

## ON USING THE FIRST DIFFERENCE IN THE STEIN–CHEN METHOD<sup>1</sup>

BY AIHUA XIA

*University of New South Wales*

This paper investigates an alternative way of using the Stein–Chen method in Poisson approximations. There are three principal bounds stated in terms of reduced Palm probabilities for general point processes. The first two are for the accuracy of Poisson random variable approximation to the distribution of the number of points in a point process with respect to the total variation metric and the Wasserstein metric, and the third is for bounding the errors of Poisson process approximation to the distribution of a point process on a general compact space with respect to a Wasserstein metric. The bounds are frequently sharper than the previous results using the Stein–Chen method when the expected number of points is large.

1. Introduction. The Stein–Chen method was introduced in Chen (1975) to work out upper bounds on the total variation distance between the distribution of a sum of dependent 0–1 random variables and a Poisson distribution. The method has then been developed for Poisson process approximation by Arratia, Goldstein and Gordon (1989) in the context of a finite carrier space and Barbour (1988), Barbour and Brown (1992a) and Brown and Xia (1995a) for a more general carrier space.

One way of understanding the Stein–Chen method is by introducing a Markov process whose equilibrium distribution is Poisson and using the properties of the Markov process to assist in the estimation of the distance. Barbour (1988) and Barbour and Brown (1992a) [see also Barbour, Holst and Janson (1992), henceforth referred to as BHJ] adapted the method so that a Poisson process is the equilibrium distribution of the adapted generator.

The idea of Stein's technique, as implemented by Barbour and Brown (1992a), is to use an auxiliary space–time Markov process to investigate the approximation of a point process by a Poisson process. The auxiliary process has equilibrium distribution which is that of a Poisson process, and the way in which the process (started at the distribution of the point process) converges to the Poisson process is used in proving the approximation result. The space-time Markov process is an immigration–death process, with immigration occurring according to the mean measure of the approximating Poisson process and deaths occurring at unit per capita death rate.

---

Received November 1996; revised April 1997.

<sup>1</sup>Research partially supported by a 1996 Australia Research Council Small Grant from the University of New South Wales.

*AMS 1991 subject classifications.* Primary 60E15; secondary 62E17, 60G55.

*Key words and phrases.* Immigration–death process, Stein–Chen method, Palm distributions, coupling.

Let  $\Gamma$  be a compact, second countable Hausdorff space. We use  $\rho$  or  $|\rho|$  to stand for the total mass of the finite measure  $\rho$  on  $\Gamma$ . Let  $\lambda$  be a finite measure on  $\Gamma$  and  $\mathcal{H}$  be the space of finite point process configurations on  $\Gamma$ . Then the space–time process can be defined by its generator on  $\mathcal{H}$  as follows

$$(1.1) \quad \begin{aligned} \mathcal{A}h(\xi) = & \int_{\Gamma} [h(\xi + \delta_{\alpha}) - h(\xi)] \lambda(d\alpha) \\ & + \int_{\Gamma} [h(\xi - \delta_{\alpha}) - h(\xi)] \xi(d\alpha) \quad \forall \xi \in \mathcal{H} \end{aligned}$$

for suitable functions  $h$  defined on  $\mathcal{H}$ .

We use  $\text{Po}(\lambda)$  to stand for the distribution of a Poisson process with mean measure  $\lambda$  and  $\text{Po}(\lambda)$  to denote a Poisson distribution with mean  $\lambda$ . The corresponding Stein equation of  $\mathcal{A}$  for a bounded function  $f$  is then given by

$$(1.2) \quad \mathcal{A}h(\xi) = f(\xi) - \text{Po}(\lambda)(f).$$

Note that if  $f$  is a function of  $|\xi|$  alone, then Stein's equation is reduced to

$$(1.3) \quad \lambda g(j+1) - jg(j) = f(j) - \text{Po}(\lambda)(f),$$

where  $g(j+1) := h(j+1) - h(j)$ ,  $j \geq 0$ . If we can find such an  $h$  satisfying (1.2), then the estimation of  $|\text{E}(f(\Xi)) - \text{Po}(\lambda)(f)|$  is transformed to the investigation of  $|\text{E}(\mathcal{A}h(\Xi))|$ , where  $\Xi$  is a *point process* on  $\Gamma$ ; that is,  $\Xi$  is a random element on a probability space with values in  $\mathcal{H}$ .

Stein's approximation is centered around the Poisson process with mean measure  $\lambda$ , and the generator equation (1.1) is designed so that  $\text{Po}(\lambda)$  is the corresponding equilibrium distribution. The solution of (1.2) is given explicitly by (2.3) of Barbour and Brown (1992a) as

$$(1.4) \quad h(\xi) = - \int_0^{\infty} [\text{E}^{\xi} f(\mathbf{Z}(t)) - \text{Po}(\lambda)(f)] dt,$$

where  $\mathbf{Z}$  is an immigration–death process on  $\Gamma$  [for each fixed time  $t$ ,  $\mathbf{Z}(t)$  is a random element with values in  $\mathcal{H}$ ] with immigration intensity  $\lambda$  and unit per capita death rate, and  $\text{E}^{\xi}$  is the conditional expectation given  $\mathbf{Z}(0) = \xi$ .

If we set  $\mathcal{F}_1 := \{1_A: A \subset \mathbb{Z}^+\}$ , where  $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$ , then the test functions in  $\mathcal{F}_1$  will define a total variation metric  $d_{TV}$  on the set of probability measures on  $\mathbb{Z}^+$ :

$$d_{TV}(Q_1, Q_2) = \sup_{f \in \mathcal{F}_1} \left| \int f dQ_1 - \int f dQ_2 \right|.$$

We also have

$$d_{TV}(Q_1, Q_2) = \inf P(X \neq Y),$$

where the infimum is taken over all possible joint distributions for random variables  $(X, Y)$  such that  $X$  has distribution  $Q_1$  and  $Y$  has distribution  $Q_2$ ,

and it is possible to construct  $(X, Y)$ , called *maximal coupling*, so that the infimum is attained; and

$$d_{TV}(Q_1, Q_2) = \frac{1}{2} \sum_{i=0}^{\infty} |Q_1(i) - Q_2(i)|$$

[see BHJ, Appendix]. Similarly, the test functions in  $\mathcal{F}_2 := \{f: Z^+ \rightarrow \mathbb{R}, |f(n+1) - f(n)| \leq 1, \forall n \in Z^+\}$  define a Wasserstein distance  $d_W$  between two probability measures on  $Z^+$ :

$$d_W(Q_1, Q_2) = \sup_{f \in \mathcal{F}_2} \left| \int f dQ_1 - \int f dQ_2 \right|.$$

Equivalently, we can write the Wasserstein metric as

$$d_W(Q_1, Q_2) = \inf E|X - Y| = \sum_{i=0}^{\infty} |Q_1[i, \infty) - Q_2[i, \infty)|,$$

where the infimum is as for the total variation metric and there also exists a *maximal coupling*  $(X, Y)$  so that the infimum is attained [see BHJ, Appendix].

The following example shows how the Stein–Chen method has been widely used. Let  $I_1, I_2, \dots, I_n$  denote a sequence of independent 0–1 random variables such that

$$P[I_i = 1] = p_i, \quad P[I_i = 0] = 1 - p_i$$

for each  $i$ , and let

$$W = \sum_{i=1}^n I_i, \quad \lambda = \sum_{i=1}^n p_i.$$

Observe from (1.3) that

$$\begin{aligned} Ef(W) - \text{Po}(\lambda)(f) &= E\{\lambda g(W+1) - Wg(W)\} \\ &= \sum_{i=1}^n p_i E\{g(W+1) - g(W_i+1)\} \\ &= \sum_{i=1}^n p_i^2 E\{g(W_i+2) - g(W_i+1)\}, \end{aligned} \tag{1.5}$$

where  $W_i = \sum_{j \neq i} I_j$ . In particular, it follows that

$$|P[W \in A] - \text{Po}(\lambda)(A)| \leq \sup_j |g_{\lambda, A}(j+1) - g_{\lambda, A}(j)| \sum_{i=1}^n p_i^2.$$

It can be shown that

$$\sup_j |g_{\lambda, A}(j+1) - g_{\lambda, A}(j)| \leq \frac{1 - e^{-\lambda}}{\lambda}. \tag{1.6}$$

Therefore,

$$(1.7) \quad d_{TV}(\mathcal{L}W, \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i=1}^n p_i^2$$

[see BHJ]. The bound is based on using the first difference of  $g$  and thus the *second difference* of  $h$  in Stein's equation. It seems to be the best possible bound since the equality in (1.6) is attained at  $A = \{1\}$  [see page 8 of BHJ]. However, a somewhat different asymptotic expansion for the probabilities  $P[W \in A]$  is given in Deheuvels and Pfeifer (1986, 1988) with

$$d_{TV}(\mathcal{L}W, \text{Po}(\lambda)) \leq \begin{cases} \left(\frac{1}{e} + 2.6 \max_{1 \leq i \leq n} p_i\right)\theta, & \text{for } \max_{1 \leq i \leq n} p_i \leq 1/4, \\ \left(\frac{1}{e} + \frac{(2\theta)^{1/2}}{1 - (2\theta)^{1/2}}\right)\theta, & \text{for } \theta < 1/2, \end{cases}$$

where  $\theta := (1/\lambda) \sum_{i=1}^n p_i^2$ . Hence, there is a room to improve the performance of the Stein–Chen method. To do so, it is necessary to look at the method differently. As a matter of fact, it is not difficult to see from (1.5) that

$$|P[W \in A] - \text{Po}(\lambda)(A)| \leq 2 \sup_j |g_{\lambda, A}(j)| \sum_i p_i d_{TV}(\mathcal{L}W, \mathcal{L}W_i).$$

This observation gives an alternative upper bound based on the use of  $\sup_j |g_{\lambda, A}(j)|$  and hence the *first difference* of  $h$  in Stein's equation. As an application of this observation, we have (see Proposition 2.5)

$$d_{TV}(\text{Bi}(n, p), \text{Po}(np)) \leq \frac{0.6844 p \sqrt{np}}{\sqrt{\lfloor np \rfloor [1 - \lfloor np \rfloor / (n - 1)]}},$$

where  $\lfloor np \rfloor$  is the integer part of  $np$ . This result is better than

$$d_{TV}(\text{Bi}(n, p), \text{Po}(np)) \leq p$$

deduced from (1.7) when  $np$  is large.

We will use Palm theory [see, e.g., Daley and Vere-Jones (1988), Kallenberg (1983) or Karr (1986)] to state the results. Intuitively, a Palm distribution of a point process  $\Xi$  at a prescribed location  $\alpha$  is the distribution of  $\Xi$  conditional on the presence of a point of  $\Xi$  at  $\alpha$ . A point process  $\Xi_\alpha$  on  $\Gamma$  is called a Palm process of  $\Xi$  at location  $\alpha$  if it has the Palm distribution of  $\Xi$  at  $\alpha$ . For the context of Poisson approximation, it is more convenient to remove an atom at  $\alpha \in \Gamma$  from the realization of  $\Xi_\alpha$  by considering the process  $\Xi_\alpha - \delta_\alpha$ , called the *reduced Palm process*, where  $\delta_\alpha$  is the Dirac measure at  $\alpha$ . Lemma 10.2 of Kallenberg (1983) then gives, for all measurable functions  $f: \Gamma \times \mathcal{H} \rightarrow [0, \infty)$ , and all Borel sets  $B \subset \Gamma$ ,

$$(1.8) \quad E\left(\int_B f(\alpha, \Xi - \delta_\alpha) \Xi(d\alpha)\right) = \int_B E f(\alpha, \Xi_\alpha - \delta_\alpha) \lambda(d\alpha),$$

where  $\lambda$  is the mean measure of  $\Xi$ . An important fact is that, for a simple point process  $\Xi$  with a diffuse mean measure  $\lambda$ ,  $\Xi$  is a Poisson process if and only if  $\mathcal{L}(\Xi_\alpha - \delta_\alpha) = \mathcal{L}(\Xi)$   $\lambda$ -a.s. [see, e.g., Daley and Vere-Jones (1988)]. Thus, if  $\mathcal{L}(\Xi_\alpha - \delta_\alpha)$  is near  $\mathcal{L}(\Xi)$  for all  $\alpha$ , then  $\mathcal{L}(\Xi)$  is nearly distributed as a Poisson process.

Using the Palm probabilities approach, the first difference seems not of much use if the point process and the approximating Poisson process have the same mean measure [see Barbour and Brown (1992a, b), Brown and Xia (1995a)]. However, Stein's constant for the first difference has been improved many times (see Remark 1.1.2 of BHJ), and it seems worth investigating the possibility of using the first difference approach with the hope of producing more accurate upper bounds. In Section 2, we first prove several results parallel to the ones obtained by using the second difference approach [see Barbour and Brown (1992b) and BHJ], and then apply the results to the Poisson random variable approximation to the sum of independent indicator random variables. Section 3 is devoted to the estimation of errors in Poisson process approximation. We will first improve Stein's constant for the first difference, and then establish a theorem using the first difference approach. The theorem is then applied to approximating a Bernoulli process and an upper bound with the same magnitude as the lower bound is obtained, indicating a good prospect in using the first difference in Stein's equation.

2. Poisson random variable approximation. Barbour and Brown (1992b) provided an explicit bound for the departure of the distribution of the number of points in a set from Poisson in terms of the average Wasserstein distance between the distribution and the reduced Palm distribution for the number of points in the set. The results in this section provide alternative upper bounds derived from using the first difference. The results are stated in terms of the reduced Palm probabilities with a general carrier space, and in the case when  $\Gamma$  is finite, they coincide with that of BHJ, page 16. However, BHJ commented that the results are less useful because of the loss of the attractive factors. We will show in this section that results of this kind are sometimes more useful than those with Wasserstein distance in the bounds since, as we will see in applications, the information of Stein's factors is contained in the total variation distance in the upper bound. Moreover, these results provide us opportunities to improve the upper bounds in estimating the errors in Poisson approximations while the Wasserstein distance-based bounds give us little chance to obtain better estimates.

The next lemma is established in Remark 3.4 of Barbour and Brown (1992a) [see also (3.3) below].

**LEMMA 2.1.** *We have  $c_1(\lambda, d_{TV}) := \sup_{j \geq 0, A \subset Z^+} |g_{\lambda, A}(j)| \leq 1 \wedge \sqrt{\frac{2}{e\lambda}}$ .*

**THEOREM 2.2.** *Suppose  $\Xi$  is a point process on  $\Gamma$  with a finite mean measure  $\lambda$ , and that, for each  $\alpha \in \Gamma$ ,  $\Xi_\alpha$  is the Palm process at  $\alpha$  for  $\Xi$ . Then, for any*

bounded measurable set  $B$  of  $\Gamma$  and  $\mu > 0$ ,

$$(2.1) \quad \begin{aligned} & d_{TV}(\mathcal{L}\Xi(B), \text{Po}(\mu)) \\ & \leq \sqrt{\frac{2}{e\mu}} \left\{ 2 \int_{\alpha \in B} d_{TV}(\mathcal{L}\Xi(B), \mathcal{L}(\Xi_\alpha - \delta_\alpha)(B)) \boldsymbol{\lambda}(d\alpha) + |\boldsymbol{\lambda}(B) - \mu| \right\}. \end{aligned}$$

In particular, if  $\mu$  is taken as  $\boldsymbol{\lambda}(B)$ , then (2.1) reduces to

$$d_{TV}(\mathcal{L}\Xi(B), \text{Po}(\boldsymbol{\lambda}(B))) \leq \frac{2\sqrt{2}}{\sqrt{e\boldsymbol{\lambda}(B)}} \int_{\alpha \in B} d_{TV}(\mathcal{L}\Xi(B), \mathcal{L}(\Xi_\alpha - \delta_\alpha)(B)) \boldsymbol{\lambda}(d\alpha).$$

PROOF. Let  $g$  be any function on  $Z^+$  satisfying  $\sup_i |g(i)| \leq c_1(\mu, d_{TV})$ . Then applying (1.3) yields

$$(2.2) \quad \begin{aligned} & |\mathbb{E}\mu g(\Xi(B) + 1) - \mathbb{E}\Xi(B)g(\Xi(B))| \\ & \leq \mathbb{E}|(\mu - \boldsymbol{\lambda}(B))g(\Xi(B) + 1)| \\ & \quad + |\mathbb{E}\boldsymbol{\lambda}(B)g(\Xi(B) + 1) - \mathbb{E}\Xi(B)g(\Xi(B))|. \end{aligned}$$

The second term of (2.1) comes from an estimate of the first term of (2.2). For the second term of (2.2), we have

$$\begin{aligned} \mathbb{E}\Xi(B)g(\Xi(B)) &= \mathbb{E}\left(\int_B g(\Xi(B))\Xi(d\alpha)\right) \\ &= \mathbb{E}\left\{\int_B g[(\Xi - \delta_\alpha)(B) + 1]\Xi(d\alpha)\right\} \\ &= \mathbb{E}\left\{\int_B g[(\Xi_\alpha - \delta_\alpha)(B) + 1]\boldsymbol{\lambda}(d\alpha)\right\} \end{aligned}$$

because the integral is a sum over the values at the atoms of  $\Xi$  times the number of points at each atom, and the last equality follows from (1.8). Note that

$$\mathbb{E}\{\boldsymbol{\lambda}(B)g(\Xi(B) + 1)\} = \int_B \mathbb{E}\{g(\Xi(B) + 1)\}\boldsymbol{\lambda}(d\alpha)$$

and the second term of (2.2) is bounded by

$$\begin{aligned} & \int_B |\mathbb{E}\{g((\Xi_\alpha - \delta_\alpha)(B) + 1)\} - \mathbb{E}\{g(\Xi(B) + 1)\}|\boldsymbol{\lambda}(d\alpha) \\ & \leq 2 \sup_i |g(i)| \int_B d_{TV}(\mathcal{L}(\Xi_\alpha - \delta_\alpha)(B), \mathcal{L}\Xi(B)) \boldsymbol{\lambda}(d\alpha). \quad \square \end{aligned}$$

As an application, we have:

**PROPOSITION 2.3.** *Let  $\{I_i, 1 \leq i \leq n\}$  be independent indicator random variables with distribution  $\mathbb{P}(I_i = 1) = 1 - \mathbb{P}(I_i = 0) = p_i, 1 \leq i \leq n$ . Set*

$W = \sum_{i=1}^n I_i$ ,  $\lambda = \sum_{i=1}^n p_i$  and  $\tau_1 = \sum_{i=1}^n p_i(1 - p_i) - \max_{1 \leq j \leq n} p_j(1 - p_j)$ ; then

$$d_{TV}(\mathcal{L}W, \text{Po}(\lambda)) \leq \sqrt{\frac{2}{e\lambda\tau_1} \sum_{i=1}^n p_i^2}.$$

To prove the proposition, we need a technical lemma.

LEMMA 2.4. *In the setting of Proposition 2.3,*

$$(2.3) \quad \max_{0 \leq i \leq n} P[W = i] \leq \frac{1}{2} \left( \sum_{i=1}^n p_i(1 - p_i) \right)^{-1/2};$$

$$(2.4) \quad E \frac{1}{W + 1} \leq \frac{1 - e^{-\lambda}}{\lambda}.$$

PROOF. Inequality (2.3) is from Lemma 1 of Barbour and Jensen (1989). To show the second inequality, we have

$$\begin{aligned} E \frac{1}{W + 1} &= \sum_{i=0}^n \frac{1}{i + 1} P[W = i] = \sum_{i=0}^n \int_0^1 s^i ds P[W = i] \\ &= \int_0^1 (E s^W) ds = \int_0^1 \prod_{i=1}^n [1 - p_i(1 - s)] ds \\ &\leq \int_0^1 \prod_{i=1}^n e^{-p_i(1-s)} ds = \int_0^1 e^{-\lambda(1-s)} ds = \frac{1 - e^{-\lambda}}{\lambda}. \quad \square \end{aligned}$$

PROOF OF PROPOSITION 2.3. Set  $\Gamma = \{1, 2, \dots, n\}$ ,  $\Xi = \sum_{\alpha \in \Gamma} I_\alpha \delta_\alpha$  and  $B = \Gamma$  in Theorem 2.2; then  $(\Xi_\alpha - \delta_\alpha)(B) = \sum_{i \neq \alpha} I_i := W_\alpha$ , and

$$(2.5) \quad \begin{aligned} d_{TV}(\mathcal{L}W, \mathcal{L}W_\alpha) &\leq p_\alpha \max_{0 \leq i \leq n-1} P[W_\alpha = i] \\ &\leq \frac{p_\alpha}{2\sqrt{\sum_{i \neq \alpha} p_i(1 - p_i)}} \leq \frac{p_\alpha}{2\sqrt{\tau_1}}. \end{aligned}$$

Now the proposition follows from Theorem 2.2.  $\square$

For Poisson random variable approximation to the binomial distribution, we have:

PROPOSITION 2.5.

$$d_{TV}(\text{Bi}(n, p), \text{Po}(np)) \leq \frac{0.6844 p \sqrt{np}}{\sqrt{\lfloor np \rfloor [1 - \lfloor np \rfloor / (n - 1)]}},$$

where  $\lfloor np \rfloor$  is the integer part of  $np$ .

PROOF. Instead of using estimate (2.5), we apply Proposition A.2.5 of BHJ, page 261, to give

$$\max_{0 \leq k \leq n-1} \text{Bi}(n-1, p)(k) = \text{Bi}(n-1, p)(\lfloor np \rfloor) \leq \frac{1}{\sqrt{2\pi \lfloor np \rfloor (1 - \lfloor np \rfloor / (n-1))}},$$

and thus

$$(2.6) \quad d_{TV}(\text{Bi}(n, p), \text{Bi}(n-1, p)) \leq \frac{p}{\sqrt{2\pi \lfloor np \rfloor (1 - \lfloor np \rfloor / (n-1))}}.$$

Now, applying Theorem 2.2 gives the proposition.  $\square$

As we can see from the proof of Proposition 2.3, the difficulty in applying Theorem 2.2 is in estimating  $d_{TV}(\mathcal{L}\Xi(B), \mathcal{L}(\Xi_\alpha - \delta_\alpha)(B))$ . In general, it looks more difficult than the Wasserstein distance  $d_W(\mathcal{L}\Xi(B), \mathcal{L}(\Xi_\alpha - \delta_\alpha)(B))$  since the maximal coupling of two probability measures with respect to the Wasserstein metric can be easily identified when one of the two probability measures is stochastically greater than the other [see Brown and Greig (1996) for more applications of this property]. For example, in the case of approximating the sum of independent indicator random variables, the obvious coupling  $(W, W_\alpha)$  defined in the proof of Proposition 2.3 is actually the maximal coupling of  $\mathcal{L}W$  and  $\mathcal{L}W_\alpha$  for the Wasserstein distance, while the maximal coupling of the same distributions for the total variation distance is not obvious, although it exists.

The second result concerns the upper bound for the departure of the distribution of the number of points in a set from Poisson with respect to the Wasserstein metric.

THEOREM 2.6. *With the assumptions and notation of Theorem 2.2,*

$$(2.7) \quad d_W(\mathcal{L}\Xi(B), \text{Po}(\mu)) \leq 2 \int_{\alpha \in B} d_{TV}(\mathcal{L}\Xi(B), \mathcal{L}(\Xi_\alpha - \delta_\alpha)(B)) \lambda(d\alpha) + |\mu - \lambda(B)|.$$

PROOF. The proof is the same as the proof of Theorem 2.2 except replace  $c_1(\mu, d_{TV})$  by  $c_1(\mu, d_W) := \sup_{j \geq 0, f \in \mathcal{F}_2} |g_{\mu, f}(j)| \leq 1$  (see Remark 3.3).  $\square$

Using (2.5) and Theorem 2.6, one obtains the following proposition.

PROPOSITION 2.7. *With the assumptions and notation of Proposition 2.3,*

$$d_W(\mathcal{L}W, \text{Po}(\lambda)) \leq \frac{\sum_{i=1}^n p_i^2}{\sqrt{\tau_1}}.$$

Similarly, applying (2.6) and Theorem 2.6 gives:

PROPOSITION 2.8.  $d_W(\text{Bi}(n, p), \text{Po}(np)) \leq (0.8np^2) / \sqrt{\lfloor np \rfloor [1 - \lfloor np \rfloor / (n-1)]}$ .

REMARK 2.9. This result is slightly better than

$$d_W(\text{Bi}(n, p), \text{Poisson}(np)) \leq 1.144 p \sqrt{np}$$

(see BHJ, page 16) when  $np$  is large.

3. Poisson process approximation. It is well understood that the total variation metric is too strong for measuring the errors of Poisson process approximation [see, e.g., Barbour and Brown (1992a) for more discussion]. Thus, we will use the Wasserstein metric introduced in Barbour and Brown (1992a) to measure the errors of process approximation. To define the metric, let  $d_0$  be a metric on  $\Gamma$  bounded by 1. Let  $\mathcal{K}$  stand for the set of Lipschitz functions  $k: \Gamma \mapsto [-1, 1]$  such that

$$s_1(k) := \sup_{\alpha \neq \beta \in \Gamma} \frac{|k(\alpha) - k(\beta)|}{d_0(\alpha, \beta)} \leq 1.$$

Barbour and Brown (1992a) defined the first Wasserstein metric  $d_1$  between finite measures on  $\Gamma$  as

$$d_1(\rho, \sigma) = \begin{cases} 1, & \text{if } \rho \neq \sigma, \\ \sigma^{-1} \sup_{k \in \mathcal{K}} \left| \int k d\rho - \int k d\sigma \right|, & \text{if } \rho = \sigma > 0. \end{cases}$$

Let  $\mathcal{F}_3$  denote the set of Lipschitz functions  $f: \mathcal{K} \mapsto [-1, 1]$  such that

$$s_2(f) := \sup_{\xi_1 \neq \xi_2 \in \mathcal{K}} \frac{|f(\xi_1) - f(\xi_2)|}{d_1(\xi_1, \xi_2)} \leq 1,$$

and define the second Wasserstein distance with respect to  $d_1$  between probability measures  $Q_1$  and  $Q_2$  over  $\mathcal{K}$  by

$$d_2(Q_1, Q_2) := \sup_{f \in \mathcal{F}_3} \left| \int f dQ_1 - \int f dQ_2 \right|.$$

Note that

$$d_2(Q_1, Q_2) = \inf_{(X, Y)} E d_1(X, Y),$$

where the infimum is over all point processes  $(X, Y)$  such that  $X$  has distribution  $Q_1$  and  $Y$  has distribution  $Q_2$  [see Rachev (1984)].

Barbour and Brown (1992a), Brown and Xia (1995a), Barbour, Brown and Xia (1995) and BHJ established upper bounds for the  $d_2$  distance between the distributions of a simple point process  $\Xi$  and a Poisson process with mean measure  $\lambda$  using different approaches. The bound may not be useful for very large  $\lambda$ , because there is an  $\ln^+ \lambda$  factor in Stein's constant. Getting rid of the  $\ln^+ \lambda$  factor has been proved impossible if one wishes to extract Stein's constants in their theory derived from the second difference estimation [see Brown and Xia (1995b)]. However, as analyzed in Xia (1997), the logarithmic factor is more likely to be superfluous because from the definition of  $d_1$ ,

$d_2(\mathcal{L}\Xi, \text{Po}(\boldsymbol{\lambda}))$  consists of two parts. The first comes from  $d_{\text{TV}}(\mathcal{L}|\Xi|, \text{Po}(\lambda))$  and the second is a quantity measuring the discrepancy of the locations of the points of  $\Xi$  and  $\text{Po}(\boldsymbol{\lambda})$ , given the two processes have the same number of points. By looking at the first difference, we shall establish a general result which gives an upper bound with the right order of convergence in a Poisson process approximation to a Bernoulli process. Before we state the main result in this section, we improve the upper bound for the first difference.

LEMMA 3.1. *We have*

$$c_1(\lambda, d_2) := \sup_{f \in \mathcal{F}_3, \xi \in \mathcal{H}, \alpha \in \Gamma} |h(\xi + \delta_\alpha) - h(\xi)| \leq 1 \wedge \left[ \sqrt{\frac{2}{e\lambda}} + \frac{1}{\lambda} \int_0^\lambda \frac{1 - e^{-y}}{y} dy \right].$$

REMARK 3.2. This estimate is better than

$$c_1(\lambda, d_2) \leq 1 \wedge 1.65\lambda^{-1/2}$$

proved in Lemma 3.3 of Barbour and Brown (1992a) when  $\lambda$  is large.

PROOF. It suffices to show that, for any  $\xi \in \mathcal{H}$  and  $\alpha \in \Gamma$ ,

$$|h(\xi + \delta_\alpha) - h(\xi)| \leq 1 \wedge \left[ \sqrt{\frac{2}{e\lambda}} + \frac{1}{\lambda} \int_0^\lambda \frac{1 - e^{-y}}{y} dy \right],$$

where

$$h(\xi) = - \int_0^\infty [E^\xi f(\mathbf{Z}(t)) - \text{Po}(\boldsymbol{\lambda})(f)] dt$$

and  $\mathbf{Z}$  is an immigration–death process over  $\Gamma$  with immigration intensity  $\boldsymbol{\lambda}$  and unit per capita death rate. Construct two immigration–death processes  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  with measures

$$\mathcal{L}(\mathbf{Z}_1) = \mathcal{L}(\mathbf{Z}|\mathbf{Z}(0) = \xi) \quad \text{and} \quad \mathcal{L}(\mathbf{Z}_2) = \mathcal{L}(\mathbf{Z}|\mathbf{Z}(0) = \xi + \delta_\alpha)$$

together, by taking independent realizations of a third process  $\mathbf{Z}_0$  with  $\mathcal{L}(\mathbf{Z}_0) = \mathcal{L}(\mathbf{Z}|\mathbf{Z}(0) = 0)$ , 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

$$\mathbf{Z}_1(t) = \mathbf{Z}_0(t) + X(t), \quad \mathbf{Z}_2(t) = \mathbf{Z}_1(t) + \delta_\alpha 1_{[E > t]}.$$

Then it follows that

$$\begin{aligned} & h(\xi + \delta_\alpha) - h(\xi) \\ (3.1) \quad &= \int_0^\infty E[f(\mathbf{Z}_1(t)) - f(\mathbf{Z}_2(t))] dt \\ &= \int_0^\infty e^{-t} \sum_{\eta \leq \xi} P(X(t) = \eta) E[f(\mathbf{Z}_0(t) + \eta) - f(\mathbf{Z}_0(t) + \eta + \delta_\alpha)] dt, \end{aligned}$$

where the notion  $\{\eta \leq \xi\}$  means  $\eta$  and  $\xi - \eta$  are both elements of  $\mathcal{H}$ . Since

$$(3.2) \quad |f(\mathbf{Z}_0(t) + \eta) - f(\mathbf{Z}_0(t) + \eta + \delta_\alpha)| \leq 1,$$

we have  $c_1(\lambda, d_2) \leq 1$  immediately. On the other hand,

$$\begin{aligned} |h(\xi + \delta_\alpha) - h(\xi)| &\leq \left| h(\xi + \delta_\alpha) - \frac{1}{\lambda} \int_{\Gamma} h(\xi + \delta_y) \boldsymbol{\lambda}(dy) \right| \\ &\quad + \left| \frac{1}{\lambda} \int_{\Gamma} h(\xi + \delta_y) \boldsymbol{\lambda}(dy) - h(\xi) \right|. \end{aligned}$$

By Lemma 2.6 of Brown and Xia (1995a), we have

$$\left| h(\xi + \delta_\alpha) - \frac{1}{\lambda} \int_{\Gamma} h(\xi + \delta_y) \boldsymbol{\lambda}(dy) \right| \leq \frac{1}{\lambda} \int_0^\lambda \frac{1 - e^{-y}}{y} dy,$$

and so it remains to show

$$(3.3) \quad \left| \frac{1}{\lambda} \int_{\Gamma} h(\xi + \delta_y) \boldsymbol{\lambda}(dy) - h(\xi) \right| \leq \sqrt{\frac{2}{e\lambda}}.$$

Noting that

$$\begin{aligned} \frac{1}{\lambda} \int_{\Gamma} h(\xi + \delta_y) \boldsymbol{\lambda}(dy) - h(\xi) &= \frac{1}{\lambda} \int_{\Gamma} \int_0^\infty e^{-t} \sum_{\eta \leq \xi} \mathbb{P}(X(t) = \eta) \\ &\quad \times \mathbb{E}[f(Z_0(t) + \eta) - f(Z_0(t) + \eta + \delta_y)] dt \boldsymbol{\lambda}(dy) \end{aligned}$$

and, for each  $k \geq 1$ ,

$$\mathbb{E}[f(Z_0(t) + \eta) | |Z_0(t)| = k] = \frac{1}{\lambda^k} \int_{\Gamma^k} f\left(\sum_{i=1}^k \delta_{x_i} + \eta\right) \boldsymbol{\lambda}(dx_1) \cdots \boldsymbol{\lambda}(dx_k),$$

by conditioning on the value of  $|Z_0(t)|$ , we have

$$\begin{aligned} &\frac{1}{\lambda} \int_{\Gamma} \mathbb{E}[f(Z_0(t) + \eta) - f(Z_0(t) + \eta + \delta_y)] \boldsymbol{\lambda}(dy) \\ &= \mathbb{P}[|Z_0(t)| = 0] f(\eta) + \sum_{k \geq 1} \{\mathbb{P}[|Z_0(t)| = k] - \mathbb{P}[|Z_0(t)| = k - 1]\} \\ &\quad \times \frac{1}{\lambda^k} \int_{\Gamma^k} f\left(\sum_{i=1}^k \delta_{x_i} + \eta\right) \boldsymbol{\lambda}(dx_1) \cdots \boldsymbol{\lambda}(dx_k). \end{aligned}$$

Subtracting  $\frac{1}{2} \{\inf_{\zeta \in \mathcal{X}} f(\zeta) + \sup_{\zeta \in \mathcal{X}} f(\zeta)\}$  from  $f$  if necessary, we may assume  $\sup_{\zeta} |f(\zeta)| \leq \frac{1}{2} s_2(f) \leq \frac{1}{2}$ , so

$$\begin{aligned} &\left| \frac{1}{\lambda} \int_{\Gamma} \mathbb{E}[f(Z_0(t) + \eta) - f(Z_0(t) + \eta + \delta_y)] \boldsymbol{\lambda}(dy) \right| \\ &\leq \frac{1}{2} \mathbb{P}[|Z_0(t)| = 0] + \frac{1}{2} \sum_{k \geq 1} |\mathbb{P}[|Z_0(t)| = k] - \mathbb{P}[|Z_0(t)| = k - 1]| \\ &= \max_{k \geq 0} \mathbb{P}[|Z_0(t)| = k] \\ &\leq \frac{1}{\sqrt{2e\lambda(1 - e^{-t})}}, \end{aligned}$$

where the equality is because  $|Z_0(t)| \sim \text{Po}(\lambda(1 - e^{-t}))$  and  $P[|Z_0(t)| = i]$  is increasing in  $i$  for  $i < \lambda(1 - e^{-t})$  and decreasing in  $i$  for  $i > \lambda(1 - e^{-t})$ ; the second inequality comes from Proposition A.2.7 of BHJ. Hence

$$\begin{aligned} \left| \frac{1}{\lambda} \int_{\Gamma} h(\xi + \delta_y) \boldsymbol{\lambda}(dy) - h(\xi) \right| &\leq \int_0^\infty e^{-t} \left[ 1 \wedge \frac{1}{\sqrt{2e\lambda(1 - e^{-t})}} \right] dt \\ &= \int_0^1 1 \wedge \frac{1}{\sqrt{2e\lambda s}} ds \\ &= \sqrt{\frac{2}{e\lambda}} - \frac{1}{2e\lambda}, \end{aligned}$$

and (3.3) follows.  $\square$

If  $f \in \mathcal{F}_2$ , then (3.2) holds. Thus, using (3.1) gives the following.

**REMARK 3.3.** We have  $c_1(\lambda, d_W) \leq 1$ .

**THEOREM 3.4.** *With the assumptions and notation of Theorem 2.2, if  $\boldsymbol{\mu}$  is any other finite nonnegative measure on  $\Gamma$ , then*

$$\begin{aligned} &d_2(\mathcal{L}\Xi, \text{Po}(\boldsymbol{\mu})) \\ (3.4) \quad &\leq 2 \int_{\alpha \in \Gamma} \{c_1(\boldsymbol{\mu}, d_2) d_{TV}(\mathcal{L}|\Xi|, \mathcal{L}|\Xi_\alpha - \delta_\alpha|) + E\phi_\alpha(|\Xi|)\} \boldsymbol{\lambda}(d\alpha) \\ &\quad + (1 - e^{-\mu}) d_1(\boldsymbol{\lambda}/\lambda, \boldsymbol{\mu}/\mu) + c_1(\boldsymbol{\mu}, d_2) |\lambda - \mu|, \end{aligned}$$

where  $\phi_\alpha(k) := d_2(\mathcal{L}(\Xi || \Xi| = k), \mathcal{L}(\Xi_\alpha - \delta_\alpha || |\Xi_\alpha - \delta_\alpha| = k))$ .

**REMARK 3.5.** One might expect smaller Stein's constants in (3.4) since previous bounds have a Stein's magic factor  $(1 + \ln^+ \lambda)/\lambda$  [see BHJ, Barbour and Brown (1992a) and Brown and Xia (1995a)]. However, the  $d_{TV}$  and  $d_2$  distances between the process and its reduced Palm probabilities are smaller than the counterparts in the previous results in Barbour and Brown (1992a) and Brown and Xia (1995a), indicating that the more attractive magic factor in BHJ, Barbour and Brown (1992a) and Brown and Xia (1995a) was introduced at the cost of a larger remainder. The following examples show that the constants in (3.4) are in fact of the right order.

**PROOF OF THEOREM 3.4.** The last two terms of (3.4) come from the direct estimation

$$(3.5) \quad d_2(\text{Po}(\boldsymbol{\lambda}), \text{Po}(\boldsymbol{\mu})) \leq (1 - e^{-\mu}) d_1(\boldsymbol{\lambda}/\lambda, \boldsymbol{\mu}/\mu) + c_1(\boldsymbol{\mu}, d_2) |\lambda - \mu|$$

[see (2.8) of Brown and Xia (1995a)], so it suffices to show that (3.4) is true under the assumption that  $\boldsymbol{\mu} = \boldsymbol{\lambda}$ .

Applying (1.8) gives

$$\begin{aligned} E(\mathcal{A}h)(\Xi) &= E \int_{\Gamma} [h(\Xi + \delta_\alpha) - h(\Xi)] \boldsymbol{\lambda}(d\alpha) + E \int_{\Gamma} [h(\Xi - \delta_\alpha) - h(\Xi)] \Xi(d\alpha) \\ &= E \int_{\Gamma} [h(\Xi + \delta_\alpha) - h(\Xi) + h(\Xi_\alpha - \delta_\alpha) - h(\Xi_\alpha)] \boldsymbol{\lambda}(d\alpha), \end{aligned}$$

and we have also

$$\begin{aligned} & E[h(\Xi + \delta_\alpha) - h(\Xi) + h(\Xi_\alpha - \delta_\alpha) - h(\Xi_\alpha)] \\ &= \sum_{k=0}^{\infty} \{E[h(\Xi + \delta_\alpha) - h(\Xi) | |\Xi| = k] \\ &\quad + E[h(\Xi_\alpha - \delta_\alpha) - h(\Xi_\alpha) | |\Xi_\alpha - \delta_\alpha| = k]\} P[|\Xi| = k] \\ &\quad + \sum_{k=0}^{\infty} E[h(\Xi_\alpha - \delta_\alpha) - h(\Xi_\alpha) | |\Xi_\alpha - \delta_\alpha| = k] \{P[|\Xi_\alpha - \delta_\alpha| = k] - P[|\Xi| = k]\}. \end{aligned}$$

The first term of (3.4) arises as an estimate of

$$\sum_{k=0}^{\infty} E[h(\Xi_\alpha - \delta_\alpha) - h(\Xi_\alpha) | |\Xi_\alpha - \delta_\alpha| = k] \{P[|\Xi_\alpha - \delta_\alpha| = k] - P[|\Xi| = k]\}.$$

For the second term of (3.4), it is sufficient to show that for two configurations  $\xi_1 = \sum_{i=1}^n \delta_{y_i}$  and  $\xi_2 = \sum_{i=1}^n \delta_{z_i}$  on  $\Gamma$ ,

$$(3.6) \quad |[h(\xi_1 + \delta_\alpha) - h(\xi_1)] - [h(\xi_2 + \delta_\alpha) - h(\xi_2)]| \leq 2d_1(\xi_1, \xi_2).$$

To prove (3.6), let  $Z_0$  be an immigration–death process over  $\Gamma$  with immigration intensity  $\lambda$  and unit per capita death rate starting with zero initial points,  $\{E_i\}_{1 \leq i \leq n+1}$  be independent standard exponential random variables and independent of  $Z_0$ . Set

$$Z_1(t) = \sum_{i=1}^n \delta_{y_i} 1_{E_i > t}, \quad Z_2(t) = \sum_{i=1}^n \delta_{z_i} 1_{E_i > t},$$

and let  $Y_1 = Z_0 + Z_1$  and  $Y_2 = Z_0 + Z_2$ ; then

$$\begin{aligned} & |[h(\xi_1 + \delta_\alpha) - h(\xi_1)] - [h(\xi_2 + \delta_\alpha) - h(\xi_2)]| \\ & \leq \int_0^\infty |E[f(Y_1(t) + \delta_\alpha 1_{E_{n+1} > t}) - f(Y_2(t) + \delta_\alpha 1_{E_{n+1} > t}) \\ & \quad - f(Y_1(t)) + f(Y_2(t))]| dt \\ (3.7) \quad & \leq \int_0^\infty e^{-t} |E[f(Y_1(t) + \delta_\alpha) - f(Y_2(t) + \delta_\alpha) \\ & \quad - f(Y_1(t)) + f(Y_2(t))]| dt. \end{aligned}$$

By Lemma 2.4 of Brown and Xia (1995a) and using the fact that  $|Z_0(t)|$  follows Poisson( $\lambda(1 - e^{-t})$ ), we have

$$\begin{aligned} & |E[f(Y_1(t)) - f(Y_2(t))]| \\ & \leq d_1(\xi_1, \xi_2) \frac{n \exp(-\lambda(1 - e^{-t}))e^{-t}}{\lambda(1 - e^{-t})} \int_0^{\lambda(1 - e^{-t})} e^s \left(1 - e^{-t} + \frac{se^{-t}}{\lambda(1 - e^{-t})}\right)^{n-1} ds \\ & \leq d_1(\xi_1, \xi_2), \end{aligned}$$

and similarly

$$|E[f(Y_1(t) + \delta_\alpha) - f(Y_2(t) + \delta_\alpha)]| \leq d_1(\xi_1, \xi_2),$$

and so (3.6) follows from (3.7).  $\square$

As an application of Theorem 3.4, we estimate the errors of the Poisson process approximation to a Bernoulli process. The reason we look at the independent case is that, if the method does not work for this case, it is very unlikely to work well for a dependent case.

**PROPOSITION 3.6.** *Let  $\Gamma = \{\alpha_1, \dots, \alpha_n\}$ , and suppose  $\{I_\alpha: \alpha \in \Gamma\}$  are independent indicator random variables with  $P(I_\alpha = 1) = 1 - P(I_\alpha = 0) = p_\alpha, \forall \alpha \in \Gamma$ . Set  $\Xi = \sum_{\alpha \in \Gamma} I_\alpha \delta_\alpha, \lambda = \sum_{\alpha \in \Gamma} p_\alpha \delta_\alpha, \tau_1 = \sum_{i=1}^n p_i(1-p_i) - \max_{1 \leq j \leq n} p_j(1-p_j)$  and  $\tau_2 = \lambda - \max_{\alpha \in \Gamma} p_\alpha$ . Then for any other finite nonnegative measure  $\mu$  on  $\Gamma$ ,*

$$d_2(\mathcal{L}\Xi, \text{Po}(\mu)) \leq \left[ \sqrt{\frac{2}{e\mu\tau_1}} + \frac{1}{\mu\sqrt{\tau_1}} \int_0^\mu \frac{1-e^{-y}}{y} dy + \frac{2(1-e^{-\tau_2})}{\tau_2} \right] \sum_{\alpha \in \Gamma} p_\alpha^2 + (1-e^{-\mu})d_1\left(\frac{\lambda}{\lambda}, \frac{\mu}{\mu}\right) + \left[ \sqrt{\frac{2}{e\mu}} + \frac{1}{\mu} \int_0^\mu \frac{1-e^{-y}}{y} dy \right] |\lambda - \mu|.$$

In particular,

$$(3.8) \quad d_2(\mathcal{L}\Xi, \text{Po}(\lambda)) \leq \left[ \sqrt{\frac{2}{e\lambda\tau_1}} + \frac{1}{\lambda\sqrt{\tau_1}} \int_0^\lambda \frac{1-e^{-y}}{y} dy + \frac{2(1-e^{-\tau_2})}{\tau_2} \right] \sum_{\alpha \in \Gamma} p_\alpha^2.$$

**REMARK 3.7.** The bound in (3.8) is of the correct order.

In fact, recalling that  $\theta = (1/\lambda) \sum_{i=1}^n p_i^2$ , we have  $d_{TV}(\text{Po}(\lambda), \mathcal{L}|\Xi|) \geq \theta/32$  [see Corollary 3.D.1 of BHJ, page 61 or Barbour and Hall (1984)], which implies

$$d_2(\mathcal{L}\Xi, \text{Po}(\lambda)) \geq \frac{\theta}{32}.$$

To prove the proposition, we need the following lemma.

**LEMMA 3.8.** *Let  $\mathcal{U}$  be any finite subset of  $\Gamma, \{X_\beta, \beta \in \mathcal{U}\}$  be independent indicator random variables,  $V = \sum_{\beta \in \mathcal{U}} X_\beta \delta_\beta$ ; then for every  $\xi \in \mathcal{H}, \alpha \in \Gamma$ , we have*

$$d_2(\mathcal{L}(\xi + V + \delta_\alpha ||V| = k), \mathcal{L}(\xi + V ||V| = k + 1)) \leq \frac{1}{|\xi| + k + 1}.$$

PROOF. We use mathematical induction to prove the lemma. The claim is obvious for  $k = 0$ . Suppose the claim is true for  $k = m$ ; then for  $k = m + 1$ ,

$$\begin{aligned}
 & d_2(\mathcal{L}(\xi + V + \delta_\alpha || V| = m + 1), \mathcal{L}(\xi + V || V| = m + 2)) \\
 &= \sup_{f \in \mathcal{F}_3} \left| E[f(\xi + V + \delta_\alpha) | V| = m + 1] \right. \\
 &\quad \left. - \frac{E|V|f(\xi + V)1_{|V|=m+2}}{E|V|1_{|V|=m+2}} \right| \\
 (3.9) \quad &= \sup_{f \in \mathcal{F}_3} \left| E[f(\xi + V + \delta_\alpha) | V| = m + 1] \right. \\
 &\quad \left. - \frac{\sum_{\beta \in \mathcal{Z}} p_\beta E f(\xi + \delta_\beta + Y_\beta) 1_{|Y_\beta|=m+1}}{\sum_{\beta \in \mathcal{Z}} p_\beta P[|Y_\beta| = m + 1]} \right| \\
 &\leq \sup_{f \in \mathcal{F}_3, \beta \in \mathcal{Z}} |E[f(\xi + V + \delta_\alpha) | V| = m + 1] \\
 &\quad - E[f(\xi + \delta_\beta + Y_\beta) | |Y_\beta| = m + 1]|,
 \end{aligned}$$

where  $Y_\beta = \sum_{\gamma \in \mathcal{Z} \setminus \{\beta\}} X_\gamma \delta_\gamma$ ,  $p_\beta = P(X_\beta = 1)$ . On the other hand,

$$\begin{aligned}
 & E[f(\xi + V + \delta_\alpha) | V| = m + 1] \\
 (3.10) \quad &= \frac{p_\beta E f(\xi + \delta_\alpha + \delta_\beta + Y_\beta) 1_{|Y_\beta|=m} + (1 - p_\beta) E f(\xi + \delta_\alpha + Y_\beta) 1_{|Y_\beta|=m+1}}{p_\beta P[|Y_\beta| = m] + (1 - p_\beta) P[|Y_\beta| = m + 1]}.
 \end{aligned}$$

However, for each  $f \in \mathcal{F}_3$ ,

$$|E[f(\xi + \delta_\alpha + Y_\beta) | |Y_\beta| = m + 1] - E[f(\xi + \delta_\beta + Y_\beta) | |Y_\beta| = m + 1]| \leq \frac{1}{|\xi| + m + 2}$$

and

$$|E[f(\xi + \delta_\alpha + \delta_\beta + Y_\beta) | |Y_\beta| = m] - E[f(\xi + \delta_\beta + Y_\beta) | |Y_\beta| = m + 1]| \leq \frac{1}{|\xi| + m + 2},$$

by induction hypothesis, and thus combining (3.9) and (3.10) gives the claim for  $k = m + 1$ .  $\square$

PROOF OF PROPOSITION 3.6. It suffices to show (3.8) because the proof of the other is similar upon using (3.5). Noting that, for each  $f \in \mathcal{F}_3$ ,

$$E[f(\Xi) | |\Xi| = k] = \frac{p_\alpha E[f(\Xi_\alpha) 1_{|\Xi_\alpha|=k}] + (1 - p_\alpha) E[f(\Xi_\alpha - \delta_\alpha) 1_{|\Xi_\alpha - \delta_\alpha|=k}]}{p_\alpha P[|\Xi_\alpha| = k] + (1 - p_\alpha) P[|\Xi_\alpha - \delta_\alpha| = k]},$$

one obtains

$$\begin{aligned}
& d_2(\mathcal{L}(\Xi||\Xi| = k), \mathcal{L}(\Xi_\alpha - \delta_\alpha||\Xi_\alpha - \delta_\alpha| = k)) \\
&= \sup_{f \in \mathcal{F}_3} |E[f(\Xi)||\Xi| = k] - E[f(\Xi_\alpha - \delta_\alpha)||\Xi_\alpha - \delta_\alpha| = k]| \\
&= \sup_{f \in \mathcal{F}_3} \frac{p_\alpha |E[f(\Xi_\alpha)||\Xi_\alpha| = k] - E[f(\Xi_\alpha - \delta_\alpha)||\Xi_\alpha - \delta_\alpha| = k]| P[|\Xi_\alpha| = k]}{P[|\Xi| = k]} \\
&= \frac{p_\alpha P[|\Xi_\alpha| = k]}{P[|\Xi| = k]} d_2(\mathcal{L}(\Xi_\alpha||\Xi_\alpha| = k), \mathcal{L}(\Xi_\alpha - \delta_\alpha||\Xi_\alpha - \delta_\alpha| = k)) \\
&\leq \frac{p_\alpha P[|\Xi_\alpha| = k]}{k P[|\Xi| = k]},
\end{aligned}$$

where the last inequality follows from Lemma 3.8. Therefore, using (2.4) gives

$$\begin{aligned}
(3.11) \quad E\phi_\alpha(|\Xi|) &= \sum_{k=1}^{\infty} \phi_\alpha(k) P[|\Xi| = k] \\
&\leq p_\alpha \sum_{k=1}^{\infty} \frac{P[|\Xi_\alpha| = k]}{k} \\
&= p_\alpha E \frac{1}{|\Xi_\alpha - \delta_\alpha| + 1} \\
&\leq \frac{p_\alpha (1 - e^{-\tau_2})}{\tau_2},
\end{aligned}$$

where  $\phi_\alpha(k) = d_2(\mathcal{L}(\Xi||\Xi| = k), \mathcal{L}(\Xi_\alpha - \delta_\alpha||\Xi_\alpha - \delta_\alpha| = k))$ . Now Proposition 3.6 follows from (2.5) and (3.11).  $\square$

Finally, it may be worth pointing out that when all of the  $p_i$ 's are equal, there is a symmetric property which enables us to produce a better estimate of the errors in Poisson process approximation [cf. Barbour and Brown (1992a), Example 3.9 and Xia (1997)].

**PROPOSITION 3.9.** *Let  $\Gamma = [0, S]$  with metric  $d_0(x, y) = |x - y| \wedge 1$ , where  $S$  is a positive integer. Let  $\mathcal{U} = \{i/n: 1 \leq i \leq nS\}$ . Suppose that  $\{I_\alpha, \alpha \in \mathcal{U}\}$  are independent 0-1 random variables with  $P(I_\alpha = 1) = 1 - P(I_\alpha = 0) = p$  for all  $\alpha \in \mathcal{U}$ . Set  $\Xi = \sum_{\alpha \in \mathcal{U}} I_\alpha \delta_\alpha$ . If  $\mu$  is the measure with constant intensity  $np$  with respect to Lebesgue measure on  $[0, S]$ , then*

$$\begin{aligned}
d_2(\mathcal{L}\Xi, \text{Po}(\mu)) &\leq \left[ \sqrt{\frac{2\mu}{e}} + \int_0^\mu \frac{1 - e^{-y}}{y} dy \right] \frac{0.8p}{\sqrt{[\mu][1 - [\mu]]/(nS - 1)}} \\
&\quad + pe^p + \frac{1 - e^{-\mu}}{2n}.
\end{aligned}$$

PROOF. Let  $m = nS$  and  $\lambda$  be the mean measure of  $\Xi$ , then  $\lambda = \mu = nSp$ . Equations (14) and (16) of Xia (1997) give

$$\begin{aligned} E_{\mathcal{A}} h(\Xi) &= \sum_{k=0}^m E \left[ \int_0^S [h(\Xi + \delta_\alpha) - h(\Xi)] \lambda(d\alpha) \Big| |\Xi| = k \right] (p(m, k) - p(m-1, k)) \\ &\quad + \frac{1}{\lambda} E \int_0^S \int_0^S [h(\Xi_\beta - \delta_\beta + \delta_\alpha) - h(\Xi_\alpha)] \lambda(d\alpha) \lambda(d\beta) \end{aligned}$$

and

$$\left| \frac{1}{\lambda} E \int_0^S \int_0^S [h(\Xi_\beta - \delta_\beta + \delta_\alpha) - h(\Xi_\alpha)] \lambda(d\alpha) \lambda(d\beta) \right| \leq pe^p,$$

where  $p(i, j) := \binom{i}{j} p^j (1-p)^{i-j}$  for  $0 \leq j \leq i$  and  $p(i, j) := 0$  if  $j > i$ . However,

$$\begin{aligned} &\left| \sum_{k=0}^m E \left[ \int_0^S [h(\Xi + \delta_\alpha) - h(\Xi)] \lambda(d\alpha) \Big| |\Xi| = k \right] (p(m, k) - p(m-1, k)) \right| \\ &\leq 2c_1(\lambda, d_2) \lambda d_{TV}(\text{Bi}(m, p), \text{Bi}(m-1, p)), \end{aligned}$$

and thus the proposition follows from (2.6) and the fact that  $d_1(\lambda/\lambda, \mu/\mu) \leq 1/2n$  [see Xia (1997)].  $\square$

## REFERENCES

- ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* 17 9-25.
- BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* 25A 175-184.
- BARBOUR, A. D. and BROWN, T. C. (1992a). Stein's method and point process approximation. *Stochastic Process. Appl.* 43 9-31.
- BARBOUR, A. D. and BROWN, T. C. (1992b). Stein-Chen method, point processes and compensators. *Ann. Probab.* 20 1504-1527.
- BARBOUR, A. D., BROWN, T. C. and XIA, A. (1995). Point processes in time and Stein's method. Unpublished manuscript.
- BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* 95 473-480.
- BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press.
- BARBOUR, A. D. and JENSEN, J. L. (1989). Local and tail approximations near the Poisson limit. *Scand. J. Statist.* 16 75-87.
- BROWN, T. C. and GREIG, D. (1996). Poisson approximation for point processes via monotone couplings. *Ann. Appl. Probab.* 545-560.
- BROWN, T. C. and XIA, A. (1995a). On metrics in point process approximation. *Stochastics Stochastics Rep.* 52 247-263.
- BROWN, T. C. and XIA, A. (1995b). On Stein-Chen factors for Poisson approximation. *Statist. Probab. Lett.* 23 327-332.
- CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* 3 534-545.
- DALEY, D. J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.
- DEHEUVELS, P. and PFEIFER, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* 14 665-678.

- DEHEUVELS, P. and PFEIFER, D. (1988). On a relationship between Uspensky's theorem and Poisson approximations. *Ann. Inst. Statist. Math.* 40 671–681.
- KALLENBERG, O. (1983). *Random Measures*. Academic Press, New York.
- KARR, A. F. (1986). *Point Processes and Their Statistical Inference*. Dekker, New York.
- RACHEV, S. T. (1984). The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory Probab. Appl.* 29 647–676.
- XIA, A. (1997). On the rate of Poisson process approximation to a Bernoulli process. *J. Appl. Probab.* To appear.

DEPARTMENT OF STATISTICS  
SCHOOL OF MATHEMATICS  
UNIVERSITY OF NEW SOUTH WALES  
SYDNEY 2052  
AUSTRALIA  
E-MAIL: xia@maths.unsw.edu.au