Stein's Method and Poisson Process Convergence

A. D. BARBOUR

Abstract

Stein's method of obtaining rates of convergence, well known in normal and Poisson approximation, is considered here in the context of approximation by Poisson point processes, rather than their one-dimensional distributions. A general technique is sketched, whereby the basic ingredients necessary for the application of Stein's method may be derived, and this is applied to a simple problem in Poisson point process approximation.

RATE OF CONVERGENCE: POISSON POINT PROCESS

1. General considerations

Stein's (1970) method of obtaining rates of convergence has found considerable application, both in the original context of convergence to the standard normal distribution, as in Chen (1987), Takahata (1983) and many other papers, and in that of convergence to the Poisson, as for example in Chen (1975) and Barbour (1982). The method is not, however, restricted in principle to these two limit distributions. Barbour (1980) used a variant of Stein's technique to obtain optimal rates of convergence to certain other limit distributions, and Chen (1978) has obtained conditions for convergence to infinitely divisible limits without rate estimates, but wider application has been hampered by two obstacles: finding a suitable Stein equation, and, once found, obtaining estimates of the smoothness of solutions to the equation. The purpose of this note is to sketch a method of overcoming these obstacles, using probabilistic rather than analytic considerations. The argument is illustrated in terms of obtaining rates of convergence to multivariate Poisson and Poisson point process limits, but the approach is more generally applicable.

The broad idea of Chen's (1975) argument for the Poisson distribution is as follows. First, for any $A \in \mathbb{Z}^{+d}$, a function $g_A : \mathbb{Z} \to \mathbb{R}$ is constructed to solve the equation

$$\lambda g(j+1) - jg(j) = I[j \in A] - \pi_\lambda(A), \qquad j \geqq 0, \tag{1.1}$$

where $\pi_\lambda(A)$ denotes the Poisson $\mathscr{P}(\lambda)$ probability of the set A. Replacing j by W, for any random variable W, it therefore follows that the difference between $\mathbb{P}[W \in A]$ and $\pi_\lambda(A)$ can be expressed as $\mathbb{E}\{\lambda g_A(W+1) - W g_A(W)\}$. Now, if $W = \sum_{i=1}^n X_i$, where the X_i are 0–1 random variables with mean p_i and $W_i := W - X_i$ is only weakly dependent on X_i, it follows that

$$\mathbb{E}\{W g_A(W)\} = \sum_{i=1}^n \mathbb{E}\{X_i g_A(W_i + 1)\} \approx \sum_{i=1}^n p_i \mathbb{E} g_A(W_i + 1),$$

and hence that, for $\lambda = \sum_{i=1}^n p_i$,

$$\mathbb{E}\{\lambda g_A(W+1) - W g_A(W)\} \approx \sum_{i=1}^n p_i \mathbb{E}\{g_A(W+1) - g_A(W_i+1)\}. \tag{1.2}$$

Both the right-hand side of (1.2) and what is lost in the approximation of equality can usually be estimated by

$$\varepsilon \sup_{j \geqq 1} |g_A(j+1) - g_A(j)| \tag{1.3}$$

for some suitable ε: for the former, $\varepsilon = \sum_{i=1}^n p_i^2$ is sufficient, but the latter estimate requires more of the structure of W. An analytic argument now shows that the inequality

$$|g_A(j+1) - g_A(j)| \leqq \min(1, \lambda^{-1}) \tag{1.4}$$

holds for all A and j, and (1.3) thus gives a means by which the total variation distance between the distribution of W and $\mathscr{P}(\lambda)$ can be conveniently bounded.

For limit distributions more complicated than the normal or Poisson, an equation analogous to (1.1) has to be found, and smoothness estimates for certain solutions to the equation analogous to (1.4) must be established. The first problem can be approached as follows. The original Stein (1970) equation for the standard normal limit,

$$g'(w) - w g(w) = f(w)$$

and Chen's (1975) counterpart (1.1) for the Poisson $\mathscr{P}(\lambda)$ limit,

$$\lambda g(j+1) - j g(j) = f(j), \qquad j \geqq 0,$$

to be solved for g when f is prescribed, are both first-order equations. This is in some sense accidental: the equations are better understood as (simple) second-order equations, in the Poisson case as

$$\lambda[x(j+1) - x(j)] + j[x(j-1) - x(j)] = f(j), \qquad j \geqq 0. \tag{1.5}$$

It is now immediate that the equation is of the form $\mathscr{A}x = f$, where $\mathscr{A}$ is the infinitesimal operator of the immigration–death process with immigration rate

λ and unit *per capita* death rate, whose equilibrium distribution is $\mathcal{P}(\lambda)$. For normal approximation, the corresponding process is an Ornstein–Uhlenbeck process. When exploiting such Stein equations, it seems that one is implicitly using the fact that the random variable being approximated can be imbedded in a Markov process in equilibrium, whose infinitesimal operator is close to that appearing in the Stein equation. Thus, in order to choose an appropriate equation for less common limits, one possibility is to look for a Markov process whose structure reflects the structure of the random elements whose distribution is to be approximated, and whose equilibrium distribution is the one required. For instance, in the case of a Poisson point process with intensity measure λ over a space $\mathcal{X}$, being used to approximate a sum of many independent and mostly null terms, the natural candidate is an immigration–death process, with unit *per capita* death rate and immigration intensity λ.

The second problem, of obtaining smoothness estimates for solutions to the equation $\mathcal{A}x = f$, can then also be approached by probabilistic arguments. For bounded f, a solution x to $\mathcal{A}x = f$ is typically given by

$$x(z) = -\int_0^\infty \mathbb{E}_z f(Z(t))dt, \tag{1.6}$$

for all f such that $\int f d\pi = 0$, where Z denotes a Markov process with infinitesimal operator $\mathcal{A}$, $\mathbb{E}_z$ its distribution conditional on $Z(0) = z$, and π its equilibrium distribution. The estimates required of x can then usually be reduced to comparing $\mathbb{E}_z f(Z(t))$ with $\mathbb{E}_{z'} f(Z(t))$ for z near z', and standard coupling arguments can be applied. This may often greatly reduce the effort required to achieve good smoothness estimates.

Note that Stein's method does not explicitly require that the Stein equation be of the form $\mathcal{A}x = f$ for some infinitesimal operator $\mathcal{A}$. In certain applications, it may be convenient to exploit a different kind of structure in the random variable whose distribution is to be approximated, and to use a different kind of equation. In such cases, the problem of finding suitable smoothness estimates can be much harder.

2. Multivariate Poisson approximation

To make these general ideas precise, we consider a version of Poisson approximation to a mixed-multinomial distribution. Let Z be the pure jump Markov process on $\mathbb{Z}^{+d}$ defined by $Z(t) := (Z_1(t), \cdots, Z_d(t))$, where, for each $1 \leqq j \leqq d$, Z_j is an immigration–death process with immigration rate λ_j and unit *per capita* death rate, and the $\{Z_j\}_{j=1}^d$ are independent. Let $\pi := \mathcal{P}(\lambda_1) \times \cdots \times \mathcal{P}(\lambda_d)$ denote the equilibrium distribution of Z, and, for any $A \subset \mathbb{Z}^{+d}$, let

$x_A : \mathbb{Z}^{+d} \to \mathbb{R}$ be given by

$$(2.1) \qquad x_A(J) := \int_0^\infty \{\mathbb{P}_J[Z(t) \in A] - \pi(A)\} dt, \qquad J \in \mathbb{Z}^{+d},$$

where $\mathbb{P}_J$ denotes the distribution given $Z(0) = J$. Note that

$$(2.2) \qquad |P_J[Z(t) \in A] - \pi(A)| \leqq \mathbb{P}_J(\tau > t),$$

where τ denotes the time of first coincidence of a process Z started in J and another started with initial distribution π (Pitman (1974)), and hence that $|x_A(J)| \leqq \mathbb{E}_J(\tau)$. For instance, coupling the J-process and the π-process by identifying the individuals immigrating after time 0 and letting those present at time 0 die independently of one another, one easily obtains the estimate

$$(2.3) \qquad |x_A(J)| \leqq \mathbb{E}_\pi \psi(|Z(0)| + |J|),$$

where $\psi(l) := \sum_{r=1}^{l} r^{-1}$ and, for $J \in \mathbb{Z}^{+d}$, $|J| := \sum_{j=1}^{d} J_j$.

We begin by showing that the functions x_A are solutions to a Stein equation for the multivariate Poisson distribution π. For $g : \mathbb{Z}^{+d} \to \mathbb{R}$, let $\Delta_j g(J) := g(J + e(j)) - g(J)$, where $e_k(j) = \delta_{jk}$.

Lemma 1. *The function x_A satisfies the equations*

$$(2.4) \qquad \sum_{j=1}^{d} \{\lambda_j \Delta_j x_A(J) - J_j \Delta_j x_A(J - e(j))\} = \pi(A) - I[J \in A], \qquad J \in \mathbb{Z}^{+d}.$$

Proof. Let $E_J(T) := \int_0^T \{\mathbb{P}[Z(t) \in A] - \pi(A)\} dt$. Then, by considering the first jump of the process Z started in J, we have

$$(2.5) \qquad \begin{aligned} E_J(T) = {} & T \exp(-q_J T)\{I[J \in A] - \pi(A)\} \\ & + \int_0^T \exp(-q_J u) \Big\{ q_J u \{I[J \in A] - \pi(A)\} \\ & + \sum_{j=1}^{d} [\lambda_j E_{J+e(j)}(T-u) + J_j E_{J-e(j)}(T-u)] \Big\} du, \end{aligned}$$

where $q_J := \sum_{j=1}^{d} (\lambda_j + J_j)$. Note also that, from (2.2),

$$(2.6) \qquad |x_A(J) - E_J(T)| \leqq \mathbb{E}_J\{(\tau - T) I[\tau > T]\}.$$

Hence, letting $T \to \infty$ in (2.5) and by dominated convergence, it follows that

$$\begin{aligned} x_A(J) = {} & q_J^{-1}\{I[J \in A] - \pi(A)\} \\ & + \sum_{j=1}^{d} [\lambda_j x_A(J + e(j)) + J_j x_A(J - e(j))] \int_0^\infty \exp(-q_J u) du, \end{aligned}$$

implying the lemma.

In order to use a Stein equation, one needs estimates of the smoothness of the solutions. We prove two such estimates.

Lemma 2. *For all* $J \in \mathbb{Z}^{+d}$ *and all* $A \subset \mathbb{Z}^{+d}$, $|\Delta_{jk} x_A(J)| \leqq 1$, *where* $\Delta_{jk} := \Delta_j \Delta_k$.

Proof. From (2.1), we have

$$\Delta_{jk} x_A(J) = \int_0^\infty \{\mathbb{P}_{J+e(j)+e(k)}[Z(t) \in A] - \mathbb{P}_{J+e(j)}[Z(t) \in A] - \mathbb{P}_{J+e(k)}[Z(t) \in A] + \mathbb{P}_J[Z(t) \in A]\} dt.$$

Define the following four coupled immigration–death processes: $Z^{(0)}$ is distributed as Z started in J, and

$$Z^{(1)}(t) := Z^{(0)}(t) + e(j) I[\tau_1 > t],$$
$$Z^{(2)}(t) := Z^{(0)}(t) + e(k) I[\tau_2 > t],$$
$$Z^{(3)}(t) := Z^{(1)}(t) + e(k) I[\tau_2 > t],$$

where τ_1 and τ_2 are independent standard negative exponential random variables, independent of $Z^{(0)}$. Thus we find that

$$\Delta_{jk} x_A(J) = \int_0^\infty \mathbb{E}\{I[Z^{(3)}(t) \in A] - I[Z^{(2)}(t) \in A] - I[Z^{(1)}(t) \in A] + I[Z^{(0)}(t) \in A]\} \, dt.$$

Now, for any $T < \infty$,

$$\begin{aligned}(2.7)\quad &\int_0^T \mathbb{E}\{I[Z^{(3)}(t) \in A] - I[Z^{(2)}(t) \in A] - I[Z^{(1)}(t) \in A] + I[Z^{(0)}(t) \in A]\} dt \\ &= \mathbb{E} \int_0^T \{I[Z^{(3)}(t) \in A] - I[Z^{(2)}(t) \in A] - I[Z^{(1)}(t) \in A] + I[Z^{(0)}(t) \in A]\} dt,\end{aligned}$$

and the integrand is bounded in modulus by 2. Furthermore, for $t \geqq \tau_1$, $Z^{(0)}(t) = Z^{(1)}(t)$ and $Z^{(2)}(t) = Z^{(3)}(t)$, whereas, for $t \geqq \tau_2$, $Z^{(0)}(t) = Z^{(2)}(t)$ and $Z^{(1)}(t) = Z^{(3)}(t)$, so that the integrand is 0 whenever $t \geqq \min(\tau_1, \tau_2)$. Hence the common absolute value in (2.7) is no greater than

$$\mathbb{E}\{2 \min(\tau_1, \tau_2, T)\} = \int_0^T 4u \exp(-2u) du \leqq 1,$$

and, letting $T \to \infty$ using (2.6), the lemma follows.

Lemma 3. *For all* $\alpha \in \mathbb{Z}^d$, $J \in \mathbb{Z}^{+d}$ *and* $A \subset \mathbb{Z}^{+d}$,

$$(2.8)\quad \left| \sum_{j,k=1}^d \alpha_j \alpha_k \Delta_{jk} x_A(J) \right| \leqq \tfrac{1}{2} \sum_{j=1}^d (\alpha_j^2/\lambda_j)\{1 + 2 \log^+ 2\lambda\} \wedge \left(\sum_{j=1}^d |\alpha_j| \right)^2,$$

where $\lambda := \sum_{j=1}^d \lambda_j$.

Proof. That $|\sum_{j,k=1}^d \alpha_j\alpha_k\Delta_{jk}x_A(J)| \leqq (\sum_{j=1}^d |\alpha_j|)^2$ for all values of λ is implied by Lemma 2. It is therefore enough to prove that

$$\left|\sum_{j,k=1}^{d} \alpha_j\alpha_k\Delta_{jk}x_A(J)\right| \leqq \tfrac{1}{2}\sum_{j=1}^{d} (\alpha_j^2/\lambda_j)\{1+2\log^+ 2\lambda\}$$

when

$$\theta := \sum_{j=1}^{d} (\alpha_j^2/\lambda_j)/2\Big(\sum_{j=1}^{d} |\alpha_j|\Big)^2 \leqq 1: \tag{2.9}$$

note also that $\theta \geqq 1/2\lambda$, so that $\theta \leqq 1$ is only possible when $\lambda \geqq \frac{1}{2}$.

From the proof of Lemma 2, it follows that

$$\begin{aligned}\Delta_{jk}x_A(J) = \int_0^\infty \exp(-2t)\{&\mathbb{P}[Z^{(0)}(t) \in A - e(j) - e(k)] \\ &- \mathbb{P}[Z^{(0)}(t) \in A - e(j)] \\ &- \mathbb{P}[Z^{(0)}(t) \in A - e(k)] + \mathbb{P}[Z^{(0)}(t) \in A]\}dt,\end{aligned} \tag{2.10}$$

where $A - e(j) := \{k \in \mathbb{Z}^{+d} : k + e(j) \in A\}$, and so on. Let $Z^{(0)}(t) = W(t) + Y(t)$, where $W(t)$ denotes those of the original $|J|$ individuals still alive at time t, and $Y(t)$ denotes those alive at t who immigrated after time 0: then $W(t)$ and $Y(t)$ are independent, $W(t) \stackrel{\mathscr{D}}{=} \prod_{j=1}^d B(J_j, e^{-t})$ and $Y(t) \stackrel{\mathscr{D}}{=} \prod_{j=1}^d \mathscr{P}(\lambda_j(t))$, where $\lambda_j(t) := (1-e^{-t})\lambda_j$. Thus it follows that

$$\begin{aligned}\sum_{j,k=1}^{d} \alpha_j\alpha_k\{&\mathbb{P}[Z^{(0)}(t) \in A - e(j) - e(k)] - \mathbb{P}[Z^{(0)}(t) \in A - e(j)] \\ &- \mathbb{P}[Z^{(0)}(t) \in A - e(k)] + \mathbb{P}[Z^{(0)}(t) \in A]\} \\ = \sum_{K \leqq J} &\mathbb{P}[W(t) = K] \sum_{L: L+K \in A} \sum_{j,k=1}^{d} \alpha_j\alpha_k p_L(t; j, k),\end{aligned} \tag{2.11}$$

where

$$\begin{aligned}p_L(t; j, k) := \mathbb{P}[Y(t) = L - e(j) - e(k)] - \mathbb{P}[Y(t) = L - e(j)] \\ - \mathbb{P}[Y(t) = L - e(k)] + \mathbb{P}[Y(t) = L]\end{aligned}$$

and $\leqq$ has its usual meaning in $\mathbb{Z}^{+d}$. Direct computation of the Poisson probabilities now shows that

$$p_L(t; j, k) = \begin{cases}\mathbb{P}[Y(t) = L]\{(L_j\lambda_j^{-1}(t) - 1)^2 - L_j\lambda_j^{-2}(t)\}, & j = k; \\ \mathbb{P}[Y(t) = L]\{(L_j\lambda_j^{-1}(t) - 1)(L_k\lambda_k^{-1}(t) - 1)\}, & j \neq k.\end{cases}$$

Hence we obtain

$$(2.12)\quad \begin{aligned}&\sum_{L:L+K\in A}\sum_{j,k=1}^{d}\alpha_j\alpha_k p_L(t;j,k)\\ &\leqq \sum_{L:L+K\in A}\mathbb{P}[Y(t)=L]\Big(\sum_{j=1}^{d}\alpha_j(L_j\lambda_j^{-1}(t)-1)\Big)^2\\ &\leqq \sum_{L\in\mathbb{Z}^{+d}}\mathbb{P}[Y(t)=L]\Big(\sum_{j=1}^{d}\alpha_j(L_j\lambda_j^{-1}(t)-1)\Big)^2\\ &=\sum_{j=1}^{d}(\alpha_j^2/\lambda_j(t)),\end{aligned}$$

and, similarly,

$$(2.13)\quad \begin{aligned}&\sum_{L:L+K\in A}\sum_{j,k=1}^{d}\alpha_j\alpha_k p_L(t;j,k)\\ &\geqq -\sum_{L:L+K\in A}\mathbb{P}[Y(t)=L]\sum_{j=1}^{d}\alpha_j^2L_j\lambda_j^{-2}(t)\geqq -\sum_{j=1}^{d}(\alpha_j^2/\lambda_j(t)).\end{aligned}$$

Thus, from (2.9)–(2.13), it follows that

$$\begin{aligned}\Big|\sum_{j,k=1}^{d}\alpha_j\alpha_k\Delta_{jk}x_A(J)\Big| &\leqq \int_0^\infty \exp(-2t)\Big\{2\Big(\sum_{j=1}^{d}|\alpha_j|\Big)^2\wedge\sum_{j=1}^{d}(\alpha_j^2/\lambda_j(t))\Big\}dt\\ &=\int_0^\infty \exp(-2t)\Big\{2\Big(\sum_{j=1}^{d}|\alpha_j|\Big)^2\wedge(1-e^{-t})^{-1}\sum_{j=1}^{d}(\alpha_j^2/\lambda_j)\Big\}dt\\ &=\tfrac{1}{2}\sum_{j=1}^{d}(\alpha_j^2/\lambda_j)\{\theta-2\log\theta\},\end{aligned}$$

in $\theta\leqq 1$, which establishes (2.8).

The final step is to show how to apply the Stein equation. We illustrate this in the simplest possible case. Suppose that $(X^{(i)})_{i=1}^n$ are independent random elements of $\mathbb{Z}^{+d}$ with distributions

$$\mathbb{P}[X^{(i)}=e(j)]=p_{ji},\qquad 1\leqq j\leqq d;\qquad \mathbb{P}[X^{(i)}=0]=1-p_i,$$

where $p_i:=\sum_{j=1}^d p_{ji}\leqq 1$. Let $S:=\sum_{i=1}^n X^{(i)}$, and set $\lambda_j:=\sum_{i=1}^n p_{ji}$, $\lambda:=\sum_{j=1}^d\lambda_j$ and $c_\lambda:=\frac{1}{2}+\log 2\lambda$. In the case $d=1$, it is known that $d_{TV}(\mathscr{L}(S),\pi)\leqq \lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^n p_i^2$ (Barbour and Hall (1984)), where $\mathscr{L}(S)$ denotes the distribution of S and d_{TV} the total variation distance between distributions. Here, we prove a rather similar result, valid for any d.

Theorem 1. *The accuracy of $\pi:=\prod_{j=1}^d\mathscr{P}(\lambda_j)$ as an approximation to the distribution of S is given by*

$$d_{TV}(\mathscr{L}(S),\pi)\leqq\sum_{i=1}^n\min\Big\{c_\lambda\sum_{j=1}^d(p_{ji}^2/\lambda_j),\,p_i^2\Big\}.$$

In particular, if $\lambda \max_{1 \leqq j \leqq d} (p_{ji}^2/\lambda_j^2) \leqq \varepsilon_i$, $1 \leqq i \leqq n$,

$$d_{TV}(\mathscr{L}(S), \pi) \leqq c_\lambda \sum_{i=1}^n \varepsilon_i.$$

Remark. With $d = 1$, $\varepsilon_i = p_i^2/\lambda$, and the theorem is weaker than the one-dimensional result, inasfar as $1 - e^{-\lambda} < c_\lambda$. There seems no reason to suppose that Lemma 3 could not be improved to show that c_λ could be replaced by a universal constant.

Proof. Let $A \subset \mathbb{Z}^{+d}$. Then, from Lemma 1, it follows that

$$\pi(A) - \mathbb{P}[S \in A] = \mathbb{E} \sum_{j=1}^d \{\lambda_j \Delta_j x_A(S) - S_j \Delta_j x_A(S - e(j))\}.$$

Now, if $S^{(i)}$ denotes $\sum_{l \neq i} X^{(l)}$, we have

$$\mathbb{E}\{X_j^{(i)} \Delta_j x_A(S - e(j))\} = \mathbb{E}\{X_j^{(i)} \Delta_j x_A(S^{(i)})\},$$

since the integrand is 0 unless $X^{(i)} = e(j)$. Since $X^{(i)}$ and $S^{(i)}$ are independent, this implies that

$$\mathbb{E}\{X_j^{(i)} \Delta_j x_A(S - e(j))\} = p_{ji} \mathbb{E}\{\Delta_j x_A(S^{(i)})\},$$

and hence that

$$\begin{aligned} \pi(A) - \mathbb{P}[S \in A] &= \sum_{i=1}^n \sum_{j=1}^d p_{ji} \mathbb{E}\{\Delta_j x_A(S) - \Delta_j x_A(S^{(i)})\} \\ &= \sum_{i=1}^n \sum_{j,k=1}^d p_{ji} p_{ki} \mathbb{E}\{\Delta_{jk} x_A(S^{(i)})\}. \end{aligned}$$

The theorem now follows from Lemma 3.

Theorem 2. Let S be the point process given by $S = \sum_{i=1}^n \delta_{Y_i}$, *where* $(Y_i)_{i=1}^n$ *are independent random elements of a space* $(Y, \mathscr{Y})$, *and, for any* $C \in \mathscr{Y}$, *define* $p_i(C) := \mathbb{P}[Y_i \in C]$, $\lambda(C) := \sum_{i=1}^n p_i(C)$. *Fixing* $B \in \mathscr{Y}$, *let* S_B *denote the restriction of S to B, and let* π_B *be the measure of a Poisson point process over B with intensity* λ. *Then*

$$d_{TV}(\mathscr{L}(S_B), \pi_B) \leqq c_{\lambda(B)} \sum_{i=1}^n \varepsilon_i,$$

where $\varepsilon_i := \lambda(B) \sup_{C \subset B} \{p_i(C)/\lambda(C)\}^2$.

Remark. If the $(Y_i)_{i=1}^n$ are identically distributed, conditional on being in B,

$$d_{TV}(\mathscr{L}(S_B), \pi_B) \leqq c_{\lambda(B)} \sum_{i=1}^n p_i^2(B)/\lambda(B).$$

In this case, the simple argument of Michel (1987) would give the result with 1 in place of $c_{\lambda(B)}$.

Proof. From Theorem 1, the probability of any finite-dimensional set, of the form $\{(N(C_1), \cdots, N(C_d)) \in A\}$, for some $d \geqq 1$, where $(C_j)_{j=1}^d$ are disjoint elements of $\mathscr{Y}|_B$, $A \subset \mathbb{Z}^{+d}$ and $N(C) := \sum_{i=1}^n I[Y_i \in C]$, can be approximated by the corresponding Poisson probability with the required accuracy. Since the finite-dimensional sets generate the whole σ-algebra associated with S, the theorem follows.

Remarks. 1. The estimate $d_{TV}(\mathscr{L}(S_B), \pi_B) \leqq \frac{1}{2} \sum_{i=1}^n p_i^2$ has been elegantly obtained by Karr and Serfling (1987), for a slightly different Poisson process limit. The present estimate is better for all large enough $\lambda(B)$.

2. For more complicated applications, it may be necessary to estimate more combinations of $\Delta_{jk} x_A(J)$ than those considered in Lemma 3.

3. In approximating point processes by Poisson point processes, it may often be desirable to approximate by a Poisson process with a slightly different intensity from that obtained by applying the above technique—for instance, one with a smoother intensity. This entails a second approximation, estimating the distance between two Poisson point processes with different intensities. In this part of the approximation, total variation distance is typically too much to ask.

4. It is also possible to write down directly a Stein equation for the Poisson point process, instead of proceeding through the finite-dimensional approximations.

Acknowledgements

It is a pleasure to thank the Department of Mathematics at Uppsala University, and the Department of Mathematical Sciences at Johns Hopkins University (ONR contract N00014-79-C-0801), for their hospitality while part of this work was undertaken, and in particular Svante Janson, Robert Serfling and Priscilla Greenwood for much helpful discussion.

References

Barbour, A. D. (1980) Equilibrium distributions for Markov population processes. *Adv. Appl. Prob.* **12,** 591–614.

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

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

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

Chen, L. H. Y. (1978) Two central limit problems for dependent random variables. *Z. Wahrscheinlichkeitsth.* **43,** 223–243.

CHEN, L. H. Y. (1987) The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. *Ann. Prob.*

KARR, A. F. AND SERFLING, R. J. (1987) Poisson approximation of Bernoulli point processes and their superpositions, via coupling.

MICHEL, R. (1987) An improved error bound for the compound Poisson approximation. *ASTIN Bull.* **17**.

PITMAN, J. W. (1974) Uniform rates of convergence for Markov chain transition probabilities. *Z. Wahrscheinlichkeitsth.* **29**, 193–227.

STEIN, C. (1970) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. 6th Berkeley Symp. Math. Statist. Prob.* **2**, 583–602.

TAKAHATA, H. (1983) On the rates in the central limit theorem for weakly dependent random fields. *Z. Wahrscheinlichkeitsth.* **64**, 445–456.