DISTANCES OF RANDOM VARIABLES AND POINT PROCESSES

by

D.J. Daley
Department of Statistics (I.A.S.)
Australian National University
Canberra 2600 Australia

ABSTRACT

The paper is basically a discussion and survey of some
literature concerning the concept of a distance function for random
variables and point processes. Work of Kal'zanov on the closeness of
renewal processes to Poisson processes is outlined, and there are
comments on various possible comparison methods for point processes
as given by Whitt.

Keywords: Point processes, renewal processes, distance functions,
metrics, comparability of processes.

Research done whilst a visitor in the Department of Statistics,
University of North Carolina, Chapel Hill, and supported by the
Office of Naval Research under Contract N00014-75-C-0809.
1. INTRODUCTION

When are two point processes close to one another? Indeed, when are two stochastic processes close to one another? These notes are basically intended as a review of some of the literature concerning the proximity of one point process to another, and necessarily therefore also includes notions of one point process being in some sense larger or thicker than another.

Since part of the difficulty in defining the distance between point processes may be seen in defining the distance between two random variables (as distinct from processes, which are indexed families of random variables), we start with some ideas related to the distance of two random variables, then review some work of Kalzanov measuring the closeness of a stationary renewal process to a Poisson process, proceed to ideas of comparison of point processes, and finally come to a (brief) section covering some ideas on distance functions for point processes.

2. DISTANCE BETWEEN RANDOM VARIABLES

The distance between two points in $\mathbb{R}^d$ is a familiar enough idea, so that the distance between realizations of two $\mathbb{R}^d$-valued random variables (r.v.s) $X$ and $Y$, defined on some common probability space $(\Omega, \mathcal{F}, P)$ say, is sensibly given by

$$|X(\omega) - Y(\omega)|$$  \hspace{1cm} (2.1)
for each \( \omega \) in \( \Omega \), with \(|\cdot|\) here denoting the usual distance function in \( \mathbb{R}^d \). As soon as we move away from just realizations of two r.v.s and seek to measure the closeness or otherwise of the r.v.s \textit{per se}, by which we mean the measurable functions \( X, Y \) mapping \( \Omega \) into \( \mathbb{R}^d \), we are forced either to perform some operation on the random function \(|X(\omega) - Y(\omega)|\), or else to have recourse to some other notion not necessarily related to the distance of the realizations of the two functions involved.

Clearly the function

\[
d_E(X,Y) \equiv \int_\Omega |X(\omega) - Y(\omega)| P(d\omega)
\]

is an obvious candidate for measuring the distance between the r.v.s \( X \) and \( Y \). Provided the r.v.s \( X \) and \( Y \) have been given, and that \( d_E(X,Y) < \infty \), the definition (2.2) is not easily disputed. However, it is not uncommon to be given not two r.v.s \( X \) and \( Y \) but their distributions in \( \mathbb{R}^d \): for the present it will be enough to consider the case of the distribution functions (d.f.s) \( F \) and \( G \) of \( \mathbb{R}^1 \)-valued r.v.s, so that for example \( F(t) = P\{X \leq t\} \). One metric on the space of d.f.s with finite first moment is the function

\[
d_{df}(F,G) \equiv \int_{-\infty}^{\infty} |F(t) - G(t)| dt .
\]

Manipulation shows that \( d_{df}(F,G) = d_E(X^*,Y^*) \) for the r.v.s \( X^* \) and \( Y^* \), with d.f.s \( F \) and \( G \) respectively, defined by

\[
X^*(\omega) = F^{-1}(U(\omega)) , \quad Y^*(\omega) = G^{-1}(U(\omega)) \quad (2.4)
\]

where \( F^{-1} \) and \( G^{-1} \) are the functional inverses of the d.f.s \( F \) and \( G \), and \( U(\omega) \) is a r.v. on \( (\Omega,F,P) \) with the uniform
distribution, so that \( P\{U(\omega) \leq t\} = t \) for \( 0 \leq t \leq 1 \). Thus, (2.3) may be expressed as

\[
d_{df}(F,G) = d_{E}(X**,Y**) = \int_{0}^{1} |F^{-1}(u) - G^{-1}(u)| \, du .
\]  

(2.5)

When the d.f.s \( F \) and \( G \) are partially ordered in the distributional sense, \( F \leq_{d} G \), meaning that \( F(t) \geq G(t) \) (all \( t \) in \( \mathbb{R} \)), then \( X** \leq_{a.s.} Y** \), and hence

\[
d_{df}(F,G) = |EX - EY| = EY - EX .
\]

If \( F \) and \( G \) cannot be ordered in this way, or if the pair \( (X,Y) \) of r.v.s is not equivalent in distribution to \( (X**,Y**) \), then \( d_{E}(X,Y) \) also includes some measure of the dispersion of \( X \) and \( Y \).

For example, when \( F \) and \( G \) are normal d.f.s with zero mean and variances \( 1 \) and \( \sigma^2 < 1 \),

\[
d_{E}(X**,Y**) = (1 - \sigma)(2/\pi)^{1/2} .
\]

At the opposite extreme from \( (X**,Y**) = (F^{-1}(U(\omega)), G^{-1}(U(\omega))) \) is the pair \( (F^{-1}(U(\omega)), G^{-1}(1-U(\omega))) \), for which it may be shown that

\[
\int_{0}^{1} |F^{-1}(u) - G^{-1}(1-u)| \, du = 2E|Z-\xi|
\]

\[
= E|X-\xi| + E|Y-\xi|
\]

(2.6)

where the r.v. \( Z \) has \( \frac{1}{2}(F(t) + G(t)) \) as its d.f. and \( \xi \) is a median of \( Z \). Intermediate between (2.6) and (2.5), when \( X \) and \( Y \) are independently distributed r.v.s with d.f.s \( F \) and \( G \),

\[E|X-Y| \]

again vanishes (like (2.6) but unlike (2.5)) if and only if \( X = Y = \) constant a.s.: (2.5) vanishes if and only if \( X =_{d} Y \).

A major difficulty in extending the definition \( d_{E} \) or \( d_{df} \) to \( \mathbb{R}^{d} \)-valued r.v.s, relates to the need of an analogue of the construction of \( (X**,Y**) \). Consider the following (cf. also Daley (1981)).
Problem 1. Define $\mathbb{R}^2$-valued r.v.s $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ distributed uniformly over a circle of unit area with centre at the origin, and a square of unit area with centre at the origin, respectively. Amongst all r.v.s with these two specified distributions, is there a construction of r.v.s $X^*, Y^*$ such that

$$d_E(X^*, Y^*) \leq d_E(X, Y)$$

A strategy that avoids the sample path problem, given two distributions, is the use of metrics on the distributions on the state space of the realizations induced by the r.v.s. Thus we have the variation metric

$$d_V(F, G) \equiv \frac{1}{2} \int_{\mathbb{R}^d} |F(dt) - G(dt)|$$  \hspace{1cm} (2.7)

for $\mathbb{R}^d$-valued r.v.s with distributions $F$ and $G$, and in particular, for integer-valued r.v.s $X$ and $Y$,

$$d_V(X, Y) = \frac{1}{2} \sum_{i} |P\{X = i\} - P\{Y = i\}|$$

$$= P\{X^+ \neq Y^+\}$$  \hspace{1cm} (2.8)

for appropriately defined r.v.s $(X^+, Y^+)$ (and this construction is in general different from $(X^*, Y^*)$). Also, for real-valued r.v.s $X$ and $Y$, we have the supremum metric

$$d_{\sup}(X, Y) \equiv \sup_t |P\{X \leq t\} - P\{Y \leq t\}|$$  \hspace{1cm} (2.9)

the Lévy metric and others.
3. DISTANCE BETWEEN POINT PROCESSES: INTRODUCTORY COMMENTS

Realizations of point processes on $\mathbb{R}$ can be described either as counting measures or by the distances of the nearest point to the right of the origin and by the lengths of the intervals between successive points enumerated respectively on each side of that point. For the present discussion we confine attention to point processes on $\mathbb{R}_+ = (0, \infty)$, and assume without essential loss of generality that every realization has $N(\mathbb{R}_+) \geq 1$ where $N$ denotes the counting measure. We use $N, N_1, N_2, \ldots$ as generic names of point processes. The two descriptions referred to above are related via

$$S_0 = 0, \quad S_n = \inf\{t: N(0, t] \geq n\}, \quad (3.1)$$

and $X_n = S_n - S_{n-1}$ ($n = 1, 2, \ldots$): $X_1$ is the forward recurrence time interval, and $\{X_n: n = 2, 3, \ldots\}$ the successive intervals between points so long as $S_n < \infty$; otherwise we may set $X_n = \infty$.

Given two point processes $N_i$ ($i = 1, 2$), with intervals $\{X_n\}$ ($i=1$) and $\{Y_n\}$ ($i=2$), the discussion of section 2 leads to consideration as "natural" distance functions for $N_1$ and $N_2$ either functions of $|N_1(t) - N_2(t)|$, or else of the sequence $\{X_n - Y_n\}$. The simplest constructions of point processes usually proceed via interval properties rather than counting properties (though as a counter-example of this conventional construction, recall Lewis and Shedler's (1979) discussion on realizations of inhomogeneous Poisson processes). Accordingly, the distance of $N_1$ and $N_2$ may be studied via functionals of $\{X_n - Y_n\}$ where the elements $(X_n^*, Y_n^*)$ are constructed successively from the sequence
\{U^*\}_{n} \text{ of independent identically distributed (i.i.d.) r.v.s uniformly distributed on } (0,1). \text{ Such a construction for renewal processes } N_1 \text{ is sensible, and also for point processes whose intervals are stochastically monotone Markov chains or even certain semi-Markov processes (see Sonderman (1980)). However, when the successive members of the sequence } \{X_n\}_{n} \text{ are not monotonically dependent on their predecessors, there may not necessarily exist a mapping linking } \{X_n^*\}_{n} \text{ and } \{Y_n^*\}_{n} \text{ that is unequivocally the optimum one.}

4. THE MARGINAL COUNTING DISTRIBUTIONS

Let \( N_1 \) be a stationary renewal process for which
\[
E N_1(0,1) = \lambda \quad \text{and the lifetime d.f. is } F, \quad \text{so that } \{X_n : n = 2, 3, \ldots\}
\]
are i.i.d. r.v.s with d.f. \( F \) and
\[
F_1(t) = P\{X_1 \leq t\} = \lambda \int_0^t (1-F(u))du. \tag{4.1}
\]
Kal'janov (1970, 1975a, 1975b) aimed to compare \( N_1 \) with the most "typical" (?) of point processes, namely, he took a Poisson process \( N_2 \) with rate parameter \( \lambda \), so that
\[
P\{N_2(0,t] = k\} = v_k(t) = (\lambda t)^k e^{-\lambda t}/k!. \tag{4.2}
\]
His comparison was based on the probabilities \( \{v_k(\cdot)\} \) and
\[
P_k(t) = P\{N_1(0,t] = k\} = P\{S_k < t < S_{k+1}\}
= \int_0^t \left( F(k-1)^*(t-u) - F^*(t-u) \right) dF_1(u) \tag{4.3a}
= \lambda \int_0^t \left( F(k-1)^*(u) - 2F^k(u) + F(k+1)^*(u) \right) du,
\]
provided \( k \geq 1 \); for \( k = 0 \),
\[ P_0(t) = 1 - F_1(t) . \quad (4.3b) \]

Kal'žanov (1975a) claimed that
\[ \delta_k \equiv \rho_1(P_k, v_k) = \sup_t |P_k(t) - v_k(t)| \leq \begin{cases} 
\rho_1 & (k = 0) , \\
(3 + 4\lambda n(1/\rho_1))\rho_1 & (k = 1, 2, \ldots) , 
\end{cases} \quad (4.4) \]

where \( \rho_1 \equiv \sup_t |F(t) - F_1(t)| \quad (4.5) \)
so that \( \rho_1 = 0 \) if and only if \( F(t) = F_1(t) = 1 - e^{-\lambda t} \) and hence \( N_1 \) is a Poisson process. He does indeed establish (4.4) for \( k = 0, 1, 2, 3, \) and shows that \( \sup_{\lambda t \leq \lambda n(1/\rho_1)} |P_k(t) - v_k(t)| \) is bounded as asserted, but his proof for \( \lambda t > \lambda n(1/\rho_1) \) is erroneous, being based on the false assertion that for non-negative r.v.s \( X \) and \( Y \) and \( a < b \),

"\( \Pr\{a \leq X + Y < b\} \leq \Pr\{a \leq X < b\} \)" . \( \quad (*) \)

**Problem 2.** Investigate whether the bounds at (4.4) hold for \( k \geq 4 \). Can the coefficient of \( \rho_1 \) be tightened?

**Problem 3.** It was also claimed by Kal'žanov (1975a) that
\[ \delta_k \leq \begin{cases} 
\lambda\rho_2 & (k = 0) , \\
(5 + 2\lambda n(1/\lambda\rho_2))\lambda\rho_2 & (k = 1, 2, \ldots) , 
\end{cases} \quad (4.6) \]

where \( \lambda\rho_2 = \lambda \int_0^\infty |F(t) - F_1(t)| dt \). Investigate whether these bounds on \( \delta_k \) hold for \( k \geq 3 \).

Since these publications may not be so readily accessible, an outline of the (correct) part of their derivation is given later in this section (see around (4.15)).
Kalčanov's motivation is clear enough; he is comparing $N_1$ with a Poisson process when the lifetime d.f. $F$ and stationary forward recurrence time d.f. $F_1$ are close, and hence $F$ is close to the exponential d.f., because as shown in Kalčanov (1970),

$$F_1(t) - (1 - e^{-\lambda t}) = \int_0^t (F_1(t-u) - F(t-u))\lambda e^{-\lambda u} du ,$$  \hspace{1cm} (4.7)

so

$$\sup_{t} |F_1(t) - (1 - e^{-\lambda t})| \leq \rho_1 (1 - e^{-\lambda t}) \leq \rho_1 ,$$  \hspace{1cm} (4.8)

$$\int_0^t |F_1(t) - (1 - e^{-\lambda t})| dt \leq \rho_2 (1 - e^{-\lambda t}) \leq \rho_2 ,$$  \hspace{1cm} (4.9)

and

$$\sup_{t} |F(t) - (1 - e^{-\lambda t})| \leq \rho_1 + \sup_{t} |F_1(t) - (1 - e^{-\lambda t})| = 2\rho_1 .$$  \hspace{1cm} (4.10)

However, the bound (assumed true) is somewhat larger than is probably likely, because it exceeds 1 for $\rho_1 \geq 0.074778$, as in the table below:

<table>
<thead>
<tr>
<th>$\rho_1$</th>
<th>$(3 + 4\ln(1/\rho_1))\rho_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.214207</td>
</tr>
<tr>
<td>0.02</td>
<td>0.372962</td>
</tr>
<tr>
<td>0.03</td>
<td>0.510787</td>
</tr>
<tr>
<td>0.04</td>
<td>0.635020</td>
</tr>
<tr>
<td>0.05</td>
<td>0.749146</td>
</tr>
<tr>
<td>0.06</td>
<td>0.855219</td>
</tr>
<tr>
<td>0.07</td>
<td>0.954593</td>
</tr>
<tr>
<td>0.074778</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Suppose the lifetime d.f. $F$ has a finite second moment. It may then be of more interest to compare the marginal probabilities or even the simpler probabilities

$$P^0_k(t) \equiv F^{k*}(t) - F^{(k+1)*}(t) \hspace{1cm} (4.11)$$
of a renewal process \( N_{1}^{0}(\cdot) \) with \( F \) as the d.f. of \( X_1 \) as well as \( X_2, X_3, \ldots \), with the corresponding probabilities \( v_{k,\gamma}(t) \) or \( v_{k,\gamma}^{0}(t) \) of a stationary renewal or renewal process respectively with gamma distributed intervals \( Y_2, Y_3, \ldots \) matching \( X_2, X_3, \ldots \) in their first two moments. Then for moderate (to large) values of \( k \), using \( \Phi \) to denote the standard normal d.f., and \( \sigma^2 = \text{var} X_n \ (n \geq 2) \),

\[
P_{k}^{0}(t) \approx \Phi((t-k\lambda^{-1})/(\sigma k^{\frac{1}{2}})) - \Phi((t-(k+1)\lambda^{-1})/(\sigma(k+1)^{\frac{1}{2}})), \quad (4.12)
\]

which is possible for all \( t \leq 0 \), and, of course, then

\[
P_{k}^{0}(t) \approx v_{k,\gamma}^{0}(t). \quad (P_{k}^{0}(t) \text{ may have discrete jumps, but the approximation will remain valid.})
\]

Indeed it is immediately obvious that, provided \( X_n \) has a finite third moment, the Berry–Esseen bound may be applied to bounding \( |P_{k}^{0}(t) - v_{k,\gamma}^{0}(t)| \) uniformly in \( t \), \( \to 0 \) as \( k \to \infty \) like \( 0(k^{-\frac{1}{2}}) \) — though large \( k \) is needed before this is necessarily smaller than one.

Recognizing from (4.3) that for \( k \geq 1 \),

\[
P_{k}(t) = \int_{0}^{t} P_{k-1}^{0}(t-u)dF_{1}(u)
\]

and that \( P_{k}^{0}(t) = \int_{0}^{\infty} (F_{k}(t) - F_{k}(t-u))dF(u) \), \quad (4.13)

another possible approach to bounding \( P_{k}(t) \) is to use the concentration function of \( F \) which behaves like \( k^{-\frac{1}{2}} \) for increasing \( k \) (see section 2.2 of Hengartner and Theodorescu (1973)). Alternatively, smoothing inequalities on transforms may be useful (cf. Paulauskas (1971), Petrov (1975)).

For approximation purposes, observe that when \( \phi(\cdot) \) denotes the standard normal probability density function,

\[
\sup_{t} |\phi(t) - \sigma^{-1}\phi(t/\sigma)| = (1-\sigma')/\sigma' \quad (4.14)
\]
where $\sigma' = \min(\sigma, \sigma^{-1})$.

Kal'yanov's (1975a) partial derivation of (4.4) comes from (4.8) and (4.10) applied to

$$P_k(t) - v_k(t) = \int_0^t P_{k-1}(t-u)dF(u) - \int_0^t v_{k-1}(t-u)dG(u) \quad (4.15)$$

which holds for $k = 2,3,\ldots$, where we have written $G(u) = 1 - e^{-\lambda u}$, and, using convolution notation,

$$P_1(t) - v_1(t) = (F_1 - F*F_1)(t) - (G - G^{\ast 2})(t).$$

(4.7) can be written as $F_1 - G = (F_1 - F)^*G$. Then

$$P_1 - v_1 = (F_1 - F)^*G - F*F_1 + G*G = (G-F)^*(F_1+G),$$

$$|P_1(t) - v_1(t)| \leq (F_1 + 2G + G^{\ast 2})\rho_1.$$ Generally,

$$P_{k+1} - v_{k+1} = (P_k - v_k)^*F + v_k^*(F-G)$$

so

$$|P_{k+1}(t) - v_{k+1}(t)| \leq \sup_{u\geq t}|P_k(u) - v_k(u)| + (G^{\ast k} + G^{(k+1)^*})*(G^0 + G)\rho_1$$

$$\leq \rho_1(F_1 + 3G + 4G^{\ast 2} + \ldots + 4G^{\ast k} + 3G^{(k+1)^*} + G^{(k+2)^*})(t).$$

The coefficient of $\rho_1$ at (4.4) $\geq 12$ for $\rho_1 \leq 0.074778$ so it holds for $k = 1,2,3$. For all $k$ and finite $t$, the coefficient of $\rho_1$ in the bound on $P_{k+1}(t) - v_{k+1}(t)$ is at most $\rho_1(F_1(t) - G(t) + 4\lambda t) \leq \rho_1(\rho_1 + 4\lambda t)$.

Kal'yanov (1975b) endeavours to bound the difference $|P^{(1)}_k(t) - P^{(2)}_k(t)|$ of the probabilities $P^{(i)}_k(t)$ of stationary renewal processes $N_i$ ($i=1,2$) in terms of the metric $\sup_{t}\|P^{(1)}(t) - F^{(2)}(t)\|$ on the lifetime d.f.s $F^{(i)}(\cdot)$, assuming these to be absolutely continuous d.f.s with bounded densities and finite second moments. The argument
resembles that outlined above, and has the same appeal to the erroneous statement (*).

Azlarov and Kalzhanov (1976a) refers to Kalzhanov (1975a) in deriving bounds on the difference of the stationary waiting time d.f. for a GI/M/k from the same d.f. in an M/M/k queue with the same arrival and service rates. Bounds are given for each of the cases of an IFR and DFR inter-arrival d.f. in GI/M/k. Details of the proof are omitted, so it is not clear whether (*) has been used. In Azlarov and Kalzhanov (1976b) the loss probabilities in pure loss GI/M/k and M/M/k systems are compared, but only the abstract in Mathematical Reviews is available.

While for example the sequence \( \{\delta_k\} \) provides some information on all the one-dimensional marginal distributions \( \{P_k\} \) as compared with those \( \{v_k\} \) of a Poisson process - or, indeed, we could consider \( \{\delta_k\} \) for any two point processes via

\[
\sup_t \left| \Pr(N_1(0,t)=k) - \Pr(N_2(0,t)=k) \right|
\]

we may prefer, as an indicator of the "distance" between the processes \( N_1 \), some function that discounts behaviour for either large \( t \) or large \( k \) or both, such as (for appropriate \( \varepsilon > 0 \) and \( z \) in \( 0 < z < 1 \))

\[
\sum_{k=0}^{\infty} z^k \int_0^{\infty} e^{-\varepsilon t} \left| \Pr(N_1(0,t)=k) - \Pr(N_2(0,t)=k) \right| dt.
\]

Alternatively, interest in large \( k \) may be measured by for example

\[
\sup_t \| P_k^{(1)}(t) - P_k^{(2)}(t) \|
\]

Trivially, since there exist point processes with distinct probability distributions whose finite-dimensional distributions agree up to some finite order \( r \) (cf. Szasz (1970) and Oakes (1974)), no single member of \( \{\delta_k\} \), nor even the entire sequence, nor any function
based on finite dimensional distributions of bounded order, will be a
metric on the space of point processes. Of course \( \sup_{t} |F^{(1)}(t) - F^{(2)}(t)| \)
is a metric on the space of renewal processes, as these are characterized
by their lifetime d.f.s.

An entry in Mathematical Reviews refers to Bol'sakov and
Rakočić (1978) as having studied in detail Poisson and close-to-Poisson
point processes. The meaning of the phrase in the context can only be
guessed at.

5. COMPARABILITY OF POINT PROCESSES

The idea of studying the \textit{distance} of one object from another,
is an endeavour by implication to \textit{compare} the two objects. Comparisons
in stochastic analysis usually refers to the partial ordering of the
objects, and consideration of the variety of definitions proposed for
the comparability of point processes may be helpful. Let \( \prec \) denote
a partial ordering on the space \( X \) of (real-valued) r.v.s on \( (\Omega, F, P) \):
Soyan (1977) surveys some of the ideas and properties of some partial
orderings on \( X \), mostly expressed in terms of partial orderings on
functions of the probability measures \( P_{X} \) induced on \( (\Omega, F, P) \) via
\( X \in X \).

For point processes, work of Schmidt (1976) that Stoyan reviews
is covered in Whitt (1981) whose survey is the source of the definitions
below; Whitt's numerical numbering of definitions is replaced by a more
suggestive notation. The definitions are interspersed with some comments;
write \( N(0,t) = N(t) \) for brevity. Stoyan (1977) has an appealing
terminology: if \( N_{1} \prec N_{2} \) for some partial ordering \( \prec \) on point processes,
say that \( N_{1} \) is \textit{--thinner} than \( N_{2} \), or, equivalently, that \( N_{2} \) is
\( \prec \text{--thicker} \) than \( N_{1} \).
Define

\[ F(x|H_t) = \Pr\{T_{N(t)+1} - t \leq x | N(s) , 0 < s \leq t \} , \]  

(5.1)

the forward recurrence time d.f. at \( t \) conditional on the entire history \( H_t \) of \( N(\cdot) \) up to time \( t \). Then say

\[ N_1 \leq_h N_2 \text{ if } F^{(1)}(x|H_t) \leq F^{(2)}(x|H_t) \]  

(5.2)

holds for all \( x > 0 \) and all histories \( H_t \). In other words, the conditional forward recurrence time is stochastically larger in \( N_1 \) than \( N_2 \) for all realizations on \((0,t)\) and for all \( t \).

In the particular case that (say) \( N_1 \) is a renewal process with lifetime d.f. \( F \), and \( N_2 \) is a possibly inhomogeneous Poisson process with \( EN_2(0,t] = \Lambda(t) \), (5.2) implies that

\[ \sup_{0 \leq u \leq t} (F(u+x) - F(u))/(1-F(u)) \leq 1 - e^{-(\Lambda(t+x) - \Lambda(t))} . \]  

(5.3)

This condition is more easily written via the representation, valid for all \( u \) for which \( F(u) < 1 \), that \( F(u) = 1 - e^{-\mu(u)} \) for some non-decreasing function \( \mu(\cdot) \), \( \mu(u) \to \infty \text{ as } u \to \infty \). Then (5.3) is equivalent to

\[ \mu(u+x) - \mu(u) \leq \Lambda(t+x) - \Lambda(t) \quad \text{ (all } 0 \leq u \leq t) , \]

and it follows that \( \Lambda(t+0) = \infty \) for \( \inf\{t: \mu(t+0) > \mu(t-0)\} \) if such set is non-empty. This would lead to a Poisson process exploding at such \( t \), as would be the case also if \( \mu(x) \to \infty \) for \( x \to x_0 < \infty \).

Eliminating these cases, we are left with the (interesting) case that

\[ \Lambda(t) \leq Ct \quad \text{ (all } t) \text{ for some finite constant } C \geq \lim_{y \to 0} (\mu(u+y) - \mu(u))/y \equiv h(u) \quad \text{ (all } u) \]  

provided the limit exists. Indeed, it suffices to write

\[ m(t) = \sup_{0 < u \leq t} h(u) , \]  

and then we may take
\[ \Lambda(t) = \int_0^t m(u) \, du. \]

This is the smallest function \( \Lambda(\cdot) \) satisfying (5.3); Miller (1979) also showed that such a nonhomogeneous Poisson process can be constructed so that \( N_2 \) will bound the renewal process \( N_1 \) with lifetime d.f.

\[ F(x) = 1 - \exp\left(-\int_0^x h(u) \, du\right) \]

in the sense that \( N_1 \preceq_h N_2 \). He also shows that an inhomogeneous Poisson process \( N_0 \) with \( \mathbb{E} N_0(0,t) = \int_0^t (\inf_{0 \leq s \leq x} h(u)) \, dx \) satisfies \( N_0 \preceq_h N_1 \).

Returning to general point processes \( N_1 \) and \( N_2 \), suppose that their realizations satisfy

\[ N_1(x,x+dx) \leq N_2(x,x+dx) \text{ a.s.} \quad (5.4) \]

for all \( 0 < x < \infty \), i.e., every jump in the counting function \( N_1(x,\omega) \) is also a jump in \( N_2(x,\omega) \), and its size in \( N_2 \) is at least as large as in \( N_1 \). When (5.4) is satisfied, write

\[ N_1 \subseteq N_2 \quad (5.5) \]

since the definition is precisely that the sets \( \{S^{(1)}_n\} \) as defined by (3.1) satisfy such an inclusion relation. If \( N_1 \) is obtained from \( N_2 \) by any thinning operation, then clearly \( N_1 \subseteq N_2 \).

If \( N_1 \preceq_h N_2 \), then (there exists a common probability space on which \( N_1 \) and \( N_2 \) may be defined such that)* \( N_1 \subseteq N_2 \), but the converse need not hold. However, if \( N_1 \subseteq N_2 \) and both \( N_1 \) are Poisson processes, then the converse will hold, because the intensity measures \( \Lambda_1 \) will satisfy \( \Lambda_1(A) \leq \Lambda_2(A) \) for all bounded Borel sets \( A \), and for

---

*This parenthesized remark is to be understood in defining \( N_1 \subseteq N_2 \), \( N_1 \preceq \text{int} N_2 \) and \( N_1 \preceq \text{int} N_2 \).
inhomogeneous Poisson processes, \( F(x|H_t) = 1 - \exp(-\Lambda(t,t+x)) \)
independent of \( H_t \). If both \( N_1 \) are doubly stochastic Poisson 
processes for which the sample realizations of the intensity functions 
satisfy \( \Lambda_1(A,\omega) \leq \Lambda_2(A,\omega) \) a.s. for all Borel sets \( A \), then while 
we have \( N_1 \subseteq N_2 \), we need not have \( N_1 \subseteq_h N_2 \) because the "history" 
\( H_t = \{N_1(s): 0 < s \leq t\} \) is not equivalent to \( \Lambda_1(\cdot,\omega) \). This statement 
could be rephrased in measure-theoretic language, but an illustration 
is probably worth more in word-value: let \( \Lambda_1(\cdot,\omega) \) have the densities 
\( \lambda_1(\cdot,\omega) \) given by

\[
\lambda_1(t) = \begin{cases} 
0 & \text{for } t \leq 1, \\
2 & \text{for } 1 < t \leq 2, \\
1 & \text{otherwise with probability .5,}
\end{cases}
\]

\[
\lambda_2(t) = \begin{cases} 
2 & \text{for all } t \\
1 & \text{for } t \leq 2, \\
 & \text{for } t > 2 \text{ with probability .5.}
\end{cases}
\]

Then 
\[
\Pr\{N_1(2, x) = 0 | N_1(0, 2) = 0\} = \frac{(e^{-2(x-2)} + e^{-1}e^{-(x-2)})}{(1+e^{-1})}
\]

\[
\Pr\{N_2(2, x) = 0 | N_2(0, 2) = 0\} = \frac{(e^{-2(x-2)} + e^{-(x-2)})}{2},
\]

so that 
\[
F^{(1)}(x|N_1(0,2)=0) > F^{(2)}(x|N_2(0,2)=0),
\]

contrary to (5.2) if we were 
to have \( N_1 \subseteq_h N_2 \).

Particularly as inputs to queueing systems, there is much 
interest in the intervals \( \{X_n\} \) following (3.1). For \( N_1 \) defined on a 
common probability space, say that 

\[
N_1 \subseteq \text{int } N_2 \text{ if } X_{n}^{(1)} \geq X_{n}^{(2)} \text{ (all } n = 1,2,\ldots) \text{ a.s. (5.6)}
\]

In other words, the points of \( N_1 \) are more spaced out than those of \( N_2 \).

If \( N_1 \) are renewal processes with lifetime d.f.s \( F, G \) for 
which \( F(x) \leq G(x) \) (all \( x \)) (in the language of Stoyan, the d.f. \( F \) is 
larger than \( G \)), then there are realizations \( \{X_n^{(i)}\} \) for which (5.6) is
necessarily satisfied. It need not be the case that if \( N_1 \subseteq N_2 \), then \( N_1 \leq_{\text{int}} N_2 \), nor conversely, although if \( N_1 \) are renewal processes satisfying \( N_1 \subseteq N_2 \), then there will be renewal processes \( N_1' \) distributed like \( N_1 \) for which \( N_1' \leq_{\text{int}} N_2' \). In terms of models for point processes, the ordering \( \leq_{\text{int}} \) arises naturally only when the process is conveniently specified by its intervals (cf. also the comments in section 3).

It will be evident that all three orderings presented so far have the property of implying that \( N_1 \prec N_2 \) yields

\[
N_1(0,t] \leq N_2(0,t] \quad (\text{all } t) \text{ a.s.,} \tag{5.7}
\]

and when (5.7) holds, we shall say that

\[
N_1 \preceq N_2. \tag{5.8}
\]

To justify the notation as suggesting that some underlying probabilistic inequality may suffice, recall the equivalence of events

\[
\{N_1(0,x_j] \leq n_j \ (j = 1, \ldots, k)\} = \{S^{(1)}_{n_j} \geq x_j \ (j = 1, \ldots, k)\},
\]

and that the multivariate analogue of the inequality \( X \preceq_d Y \) between r.v.s \( X \) and \( Y \) is that

\[
\text{Ef}(X_1, \ldots, X_k) \leq \text{Ef}(Y_1, \ldots, Y_k) \tag{5.9}
\]

for all functions \( f \) increasing in each argument such that the expectations exist. If for all such \( f \) we have

\[
\text{Ef}(\{S^{(1)}_n\}) \geq \text{Ef}(\{S^{(2)}_n\}) \tag{5.10}
\]

then by an extension of a result of Strassen (see Chapter 17 of Marshall and Olkin (1979), or, for more detail, Kamae, Krengel and O'Brien (1977)) there is a probability space on which
\[ S^{(1)}_n \geq S^{(2)}_n \quad (\text{all } n) \quad \text{a.s.,} \quad \quad (5.11) \]

and hence (5.7) holds also.

Since (5.11) for \( n = 1 \) implies that \( X^{(1)}_1 \geq X^{(2)}_1 \) a.s., it follows immediately that for renewal processes, \( N^{(1)}_1 \leq_p N^{(2)}_2 \) implies \( N^{(1)}_1 \leq_{\text{int}} N^{(2)}_2 \), but this result need not be true in general. For example, \( N^{(1)}_1 \) may be a renewal process with lifetime d.f. \( F \), and \( N^{(2)}_2 \) an alternating renewal process with lifetime d.f.s \( G^{(1)}_1 \) and \( G^{(2)}_2 \) for which for all \( x \), \( G^{(1)}_1(x) \geq F(x) \geq G^{(2)}_2(x) \) and

\[ \int_0^x G^{(1)}_1(x-u) dG^{(2)}_2(u) \geq F^{2*}(x) ; \quad \text{this is an easy construction} \]
(take \( F(x) = G^{(2)}_2(x+a) \), \( G^{(1)}_1(x) = G^{(2)}_2(x+a+b) \) for \( 0 < b < a \)).

The last inequality Whitt lists is just the marginal distribution property that

\[ N^{(1)}_1(t) \leq_d N^{(2)}_2(t) \quad (\text{all } t) \quad . \quad \quad (5.12) \]

The quasi-Poisson processes of Szasz (1970) or Oakes (1974) afford simple examples of processes satisfying \( N^{(1)}_1 \leq_d N^{(2)}_2 \) but none of the other partial orderings. (As Whitt notes, (5.12) is not in fact a partial ordering on the space of point processes, because, as for example with the quasi-Poisson processes, we can have \( N^{(1)}_1 \leq_d N^{(2)}_2 \) and \( N^{(2)}_2 \leq_d N^{(1)}_1 \) without \( N^{(1)}_1 \) having a common (joint) distribution.) Note that if (5.12) holds, then, by the equivalence as below (5.8),

\[ S^{(1)}_n \leq_d S^{(2)}_n \quad (\text{all } n) \quad . \]

Schmidt (1976) formulated the definition (5.12), unnecessarily restricted to stationary point processes, and also formulated a distributional version of \( \leq_{\text{int}} \) for stationary point processes via the inequality

\[ E_0 f(\{X^{(1)}_n\}) \geq E_0 f(\{X^{(2)}_n\}) \quad \quad (5.13) \]
with \( f \) as at (5.13) and \( \mathbb{E}_0 \) denoting expectation with respect to the Palm probabilities, for which \( \{x_n^{(1)}\} \) are then stationary sequences. By the extension of the Strassen result, there then exist point processes \( N_1^0 \) for which \( N_1^0 \leq \text{int} \, N_2^0 \), these processes having the Palm distribution of the stationary processes \( N_1 \).

Note that if \( N_1 \) are renewal processes for which \( N_1 \leq_d N_2 \), then since

\[
\Pr\{N_1(t) = 0\} \geq \Pr\{N_2(t) = 0\},
\]

we must also have \( N_1 \leq \text{int} \, N_2 \). But if the \( N_1 \) are stationary renewal processes, neither \( \leq \text{int} \) nor \( \leq_d \) need imply the other, as Schmidt showed.

In subsequent work, Whitt (1981b) has taken a more pragmatic view and compared some point processes numerically by studying how different processes may have different effects on a queueing system when fed as inputs into the system. More work of this nature has been done by Whitt and his colleagues.

### 6. DISTANCES BETWEEN POINT PROCESSES

For the time being we pass over the comparisons \( \leq_h \) and \( \subset \) for suggesting distances between point processes. Coming to \( \leq_{\text{int}} \), and recalling (2.3), we could use

\[
d_{\text{int,1}}(N_1, N_2) \equiv \sup_{n \geq 1} \int_0^\infty \left| \Pr\{S_n^{(1)} > t\} - \Pr\{S_n^{(2)} > t\} \right| dt \quad (6.1)
\]

where the factor \( n^{-1} \) has been introduced to average out the effect of \( S_n^{(1)} - S_n^{(2)} \) being asymptotically like \( n(EX_1^{(1)} - EX_1^{(2)}) \) if the sequences \( \{x_n^{(1)}\} \) are stationary. However, if \( N_1 \leq_d N_2 \) and \( N_2 \leq_d N_1 \), as at (5.11) and below, we should have \( d_{\text{int,1}}(N_1, N_2) = 0 \), for (6.1) depends...
only on the marginal distributions of $N^{(1)}(t)$, not the joint
distribution over a set of values of $t$.

$$d_E(N_1, N_2) = \sup_{t > 0} |E[N_1(t) - E[N_2(t)]/(t+1)$$  \hspace{1cm} (6.2)$$

has a similar drawback. However

$$d_{E^*}(N_1, N_2) = \sup_{t > 0} E|N_1(t) - N_2(t)|/(t+1)$$  \hspace{1cm} (6.3)$$
does not have this drawback: instead we are confronted with the
problem of requiring $N_1$ and $N_2$ to be defined on the same
probability space, and, as noted in section 2, the appropriate
construction is far from evident as soon as we dispense with renewal
processes. As an alternative to (6.3), and one that is better suited
to processes constructed by the interval properties, we have

$$d_{int}(N_1, N_2) = \sup_{n \geq 1} n^{-1}E|S^{(1)}_n - S^{(2)}_n|$$  \hspace{1cm} (6.4)$$

For renewal processes, defined via sequences $\{(X^{(1)*}_n, X^{(2)*}_n)\}$
as in section 3, $d_{int} = d_{int,1}$, although the proof is not immediately
obvious.

Since renewal processes $N_1$ are characterized by their
renewal functions $U_1(\cdot)$, we could use, as an alternative to (6.2)
(but only for renewal processes) the function

$$\sup_{t \geq 0} |U_1(t) - U_2(t)|/(t+1)$$  \hspace{1cm} (6.5)$$

Matthes, Kerstan and Mecke (1978) (e.g. in their Sections
1.9 and 1.12) use the variation distance for point processes, and
T.C. Brown (private correspondence) has written of using compensators
to obtain total variation results for the distance of point processes.
from Poisson processes. I have been told of an unpublished review by R.L. Farrell that may have relevance to this section. Mark Brown (1981 and correspondence) has compared renewal processes with decreasing failure rate (DFR): in the notation of section 2, he has shown that for distributions with the weaker property of increasing mean residual life,

\[ \max(\rho_1(F,F_1), \rho_1(F,G), \rho_1(F_1,G)) \leq 1 - 2(EX)^2/EX^2; \quad (6.6) \]

for DFR d.f.s it is easily checked from a sketch of the tails of the d.f.s involved and the results of section 2, that the maximum on the left-hand side of (6.6) is in fact \( \rho_1(F,F_1) \). He has also studied metrics in terms of the hazard functions being close.

We have not attempted to assess the effect of common operations on point processes (thinning, superposition, translation, cluster-formation, etc.) on the distance functions given above.

ACKNOWLEDGEMENT I am greatly indebted to Ross Leadbetter for his interest.

REFERENCES


Probab. 14 (1977), 7-12).

PETROV, V.V. (1975) Sums of independent Random Variables
Springer-Verlag, New York. (= Translation from original

SCHMIDT, V. (1976) On the non-equivalence of two criteria of
comparability of stationary point processes. Zast. Mat.
15, 33-37.


STOYAN, D. (1977) Qualitative Eigenschaften und Abschätzungen
stochastischer Modelle. Akademie-Verlag, Berlin.


------ (1981b) Approximating a point process by a renewal process,

Distances of Random Variables and Point Processes

D.J. Daley

Office of Naval Research
Statistics and Probability Program (Code 436)
Arlington, VA 22217

March 1981

point processes, renewal processes, distance functions, metrics, comparability of processes

The paper is basically a discussion and survey of some literature concerning the concept of a distance function for random variables and point processes. Work of Kalčanov on the closeness of renewal processes to Poisson processes is outlined, and there are comments on various possible comparison methods for point processes as given by Whitt.