The Annals of Applied Probability
1996, Vol. 6, No. 2, 545–560

POISSON APPROXIMATION FOR POINT PROCESSES VIA MONOTONE COUPLINGS

BY TIMOTHY C. BROWN AND DARRYL GREIG

University of Melbourne

Monotonicity properties of certain classes of point processes with respect to the Palm measure are exploited to derive upper and lower bounds on the total variation distance away from Poisson of these processes. The results obtained are applied to new better than used and new worse than used renewal processes and to a Cox process with rates given by a two state Markov chain.

1. Introduction. Poisson approximation using the Stein–Chen method has been the subject of much recent work. This method, introduced in Chen (1975), produces remarkably accurate results for a wide range of processes; see Barbour, Holst and Janson (1992) for an encyclopedia of applications. The bounds that we produce here are based on an inequality in Theorem 3.1 of Barbour and Brown (1992) which bounds the total variation distance of the number of points in a point process, N, from Poisson by the average Wasserstein distance of N from its reduced Palm distribution. This bound is a point process generalization of the bound which is often called the *coupling method* in discrete time: coupling is also an important tool in the examples studied here, but the bounds obtained are independent of these couplings, requiring only the moments of the process to be calculated. Note that the reduced Palm distribution is that of $N - 1$ given the existence of a point at a particular location, and the Poisson process is characterized by the reduced Palm distribution being identically that of the process itself.

In Barbour, Holst and Janson [(1992), pages 24–26, 60–63] concepts of positive and negative relationships between indicator random variables are used to provide upper and lower bounds for the total variation distance between a sum of indicators and the Poisson distribution. Although these results do not immediately generalize to point processes on the continuum, much of the underlying framework provides a helpful basis for deriving general point process results. We demonstrate that if a given point process satisfies either a positive or negative ordering condition with respect to its Palm process, then the computation of bounds on the total variation distance away from Poisson is greatly simplified, being expressed only in terms of the first four cumulants of the process. The new definitions required for this also afford a further benefit, namely, that the treatment of the positive and

Received April 1994; revised April 1995.
AMS 1991 *subject classifications*. Primary 60G55; secondary 60K05, 60J25.
Key words and phrases. Poisson approximation, monotone couplings, Cox process, NBU, NWU renewal processes.

negative cases requires only a sign change, whereas the positive case is more complicated in Barbour, Holst and Janson. The upper and lower bounds under both kinds of monotonicity involve the variance to mean ratio of the number of points minus 1; this quantity is an upper bound, while the lower bound involves the fourth cumulant in a multiplying factor (Theorem 2.1). In cases that are studied here the multiplying factor is bounded over large ranges of the parameter space for the process, and thus the bounds are tight up to constants. In particular, in the cases here, "two moments suffice" for tight bounds as well as for convergence [Arratia, Goldstein and Gordon (1989)]. While complicated calculations are needed to compute the bounds, the results are pleasingly simple.

Let $\mathscr{P}$ be the set of probability measures on $\mathbb{N} = \{0, 1, 2, \ldots\}$, and for any random variable X, denote its probability law by $\mathscr{L}X$. Define the *total variation distance* d_{TV} between two measures $\mathsf{P}, \mathsf{Q} \in \mathscr{P}$ by

$$d_{\mathrm{TV}}(\mathsf{P}, \mathsf{Q}) = \tfrac{1}{2} \sum_{n=0}^{\infty} |\mathsf{P}(n) - \mathsf{Q}(n)| = \sup_{A \subset \mathbb{N}} |\mathsf{P}(A) - \mathsf{Q}(A)|, \tag{1}$$

and the *Wasserstein distance* d_{W} by

$$d_{\mathrm{W}}(\mathsf{P}, \mathsf{Q}) = \sum_{n=1}^{\infty} \left| \sum_{m=n}^{\infty} \mathsf{P}(m) - \sum_{m=n}^{\infty} \mathsf{Q}(m) \right| = \inf \mathbb{E}|X - Y|, \tag{2}$$

where the infimum is taken over all possible joint distributions of (X, Y) such that $\mathscr{L}X = \mathsf{P}$ and $\mathscr{L}Y = \mathsf{Q}$. The probability law of a Poisson random variable with mean μ is denoted by $\mathrm{Poisson}_{\mu}$.

Let $\mathscr{X}$ be the state space for point processes on $\mathbb{R}^+$. As usual, this can be considered as a space of increasing nonnegative right-continuous functions with value 0 at 0, or a space of nonnegative integer-valued measures. Both representations will be convenient and the context should make it clear which is being used.

Suppose $\{N_t\}_{t \geq 0}$ is a point process on the line and that, for $s \geq 0$, N^s is a point process with the Palm distribution of N conditional on a point at s. The defining equations for these processes N^s are that, for measurable $f: [0, \infty) \times \mathscr{X} \to \mathbb{R}^+$ and π the mean measure of N,

$$\mathbb{E}\left\{ \int_0^{\infty} f(s, N) N(ds) \right\} = \int_0^{\infty} \mathbb{E}\{ f(s, N^s) \} \pi(ds). \tag{3}$$

See Kallenberg [(1983), page 84] for related theory. It will be convenient to use the reduced Palm process $N^s - \delta_s$ which satisfies

$$\mathbb{E}\left\{ \int_0^{\infty} f(s, N - \delta_s) N(ds) \right\} = \int_0^{\infty} \mathbb{E}\{ f(s, N^s - \delta_s) \} \pi(ds). \tag{4}$$

Intuitively the Palm distributions of a point process N are the distributions of N conditional on the presence of a point at a prescribed location. Furthermore, the reduced Palm distribution at s of N is the distribution of N conditional on the presence of a point at s, with the atom at s removed.

Equation (3) is thus a manifestation of the law of total probability, at least in the case when N is simple.

In the examples considered in this paper the point process *is* simple and also stationary. It is therefore convenient to use the formulation of Palm probabilities given by Baccelli and Brémaud (1987). This relies on stationarity of the whole probability space $(\Omega, \mathscr{F}, \mathbb{P})$ under translations. These are represented by a measurable semigroup of automorphisms $\{\theta_t\}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ such that for all t, $\mathbb{P} \circ \theta_t^{-1} = \mathbb{P}$, and furthermore for C a Borel subset of $\mathbb{R}$,

$$N(\theta_t \omega, C) = N(\omega, C + t).$$

That is, $N \circ \theta_t$ has the points of N translated by t to the left. The Palm probability $\mathbb{P}^s$ in this context may be defined for *any* event $A \in \mathscr{F}$ by

$$\mathbb{P}^s(A) = \frac{1}{\mu} \mathbb{E}\left\{ \int_0^1 \mathrm{I}[A] \circ \theta_{u-s} N(du) \right\}, \tag{5}$$

where $\mu = \mathbb{E} N_1$. The Palm process N^s is then a point process with law given by the law of N under $\mathbb{P}^s$, but it will be important in Section 3 that (5) is used for another process X to construct a process X^s whose law is that of X under $\mathbb{P}^s$.

We may now state Theorem 3.1 of Barbour and Brown (1992).

THEOREM 1.1. *Let $\{N_t\}_{t \geq 0}$ be a simple point process defined as above and let $\{(N^s - \delta_s)_t\}_{t \geq 0}$ be its reduced Palm process. For some fixed $t > 0$, let $\lambda = \mathbb{E} N_t$. Then*

$$d_{\mathrm{TV}}(\mathscr{L} N_t, \mathrm{Poisson}_\lambda) \leq \frac{1 - e^{-\lambda}}{\lambda} \int_0^t d_{\mathrm{W}}(\mathscr{L} N_t, \mathscr{L}(N^s - \delta_s)_t) \pi(ds),$$

where π is the mean measure of N.

2. Of PIGS and PILS. The definition (2) of Wasserstein distance suggests a class of point processes for which a straightforward application of the upper bounds in Theorem 1.1 is possible. We say that a point process has the property *Palm is greater* (*less*) *stochastically* [PIGS (PILS)] if, for $0 \leq s \leq t$, the distribution of $(N^s - \delta_s)_t$ is stochastically greater (less) than that of N_t. That is, for all $m \geq 1$,

$$\mathbb{P}\big((N^s - \delta_s)_t \geq m\big) \geq (\leq) \mathbb{P}(N_t \geq m). \tag{6}$$

These properties are quite strong, but the benefits in simplicity of Poisson approximation are substantial and the examples illustrate that there are many processes which satisfy these properties. Moreover, these properties are related to the positive and negative dependence of Barbour, Holst and Janson (1992) in the discrete case, and there are many examples of that dependence in Barbour, Holst and Janson.

THEOREM 2.1. *Let* $\{N_t\}_{t\geq 0}$ *be a point process and, for some fixed* $t > 0$, *let* $\lambda = \mathbb{E}N_t$. *Let* $\kappa_4(N_t)$ *denote the fourth cumulant of* N_t, *which is assumed finite. If* N *has the PIGS property, let*

$$\eta = \frac{\kappa_4(N_t)}{\lambda} - 1, \qquad \varepsilon = \frac{\operatorname{var} N_t}{\lambda} - 1, \qquad \psi = \max\left(0, \frac{\eta}{\lambda\varepsilon}\right) + 3\varepsilon;$$

if N *has the PILS property, let*

$$\eta = 1 - \frac{\kappa_4(N_t)}{\lambda}, \qquad \varepsilon = 1 - \frac{\operatorname{var} N_t}{\lambda}, \qquad \psi = \max\left(0, \frac{\eta}{\lambda\varepsilon} - 3\varepsilon\right).$$

In either case,

$$\frac{\varepsilon}{11 + 3\psi} \leq d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \leq (1 - e^{-\lambda})\varepsilon. \tag{7}$$

PROOF. Let N_t have the property PIGS: the proof for PILS only requires sign changes. Now, using the PIGS property (6), the upper bound in Theorem 1.1 may be simplified, since

$$\begin{aligned} d_{\mathrm{W}}(\mathscr{L}N_t, \mathscr{L}(N^s - \delta_s)_t) &= \sum_{m=1}^{\infty} \left|\mathbb{P}(N_t \geq m) - \mathbb{P}((N^s - \delta_s)_t \geq m)\right| \\ &= \sum_{m=1}^{\infty} \left\{\mathbb{P}((N^s - \delta_s)_t \geq m) - \mathbb{P}(N_t \geq m)\right\} \\ &= \mathbb{E}(N^s - \delta_s)_t - \lambda. \end{aligned}$$

Hence,

$$d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \leq \frac{(1 - e^{-\lambda})}{\lambda}\left(\int_0^t \mathbb{E}(N^s - \delta_s)_t \pi(ds) - \lambda^2\right) \tag{8}$$

and, from (4), with $f(s, \xi) = \xi_t$ for $s \in [0, t]$,

$$\begin{aligned} \int_0^t \mathbb{E}(N^s - \delta_s)_t \pi(ds) &= \mathbb{E}\left\{\int_0^t (N_t - 1)N(ds)\right\} \\ &= \mathbb{E}\{N_t^2\} - \mathbb{E}N_t, \end{aligned}$$

so

$$d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \leq \frac{(1 - e^{-\lambda})}{\lambda}(\operatorname{var} N_t - \lambda),$$

as required.

To derive the lower bound we follow an argument similar to that of Barbour, Holst and Janson [(1992), Theorem 3.D], with appropriate changes for the different context here. Let $\theta > 0$ be fixed and define $f\colon \mathbb{R} \to \mathbb{R}$ by $f(z) = (z - \lambda)\exp(-(z - \lambda)^2/\theta\lambda)$. Then by Lemma 3.2.1 of Barbour, Holst and Janson (1992) the following holds for all $y \geq x$:

$$f(y + \lambda) - f(x + \lambda) \geq y - x - \frac{y^3 - x^3}{\theta\lambda}. \tag{9}$$

Also the derivation of Barbour, Holst and Janson [(1992), equation (3.2.7)] remains unchanged for continuous time. Thus, for $\theta \geq e$,

$$d_{\mathrm{TV}}(\mathscr{L}(N_t), \mathrm{Poisson}_\lambda) \geq \frac{\left|\mathbb{E}\{\lambda f(N_t + 1) - N_t f(N_t)\}\right|}{2\lambda(2e^{-3/2} + \theta e^{-1})}. \tag{10}$$

Now from (4) it follows that

$$\begin{aligned}\mathbb{E}\{N_t f(N_t) - \lambda f(N_t + 1)\} &= \mathbb{E}\left\{\int_0^t f(N_t) N(ds)\right\} - \int_0^t \mathbb{E}\{f(N_t + 1)\}\pi(ds) \\ &= \int_0^t \mathbb{E}\big\{f\big((N^s - \delta_s)_t + 1\big) - f(N_t + 1)\big\}\pi(ds).\end{aligned}$$

Let F and F_s be the distribution functions of N_t and $(N^s - \delta_s)_t$, respectively, and for $y \in [0, 1]$, let $F^{-1}(y) = \inf\{x\colon F(x) > y\}$. If U has a uniform distribution on $[0, 1]$, then $W = F^{-1}(U)$ has the same distribution as N_t, and similarly $V_s = F_s^{-1}(U)$ has the same distribution as $(N^s - \delta_s)_t$. Moreover, by the PIGS property, $V_s \geq W$. Hence

$$\begin{aligned}&\int_0^t \mathbb{E}\big\{f\big((N^s - \delta_s)_t + 1\big) - f(N_t + 1)\big\}\pi(ds) \\ &\quad = \int_0^t \mathbb{E}\{f(V_s + 1) - f(W + 1)\}\pi(ds) \\ &\quad \geq \int_0^t \mathbb{E}\left\{(V_s + 1 - \lambda) - (W + 1 - \lambda) \right. \\ &\qquad \left. - \frac{(V_s + 1 - \lambda)^3 - (W + 1 - \lambda)^3}{\theta\lambda}\right\}\pi(ds)\end{aligned}$$

by (9). However, by applying (4) to terms involving $(N^s - \delta_s)$, the right-hand side of the last inequality becomes

$$\begin{aligned}&\int_0^t \mathbb{E}\left\{(N^s - \delta_s)_t - N_t - \frac{\big((N^s - \delta_s)_t + 1 - \lambda\big)^3 - (N_t + 1 - \lambda)^3}{\theta\lambda}\right\}\pi(ds) \\ &\quad = \operatorname{var} N_t - \lambda - \frac{1}{\theta\lambda}\left\{\mathbb{E}\big\{N_t(N_t - \lambda)^3\big\} - \lambda\mathbb{E}\{N_t + 1 - \lambda\}^3\right\} \\ &\quad = \operatorname{var} N_t - \lambda - \frac{1}{\theta\lambda}\left\{\mathbb{E}\big\{(N_t - \lambda)^4\big\} - 3(\operatorname{var} N_t)^2 - \lambda \right. \\ &\qquad \left. + 3(\operatorname{var} N_t - \lambda)\operatorname{var} N_t\right\} \\ &\quad = \lambda\varepsilon\left(1 - \frac{(\eta/\lambda\varepsilon) + 3 + 3\varepsilon)}{\theta}\right),\end{aligned}$$

so

$$d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \geq \frac{\lambda\varepsilon(1 - ((\eta/\lambda\varepsilon) + 3 + 3\varepsilon)/\theta)}{2\lambda(2e^{-3/2} + \theta e^{-1})}$$
$$\geq \frac{\varepsilon(1 - (\psi + 3)/\theta)}{(4e^{-3/2} + 2\theta e^{-1})}.$$

This lower bound achieves its maximum at (approximately) $\theta = 2\psi + 6$, so by choosing θ to be this value, it follows from (10) that

$$d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \geq \frac{\varepsilon}{8e^{-3/2} + 24e^{-1} + 8e^{-1}\psi} \geq \frac{\varepsilon}{11 + 3\psi}. \qquad \square$$

Those processes for which the PIGS or PILS properties can be established are thereby shown to achieve the best upper bounds obtainable under the application of the Stein–Chen method in Theorem 3.1 of Barbour and Brown (1992), since Theorem 2.1 evaluates the upper bound expression exactly.

We also note here that if a process may be shown to be PIGS or PILS, then Theorem 2.1 reduces the computation of the upper and lower bounds away from Poisson to the computation of the first four cumulants of N_t for some fixed $t > 0$. It has to be said that for many point processes this will prove to be a rather nontrivial problem, and perhaps intractable in some cases. In the next two sections of the paper we present some applications of Theorem 2.1 to specific examples and include some techniques for computing cumulants that may be more generally applied. In the next section we are able to derive an exact form for the first four cumulants of a Cox process, which appears to be extendible to general finite state Markov processes, for cumulants of any order. On the other hand, Section 4 involves some processes for which a direct computation of the cumulants does not seem feasible, and so some asymptotic results are derived which demonstrate the behavior of the Poisson approximation as $t \to \infty$.

It is instructive to compare Theorem 2.1 with Theorems 3D and 3E of Barbour, Holst and Janson (1992), whose proofs have much in common with the above proof. In the case of PILS, the result is of the same form. Indeed, if the indicator random variables $(I_\beta,\ \beta \in \{0, 1, \ldots, t\})$ are negatively related and for $C \subseteq \{0, 1, \ldots, t\}$,

$$N(C) = \sum_{\beta \in C} N_\beta,$$

then it is easy to see [using (2.1.3) of Barbour, Holst and Janson] that N is PILS.

In the case of PIGS, Theorem 2.1 is simpler and of a similar form to that of the PILS case, whereas Theorem 3E of Barbour, Holst and Janson has an extra term in the upper bound. This term appears because the positive relationship used there implies that

$$\mathbb{P}\big((N^s - \delta_s)_t \geq m\big) \geq \mathbb{P}\big((N - \delta_s)_t \geq m\big),$$

but not necessarily

$$\mathbb{P}\big((N^s - \delta_s)_t \geq m\big) \geq \mathbb{P}(N_t \geq m),$$

which is the PIGS property. On the other hand, it is not necessarily true that the PIGS property implies the positive relationship of Barbour, Holst and Janson, being formulated for continuous processes rather than indicators.

3. A two state Cox process which is PIGS. Let $\{\alpha, \beta, a, b\} \subset (0, \infty)$, $a < b$. Let $\{X_t\}_{t \geq 0}$ be a stationary alternating continuous time Markov process with states α and β and associated rates a and b, respectively. The process $\{N_t\}_{t \geq 0}$ is defined in the following way: when X_t is in state ξ for $\xi \in \{\alpha, \beta\}$, then N evolves as a Poisson process with rate ξ. Because N can be defined using X to change time randomly for an independent standard Poisson process, N is an example of a Cox process [see Snyder (1977) and Daley and Vere-Jones (1988)]. Throughout this section, let $\mu = \mathbb{E}N_{[0,1]}$ and assume that λ, ε, ψ and η are defined as in Theorem 2.1.

LEMMA 3.1. *For $t > 0$, the random variable N_t has first, second and fourth cumulants given by*

$$\kappa_1 = \lambda = \mu t = \frac{(\beta a + \alpha b)t}{(a+b)},$$

$$\kappa_2 = \mu t + \frac{2\delta}{(a+b)}\left\{t + \frac{(\exp(-(a+b)t) - 1)}{(a+b)}\right\},$$

$$\kappa_4 = \mu t + \frac{2\delta}{(a+b)}\left\{\nu_1\left\{t + \frac{(\exp(-(a+b)t) - 1)}{(a+b)}\right\} + \nu_2 t(\exp(-(a+b)t) - 1) + \nu_3 t^2 \exp(-(a+b)t) + \nu_4(\exp(-(a+b)t) - 1)^2\right\},$$

where

$$\delta = \frac{ab(\beta - \alpha)^2}{(a+b)^2},$$

$$\nu_1 = 7 + \frac{36(b-a)(\beta-\alpha)}{(a+b)^2} + \frac{36(a^2 - 3ab + b^2)(\beta-\alpha)^2}{(a+b)^4},$$

$$\nu_2 = \frac{24(a^2 - 3ab + b^2)(\beta-\alpha)^2}{(a+b)^4} + \frac{18(b-a)(\beta-\alpha)}{(a+b)^2},$$

$$\nu_3 = \frac{6(b-a)^2(\beta-\alpha)^2}{(a+b)^3},$$

$$\nu_4 = -\frac{6\delta}{(a+b)^3}.$$

PROOF (Sketch). We find an expression for the moment generating function of N_t by conditioning on the state of N_t at time t, that is,

$$\mathbb{E}\exp(\theta N_t) = \mathbb{E}\{\exp(\theta N_t)\mathrm{I}[X_t = \alpha]\} + \mathbb{E}\{\exp(\theta N_t)\mathrm{I}[X_t = \beta]\}. \tag{11}$$

Let $\mathscr{F}_t = \sigma(N_z, X_z,\ z \le t)$. Then the conditional intensity of N (which we denote by λ_s) relative to $\mathscr{F}$ is X_-. For the process $\mathrm{I}[X_s = \alpha]$ we can also define a conditional intensity ϕ_s via

$$\mathbb{E}\left\{\int Y_s d(\mathrm{I}[X_s = \alpha])\right\} = \int \mathbb{E}\{Y_s \phi_s\}\, ds$$

for bounded previsible Y. Thus the conditional intensity is $\phi_s = b\mathrm{I}[X_{s-} = \beta] - \alpha\mathrm{I}[X_{s-} = \alpha]$. Furthermore, $\Delta N_s\, \Delta X_s = 0$ a.e. Hence,

$$\begin{aligned}
&\mathbb{E}\{\exp(\theta N_t)\mathrm{I}[X_t = \alpha]\} - \mathbb{P}(X_0 = \alpha) \\
&\quad = \mathbb{E}\left\{\int_0^t \exp(\theta N_{s-})\mathrm{I}[X_{s-} = \alpha](\exp(\theta) - 1)\, dN_s\right\} \\
&\qquad + \mathbb{E}\left\{\int_0^t \exp(\theta N_{s-})\, d(\mathrm{I}[X_s = \alpha])\right\} \\
&\quad = \int_0^t \mathbb{E}\{\exp(\theta N_s)\mathrm{I}[X_s = \alpha]\lambda_s\}(\exp(\theta) - 1)\, ds \\
&\qquad + \int_0^t \mathbb{E}\{\exp(\theta N_s)(b\mathrm{I}[X_s = \beta] - a\mathrm{I}[X_s = \alpha])\}\, ds \\
&\quad = \int_0^t \mathbb{E}\{\exp(\theta N_s)\mathrm{I}[X_s = \alpha]\}((\exp(\theta) - 1)\alpha - a)\, ds \\
&\qquad + \int_0^t \mathbb{E}\{\exp(\theta N_s)\mathrm{I}[X_s = \beta]\} b\, ds,
\end{aligned}$$

where the first equality follows from integration by parts [Brémaud (1981), Appendix A.4.2] and the second from the previsibility of the argument and Fubini's theorem (noting that N is bounded above by a Poisson random variable with mean $\sup\{\alpha, \beta\}$).

An analogous expression for $\mathbb{E}\{\exp(\theta N_t)\mathrm{I}[X_t = \beta]\}$ may be derived in a similar fashion. Then, letting $x_\xi(t) = \mathbb{E}\{\exp(\theta N_t)\mathrm{I}[N_t = \xi]\}$ for $\xi \in \{\alpha, \beta\}$,

$$\begin{bmatrix} x'_\alpha(t) \\ x'_\beta(t) \end{bmatrix} = \begin{bmatrix} (e^\theta - 1)\alpha - a & b \\ a & (e^\theta - 1)\beta - b \end{bmatrix} \begin{bmatrix} x_\alpha(t) \\ x_\beta(t) \end{bmatrix}. \tag{12}$$

The solution of this equation is then

$$\begin{bmatrix} x_\alpha(t) \\ x_\beta(t) \end{bmatrix} = \mathbf{u}_1 \exp(\zeta_1 t) + \mathbf{u}_2 \exp(\zeta_2 t),$$

where $(\mathbf{u}_i, \zeta_i)$ is the ith eigenvector–eigenvalue pair of the matrix in (12) and $\mathbf{u}_1, \mathbf{u}_2$ satisfy the extra constraint that

$$\mathbf{u}_1 + \mathbf{u}_2 = \begin{bmatrix} b/(a+b) \\ a/(a+b) \end{bmatrix}.$$

Assume that $\zeta_1 > \zeta_2$ and let f_i be the sum of the terms in the ith eigenvector. Then by (11),

$$\mathbb{E}\{\exp(\theta N_t)\} = \exp(\zeta_1 t)\big(f_1 + f_2 \exp(-(\zeta_1 - \zeta_2)t)\big),$$

so the cumulant generating function is

$$\log \mathbb{E}\{\exp(\theta N_t)\} = \zeta_1 t + \log\big(f_1 + f_2 \exp(-(\zeta_1 - \zeta_2)t)\big). \tag{13}$$

Note that both $\mathbf{u}_i$ and ζ_i are functions of θ. The cumulants are therefore obtained by calculating the eigenvalues and eigenvectors and differentiating (13) with respect to θ and evaluating the resulting expression at $\theta = 0$. Details are available from the authors on request. □

For any simple point process driven by a finite state Markov process an equation similar to (11) may be written down. If the appropriate conditional intensities can be derived, one would expect a matrix equation like (12) and, consequently, a cumulant generating function (CGF) of the form of (13). Note also that the form of this CGF implies that, as $t \to \infty$, any such point process will have its cumulants dominated by a linear term involving the derivatives of the largest eigenvalue of the matrix obtained.

THEOREM 3.1. *The process N_t has the PIGS property.*

PROOF. The theorem is proved by specifying a coupling of N_t with its reduced Palm process $(N^s - \delta_s)_t$, for some $0 \le s \le t$, that demonstrates the necessary monotonicity requirements. Let $\mathbb{P}^s$ be the Palm probability conditional on a point occurring at time s. Furthermore let X^s be a process which has the law of X under $\mathbb{P}^s$.

To begin with we couple X with X^s. Papangelou's formula [Brémaud, Kannurpatti and Mazumdar (1992)] implies that $\mathbb{P}^s(X_{s+} = \alpha) = \mathbb{P}^s(X_{s-} = \alpha) = \alpha b/(\alpha b + \beta a)$, and it follows from standard Markov chain theory that $\mathbb{P}(X_s = \alpha) = b/(a + b)$. Let U be a uniform $[0, 1]$ random variable. Then the values of X_s and X_s^s are assigned according to the position of U in the partition of $[0, 1]$ given in Figure 1. Note that the positions of $b/(a + b)$ and $\alpha b/(\alpha b + \beta a)$ are as shown in the diagram. So, for example, if $\alpha > \beta$, then with probability $b/(a + b)$, $X_s = X_s^s = \alpha$, with probability $\beta a/(\alpha b + \beta a)$,

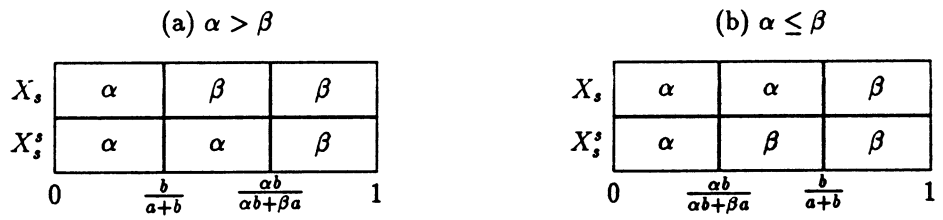

FIG. 1.

$X_s = X_s^s = \beta$ and with probability $\alpha b/(\alpha b + \beta a) - b/(a + b)$, $X_s = \beta$, $X_s^s = \alpha$. Clearly this coupling preserves the correct marginal probabilities for N and $(N^s - \delta_s)$ at time s.

If we condition on the event $X_s = X_s^s$ we can take $X_t = X_t^s$ for all t and N_t and $(N^s - \delta_s)_t$ will be stochastically indistinguishable. In this instance, therefore, we define $(N^s - \delta_s)_t = N_t$ in both cases of Figure 1.

Now assume that $\alpha > \beta$. If $X_s \neq X_s^s$, then the coupling of Figure 1(a) ensures that $X_s = \beta$ and $X_s^s = \alpha$. Let T_i, $i \in \mathbb{N}$, be the time of the ith transition of X after time s. We define the analogous times $\{T_j^s\}_{j \in \mathbb{N}}$ for X^s in the following way.

Let $A \sim \exp(a)$ be independent of X. If $A > T_1$, then let $T_i^s = T_{i+1}$ for $i \geq 1$. If $A \leq T_1$, then let $T_1^s = A$ and $T_i^s = T_{i-1}$ for $i \geq 2$. Note that X and X^s are coupled after the first transition time $T = T_1 \wedge T_1^s$, and consequently so are N and $(N^s - \delta_s)$. That is, $\forall\ u > T$, $N_{[T,u)} = (N^s - \delta_s)_{[T,u)}$.

Assume that the points of N in the interval $[s, T)$ are determined by a sequence of $\exp(\beta)$ random variables, $\{B_i\}$. Then the points of $(N^s - \delta_s)$ in the same interval may be defined by the sequence $\{(\beta/\alpha)B_i\}$. Since $\alpha > \beta$, it follows that $(N^s - \delta_s)_{[s,u)} \geq N_{[s,u)}\ \forall\ s \leq u \leq T$.

For $u \leq s$ we need only to notice that N_t is reversible, and hence the above coupling may be used in reversed time. Hence, if $\alpha > \beta$, we have that $(N^s - \delta_s)_t \geq N_t$ for all $0 \leq s \leq t$.

The coupling is essentially the same in the case $\beta \geq \alpha$, with $X_s^s = \beta$, $X_s = \alpha$ the only nontrivial outcome from the initial uniform $[0, 1]$ realization; the rest of the argument follows mutatis mutandis. Intuitively, $(N^s - \delta_s)$ will still dominate N since the greater of β and α is again the state of $(N^s - \delta_s)$ at time s when a mismatch occurs. □

COROLLARY 3.1. *Let δ and ν_i be defined as in Lemma* 3.1. *Then for N_t we have*

$$\varepsilon = \frac{2\delta}{\mu t(a+b)} g(t), \quad \text{where} \quad g(t) = \left\{ t + \frac{\exp(-(a+b)t) - 1}{(a+b)} \right\}.$$

Let

$$c_1 = |\nu_1| + |\nu_2| + \frac{3e^{-3}\nu_3}{(a+b)}, \qquad c_2 = 11 + \frac{3c_1(a+b)}{2\mu} + \frac{18\delta}{\mu(a+b)},$$

$$c_3 = |\nu_1| + 7|\nu_2| + \frac{7\nu_3}{(a+b)}, \qquad c_4 = \frac{33\mu}{(a+b)} + 3c_3 + \frac{81\delta}{(a+b)^2}.$$

Then, if $t > 3/(a+b)$,

$$\frac{\varepsilon}{c_2} \leq d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \leq (1 - e^{-\mu t})\varepsilon \tag{14}$$

and, in particular, if $t \geq 3c_1/\mu + 1/(a+b)$ and $(\beta - \alpha)^2 \leq (a+b)^2 \times (\beta a + \alpha b)/(18ab)$,

$$\frac{\varepsilon}{13} \leq d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \leq \varepsilon. \tag{15}$$

Also for $t \le 3/(a+b)$,

$$\frac{2}{7c_4}\,\delta t^2 \le d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \le \delta t^2. \tag{16}$$

NOTE. Equations (14)–(16) show that the upper and lower bounds are tight up to a constant order both for small and large t. Note, however, that the bounds are of a different character in the two different cases: those in (14) and (15) being essentially constant in t, while those for small t are quadratic in t.

PROOF OF COROLLARY 3.1. The value of ε follows directly from the definitions of Theorem 2.1 and the results of Lemma 3.1. Furthermore,

$$\begin{aligned}\frac{\eta}{\lambda\varepsilon} = \frac{1}{\mu}\Bigg\{ &\frac{\nu_1}{t} + \frac{\nu_2(\exp(-(a+b)t) - 1)}{g(t)} \\ &+ \frac{\nu_3 t\exp(-(a+b)t)}{g(t)} + \frac{\nu_4(\exp(-(a+b)t) - 1)^2}{tg(t)}\Bigg\}.\end{aligned} \tag{17}$$

Clearly these functions may be directly applied to the results of Theorem 2.1 and the most accurate bounds allowed by that theorem obtained. However, a careful analysis of (17) provides the more appealing and easily applied bounds given above.

In the first instance, let $t > 3/(a+b)$. Note that $t - 1/(a+b) < g(t) \le t$. It follows from (17) and the fact that $\nu_4 < 0$ that

$$\begin{aligned}\frac{\eta}{\lambda\varepsilon} &\le \frac{1}{\mu g(t)}\{|\nu_1| + |\nu_2| + \nu_3 t\exp(-(a+b)t)\} \\ &\le \frac{1}{\mu(t - 1/(a+b))}\left\{|\nu_1| + |\nu_2| + \frac{3e^{-3}\nu_3}{(a+b)}\right\} = \frac{c_1}{\mu(t - 1/(a+b))}\end{aligned}$$

and so

$$d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \ge \varepsilon\left\{11 + \frac{3c_1}{\mu(t - 1/(a+b))} + 9\varepsilon\right\}^{-1} \ge \frac{\varepsilon}{c_2}.$$

The lower bound in (15) follows from noticing that under the conditions there, $3c_1\{\mu(t - 1/(a+b))\}^{-1} \le 1$ and $9\varepsilon \le 18\delta/(\mu(a+b)) \le 1$.

Now assume $0 < t \le 3/(a+b)$. By expanding the exponential term of $g(t)$ we have that

$$\begin{aligned}&\frac{(a+b)t^2}{2}\left\{1 - \frac{(a+b)t}{3} + \frac{(a+b)^2t^2}{12} - \frac{(a+b)^3t^3}{60}\right\} \\ &\qquad \le g(t) \le \frac{(a+b)t^2}{2}\end{aligned} \tag{18}$$

Since the bracketed function in the lower bound is decreasing, it achieves its minimum at $t = 3/(a + b)$, so

$$g(t) \geq \frac{3(a+b)t^2}{20} \geq \frac{(a+b)t^2}{7}. \tag{19}$$

Thus

$$\begin{aligned}\frac{\eta}{\lambda\varepsilon} &\leq \frac{1}{\mu}\left\{\frac{|\nu_1|}{t} + \frac{7|\nu_2|(1-\exp(-(a+b)t))}{(a+b)t^2} + \frac{7\nu_3\exp(-(a+b)t)}{(a+b)t}\right\} \\ &\leq \frac{1}{\mu t}\left\{|\nu_1| + 7|\nu_2| + \frac{7\nu_3}{a+b}\right\}\end{aligned}$$

and again applying Theorem 2.1,

$$\begin{aligned}d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) &\geq \frac{\mu t\varepsilon}{11\mu t + 3c_3 + 9\varepsilon\mu t} \\ &= \frac{2\,\delta g(t)}{(a+b)(11\mu t + 3c_3 + 9\varepsilon\mu t)} \\ &\geq \frac{2\,\delta g(t)}{(a+b)c_4} \geq \frac{2}{7c_4}\,\delta t^2,\end{aligned}$$

where the last equality follows from (19).

The upper bound in this case comes directly from the expansion of the exponential term in the upper bound of (7):

$$d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_\lambda) \leq (1 - e^{-\mu t})\varepsilon \leq \mu t\frac{2\,\delta g(t)}{(a+b)\mu t} = \frac{2\,\delta g(t)}{a+b} \leq \delta t^2,$$

where the last inequality follows from (18). □

4. A monotone coupling for NBU and NWU renewal processes. The results of Theorem 2.1 may also be applied to two classes of renewal processes that are characterized by a weak ordering condition on the renewal distribution.

Let N be a renewal process with interpoint distribution F concentrated on $(0, \infty)$. Let $s \geq 0$ be fixed and let D_s and B_s be the forward and backward recurrence times of N at s. That is,

$$D_s = \inf\{y\colon N(s, s+y] > 0\}, \qquad B_s = \inf\{x\colon N(s-x, s] > 0\}.$$

Now $\{D_s\colon s \geq 0\}$ is a Markov process such that at time s, D_s only depends on the past of the renewal process through B_s. Thus, if Y is some F-random variable,

$$\begin{aligned}F_x(y) &=_{\mathrm{def}} \mathbb{P}(D_s \leq y \mid B_s = x) = \mathbb{P}(Y \leq y \mid Y > x) \\ &= \frac{F(x+y) - F(x)}{1 - F(x)}.\end{aligned}$$

For any distribution function H, we define its inverse H^{-1} by

$$H^{-1}(y) = \inf\{x: H(x) \geq y\}, \qquad y \in [0,1].$$

We may now define a coupling of N with its reduced Palm process $(N^s - \delta_s)$. Let U_1, U_2 be independent $U[0,1]$ random variables and let $\{Y_i\}_{i \in \mathbb{N}}$ be a sequence of independent F random variables. Let $X_1 = F^{-1}(U_1)$, $B_s = G^{-1}(U_1)$, $X_2 = F^{-1}(U_2)$ and $D_s = F_{B_s}^{-1}(U_2)$. Let $\{T_i\}_{i \in \mathbb{N}}$ be the renewal times of N with respect to s. That is, T_i is the time of the ith renewal of N after time s and T_{-i} is the time of the ith renewal of N counting back from s. Let $\{T_i^s\}_{i \in \mathbb{N}}$ be the analogous times for $(N^s - \delta_s)$.

Let $T_1 = s + D_s$, $T_1^s = s + X_2$, $T_{-1} = s - B_s$, $T_{-1}^s = s - X_1$ and, for $i > 1$, define the times by $T_i = T_{i-1} + Y_{i-1}$, $T_i^s = T_{i-1}^s + Y_{i-1}$, $T_{-i} = T_{-i+1} - Y_{-i+1}$ and $T_{-i}^s = T_{-i+1}^s - Y_{-i+1}$. Note that both N and $(N^s - \delta_s)$ have the correct marginal distributions under this coupling and that to establish (say) the PILS property for N_t it is certainly sufficient to show that for all $i \geq 1$, $T_i \wedge t \leq T_i^s \wedge t$ and $T_{-i} \vee 0 \geq T_{-i}^s \vee 0$.

A distribution F is said to be *new better than used* (NBU) if

$$\frac{F(x+\tau) - F(\tau)}{1 - F(\tau)} \geq F(x) \quad \text{for all } x \geq 0, \tau > 0.$$

The class *new worse than used* (NWU) is defined by the obvious reversal of inequality. These distributions belong to the so-called aging and antiaging hierarchies of distributions from the increasing failure rate (IFR) and increasing failure rate average (IFRA) classes, and NWU contains the decreasing failure rate (DFR) and decreasing failure rate average (DFRA) classes. See Barlow and Proschan (1975) and Stoyan and Daley (1983) for definitions and examples.

THEOREM 4.1. *Let $\{N_t\}_{t \geq 0}$ be a stationary renewal process with renewal distribution F. If F is NBU (NWU), then N_t has the PILS (PIGS) property and*

$$\frac{\varepsilon}{11 + 3\psi} \leq d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_{t/\mu}) \leq (1 - e^{-t/\mu})\,\varepsilon,$$

where $\mu = \int_0^\infty (1 - F(x))\,dx$ and ε and ψ are defined as in Theorem 2.1 with $\lambda = t/\mu$.

PROOF. We shall establish PILS for NBU only and note that the PIGS case (NWU) follows from the obvious sign changes.

If N is stationary, then standard renewal theory [see Daley and Vere-Jones (1988)] gives the marginal distribution of both D_s and B_s for N as

$$G(t) = \frac{1}{\mu} \int_0^t (1 - F(x))\,dx.$$

As stated above, N_t will be PILS, if for all $i \geq 1$, $T_i \wedge t \leq T_i^s \wedge t$ and $T^{-i} \vee 0 \geq T^s_{-i} \vee 0$. However, from the coupling, this reduces to demonstrat-

ing that $B_s \le X_1$ and $D_s \le X_2$ or, equivalently, that for all $x, y \ge 0$, $G(x) \ge F(x)$ and $F_y(x) \ge F(x)$.

The second of these inequalities is exactly the definition of F NBU. The first follows by noticing that the NBU property may be restated as $(1 - F(x + \tau)) \le (1 - F(x))(1 - F(\tau))$. Thus,

$$(1 - G(x)) = \frac{1}{\mu}\int_x^\infty (1 - F(\tau))\, d\tau = \frac{1}{\mu}\int_0^\infty (1 - F(\tau + x))\, d\tau \le (1 - F(x)).$$

The bounds are then a direct application of Theorem 2.1. □

Clearly the usefulness of this corollary depends on our ability to compute the cumulants of the specific renewal process under review. In particular, the behavior of ψ is of interest since if ψ can be bounded above by a constant, then the correct order of magnitude of the total variation distance is shown to be ε for all t.

For a specific renewal distribution F, an exact computation of cumulants may be possible; however, in cases where this is not possible, then asymptotic bounds (as $t \to \infty$) may still be obtained. In fact, Cox and Miller (1965) demonstrated that the first four cumulants of a general renewal process N_t are asymptotically

$$\frac{t}{k_1}, t\frac{k_2}{k_1^3}, t\left(\frac{3k_2^2}{k_1^5} - \frac{k_3}{k_1^4}\right), t\left(\frac{k_4}{k_1^5} + \frac{15k_2^3}{k_1^7} - \frac{10k_2k_3}{k_1^6}\right),$$

where k_i represents the ith cumulant of the distribution F.

The upper and lower bounds of Theorem 4.1 may now be evaluated directly in terms of the cumulants of the renewal distribution. Let δ be defined as $k_2/k_1^2 - 1$ in the PIGS case and $1 - k_2/k_1^2$ in the PILS case. Then

$$\limsup_{t\to\infty} d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_{t/\mu}) \le \delta. \tag{20}$$

Furthermore,

$$\begin{aligned} \liminf_{t\to\infty} d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_{t/\mu}) &\ge \frac{\delta}{9\delta + 11}, \\ \liminf_{t\to\infty} d_{\mathrm{TV}}(\mathscr{L}N_t, \mathrm{Poisson}_{t/\mu}) &\ge \frac{\delta}{11} \end{aligned} \tag{21}$$

for the PIGS and PILS cases, respectively.

From this argument we see that if, for a given PIGS or PILS process, it can be shown that as $t \to \infty$ the cumulants of the process are of the same order of magnitude in t [or at least that $\eta/\lambda\varepsilon$ is $o(1)$], then $\psi \to 3\varepsilon$ in the PIGS case and $\psi \to 0$ in the PILS case. Hence the upper and lower bounds are of the same order of magnitude for sufficiently small ε and may be written in the form of (20) and (21).

The PIGS property for processes with DFR or increasing mean residual life (IMRL) interarrival distributions may be immediately inferred from an elegant coupling due to Brown [Theorem 1 of Brown (1980)] in which the Palm process and stationary process are shown to differ by only a finite number of points over the whole real line. The coupling used for Theorem 4.1 does not have Brown's strong pointwise matching property and thus admits a wider class of distributions.

5. Conclusions and comments. We have given general conditions under which both upper and lower bounds of the same order of magnitude on the total variation distance of a point process away from Poisson may be immediately written down. However, as is clear from the examples of Sections 3 and 4, these conditions are not always trivial to verify for a given process. The definition of PIGS and PILS given in (6) seems to require quite fine knowledge of the probability mass function of both the process and its Palm process, which is not at all easy to come by in many instances. Thus one is left to coupling arguments which may prove rather involved for complex processes, although we reiterate that the bounds obtained are dependent only on the existence of a satisfactory coupling and not on the specific coupling chosen.

Of course, the results are particularly interesting when both upper and lower bounds can be shown to be of the same order of magnitude, at least asymptotically in t, as in the examples here.

Note that the point processes in these examples are stationary, whereas neither Theorem 1.1 nor 2.1 requires stationarity. If a point process is assumed to be stationary, then one may use the Baccelli and Brémaud formulation of Palm probability described in the Introduction to compute the necessary probabilities. In the case of the Cox process this allows a suitable coupling of X with X^s from which the PIGS property follows. On the other hand, the stationarity in Theorem 4.1 serves only to link the result with the well known NBU and NWU renewal processes. The PIGS and PILS properties for nonstationary renewal processes should follow directly from a similar (and, one would imagine, stronger) condition on the residual life and age distributions of a nonstationary renewal process.

It is not known at this stage just how extensive the class of PIGS and PILS processes are, although the NBU and NWU renewal processes examined in Section 4 are a large and well known class. Barbour and Brown (1996) have shown that the process counting transitions between queues in a Jackson network is also PIGS, and work is currently in progress to establish lower bounds complementing the upper bounds given.

Acknowledgments. The authors would like to thank an Associate Editor and both of the referees for their comments, which have improved the presentation of the paper. We would also like to acknowledge some helpful communication with Nick Duffield regarding the proof of Lemma 3.1.

REFERENCES

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximation: the Chen–Stein method. *Ann. Probab.* **17** 9–25.

BACCELLI, F. and BRÉMAUD, P. (1987). *Palm Probabilities and Stationary Queues. Lecture Notes in Statist.* **41**. Springer, Berlin.

BARBOUR, A. D. and BROWN, T. C. (1992). The Stein–Chen method, point processes and compensators. *Ann. Probab.* **20** 1504–1527.

BARBOUR, A. D. and BROWN, T. C. (1996). Approximate versions of Melamed's theorem. *J. Appl. Probab.* **33** 472–489.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press.

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York.

BRÉMAUD, P. (1981). *Point Processes and Queues*. Springer, New York.

BRÉMAUD, P., KANNURPATTI, R. and MAZUMDAR, R. (1992). Event and time averages: a review. *Adv. in Appl. Probab.* **24** 377–411.

BROWN, M. (1980). Bounds, inequalities, and monotonicity properties for some specialized renewal processes. *Ann. Probab.* **8** 227–240.

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

COX, D. R. and MILLER, H. D. (1965). *The Theory of Stochastic Processes*. Chapman and Hall, London.

DALEY, D. J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.

KALLENBERG, O. (1983). *Random Measures*, 3rd ed. Academic Press, New York.

SNYDER, D. L. (1977). *Random Point Processes*. Wiley, New York.

STOYAN, D. and DALEY, D. J. (1983). *Comparison Methods for Queues and Other Stochastic Models*. Wiley, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MELBOURNE
PARKVILLE 3052
AUSTRALIA
E-MAIL: tim@mugga.stats.mu.oz.au