence the β in (3.4) follows from (6.4), $\iint_{\Gamma^2} \lambda^{[2]}(dx,dy) = \mathbb{E}|\Xi|(|\Xi|-1)$ and the following observation:

$$\begin{aligned}
\iint_{\Gamma^2} \mathbb{E}(\Xi_{xy}(A_{xy}) - \Xi(A_{xy}))\lambda^{[2]}(dx,dy) &= \iint_{\Gamma^2} \mathbb{E}(|\Xi_{xy}| - |\Xi|)\lambda^{[2]}(dx,dy) \\
&= \mathbb{E}(|\Xi| - 2 - |\lambda|)(|\Xi| - 1)|\Xi|.
\end{aligned}$$

Following the same steps as the estimation of (6.14), with 'stepping stones' $\Xi_x|_{B_x^c}$ and $\hat{\Xi}|_{B_x^c}$ for ϕ_1, $\Xi|_{B_x^c}$ and $\hat{\Xi}|_{B_x^c}$ for ϕ_2, $\Xi_{xy}|_{B_{xy}^c}$ and $\hat{\Xi}|_{B_{xy}^c}$ for ϕ_3 and ϕ_4, and $\Xi|_{B_{xy}^c}$ and $\hat{\Xi}|_{B_{xy}^c}$ for ϕ_5, we obtain

$$|\phi_1| \le \int_\Gamma \mathbb{E}\epsilon_{1,x}(\Xi_x)\lambda(dx);$$

$$|\phi_2| \le \int_\Gamma \mathbb{E}\epsilon_{1,x}(\Xi)\lambda(dx);$$

$$|\phi_3| \le \beta \iint_{\Gamma^2} \mathbb{E}\epsilon_{1,x,y}(\Xi_{xy})\lambda^{[2]}(dx, dy);$$

$$|\phi_4| \le \beta \iint_{\Gamma^2} \mathbb{E}\epsilon_{2,x,y}(\Xi_{xy})\lambda^{[2]}(dx, dy);$$

$$|\phi_5| \le \beta \iint_{\Gamma^2} \mathbb{E}\epsilon_{1,x,y}(\Xi)\lambda^{[2]}(dx, dy),$$

which, together with (7.3), in turn imply (7.1). This completes the proof of Theorem 3.3. □

Acknowledgements

This work was supported by the Belz fund from the University of Melbourne (AX) and NSFC10901008 and the National Excellent Ph.D. Thesis fund 200722 from Peking University (FZ). We would like to thank Prof. A.D. Barbour for a suggestion leading to the example in Remark 4.2.

References

[1] D. Aldous, Probability Approximations via the Poisson Clumping Heuristic, Springer, New York, 1989.
[2] R. Arratia, L. Goldstein, L. Gordon, Two moments suffice for Poisson approximations: the Chen–Stein method, Ann. Probab. 17 (1989) 9–25.
[3] A.D. Barbour, Stein's method and Poisson process convergence, J. Appl. Probab. 25 (A) (1988) 175–184.
[4] A.D. Barbour, T.C. Brown, Stein's method and point process approximation, Stochastic Process. Appl. 43 (1992) 9–31.
[5] A.D. Barbour, T.C. Brown, A. Xia, Point processes in time and Stein's method, Stoch. Stoch. Rep. 65 (1998) 127–151.
[6] A.D. Barbour, L.H.Y. Chen, W. Loh, Compound Poisson approximation for nonnegative random variables via Stein's method, Ann. Probab. 20 (1992) 1843–1866.
[7] A.D. Barbour, P. Hall, On the rate of Poisson convergence, Math. Proc. Cambridge Philos. Soc. 95 (1984) 473–480.
[8] A.D. Barbour, L. Holst, S. Janson, Poisson Approximation, Oxford Univ. Press, 1992.
[9] A.D. Barbour, J.L. Jensen, Local and tail approximations near the Poisson limit, Scand. J. Stat. 16 (1989) 75–87.
[10] A.D. Barbour, M. Månsson, Compound Poisson process approximation, Ann. Probab. 30 (2002) 1492–1537.
[11] A.D. Barbour, S. Utev, Solving the Stein equation in compound Poisson approximation, Adv. Appl. Prob. 30 (1998) 449–475.
[12] A.D. Barbour, S. Utev, Compound Poisson approximation in total variation, Stochastic Process. Appl. 82 (1999) 89–125.
[13] A.D. Barbour, A. Xia, Poisson perturbations, ESAIM Probab. Stat. 3 (1999) 131–150.
[14] A.D. Barbour, A. Xia, Normal approximation for random sums, Adv. Appl. Prob. 38 (2006) 693–728.
[15] T.C. Brown, Some Poisson approximations using compensators, Ann. Probab. 11 (1983) 726–744.
[16] T.C. Brown, G.V. Weinberg, A. Xia, Removing logarithms from Poisson process error bounds, Stochastic Process. Appl. 87 (2000) 149–165.
[17] T.C. Brown, A. Xia, Steins method and birth–death processes, Ann. Probab. 29 (2001) 1373–1403.
[18] V. Čekanavičius, Asymptotic expansions in the exponent: a compound Poisson approach, Adv. Appl. Prob. 29 (1997) 374–387.

[19] L.H.Y. Chen, Q.M. Shao, Normal approximation under local dependence, Ann. Probab. 32 (2004) 1985–2028.
[20] L.H.Y. Chen, A. Xia, Stein's method, Palm theory and Poisson process approximation, Ann. Probab. 32 (2004) 2545–2569.
[21] P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, Springer-Verlag, Berlin, 1997.
[22] J. Jacod, P. Mano, Une evaluation de la distance entre les lois d'une semimartingale et d'un processus a accroissements independants, Stochastics 25 (1988) 87–124.
[23] O. Kallenberg, Random Measures, Academic Press, London, 1983.
[24] J. Kruopis, Precision of approximations of the generalized Binomial distribution by convolutions of Poisson measures, Lith. Math. J. 26 (1986) 37–49.
[25] M. Nikunen, E. Valkeila, A Prohorov bound for a Poisson process and an arbitrary counting process with some applications, Stoch. Stoch. Rep. 37 (1991) 133–151.
[26] E.L. Presman, Approximation of binomial distributions by infinitely divisible ones, Theory Probab. Appl. 28 (1983) 393–403.
[27] S.T. Rachev, Probability Metrics and the Stability of Stochastic Models, John Wiley & Sons, 1991.
[28] A. Röllin, Approximation of sum of conditionally independent variables by the translated Poisson distribution, Bernoulli 11 (2005) 1115–1128.
[29] P.S. Ruzankin, On the rate of Poisson process approximation to a Bernoulli process, J. Appl. Probab. 41 (2004) 271–276.
[30] D. Schuhmacher, A. Xia, A new metric between distributions of point processes, Adv. Appl. Prob. 40 (2008) 651–672.
[31] X. Wang, A. Xia, On negative binomial approximation to k-runs, J. Appl. Probab. 45 (2008) 456–471.
[32] A. Xia, A note on the Prohorov distance between a counting process and a Poisson process, Stoch. Stoch. Rep. 45 (1993) 61–77.
[33] A. Xia, On using the first difference in the Stein–Chen method, Ann. Appl. Probab. 7 (1997) 899–916.
[34] A. Xia, in: A.D. Barbour, L.H.Y. Chen (Eds.), An Introduction to Stein's Method, World Scientific Press, Singapore, 2005, pp. 115–181.
[35] A. Xia, F. Zhang, A polynomial birth–death point process approximation to the Bernoulli process, Stochastic Process. Appl. 118 (2008) 1254–1263.