Stochastic Processes and their Applications 43 (1992) 9-31
North-Holland

Stein's method and point process approximation

A.D. Barbour
University of Zürich, Switzerland

T.C. Brown
University of Western Australia, Nedlands, Australia

Received 2 July 1990
Revised 27 August 1991

The Stein-Chen method for Poisson approximation is adapted into a form suitable for obtaining error estimates for the approximation of the whole distribution of a point process on a suitable topological space by that of a Poisson process. The adaptation involves consideration of an immigration-death process on the topological space, whose equilibrium distribution is that of the approximating Poisson process; the Stein equation has a simple interpretation in terms of the generator of the immigration-death process. The error estimates for process approximation in total variation do not have the 'magic' Stein-Chen multiplying constants, which for univariate approximation tend to zero as the mean gets larger, but examples, including Bernoulli trials and the hard-core model on the torus, show that this is not possible. By choosing weaker metrics on the space of distributions of point processes, it is possible to reintroduce these constants. The proofs actually yield an improved estimate for one of the constants in the univariate case.

AMS 1980 Subject Classification (1985 Revision): 60G55.

Poisson approximation * point process * Stein-Chen method * coupling * Palm probability * immigration-death process * hard-core model

1. Introduction

The aim of the paper is to estimate the error when approximating a point process on a general space by a Poisson process. We accomplish this with respect to the total variation metric and also to other, weaker metrics, using the Stein-Chen method. In order to describe the results further, it is convenient to start with the case when the carrier space is finite.

Correspondence to: Dr. A.D. Barbour, Institut für Angewandte Mathematik, Universität Zurich, Rämistrasse 74, CH-8001, Zürich, Switzerland.
We wish to thank the referee for many pertinent observations, especially in connection with the hard-core example. We are also grateful for financial support from the Schweiz. Nationalfonds under Grant Nr. 21-25579.88, and from the Australian Research Council.

0304-4149/92/$05.00

Stein's method of obtaining estimates of the error in distributional approximations was first developed in the context of Poisson approximation by Chen (1975). His main theorem can be described as follows. Let Γ be a finite set of indices, and let $(I_\alpha, \alpha \in \Gamma)$ be indicator random variables; set $\pi_\alpha = \mathbb{E} I_\alpha$, $W = \sum_{\alpha \in \Gamma} I_\alpha$ and $\lambda = \sum_{\alpha \in \Gamma} \pi_\alpha$. Suppose that, for each α, a subset $N_\alpha \subset \Gamma$ is given such that $N_\alpha \ni \alpha$, and set

$$Z_\alpha = \sum_{\beta \in N_\alpha \setminus \alpha} I_\beta, \qquad \eta_\alpha = \mathbb{E}\left|\mathbb{E}\{I_\alpha \,|\, (I_\beta, \beta \notin N_\alpha)\} - \pi_\alpha\right|.$$

Let d_{TV} denote the total variation metric on probability distributions, so that, if $\mathscr{P}$ and $\mathscr{Q}$ are probability distributions on some space, then

$$d_{\mathrm{TV}}(\mathscr{P}, \mathscr{Q}) = \sup|\mathscr{P}(A) - \mathscr{Q}(A)|,$$

where the supremum is over all measurable subsets.

Theorem 1.1. *There exist constants $c_1(\lambda)$ and $c_2(\lambda)$ such that*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda)) \leqslant c_1(\lambda) \sum_{\alpha \in \Gamma} \eta_\alpha + c_2(\lambda) \sum_{\alpha \in \Gamma} (\pi_\alpha^2 + \pi_\alpha \mathbb{E} Z_\alpha + \mathbb{E}(I_\alpha Z_\alpha)). \qquad \square$$

The constants c_1 and c_2 are shown in Barbour and Eagleson (1983) to satisfy the inequalities

$$c_1(\lambda) \leqslant 1 \wedge 1.4\lambda^{-1/2}, \qquad c_2(\lambda) \leqslant \lambda^{-1}(1 - \mathrm{e}^{-\lambda}).$$

Another way of using the Stein–Chen method to bound the error in Poisson approximation is to use coupling, as in Barbour and Holst (1989) and in Stein (1986, pp. 92–93). Suppose that, for each α, random variables U_α and V_α can be constructed on the same probability space in such a way that

$$\mathscr{L}(U_\alpha) = \mathscr{L}(W), \qquad \mathscr{L}(V_\alpha + 1) = \mathscr{L}(W \,|\, I_\alpha = 1).$$

Then the following result can be proved.

Theorem 1.2. *Whatever the choice of couplings $((U_\alpha, V_\alpha), \alpha \in \Gamma)$,*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda)) \leqslant c_2(\lambda) \sum_{\alpha \in \Gamma} \pi_\alpha \mathbb{E}|U_\alpha - V_\alpha|,$$

where c_2 is as before. $\square$

Theorem 1.1 is typically more suitable when dependence is in some sense local, and the N_α's can be chosen to contain those indices β where I_β is relatively strongly dependent on I_α: Theorem 1.2 is useful when the dependence among the I_α's is symmetric.

There are two cases of particular interest. If the couplings (U_α, V_α) are realized in such a way that $U_\alpha \geq V_\alpha$ a.s. for all α, the indicators $(I_\alpha, \alpha \in \Gamma)$ are said to be negatively related, and the estimate in Theorem 1.2 reduces to $c_2(\lambda)\{\lambda - \text{Var } W\}$. If slightly finer couplings $(U'_\alpha, I'_\alpha, V_\alpha)$ can be realized in such a way that, for each $\alpha \in \Gamma$,

$$\mathscr{L}((U'_\alpha, I'_\alpha)) = \mathscr{L}((W - I_\alpha, I_\alpha)),$$

$$\mathscr{L}(V_\alpha + 1) = \mathscr{L}(W \mid I_\alpha = 1), \qquad U'_\alpha \leq V_\alpha \quad \text{a.s.},$$

then the indicators $(I_\alpha, \alpha \in \Gamma)$ are said to be positively related, and the estimate in Theorem 1.2 is no greater than $c_2(\lambda)\{\text{Var } W - \lambda + 2\sum_{\alpha \in \Gamma} \pi_\alpha^2\}$.

In the context of a finite Γ, some process generalization has already been made: in Arratia, Goldstein and Gordon (1989), where an analogue of Theorem 1.1 for processes is proved, and in Barbour (1988). Here, in Section 2, it is shown that both Theorems 1.1 and 1.2 have process counterparts. However, a major difference between the results for processes and those for random variables is that the constants $c'_1(\lambda)$ and $c'_2(\lambda)$ for process approximation no longer decrease towards zero with increasing λ. This is an inevitable consequence of using the total variation metric, since, if the $(I_\alpha, \alpha \in \Gamma)$ are independent, either theorem gives a best estimate of $c'_2(\lambda) \sum_{\alpha \in \Gamma} \pi_\alpha^2$, whereas the event $A = \{\text{there exists a multiple point}\}$ has zero probability for the process $(I_\alpha, \alpha \in \Gamma)$ and a probability of order $\sum_{\alpha \in \Gamma} \pi_\alpha^2 \wedge 1$ for the Poisson process. In Section 3, by choosing smoother metrics on the space of point process distributions, it is shown that constants $c''_1(\lambda)$ and $c''_2(\lambda)$, with much the same dependence on λ as $c_1(\lambda)$ and $c_2(\lambda)$, can be recovered.

The approach used here does not, as in the previous papers, proceed by way of finite-dimensional distributions. Instead, the method used in Barbour (1988) is formulated directly in the process setting. This makes the argument more transparent, and, in particular, leads to simple probabilistic formulae for the various constants c_1 and c_2. One unexpected consequence is that the estimate of $c_1(\lambda)$ in Theorem 1.1 is slightly improved, to $c_1(\lambda) \leq 1 \wedge \lambda^{-1/2}\sqrt{2/e}$.

In constructing approximations by a Poisson process, it is natural to take Γ to be a compact, second countable Hausdorff space and to replace the process $(I_\alpha)_{\alpha \in \Gamma}$ by a point process Ξ on Γ. This gives great generality to the results and permits the consideration of approximation of arbitrary point processes in Euclidean or other spaces by a Poisson process. The only price to be paid for the generality is some technical complication, though the proofs are not very different from those which would be needed in the finite case. For general definitions and notation see Kallenberg (1976); we recall that Ξ is a *point process* if it is a non-negative, integer valued random measure on Γ and that it is *simple* if $\Xi(\{\alpha\}, \omega) \leq 1$ for all $\alpha \in \Gamma$ and $\omega \in \Omega$. Since Γ is assumed to be compact, it is natural also to assume that Ξ is almost surely finite, which we do throughout. We also assume that the processes have finite mean measures. To apply the results of the paper to processes defined over the whole of a Euclidean space, it may thus be necessary to approximate the process on an increasing sequence of compacta.

In Barbour and Brown (1990), approximations for the distribution of the total number of points $\Xi(\Gamma)$ were considered. Papangelou (1980) considered the problem of total variation approximation of the whole point process by a mixed sample process, of which a Poisson process is a special case. Apart from this difference, his bound contains as one term the distance of the distribution of the total number of points from the corresponding distribution for the mixed sample process, whereas the result here does not. Papangelou's result is most directly connected to Theorem 2.6.

2. Total variation approximation

In this section, we reformulate the multivariate Stein equation of Barbour (1988) in process form, and show how it may be used to establish bounds on the total variation distance between the distribution of the point process Ξ on Γ and the distribution $\mathrm{Po}(\boldsymbol{\pi})$ of a Poisson process over Γ with mean measure $\boldsymbol{\pi}$, assumed *finite*. Let δ_α denote the point mass at α and $\pi = \boldsymbol{\pi}(\Gamma)$. Let $\mathscr{X}$ denote the space of *finite* point process configurations on Γ, so that $\mathscr{X}$ is the space of finite, non-negative, integer valued measures on Γ. If ξ denotes a typical element of $\mathscr{X}$, the generator $\mathscr{A}$ defined by

$$(\mathscr{A}h)(\xi) = \int_\Gamma [h(\xi+\delta_\alpha) - h(\xi)]\boldsymbol{\pi}(\mathrm{d}\alpha) + \int_\Gamma [h(\xi-\delta_\alpha) - h(\xi)]\xi(\mathrm{d}\alpha), \quad (2.1)$$

for a suitable class of functions h, is that of the immigration-death process Z on Γ with immigration intensity $\boldsymbol{\pi}$ and with unit per capita death rate, and Z has equilibrium distribution $\mathrm{Po}(\boldsymbol{\pi})$. The corresponding Stein equation is the equation

$$(\mathscr{A}h)(\xi) = f(\xi) - \mathrm{Po}(\boldsymbol{\pi})(f), \quad (2.2)$$

which makes sense at least for bounded functions $f: \mathscr{X} \to \mathbb{R}$, where $\mu(f)$ denotes $\int f\,\mathrm{d}\mu$ (all functions introduced are assumed to be measurable). The first step is to construct a solution of (2.2). We use the notation $\mathbb{P}^\mu$ to denote the distribution of Z when the initial distribution is μ, and $\mathbb{P}^\xi$ as shorthand for $\mathbb{P}^{\delta_\xi}$; $\mathbb{E}^\mu$ and $\mathbb{E}^\xi$ are defined similarly.

Proposition 2.1. *For any bounded $f: \mathscr{X} \to \mathbb{R}$, the function $h: \mathscr{X} \to \mathbb{R}$ given by*

$$h(\xi) = -\int_0^\infty [\mathbb{E}^\xi f(Z(t)) - \mathrm{Po}(\boldsymbol{\pi})(f)]\,\mathrm{d}t \quad (2.3)$$

is well defined.

Proof. Consider the simple coupling of an immigration-death process Z under $\mathbb{P}^\xi$ and a similar process $\tilde{Z}$ under $\mathbb{P}^{\mathrm{Po}(\boldsymbol{\pi})}$, taking $Z = Z_0 + D$, $\tilde{Z} = Z_0 + \tilde{D}$, where Z_0, D and $\tilde{D}$ are independent, Z_0 denotes the immigration-death process under $\mathbb{P}^0$ with no initial particles, and D and $\tilde{D}$ are pure death processes with unit per capita death rate such that $D(0)=\xi$ and $\tilde{D}(0)\sim \mathrm{Po}(\pi)$. Then $Z(t)=\tilde{Z}(t)$ for all $t\geqslant\tau$, where

$$\tau=\inf\{u\geqslant 0\colon D(u)=\tilde{D}(u)=0\}.$$

Hence the integral in (2.3) is bounded by

$$\int_0^\infty |\mathbb{E}^\xi f(Z(t))-\mathbb{E}f(\tilde{Z}(t))|\,\mathrm{d}t\leqslant 2\int_0^\infty \|f\|\mathbb{P}[\tau>t]\,\mathrm{d}t$$
$$=2\|f\|\mathbb{E}\tau\leqslant 2\|f\|\mathbb{E}\psi(|\xi|+|\tilde{D}(0)|)<\infty,$$

where $\|f\|=\sup_\xi|f(\xi)|$, $\psi(l)=\sum_{r=1}^{l}1/r$, and the last inequality follows because $|\tilde{D}(0)|\sim\mathrm{Po}(\pi)$. □

It is shown in Proposition 2.3 below that the function h so defined is a solution of the Stein equation (2.2). In order to then use the Stein equation to establish Poisson process approximation, it is necessary to have a result which expresses the smoothness of its solution h, given in (2.3), in terms of properties of f. This is the substance of the following lemma, which is also needed in the proof of Proposition 2.3.

Lemma 2.2. *If h is defined by* (2.3), *and if* $f(\xi)=I[\xi\in A]$,

(i) $$\Delta_1 h=\sup_{\xi\in\mathscr{X},\alpha\in\Gamma}|h(\xi+\delta_\alpha)-h(\xi)|\leqslant 1;$$

(ii) $$\Delta_2 h=\sup_{\xi\in\mathscr{X};\alpha,\beta\in\Gamma}|h(\xi+\delta_\alpha+\delta_\beta)-h(\xi+\delta_\alpha)-h(\xi+\delta_\beta)+h(\xi)|\leqslant 1.$$

Proof. For part (i), note that, from the definition (2.3) of h,

$$h(\xi+\delta_\alpha)-h(\xi)=\int_0^\infty [\mathbb{E}^\xi f(Z(t))-\mathbb{E}^{\xi+\delta_\alpha}f(Z(t))]\,\mathrm{d}t,$$

where Z is the immigration-death process on Γ with immigration intensity $(\pi_\beta,\beta\in\Gamma)$ and with unit per capita death rate. Let Z be realized under $\mathbb{P}^\xi$, and let E be an independent negative exponential random variable with unit mean. Then the process Z' defined by $Z'(t)=Z(t)+\delta_\alpha I[E>t]$ has distribution $\mathbb{P}^{\xi+\delta_\alpha}$. Thus it follows that

$$h(\xi+\delta_\alpha)-h(\xi)=\int_0^\infty \mathbb{E}^\xi[f(Z(t))-f(Z(t)+\delta_\alpha)]\,\mathrm{e}^{-t}\,\mathrm{d}t,$$

and the estimate $\Delta_1 h\leqslant 1$ is immediate from $|f(\xi)-f(\xi+\delta_\alpha)|\leqslant 1$.

A similar coupling, introducing two independent negative exponential random variables E_α and E_β with unit means, can be used to link $\mathbb{P}^\xi,\mathbb{P}^{\xi+\delta_\alpha},\mathbb{P}^{\xi+\delta_\beta}$ and $\mathbb{P}^{\xi+\delta_\alpha+\delta_\beta}$ by means of processes Z, Z_α, Z_β and $Z_{\alpha\beta}$, where Z has distribution $\mathbb{P}^\xi$,

$$Z_\alpha(t)=Z(t)+\delta_\alpha I[E_\alpha>t],\qquad Z_\beta(t)=Z(t)+\delta_\beta I[E_\beta>t],$$

and

$$Z_{\alpha\beta}(t)=Z(t)+\delta_\alpha I[E_\alpha>t]+\delta_\beta I[E_\beta>t].$$

Then it follows that

$$\begin{aligned}&h(\xi+\delta_\alpha+\delta_\beta)-h(\xi+\delta_\alpha)-h(\xi+\delta_\beta)+h(\xi)\\&=-\int_0^\infty\{\mathbb{E}^{\xi+\delta_\alpha+\delta_\beta}f(Z(t))-\mathbb{E}^{\xi+\delta_\alpha}f(Z(t))-\mathbb{E}^{\xi+\delta_\beta}f(Z(t))+\mathbb{E}^{\xi}f(Z(t))\}\,\mathrm{d}t\\&=-\int_0^\infty\mathbb{E}^{\xi}\{f(Z(t)+\delta_\alpha+\delta_\beta)-f(Z(t)+\delta_\alpha)\\&\qquad -f(Z(t)+\delta_\beta)+f(Z(t))\}\,\mathrm{e}^{-2t}\,\mathrm{d}t,\end{aligned}$$

giving $\Delta_2 h\leqslant 1$. □

Proposition 2.3. *The function h defined in Proposition* 2.1 *satisfies equation* (2.2).

Proof. Let $h_t(\xi)=-\int_0^t[\mathbb{E}^\xi f(Z(u))-\mathrm{Po}(\boldsymbol{\pi})(f)]\,\mathrm{d}u$. The first time at which a particle is born or dies has an exponential distribution with parameter $q_\xi=\pi+\xi(\Gamma)$. Hence

$$\begin{aligned}h_t(\xi)=&-(f(\xi)-\mathrm{Po}(\boldsymbol{\pi})(f))\,\mathrm{e}^{-q_\xi t}t\\&+\int_0^t\mathrm{e}^{-q_\xi u}\Bigg[-q_\xi u(f(\xi)-\mathrm{Po}(\boldsymbol{\pi})(f))+\int_\Gamma h_{t-u}(\xi+\delta_\alpha)\boldsymbol{\pi}(\mathrm{d}\alpha)\\&+\int_\Gamma h_{t-u}(\xi-\delta_\alpha)\xi(\mathrm{d}\alpha)\Bigg]\,\mathrm{d}u.\end{aligned}$$

From the proof of Proposition 2.1, the functions $h_s(\xi)$ are uniformly bounded in s for each ξ and, from Lemma 2.2(i), $h_s(\xi+\delta_\alpha)$ and $h_s(\xi-\delta_\alpha)$ are therefore uniformly bounded in s and α. Hence we may let $t\to\infty$ and apply bounded convergence to give

$$\begin{aligned}h(\xi)=\frac{1}{q_\xi}\Bigg\{&-[f(\xi)-\mathrm{Po}(\boldsymbol{\pi})(f)]+\int_\Gamma h(\xi+\delta_\alpha)\boldsymbol{\pi}(\mathrm{d}\alpha)\\&+\int_\Gamma h(\xi-\delta_\alpha)\xi(\mathrm{d}\alpha)\Bigg\}.\end{aligned}$$

The proposition follows by rearrangement of the equation. □

In order to obtain an analogue of Theorem 1.1, it is necessary to introduce neighbourhoods N_α of each point α in Γ. These are assumed to be measurable subsets of Γ with $N_\alpha\ni\alpha$. It is also assumed that the mappings

(a) $\mathscr{X}\times\Gamma\to[0,\infty)$: $(\xi,\alpha)\mapsto\xi(N_\alpha)$

and

(b) $\mathscr{X}\times\Gamma\to\mathscr{X}$: $(\xi,\alpha)\mapsto\xi$ restricted to N_α^{c}

are product measurable. Since Γ is a metric space (see Kallenberg, 1976, p. 1), a natural choice for N_α is the closed ball of radius r centred at α.

We next show that such a choice satisfies (a) and (b). For $\varepsilon > 0$, let $f_\alpha^\varepsilon(\beta)$ be $\phi(\varepsilon^{-1} d(\beta, N_\alpha))$, where d is a metric on Γ and $\phi(t) = \max\{1-t, 0\}$. Then f_α^ε converges pointwise to the indicator of N_α. For fixed α and ξ, it therefore follows by dominated convergence that

$$\int_\Gamma f_\alpha^\varepsilon(\beta)\xi(\mathrm{d}\beta) \to \xi(N_\alpha),$$

and that, for a bounded continuous function $g : \Gamma \to \mathbb{R}$,

$$\int_\Gamma g f_\alpha^\varepsilon(\beta)\xi(\mathrm{d}\beta) \to \int_{N_\alpha} g \,\mathrm{d}\xi.$$

Thus it suffices to demonstrate the measurability, for fixed ε, of the mappings

$$(\xi, \alpha) \mapsto \int_\Gamma f_\alpha^\varepsilon \,\mathrm{d}\xi \quad \text{and} \quad (\xi, \alpha) \mapsto f_\alpha^\varepsilon \xi,$$

where $f_\alpha^\varepsilon \xi$ is the measure on Γ giving $f_\alpha^\varepsilon \xi(B) = \int_B f_\alpha^\varepsilon \,\mathrm{d}\xi$ for any Borel set B. This is done by demonstrating continuity of the maps with respect to the vague topology. Suppose that $\alpha_n \to \alpha$ and $\xi_n \to \xi$, and that $\{\beta_n\}$ is any sequence converging to some β. Then

$$|f_{\alpha_n}^\varepsilon(\beta_n) - f_\alpha^\varepsilon(\beta)| \le \varepsilon^{-1} |d(\beta_n, N_{\alpha_n}) - d(\beta, N_\alpha)| \to 0,$$

because N_α is the closed ball of radius r and centre α. Continuity now follows from Kallenberg (1976, A 7.3), and the definition of vague convergence.

To formulate the analogue of η_α in Theorem 1.1 for a point process, we make some additional assumptions. First, it is supposed that there is a fixed measure ν on Γ. In applications with Γ a finite set, ν would be counting measure, and on a compact subset of Euclidean space it would be Lebesgue measure. Next, it is supposed that, for $n = 0, 1, 2, \ldots$, the Janossi densities $j_n : \Gamma^n \to [0, \infty)$ with respect to ν^n exist for the process Ξ. This means that, for any non-negative measurable function $f : \mathscr{X} \to \mathbb{R}$,

$$\mathbb{E}(f(\Xi)) = \sum_{n \ge 0} \int_{\Gamma^n} (n!)^{-1} f\left(\sum_{i=1}^n \delta_{\alpha_i}\right) j_n(\alpha_1, \ldots, \alpha_n) \nu^n(\mathrm{d}\alpha_1, \ldots, \mathrm{d}\alpha_n). \quad (2.4)$$

Informally, $(n!)^{-1} j_n(\alpha_1, \ldots, \alpha_n)\nu^n(\mathrm{d}\alpha_1, \ldots, \mathrm{d}\alpha_n)$ is the probability of Ξ producing points near $\alpha_1, \ldots, \alpha_n$; the term with $n = 0$ is interpreted as $j_0 f(\emptyset)$. We may use these densities to produce the first moment measure π of the process: this has density $\mu(\alpha)$ given by

$$\mu(\alpha) = \sum_{n \ge 0} \int_{\Gamma^n} (n!)^{-1} j_{n+1}(\alpha, \alpha_1, \ldots, \alpha_n) \nu^n(\mathrm{d}\alpha_1, \ldots, \mathrm{d}\alpha_n) \quad (2.5)$$

(Daley and Vere-Jones, 1988, Lemma 5.4.III). The interest here is that, in like fashion, the densities can be used to define the probability of a point being near α, given the configuration Ξ^α of Ξ outside N_α. Let m be fixed and $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_m)\in(N_\alpha^c)^m$. For any α, define

$$g(\alpha,\boldsymbol{\beta})=\frac{\sum_{r\geq 0}\int_{N_\alpha^r} j_{m+r+1}(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma})(r!)^{-1}\nu^r(\mathrm{d}\boldsymbol{\gamma})}{\sum_{s\geq 0}\int_{N_\alpha^s} j_{m+s}(\boldsymbol{\beta},\boldsymbol{\eta})(s!)^{-1}\nu^s(\mathrm{d}\boldsymbol{\eta})} \tag{2.6}$$

(the term with $r=0$ is interpreted as $j_{m+1}(\alpha,\boldsymbol{\beta})$ and the term with $s=0$ similarly), so that g is the density of a point at α given that Ξ outside N_α is $\sum_{i=1}^m \delta_{\beta_i}$. If the denominator is zero, $g(\alpha,\boldsymbol{\beta})$ is interpreted as zero also. For any ξ in $\mathscr{X}$ concentrated on N_α^c, we may write $\xi=\sum_{i=1}^m \delta_{\beta_i}$ for some m and $\beta_1,\ldots,\beta_m\in N_\alpha^c$. Because of this, and because $\{j_n\}$ is a sequence of symmetric functions, it makes sense to write $g(\alpha,\xi)$ for $g(\alpha,\boldsymbol{\beta})$. It is also convenient, for any $\psi\in\mathscr{X}$, to write $j_n(\psi)$ for $j_n(\alpha_1,\ldots,\alpha_n)$ where $\psi=\sum_{i=1}^n\delta_{\alpha_i}$. Finally, for $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_n)\in\Gamma^n$, we write $\delta_{\boldsymbol{\alpha}}$ for ψ given in the previous sentence.

The important property of g is that, for any non-negative or bounded measurable $h:\mathscr{X}\to\mathbb{R}$, we have

$$\mathbb{E}\left(\int_\Gamma h(\Xi^\alpha)\Xi(\mathrm{d}\alpha)\right)=\mathbb{E}\left(\int_\Gamma h(\Xi^\alpha)g(\alpha,\Xi^\alpha)\nu(\mathrm{d}\alpha)\right). \tag{2.7}$$

To see this, note that the integrand on the right-hand side is a measurable function of Ξ, and, according to (2.4), we may write the right side as

$$\int_\Gamma\sum_{n\geq 0}\sum_{s\geq 0}\int_{(N_\alpha^c)^n}\int_{(N_\alpha)^s} h(\delta_{\boldsymbol{\beta}})g(\alpha,\boldsymbol{\beta})j_{n+s}(\delta_{\boldsymbol{\beta}}+\delta_{\boldsymbol{\eta}})(n!s!)^{-1}$$
$$\times\nu^s(\mathrm{d}\boldsymbol{\eta})\nu^n(\mathrm{d}\boldsymbol{\beta})\nu(\mathrm{d}\alpha),$$

which can be re-expressed as

$$\int_\Gamma\sum_{n\geq 0}\int_{(N_\alpha^c)^n}\sum_{r\geq 0}\int_{(N_\alpha)^r} h(\delta_{\boldsymbol{\beta}})j_{n+r+1}(\delta_\alpha+\delta_{\boldsymbol{\beta}}+\delta_{\boldsymbol{\gamma}})(n!)^{-1}(r!)^{-1}$$
$$\times\nu^r(\mathrm{d}\boldsymbol{\gamma})\nu^n(\mathrm{d}\boldsymbol{\beta})\nu(\mathrm{d}\alpha),$$

upon integrating over $\boldsymbol{\eta}$, then summing over s and using (2.6). Note that all the changes of order of integration are justified by the boundedness or non-negativity of h and the norming of $\{j_n\}$. Taking the sums first, using the fact that $\int_\Gamma h(\Xi^\alpha)\Xi(\mathrm{d}\alpha)$ is zero for the zero measure, and applying (2.4) gives (2.7).

We are now in a position to prove the analogue of Theorem 1.1.

Theorem 2.4. *Suppose that Ξ is a simple point process on Γ with mean measure $\boldsymbol{\pi}$. Suppose also that Ξ has Janossi densities $\{j_n\}$, and that $\{N_\alpha\}_{\alpha\in\Gamma}$ is a neighbourhood* *structure satisfying* (a) *and* (b) *above. Then, for any finite measure* λ,

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda})) \leqslant \mathbb{E}\left\{\int_\Gamma \Xi(N_\alpha\backslash\{\alpha\})\Xi(\mathrm{d}\alpha)\right\}$$
$$+\int_\Gamma \mathbb{E}\{\Xi(N_\alpha)\}\mu(\alpha)\nu(\mathrm{d}\alpha)$$
$$+\int_\Gamma \mathbb{E}|g(\alpha, \Xi^\alpha)-\mu(\alpha)|\nu(\mathrm{d}\alpha)+\|\boldsymbol{\pi}-\boldsymbol{\lambda}\|, \tag{2.8}$$

where μ *and* g *are defined in* (2.5) *and* (2.6), *and* $\|\cdot\|$ *denotes the total variation distance between measures, which is twice* d_{TV} *when applied to probability measures. Further, if* Γ *is finite, the bound is the same as in Theorem* 1.1, *but without the constants* $c_1(\lambda)$ *and* $c_2(\lambda)$.

Proof. The argument is an elaboration of Chen's argument for Theorem 1.1. We first assume that $\boldsymbol{\pi}=\boldsymbol{\lambda}$. It is then enough to estimate $|\mathbb{E}(\mathscr{A}h)(\Xi)|$, where $\mathscr{A}$ is defined by (2.1) and h is given by (2.3) with $f(\xi)=I[\xi\in A]$, for any $A\subset\mathscr{X}$. First, observe that

$$\left|\mathbb{E}\int_\Gamma [h(\Xi-\delta_\alpha)-h(\Xi)-h(\Xi^\alpha)+h(\Xi^\alpha+\delta_\alpha)]\Xi(\mathrm{d}\alpha)\right|$$
$$\leqslant \mathbb{E}\left(\int_\Gamma \Xi(N_\alpha\backslash\{\alpha\})\Xi(\mathrm{d}\alpha)\right), \tag{2.9}$$

because, in view of Lemma 2.2(ii), the modulus of the integrand is bounded by $\Xi(N_\alpha\backslash\{\alpha\})$. In the same way, and by the Fubini–Tonelli theorem,

$$\left|\mathbb{E}\int_\Gamma [h(\Xi^\alpha)-h(\Xi^\alpha+\delta_\alpha)-h(\Xi)+h(\Xi+\delta_\alpha)]\boldsymbol{\pi}(\mathrm{d}\alpha)\right|$$
$$\leqslant \mathbb{E}\int_\Gamma \Xi(N_\alpha)\boldsymbol{\pi}(\mathrm{d}\alpha)=\int_\Gamma \mathbb{E}\{\Xi(N_\alpha)\}\boldsymbol{\pi}(\mathrm{d}\alpha). \tag{2.10}$$

Thus, in order to estimate $|\mathbb{E}(\mathscr{A}h)(\Xi)|$, it remains only to observe that, from (2.7),

$$\left|\mathbb{E}\int_\Gamma [h(\Xi^\alpha+\delta_\alpha)-h(\Xi^\alpha)]\{\Xi(\mathrm{d}\alpha)-\boldsymbol{\pi}(\mathrm{d}\alpha)\}\right|$$
$$=\left|\mathbb{E}\int_\Gamma [h(\Xi^\alpha+\delta_\alpha)-h(\Xi^\alpha)]\{g(\alpha, \Xi^\alpha)-\mu(\alpha)\}\nu(\mathrm{d}\alpha)\right|$$
$$\leqslant \int_\Gamma \mathbb{E}|g(\alpha, \Xi^\alpha)-\mu(\alpha)|\nu(\mathrm{d}\alpha), \tag{2.11}$$

whence, combining (2.1), (2.9), (2.10) and (2.11), the bound (2.8) is proved in the case $\boldsymbol{\lambda}=\boldsymbol{\pi}$.

In order to extend the result to a general choice of $\boldsymbol{\lambda}$, it is enough to estimate the total variation distance between $\mathrm{Po}(\boldsymbol{\pi})$ and $\mathrm{Po}(\boldsymbol{\lambda})$. This can also be done using the Stein–Chen approach, as follows. Let Ξ be a Poisson process with mean measure $\boldsymbol{\lambda}$. Then, for any bounded measurable h,

$$\mathbb{E}\int_\Gamma [h(\Xi-\delta_\alpha)-h(\Xi)]\Xi(\mathrm{d}\alpha)=\mathbb{E}\int_\Gamma [h(\Xi_\alpha-\delta_\alpha)-h(\Xi_\alpha)]\boldsymbol{\lambda}(\mathrm{d}\alpha),$$

where Ξ_α is the Palm process (see (2.15) below) for Ξ at α. Here, Ξ_α is a Poisson process with mean measure $\boldsymbol{\lambda}$, with the addition of a deterministic atom at α. Hence the integral on the right-hand side is

$$\int_\Gamma \mathbb{E}[h(\Xi)-h(\Xi+\delta_\alpha)]\boldsymbol{\lambda}(\mathrm{d}\alpha),$$

and thus

$$|\mathbb{E}(\mathscr{A}h)(\Xi)|=\left|\int_\Gamma \mathbb{E}[h(\Xi+\delta_\alpha)-h(\Xi)]\{\boldsymbol{\pi}(\mathrm{d}\alpha)-\boldsymbol{\lambda}(\mathrm{d}\alpha)\}\right|.$$

The latter expression is bounded by $\|\boldsymbol{\pi}-\boldsymbol{\lambda}\|$ since, from Lemma 2.2(i), the absolute value of the integrand is bounded by one. This completes the proof of the bound (2.8) for general $\boldsymbol{\lambda}$.

Suppose now that Γ is finite, and that $I_\alpha=\Xi\{\alpha\}$ for each $\alpha\in\Gamma$. Take ν to be counting measure, so that $\mu(\alpha)=\pi_\alpha=\mathbb{E}I_\alpha$. Then, for $0\leqslant n\leqslant|\Gamma|$ and $\boldsymbol{\alpha}\in\Gamma^n$ with $\alpha_i\neq\alpha_j$ for all $i\neq j$,

$$j_n(\boldsymbol{\alpha})=\mathbb{E}\{I_{\alpha_1}\cdots I_{\alpha_n}(1-I_{\beta_1})\cdots(1-I_{\beta_m})\},$$

where $m=|\{\alpha_1,\ldots,\alpha_n\}^c|$ and $\{\beta_1,\ldots,\beta_m\}=\{\alpha_1,\ldots,\alpha_n\}^c$. The transfer of notation to that of Theorem 1.1 is now straightforward. □

The first three terms in (2.8) all have natural interpretations. The first measures the extent of local dependence, the second how 'large' the neighbourhoods have been chosen and the third how much dependence there is between what happens at α and what happens outside its neighbourhood. Clearly, there is a trade-off between the size of the second and third terms, depending on how large the neighbourhoods are chosen to be.

An example follows, to illustrate in a simple non-discrete case how the bounds can be calculated. Suppose that Γ is the d-dimensional torus $[0,1]^d$, and let κ_d be the volume of the unit ball in d dimensions. Consider Ξ to be the hard-core point process which is uniform over Γ, has hard-core radius r and mean total number of points λ. The process may be specified by the Janossi density j_n with respect to Lebesgue measure ν on Γ given by

$$j_n(\boldsymbol{\alpha})=c\kappa^n I\left[\bigcap_{i\neq j}|\alpha_i-\alpha_j|>r\right],$$

where c, the partition function of statistical physics, and κ are numbers depending on λ and r chosen in such a way that

$$\sum_{n\geq 0}\int_{\Gamma^n}(n!)^{-1}j_n(\boldsymbol{\alpha})\nu^n(\mathrm{d}\boldsymbol{\alpha})=1$$

and

$$\int_\Gamma \mu(\alpha)\nu(\mathrm{d}\alpha)=\lambda.$$

Note that, in this example, μ takes the value λ everywhere. The case $r=0$ is the homogeneous Poisson process of rate $\kappa=\lambda$, and we expect good approximation of the hard-core process by the homogeneous Poisson process $\mathrm{Po}(\boldsymbol{\pi})$, where $\boldsymbol{\pi}=\lambda\nu$, if r is small.

To demonstrate this, we take the closed ball $B_\alpha(r)$ of radius r centred at α for N_α, making the first term in (2.8) equal to zero. The second term is simply computed to be $\lambda^2\kappa_d r^d$. In order to estimate the third term, consider formula (2.6) for $g(\alpha,\boldsymbol{\beta})$ when $\boldsymbol{\beta}$ is such that $|\beta_j-\beta_i|>r$ for all i and j, noting that the value of $g(\alpha,\boldsymbol{\beta})$ for other values of $\boldsymbol{\beta}$, conventionally taken to be zero, does not contribute to the estimate. The numerator reduces to the term with $r=0$, which always takes the value κ^{m+1}. The denominator involves the term with $s=1$ as well as that with $s=0$, and is equal to $\kappa^m\{1+\kappa\nu(A_\alpha(\boldsymbol{\beta}))\}$, where

$$A_\alpha(\boldsymbol{\beta})=\{\alpha'\in N_\alpha:|\beta_i-\alpha'|>r \text{ for all } i\};$$

note that $0\leq\nu(A_\alpha(\boldsymbol{\beta}))\leq\kappa_d r^d$. Thus

$$\kappa\{1+\kappa\cdot\kappa_d r^d\}^{-1}\leq g(\alpha,\boldsymbol{\beta})\leq\kappa, \tag{2.12}$$

and furthermore

$$\begin{aligned}&\mathbb{P}[g(\alpha,\Xi^\alpha)\neq\kappa\{1+\kappa\cdot\kappa_d r^d\}^{-1}]\\&\quad=\mathbb{P}[\Xi^\alpha\{B_\alpha(2r)\}\geq 1]\leq\mathbb{E}\Xi\{B_\alpha(2r)\}=\lambda\kappa_d(2r)^d.\end{aligned} \tag{2.13}$$

Hence $\lambda=\mu(\alpha)=\mathbb{E}g(\alpha,\Xi^\alpha)$ satisfies

$$\kappa\{1+\kappa\cdot\kappa_d r^d\}^{-1}\leq\lambda\leq\kappa\{1+\kappa\cdot\kappa_d r^d\}^{-1}\{1+\lambda\kappa\kappa_d^2(2r^2)^d\}, \tag{2.14}$$

and thus, from (2.12)–(2.14), the third term in (2.6) is no greater than

$$2\kappa\{1+\kappa\cdot\kappa_d r^d\}^{-1}\cdot\lambda\kappa\kappa_d^2(2r^2)^d\leq 2^{d+1}\lambda^3(\kappa_d r^d)^2\{1-\lambda\kappa_d r^d\}^{-1}.$$

In summary, Theorem 2.4 implies that

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi),\mathrm{Po}(\boldsymbol{\pi}))\leq\lambda^2\kappa_d r^d\{1+\mathrm{O}(\lambda r^d)\},$$

uniformly in $\lambda\kappa_d r^d\leq\frac{1}{2}$.

This bound is sharp in the range where it is small. To see this, consider the event that at least two points in the approximating Poisson process are at a distance r or less apart. Given that there are n points in the approximating process, they are independently distributed over Γ according to ν. Hence, upon conditioning on the position of one point, the probability of a given pair being at most distance r apart is $\kappa_d r^d$, and these events are pairwise independent (see, for example, Brown and Eagleson, 1984). Thus, given the number N of points of the process, the Bonferroni inequality implies that the probability of at least one pair being less than r apart is greater than or equal to

$$\binom{N}{2}\kappa_d r^d - \frac{1}{2}\binom{N}{2}\left[\binom{N}{2}-1\right](\kappa_d r^d)^2.$$

Taking the expectation, it follows that the probability of at least one pair of points from the Poisson process $\mathrm{Po}(\boldsymbol{\pi})$ with rate λ being no more than r apart is at least

$$\tfrac{1}{2}\lambda^2\kappa_d r^d - \tfrac{1}{8}(\lambda^4+4\lambda^3)\kappa_d^2 r^{2d}.$$

For the hard-core process, the probability of the same event is zero. Hence we have proved the following corollary.

Corollary 2.5. *Let Ξ be the hard-core process with radius r on the d-dimensional torus $[0,1]^d$, and let κ_d denote the volume of the unit sphere in d-dimensions. Then, if $\lambda = \mathbb{E}\{\Xi([0,1]^d)\}$ and ν is Lebesgue measure,*

$$\tfrac{1}{2}\lambda^2\kappa_d r^d\{1-\mathrm{O}((\lambda \vee \lambda^2)r^d)\} \leqslant d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\lambda\nu)) \leqslant \lambda^2\kappa_d r^d\{1+\mathrm{O}(\lambda r^d)\},$$

uniformly in $\lambda\kappa_d r^d \leqslant \frac{1}{2}$. □

It is worth noting that, for $\lambda \geqslant 1$, the upper bound is small if and only if r is of smaller order than $\lambda^{-2/d}$. In this case, the lower bound is of the same order, since the difference is half the upper bound minus a quantity which is approximately the square of the upper bound.

The analogue of Theorem 1.2 is more straightforward, and does not involve the assumptions invoked previously concerning Janossi densities. Instead, we use the general theory of Palm probabilities: see, for example, Kallenberg (1976, p. 69). If Ξ is a point process on Γ, then a point process Ξ_α, for some $\alpha \in \Gamma$, has the Palm distribution associated with Ξ at α if, for any measurable function $f: \Gamma \times \mathscr{X} \to [0,\infty)$,

$$\mathbb{E}\left(\int_B f(\alpha, \Xi)\Xi(\mathrm{d}\alpha)\right) = \mathbb{E}\left(\int_B f(\alpha, \Xi_\alpha)\boldsymbol{\pi}(\mathrm{d}\alpha)\right), \tag{2.15}$$

where $\boldsymbol{\pi}$ is the mean measure of Ξ. The point process $\Xi_\alpha - \delta_\alpha$ is called the reduced Palm process.

Theorem 2.6. *Suppose that Ξ is a finite point process on Γ, and that, for each $\alpha \in \Gamma$, a finite point process Ξ_α on Γ has been realized on the same probability space, in such a way that Ξ_α has the distribution of the Palm process at α for Ξ. Then, if $\boldsymbol{\pi}$ is the mean measure of Ξ and $\boldsymbol{\lambda}$ is any other finite non-negative measure on Γ,*

$$d_{\mathrm{TV}}(\Xi, \mathrm{Po}(\boldsymbol{\lambda})) \leqslant \int_\Gamma \mathbb{E}\|\Xi - (\Xi_\alpha - \delta_\alpha)\|\boldsymbol{\pi}(\mathrm{d}\alpha) + \|\boldsymbol{\lambda} - \boldsymbol{\pi}\|. \tag{2.16}$$

Remark 2.7. As can be seen from the proof, it is not necessary to define all the Ξ_α's on the same space: it is enough to have, for each $\alpha \in \Gamma$, copies of Ξ and Ξ_α on the same space.

Proof of Theorem 2.6. First, take $\boldsymbol{\lambda} = \boldsymbol{\pi}$. From equation (2.15), for any bounded measurable $h : \mathscr{X} \to \mathbb{R}$,

$$\mathbb{E}(\mathscr{A}h)(\Xi) = \mathbb{E}\int_\Gamma [h(\Xi + \delta_\alpha) - h(\Xi) + h(\Xi_\alpha - \delta_\alpha) - h(\Xi_\alpha)]\boldsymbol{\pi}(\mathrm{d}\alpha).$$

Now, if $\phi, \psi \in \mathscr{X}$ are written in the form

$$\phi = \sum_{i=1}^n x_i \delta_{\gamma_i}, \qquad \psi = \sum_{i=1}^n y_i \delta_{\gamma_i},$$

where n denotes the number of *distinct* atoms in $\phi + \psi$ and $\gamma_1, \ldots, \gamma_n$ an enumeration of them, let $\xi = \phi \wedge \psi$ denote the configuration

$$\xi = \sum_{i=1}^n (x_i \wedge y_i)\delta_{\gamma_i}.$$

Then it follows from the definition of $\Delta_2 h$ that, for any $\alpha \in \Gamma$,

$$|[h(\psi + \delta_\alpha) - h(\psi)] - [h(\xi + \delta_\alpha) - h(\xi)]| \le \|\psi - \xi\| \Delta_2 h,$$

and the same inequality holds with ϕ for ψ. Hence

$$|[h(\psi + \delta_\alpha) - h(\psi)] + [h(\phi) - h(\phi + \delta_\alpha)]|$$
$$\le \Delta_2 h(\|\psi - \xi\| + \|\phi - \xi\|) = \Delta_2 h \|\psi - \phi\|,$$

because of the definition of ξ. The theorem now follows, in the case $\boldsymbol{\lambda} = \boldsymbol{\pi}$, from Lemma 2.2(ii), by taking any Borel set $A \subset \mathscr{X}$ and letting h be given by (2.3) with $f(\xi) = I[\xi \in A]$. For general $\boldsymbol{\lambda}$, use the same argument as was used to conclude the proof of Theorem 2.4. □

Remark 2.8. If Γ is finite, Ξ is simple and $I_\alpha = \Xi\{\alpha\}$, the first term of the estimate is just $\sum_{\alpha \in \Gamma} \pi_\alpha \mathbb{E}\|\Xi - (\Xi_\alpha - \delta_\alpha)\|$. For negatively related indicators $(I_\alpha, \alpha \in \Gamma)$,

$$\sum_{\alpha \in \Gamma} \pi_\alpha \mathbb{E}\|\Xi - (\Xi_\alpha - \delta_\alpha)\| = \lambda - \operatorname{Var} W,$$

and for positively related indicators

$$\sum_{\alpha \in \Gamma} \pi_\alpha \mathbb{E}\|\Xi - (\Xi_\alpha - \delta_\alpha)\| = \operatorname{Var} W - \lambda + 2\sum_{\alpha \in \Gamma} \pi_\alpha^2.$$

Remark 2.9. An example in which the bound of Theorem 2.6 can be explicitly calculated is the Cox process Ξ directed by a finite random measure A. Arguing conditionally on A, the first term reduces to zero, leaving the estimate

$$d_{\mathrm{TV}}(\Xi, \mathrm{Po}(\boldsymbol{\pi})) \le \mathbb{E}\|A - \boldsymbol{\pi}\|.$$

The estimates of Theorems 1.1 and 1.2 for the distribution of the total number of points have already been evaluated in many settings: see, for example, Arratia, Goldstein and Gordon (1989) and Barbour, Holst and Janson (1992). Although Theorems 2.4 and 2.6 can now be used to obtain immediate process analogues, the estimates derived are not so good for large λ, because of the factors $c_1(\lambda)$ and $c_2(\lambda)$ containing negative powers of λ, which are present in Theorems 1.1 and 1.2 but not in Theorems 2.4 and 2.6. As shown by Corollary 2.5 and the Bernoulli trials example in the Introduction, it is not in general possible to find analogous factors to improve the process bounds for total variation approximation.

3. Approximation in the d_2 metric

As is shown in the previous section, there is no possibility of improving the order of the bounds on total variation distance given in Theorems 2.4 and 2.6 by exploiting the fact that λ is large. However, the total variation metric is extremely strong, and is completely unsuitable, for instance, if one wishes to approximate a process on a lattice in $\mathbb{R}^k$ by a Poisson process with a continuous intensity over $\mathbb{R}^k$. There is therefore the hope that, by choosing a weaker metric than total variation, one might also be able to introduce factors like λ^{-1} into the bounds on process approximation, and at the same time make the approximation of a discrete process by a continuous process feasible. This is the aim of the present section.

Suppose now that d_0 is a metric on Γ bounded by 1. Define metrics between pairs of configurations in $\mathscr{X}$ and between pairs of probability measures over $\mathscr{X}$ as follows. Let $\mathscr{K}$ denote the set of functions $k:\Gamma\to\mathbb{R}$ such that

$$s_1(k)=\sup_{y_1\neq y_2\in\Gamma}|k(y_1)-k(y_2)|/d_0(y_1,y_2)<\infty, \tag{3.1}$$

and define a distance d_1 between finite measures $\boldsymbol{\rho}$, $\boldsymbol{\sigma}$ over Γ by

$$d_1(\boldsymbol{\rho},\boldsymbol{\sigma})=\begin{cases}1, & \text{if } \boldsymbol{\rho}(\Gamma)\neq\boldsymbol{\sigma}(\Gamma),\\ m^{-1}\sup\limits_{k\in\mathscr{K}}\left\{\left|\int k\,\mathrm{d}\boldsymbol{\rho}-\int k\,\mathrm{d}\boldsymbol{\sigma}\right|\Big/ s_1(k)\right\}, & \text{if } \boldsymbol{\rho}(\Gamma)=\boldsymbol{\sigma}(\Gamma)=m>0.\end{cases} \tag{3.2}$$

Similarly, let $\mathscr{F}$ denote the set of functions $f:\mathscr{X}\to\mathbb{R}$ such that

$$s_2(f)=\sup_{\xi_1\neq\xi_2\in\mathscr{X}}|f(\xi_1)-f(\xi_2)|/d_1(\xi_1,\xi_2)<\infty, \tag{3.3}$$

and define a distance d_2 between probability measures over $\mathscr{X}$ by

$$d_2(Q,R)=\frac{1}{s_2(f)}\sup_{f\in\mathscr{F}}\left|\int f\,\mathrm{d}Q-\int f\,\mathrm{d}R\right|. \tag{3.4}$$

The metric d_1, when restricted to probability measures over (Γ, d_0), is variously known as the Dudley, Fortet-Mourier, Kantorovich $D_{1,1}$ or Wasserstein metric (see

Rachev, 1984) induced by d_0. When considered as a distance between configurations $\xi_1, \xi_2 \in \mathscr{X}$ with $|\xi_1| = |\xi_2| = n$, $d_1(\xi_1, \xi_2)$ can be interpreted in dual form as

$$\min_{\pi \in S_n} \left\{ n^{-1} \sum_{i=1}^{n} d_0(y_{1i}, y_{2\pi(i)}) \right\},$$

the average distance between the points $(y_{11}, \ldots, y_{1n})$ of ξ_1 and $(y_{21}, \ldots, y_{2n})$ of ξ_2 under the closest matching (Rachev, 1984, Section 2.2). The metric d_2 is then the corresponding metric induced by d_1 over the probability measures on $\mathscr{X}$. Note that the distances d_1 and d_2, like d_0, are bounded by 1.

When Γ is finite and the approximating process is on Γ, the simplest choice for d_0 is the discrete metric. With this choice, the d_1-distance between two configurations with different numbers of points is, as always, one, and that between two configurations with the same number of points is the proportion of points not common to both configurations, the 'relative variation'. However, the relative variation distance is still too strong, if one wishes to compare an intensity measure concentrated on the points $\{jn^{-1}, 1 \leq j \leq n\}$ with an intensity absolutely continuous with respect to Lebesgue measure on $[0, 1]$, and the same is true for configurations arising from such intensities. Thus, when using a Poisson process with continuous intensity to approximate one with a discrete intensity, it is more sensible to take Γ to be $[0, 1]$, or, more generally, if a natural metric d on Γ is given, d_0 can be taken to be $d \wedge 1$. The d_1 distance then reflects an average distance between configurations when the points are optimally paired.

As in Theorems 2.4 and 2.6, the estimates obtained below for the error in approximating $\mathscr{L}(\Xi)$ by $\mathrm{Po}(\boldsymbol{\lambda})$ with respect to the d_2 metric consist of two parts. The first comes from approximating Ξ by the Poisson process with the same mean measure, and the second from approximating one Poisson process by a possibly different one. For the first, the estimate is the same for all d_2 metrics. This is because the argument involves comparison of the values of functions $f: \mathscr{X} \to \mathbb{R}$ at configurations ξ_1 and ξ_2 which either have different numbers of points, in which case $d_1(\xi_1, \xi_2) = 1$ for all d_1, or where ξ_1 and ξ_2 are identical except for the positions of one or two points, which are then unrestricted in position, so that the choice of d_0 again has no influence on the largest value then possible for $d_1(\xi_1, \xi_2)$. Thus, when Γ is finite, the strongest result for this comparison comes from taking d_0 to be the discrete metric, and it is only in the second part, when comparing $\mathrm{Po}(\boldsymbol{\pi})$ with $\mathrm{Po}(\boldsymbol{\lambda})$, that the particular choice of d_0 influences the estimate obtained.

In order to investigate process approximation in terms of the d_2-metrics, we start with some technical results.

Lemma 3.1. *Let $(Z_t)_{t \geq 0}$ be a one-dimensional immigration-death process with immigration rate λ and unit per capita death rate, with $\mathbb{P}[Z_0 = k] = 1$. Then*

$$\int_0^\infty e^{-t} \mathbb{E}\{(Z_t + 1)^{-1}\}\, dt \leq \left(\frac{1}{\lambda} + \frac{1}{k+1} \right) (1 - e^{-\lambda}).$$

Proof. Write Z_t as the sum of independent random variables X_t and Y_t, where X_t denotes the number of the initial k individuals still alive at time t, so that $X_t \sim \mathrm{Bi}(k, \mathrm{e}^{-t})$, and $Y_t \sim \mathrm{Po}(\lambda(1-\mathrm{e}^{-t}))$ counts the individuals alive at t who arrived after time zero. Thus

$$\mathbb{E}\{(Z_t+1)^{-1}\} \leq \mathbb{E}(X_t+1)^{-1} = \frac{\mathrm{e}^t}{k+1}[1-(1-\mathrm{e}^{-t})^{k+1}]$$

and

$$\mathbb{E}\{(Z_t+1)^{-1}\} \leq \mathbb{E}(Y_t+1)^{-1} = \frac{1}{\lambda(1-\mathrm{e}^{-t})}[1-\mathrm{e}^{-\lambda(1-\mathrm{e}^{-t})}]. \tag{3.5}$$

Hence, letting t_0 be such that $\mathrm{e}^{-t_0} = \lambda/(\lambda+k+1)$, we have

$$\begin{aligned}&\int_0^\infty \mathrm{e}^{-t}\mathbb{E}\{(Z_t+1)^{-1}\}\,\mathrm{d}t \\ &\leq \int_0^{t_0} \frac{1}{k+1}[1-(1-\mathrm{e}^{-t_0})^{k+1}]\,\mathrm{d}t + \int_{t_0}^\infty \frac{\mathrm{e}^{-t}}{\lambda(1-\mathrm{e}^{-t})}(1-\mathrm{e}^{-\lambda})\,\mathrm{d}t \\ &\leq (1-\mathrm{e}^{-\lambda})\left\{\frac{1}{k+1}\log(1+\lambda^{-1}(k+1)) + \frac{1}{\lambda}\log(1+\lambda(k+1)^{-1})\right\} \\ &\leq (1-\mathrm{e}^{-\lambda})\left(\frac{1}{\lambda}+\frac{1}{k+1}\right),\end{aligned}$$

as required. □

The next lemma is used to control the error incurred by approximating with a Poisson process with intensity $\boldsymbol{\lambda}$ which may not be the same as $\boldsymbol{\pi}$.

Lemma 3.2. *Let $\boldsymbol{\lambda}$ be a finite measure over Γ, and let $h: \mathscr{X} \to \mathbb{R}$ be given by (2.3), where $f: \mathscr{X} \to \mathbb{R}$ is any function in $\mathscr{F}$. Let $\boldsymbol{\pi}$ be another finite measure over Γ with $\boldsymbol{\pi}(\Gamma) = \boldsymbol{\lambda}(\Gamma) = \lambda$. Then, for any $\xi \in \mathscr{X}$,*

$$\begin{aligned}&\left|\int_\Gamma [h(\xi+\delta_y)-h(\xi)](\boldsymbol{\pi}(\mathrm{d}y)-\boldsymbol{\lambda}(\mathrm{d}y))\right| \\ &\leq s_2(f)(1-\mathrm{e}^{-\lambda})(1+\lambda(|\xi|+1)^{-1})d_1(\boldsymbol{\pi},\boldsymbol{\lambda}).\end{aligned}$$

Proof. Let $h_\xi: \Gamma \to \mathbb{R}$ be given by

$$h_\xi(y) = h(\xi+\delta_y) - h(\xi).$$

Then, from the definition of d_1,

$$\left|\int_\Gamma [h(\xi+\delta_y)-h(\xi)](\boldsymbol{\pi}(\mathrm{d}y)-\boldsymbol{\lambda}(\mathrm{d}y))\right| \leq s_1(h_\xi)\lambda d_1(\boldsymbol{\pi},\lambda).$$

Now, by taking $(Z_t)_{t\geq 0}$ to be the immigration-death process with $Z_0=\xi$, and realizing $\mathbb{P}^{\xi+\delta_y}$ and $\mathbb{P}^{\xi+\delta_z}$ together by means of

$$Z_{1t}=Z_t+\delta_y I[E>t], \qquad Z_{2t}=Z_t+\delta_z I[E>t],$$

where the standard exponential random variable E is independent of Z, we find that

$$\begin{aligned}|h_\xi(y)-h_\xi(z)| &= \left|\int_0^\infty e^{-t}\mathbb{E}^\xi[f(Z_t+\delta_y)-f(Z_t+\delta_z)]\,dt\right| \\ &\leq s_2(f)\int_0^\infty e^{-t}\mathbb{E}^\xi d_1(Z_t+\delta_y, Z_t+\delta_z)\,dt \\ &\leq s_2(f)d_0(y,z)\int_0^\infty e^{-t}\mathbb{E}^\xi\{(|Z_t|+1)^{-1}\}\,dt.\end{aligned}$$

Applying Lemma 3.1 to the one-dimensional immigration-death process $|Z_t|$, we thus have

$$s_1(h_\xi)\leq s_2(f)(1-e^{-\lambda})(\lambda^{-1}+(|\xi|+1)^{-1}),$$

from which the lemma follows. □

The next two lemmas are the analogues of Lemma 2.2(i) and (ii). Because the class of functions h under consideration is smaller than that needed for total variation approximation, the smoothness estimates are better.

Lemma 3.3. *Under the conditions of Lemma* 3.2,

$$\Delta_1 h\leq s_2(f)(1\wedge 1.65\lambda^{-1/2}).$$

Proof. What is required is to estimate $|h(\xi+\delta_y)-h(\xi)|$ for any $\xi\in\mathscr{X}$, $y\in\Gamma$, where

$$h(\xi)=-\int_0^\infty [\mathbb{E}^\xi f(Z(t))-\mathrm{Po}(\boldsymbol{\pi})(f)]\,dt$$

and Z is an immigration-death process over Γ with intensity $\boldsymbol{\lambda}$ and unit per capita death rate. As for the proof of Lemma 2.2(i), construct processes Z_1 and Z_2 with the measures $\mathbb{P}^\xi$ and $\mathbb{P}^{\xi+\delta_y}$ together, by taking independent realizations of a third process Z_0 under $\mathbb{P}^0$, a pure death process X with unit per capita death rate starting with $X(0)=\xi$, and a standard exponential random variable E, and then setting

$$Z_1(t)=Z_0(t)+X(t), \qquad Z_2(t)=Z_1(t)+\delta_y I[E>t].$$

It then follows that

$$\begin{aligned}&h(\xi+\delta_y)-h(\xi)\\ &=\int_0^\infty e^{-t}\sum_{\eta\leq\xi}\mathbb{P}[X(t)=\eta]\mathbb{E}[f(Z_0(t)+\eta)-f(Z_0(t)+\eta+\delta_y)]\,dt, \qquad (3.6)\end{aligned}$$

where the notation $\{\eta \leqslant \xi\}$ implies that η and $\xi - \eta$ are both elements of $\mathscr{X}$. The inequality $|h(\xi+\delta_y)-h(\xi)| \leqslant s_2(f)$ is now immediate from (3.2) and (3.3).

For the λ-dependent bound, by conditioning on the value of $|Z_0(t)|$, we have

$$\begin{aligned} &\mathbb{E}[f(Z_0(t)+\eta)-f(Z_0(t)+\eta+\delta_y)] \\ &\quad = \mathbb{P}[|Z_0(t)|=0]f(\eta) \\ &\qquad + \sum_{k\geqslant 0} \{\mathbb{P}[|Z_0(t)|=k+1]\mathbb{E}[f(Z_0(t)+\eta)\,|\,|Z_0(t)|=k+1] \\ &\qquad\qquad -\mathbb{P}[|Z_0(t)|=k]\mathbb{E}[f(Z_0(t)+\eta+\delta_y)\,|\,|Z_0(t)|=k]\}. \end{aligned} \tag{3.7}$$

The latter sum is simplified by observing that

$$\begin{aligned} &|\mathbb{E}[f(Z_0(t)+\eta)\,|\,|Z_0(t)|=k+1]-\mathbb{E}[f(Z_0(t)+\eta+\delta_y)\,|\,|Z_0(t)|=k]| \\ &\quad = \left|\lambda^{-1}\int_\Gamma \mathbb{E}[f(Z_0(t)+\eta+\delta_z)-f(Z_0(t)+\eta+\delta_y)\,|\,|Z_0(t)|=k]\boldsymbol{\lambda}(\mathrm{d}z)\right| \\ &\quad \leqslant s_2(f)/(k+1+|\eta|), \end{aligned} \tag{3.8}$$

since

$$\begin{aligned} |f(\xi+\delta_z)-f(\xi+\delta_y)| &\leqslant s_2(f)d_1(\xi+\delta_z, \xi+\delta_y) \\ &\leqslant s_2(f)(|\xi|+1)^{-1}d_0(z,y) \leqslant s_2(f)(|\xi|+1)^{-1}. \end{aligned}$$

Note also that, since $\frac{1}{2}\{\inf_\xi f(\xi)+\sup_\xi f(\xi)\}$ may be subtracted from f without altering (3.6), we may take $\sup_\xi |f(\xi)| \leqslant \frac{1}{2}$, and hence

$$|\mathbb{E}[f(Z_0(t)+\eta)\,|\,|Z_0(t)|=k]| \leqslant \tfrac{1}{2}s_2(f). \tag{3.9}$$

Thus

$$\begin{aligned} &|\mathbb{E}[f(Z_0(t)+\eta)-f(Z_0(t)+\eta+\delta_y)]| \\ &\quad \leqslant s_2(f)\Bigg[\mathbb{E}\{(|Z_0(t)|+1)^{-1}\}+\tfrac{1}{2}\sum_{k\geqslant 1}|\mathbb{P}[|Z_0(t)|=k]-\mathbb{P}[|Z_0(t)|=k-1]| \\ &\qquad\qquad + \tfrac{1}{2}\mathbb{P}[|Z_0(t)|=0]\Bigg] \\ &\quad \leqslant s_2(f)[\mathbb{E}\{(|Z_0(t)|+1)^{-1}\}+\max_{k\geqslant 0}\mathbb{P}[|Z_0(t)|=k]] \\ &\quad \leqslant s_2(f)[\lambda_t^{-1}(1-\mathrm{e}^{-\lambda_t})+\{1\wedge(2\,\mathrm{e}\lambda_t)^{-1/2}\}], \end{aligned} \tag{3.10}$$

where $\lambda_t = \mathbb{E}|Z_0(t)| = \lambda(1-\mathrm{e}^{-t})$; see Barbour, Holst and Janson (1992, Proposition A.2.7) for the final estimate. Since also a direct estimate yields

$$|\mathbb{E}[f(Z_0(t)+\eta)-f(Z_0(t)+\eta+\delta_y)]| \leqslant s_2(f),$$

we arrive at the formula

$$[s_2(f)]^{-1}\int_0^\infty \mathrm{e}^{-t}|\mathbb{E}[f(Z_0(t)+\eta)-f(Z_0(t)+\eta+\delta_y)]|\,\mathrm{d}t$$

$$\leqslant \int_0^\tau \mathrm{e}^{-t}\,\mathrm{d}t+\int_\tau^\infty \mathrm{e}^{-t}\left\{\frac{1}{\lambda(1-\mathrm{e}^{-t})}+\frac{1}{\sqrt{2\,\mathrm{e}\lambda(1-\mathrm{e}^{-t})}}\right\}\mathrm{d}t, \tag{3.11}$$

valid for any $\eta\leqslant\xi$, where τ is chosen so that $\mathrm{e}^{-\tau}=1-\lambda^{-1}$. Computation of the integrals yields the result

$$\lambda^{-1}\left\{1+\log\lambda+\sqrt{\frac{2}{\mathrm{e}}}(\sqrt{\lambda}-1)\right\}\leqslant 1.65\lambda^{-1/2},$$

and this, with (3.6), implies the lemma. □

Remark 3.4. Note that, if f is an indicator function of $|\xi|$ alone, the inequality (3.8) is not needed, since the difference being considered is zero, and also that $s_2(f)=1$. Hence

$$|\mathbb{E}[f(Z_0(t)+\eta)-f(Z_0(t)+\eta+\delta_y)]|\leqslant\left\{1\wedge\max_{k\geqslant 0}\mathbb{P}[|Z_0(t)|=k]\right\}$$

by the argument of Lemma 3.3, leading to the estimate

$$\Delta_1 h\leqslant 1\wedge\sqrt{\frac{2}{\mathrm{e}\lambda}}\approx 1\wedge 0.858\lambda^{-1/2}. \tag{3.12}$$

However, estimates of $\Delta_1 h$ for functions f which are indicator functions of $|\xi|$ alone are the same as estimates of $c_1(\lambda)$. Hence it follows that the estimate of $c_1(\lambda)$ can be sharpened to $1\wedge\sqrt{2/(\mathrm{e}\lambda)}$.

Lemma 3.5. *Under the conditions of Lemma* 3.2,

$$\Delta_2 h\leqslant\left\{1\wedge\frac{5}{2\lambda}\left(1+2\log^+\left(\frac{2\lambda}{5}\right)\right)\right\}s_2(f).$$

Proof. Adapting the proof of Lemma 2.2(ii) much as the proof of Lemma 2.2(i) was adapted to prove Lemma 3.3, it follows that

$$\begin{aligned} &h(\xi+\delta_y+\delta_z)-h(\xi+\delta_z)-h(\xi+\delta_y)+h(\xi)\\ &=-\int_0^\infty \mathrm{e}^{-2t}\sum_{\eta\leqslant\xi}\mathbb{P}[X(t)=\eta]\mathbb{E}\{f(Z_0(t)+\eta+\delta_y+\delta_z)\\ &\qquad -f(Z_0(t)+\eta+\delta_y)-f(Z_0(t)+\eta+\delta_z)\\ &\qquad +f(Z_0(t)+\eta)\}\,\mathrm{d}t, \end{aligned} \tag{3.13}$$

from which $\Delta_2 h\leqslant s_2(f)$ is immediate.

For the main estimate, observe that

$$\begin{aligned}
&\mathbb{E}[f(Z_0(t)+\eta+\delta_y+\delta_z)-f(Z_0(t)+\eta+\delta_y)-f(Z_0(t)+\eta+\delta_z)+f(Z_0(t)+\eta)] \\
&\quad=\sum_{k\geqslant 0}\{\mathbb{P}[|Z_0(t)|=k]\mathbb{E}[f(Z_0(t)+\eta+\delta_y+\delta_z)\,|\,|Z_0(t)|=k] \\
&\qquad\qquad -\mathbb{P}[|Z_0(t)|=k+1]\mathbb{E}[f(Z_0(t)+\eta+\delta_y)+f(Z_0(t)+\eta+\delta_z)\,|\,|Z_0(t)|=k+1] \\
&\qquad\qquad +\mathbb{P}[|Z_0(t)|=k+2]\mathbb{E}[f(Z_0(t)+\eta)\,|\,|Z_0(t)|=k+2]\} \\
&\qquad -\mathbb{P}[|Z_0(t)|=0]\{f(\eta+\delta_y)+f(\eta+\delta_z)-f(\eta)\} \\
&\qquad +\mathbb{P}[|Z_0(t)|=1]\mathbb{E}[f(Z_0(t)+\eta)\,|\,|Z_0(t)|=1].
\end{aligned}$$

Thus, using (3.8) and (3.9), it follows that

$$\begin{aligned}
&|\mathbb{E}[f(Z_0(t)+\eta+\delta_y+\delta_z)-f(Z_0(t)+\eta+\delta_y)-f(Z_0(t)+\eta+\delta_z)+f(Z_0(t)+\eta)]| \\
&\quad\leqslant s_2(f)\Bigg[\tfrac{1}{2}\sum_{k\geqslant 0}|\mathbb{P}[|Z_0(t)|=k]-2\mathbb{P}[|Z_0(t)|=k-1]+\mathbb{P}[|Z_0(t)|=k-2]| \\
&\qquad\qquad +4\mathbb{E}\{(|Z_0(t)|+1)^{-1}\}\Bigg].
\end{aligned}$$

Now, if $P\sim \mathrm{Po}(\nu)$,

$$\begin{aligned}
&\mathbb{P}[P=k]-2\mathbb{P}[P=k-1]+\mathbb{P}[P=k-2] \\
&\quad=\mathbb{P}[P=k]\{(1-\nu^{-1}k)^2-\nu^{-2}k\},
\end{aligned}$$

for all $k\geqslant 0$, and hence

$$\begin{aligned}
&\sum_{k\geqslant 0}|\mathbb{P}[|Z_0(t)|=k]-2\mathbb{P}[|Z_0(t)|=k-1]+\mathbb{P}[|Z_0(t)|=k-2]| \\
&\quad\leqslant 2\{\lambda(1-\mathrm{e}^{-t})\}^{-1}.
\end{aligned}$$

Thus

$$\begin{aligned}
&|\mathbb{E}[f(Z_0(t)+\eta+\delta_y+\delta_z)-f(Z_0(t)+\eta+\delta_y) \\
&\qquad -f(Z_0(t)+\eta+\delta_z)+f(Z_0(t)+\eta)]| \\
&\quad\leqslant\{2\wedge 5/\lambda(1-\mathrm{e}^{-t})\}s_2(f),
\end{aligned}$$

and substitution of this expression into (3.13) leads to the inequality

$$\Delta_2 h\leqslant\left(\frac{25}{4\lambda^2}+\frac{5}{\lambda}\log\left(\frac{2\lambda}{5}\right)\right)s_2(f)\leqslant\frac{5}{2\lambda}\left(1+2\log^+\left(\frac{2\lambda}{5}\right)\right)s_2(f),$$

for all $\lambda\geqslant\frac{5}{2}$, completing the proof of the lemma. □

We are now in a position to prove theorems describing the accuracy, with respect to the new metric, of the approximation of the distribution of Ξ by a Poisson process.

Theorem 3.6. *With the assumptions and notation of Theorem* 2.4,

$$d_2(\mathscr{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda}))$$
$$\leqslant \left\{1 \wedge \frac{5}{2\lambda}\left(1+2\log^+\left(\frac{2\lambda}{5}\right)\right)\right\}$$
$$\times\left[\mathbb{E}\int_\Gamma \Xi(N_\alpha\backslash\{\alpha\})\Xi(\mathrm{d}\alpha)+\int_\Gamma \mathbb{E}\Xi(N_\alpha)\boldsymbol{\pi}(\mathrm{d}\alpha)\right]$$
$$+\{1\wedge 1.65\lambda^{-1/2}\}\int_\Gamma \mathbb{E}|g(\alpha, \Xi^\alpha)-\mu(\alpha)|\nu(\mathrm{d}\alpha)$$
$$+(1-\mathrm{e}^{-\lambda})(2-\mathrm{e}^{-\lambda})d_1(\boldsymbol{\pi}, \boldsymbol{\lambda}).$$

If Γ is finite and $\boldsymbol{\pi}=\boldsymbol{\lambda}$, the bound is the same as that of Theorem 1.1, *but with $c_1(\lambda)$ replaced by $\{1\wedge 1.65\lambda^{-1/2}\}$ and $c_2(\lambda)$ replaced by $\{1\wedge(5/(2\lambda))(1+2\log^+(\frac{2}{5}\lambda))\}$.*

Proof. Let f be a function satisfying $s_2(f)<\infty$, and let h be given by (2.3). Then, by the argument used in proving Theorem 2.4, we have

$$|\mathbb{E}(\mathscr{A}h)(\Xi)|\leqslant \Delta_2 h\left[\mathbb{E}\int_\Gamma \Xi(N_\alpha\backslash\{\alpha\})\Xi(\mathrm{d}\alpha)+\int_\Gamma \mathbb{E}\Xi(N_\alpha)\boldsymbol{\pi}(\mathrm{d}\alpha)\right]$$
$$+\Delta_1 h\int_\Gamma \mathbb{E}|g(\alpha, \Xi^\alpha)-\mu(\alpha)|\nu(\mathrm{d}\alpha). \qquad (3.14)$$

We are now able to use Lemmas 3.3 and 3.5 to bound $\Delta_1 h$ and $\Delta_2 h$, obtaining for $d_2(\mathscr{L}(\Xi), \mathrm{Po}(\boldsymbol{\pi}))$ the bound given in the theorem, with the last term zero. We then use the fact that

$$\mathrm{Po}(\boldsymbol{\lambda})(f)-\mathrm{Po}(\boldsymbol{\pi})(f)=\mathrm{Po}(\boldsymbol{\lambda})(\mathscr{A}h)$$
$$=\mathbb{E}\left\{\int_\Gamma [h(\Xi'+\delta_y)-h(\Xi')](\boldsymbol{\pi}(\mathrm{d}y)-\boldsymbol{\lambda}(\mathrm{d}y))\right\},$$

where Ξ' has distribution $\mathrm{Po}(\boldsymbol{\lambda})$, which from Lemma 3.2 gives

$$d_2(\mathrm{Po}(\boldsymbol{\lambda}), \mathrm{Po}(\boldsymbol{\pi}))\leqslant d_1(\boldsymbol{\pi}, \boldsymbol{\lambda})(1-\mathrm{e}^{-\lambda})\left(1+\lambda\sum_{j\geqslant 0}\mathrm{e}^{-\lambda}\lambda^j/(j+1)!\right).$$

The theorem now follows. □

In the hard-core example which followed Theorem 2.4, it would be natural to take d_0 to be Euclidean distance trimmed at 1. This leads to the estimate

$$d_2(\mathscr{L}(\Xi), \mathrm{Po}(\lambda\nu))$$
$$\leqslant\left\{1\wedge\frac{5}{2\lambda}\left(1+2\log^+\left(\frac{2\lambda}{5}\right)\right)\right\}\lambda^2\kappa_d r^d$$
$$+\{1\wedge 1.65\lambda^{-1/2}\}2^{d+1}\lambda^3(\kappa_d r^d)^2(1-\lambda\kappa_d r^d)^{-1},$$

so that good Poisson approximation in the d_2-metric is obtained under less restrictive circumstances than was possible for total variation approximation.

Theorem 3.7. *Under the conditions of Theorem* 2.6,

$$d_2(\mathscr{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda})) \leq \left\{1 \wedge \frac{5}{2\lambda}\left(1+2\log^+\left(\frac{2\lambda}{5}\right)\right)\right\} \int_\Gamma \mathbb{E}\|\Xi - (\Xi_\alpha - \delta_\alpha)\| \boldsymbol{\pi}(\mathrm{d}\alpha)$$
$$+(1-\mathrm{e}^{-\lambda})(2-\mathrm{e}^{-\lambda}) d_1(\boldsymbol{\pi}, \boldsymbol{\lambda}).$$

Proof. The proof of Theorem 2.6 is adapted in the same way that the proof of Theorem 2.4 was adapted to prove Theorem 3.6. □

Remark 3.8. It may well be that the factor $\log^+ \lambda$ is in fact superfluous in Theorems 3.6 and 3.7. However, since the d_2-distance is larger than the total variation distance between $\mathscr{L}(|\Xi|)$ and $\mathrm{Po}(\lambda)$, the first part of the estimate cannot be improved much further.

Example 3.9. If $\Gamma = \{n^{-1}i\colon 1 \leq i \leq rn\}$ and $(I_\alpha, \alpha \in \Gamma)$ are independent with $\pi_\alpha = p$ for each α, it follows from Theorems 3.6 and 3.7 that, taking $d_0(x, y) = |x-y| \wedge 1$,

$$d_2(\mathscr{L}(\Xi), \boldsymbol{\mu}) \leq \tfrac{5}{2}(1+2\log^+(\tfrac{2}{5}rnp))p + 3/4n,$$

where $\boldsymbol{\mu}$ is the measure of the Poisson process with constant intensity np with respect to Lebesgue measure on $[0, r]$. Note that the bound depends on r only through the logarithmic factor.

Note that the methods used in this paper can also be used to give bounds for the total variation approximation of the total number of points $|\Xi|$ by $\mathrm{Po}(\lambda)$: indeed, d_2-bounds are already bounds for this approximation too, but the more restricted set of test functions to be checked for total variation approximation of $|\Xi|$ enable the multiplying factors to be improved a little. The final theorem gives an example of what can be achieved, and complements the results of Barbour and Brown (1992).

Theorem 3.10. *With the assumptions and conditions of Theorem* 2.4,

$$d_{\mathrm{TV}}(\mathscr{L}(|\Xi|), \mathrm{Po}(\lambda))$$
$$\leq \lambda^{-1}(1-\mathrm{e}^{-\lambda})\left[\mathbb{E}\int_\Gamma \Xi(N_\alpha \backslash \{\alpha\})\Xi(\mathrm{d}\alpha) + \int_\Gamma \mathbb{E}\Xi(N_\alpha)\boldsymbol{\pi}(\mathrm{d}\alpha)\right]$$
$$+\left(1 \wedge \sqrt{\frac{2}{\mathrm{e}\lambda}}\right)\left[\int_\Gamma \mathbb{E}|g(\alpha, \Xi^\alpha) - \mu(\alpha)|\nu(\mathrm{d}\alpha) + |\lambda - \boldsymbol{\pi}(\Gamma)|\right].$$

Proof. The estimates for $\Delta_1 h$ and $\Delta_2 h$ come from Remark 3.4 and from Barbour and Eagleson (1983), and the last term differs from that of Theorem 3.6 again because only the distribution of the total number of points is being considered. □

References

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

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

A.D. Barbour and T.C. Brown, The Stein–Chen method, point processes and compensators, to appear in: Ann. Probab. (1992).

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

A.D. Barbour and L. Holst, Some applications of the Stein–Chen method for proving Poisson convergence, Adv. Appl. Probab. 21 (1989) 74–90.

A.D. Barbour, L. Holst and S. Janson, Poisson Approximation (Oxford Univ. Press, Oxford, 1992).

T.C. Brown and G.K. Eagleson, A useful property of some symmetric statistics, Amer. Statist. 38 (1984) 63–65.

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

D.J. Daley and D. Vere-Jones, An Introduction to the Theory of Point Processes (Springer, New York, 1988).

O. Kallenberg, Random Measures (Academic Press, London, 1976).

F. Papangelou, On the convergence of interacting particle systems to the Poisson process, Z. Wahrsch. Verw. Gebiete 52 (1980) 235–249.

S.T. Rachev, The Monge–Kantorovich mass transference problem and its stochastic applications, Theory Probab. Appl. 29 (1984) 647–676.

C. Stein, Approximate Computation of Expectations. IMS Lecture Notes No. 7 (Inst. Math. Statist., Hayward, 1986).