The Annals of Probability
1975, Vol. 3, No. 6, 909–929

SPECIAL INVITED PAPER

THE IDENTIFICATION OF POINT PROCESS SYSTEMS[1]

BY DAVID R. BRILLINGER

The University of California, Berkeley

A point process system is a random operator assigning a nonnegative integer-valued measure to a random nonnegative integer-valued measure. We define certain parameters for such a system and discuss the problem of estimating these parameters. We also consider the related problem of measuring the degree of association of two point processes.

1. Introduction and summary. A (stochastic) *point process* M is a random nonnegative integer-valued measure. If a point process M influences an apparatus $\mathscr{S}$ (perhaps real, perhaps conceptual, typically incorporating stochastic features), to give rise to another point process N, we write

$$N = \mathscr{S}[M]$$

and say that the point process N is the output of the *system* $\mathscr{S}$ operating on the input process M. We write $M(A)$ to denote the measure of the time interval A for a realization of the input process and $N(A)$ the corresponding measure for N. In practice $M(A)$ refers to the number of occurrences in A of some phenomenon of interest and $N(A)$ to the corresponding number of occurrences of some second phenomenon. We illustrate with two examples, one specific, the other more vague.

EXAMPLE 1. Let M have single points (corresponding to isolated occurrences) located at σ_j, $j = 0, \pm 1, \cdots$ and suppose that γ_j are real-valued random variables. Then

$$N(A) = \#\{j : \sigma_j + \gamma_j \text{ in } A\}$$

(i.e., $N(A)$ denotes the number of points σ_j which when moved by γ_j lie in the set A) defines a point process system. This particular system is called a *random translation* or *motion*.

EXAMPLE 2. The pulse discharges of many nerve cells have large amplitudes and are of short duration, so that they can be conveniently described as a point process. If we take two nerve cells that have a certain physiological configuration (e.g. proximity, or electrically connected), then it may be the case that the

Received January 3, 1975.

[1] Presented at the 1974 Annual Meeting of the Institute of Mathematical Statistics held at Edmonton, Alberta, August 13–16. Part of the research carried out as a Miller Research Professor, part with the support of NSF Grant GP-31411.

AMS 1970 *subject classifications.* Primary 60G10, 10K35, 62M10, 93B30.

Key words and phrases. Identification, system, point process, stationary.

point process M of pulses from one cell influence the point process N of pulses emitted by another cell. Beyond postulating that $N = \mathscr{S}[M]$, we may have little notion of the system operator $\mathscr{S}$ until investigation is undertaken. We discuss such a problem in Section 7.

We say that the system is *deterministic* if $\mathscr{S}$ incorporates no random feature. Its input, M, may of course still be random. We say that the system is *time invariant* when the bivariate process (M, N) is stationary for stationary M.

By the problem of the *identification* of a point process system we shall mean that of determining characteristics of the system from observations of inputs and corresponding outputs. In the case that the system, $\mathscr{S}$, is stochastic, the most that we can hope for is to determine average quantities or parameters that characterize the statistical properties of $\mathscr{S}$. Complete identification is not possible in general.

In Section 2 we define certain parameters of stochastic point processes. In Section 3 we set down a number of useful parameters for point process systems and indicate how they might be estimated. In Section 4 we discuss the related problem of measuring the degree of association of two point processes. In Section 5 we consider the identification of systems having multidimensional input or output. The problem of identification is sometimes taken to be that of determining an estimate of a finite dimensional parameter that characterizes the behavior of a process or system. In Section 6 we present one approach to this problem. The final Section, 7, presents some results concerning the identification of the point process system corresponding to a nerve cell with a single input nerve fibre.

We do not discuss the interesting problems of "on line" (or recursive) identification, of the identification of systems with feedback, nor of special procedures for realizable systems. We do not give specific references to well-known results. These may be found in Bartlett (1963), Cox and Lewis (1966), Lewis (1972).

2. Stochastic point process parameters. Before discussing specific identification procedures, we must first introduce certain parameters that describe stochastic point processes. We will restrict ourselves to parameters of stationary bivariate processes with isolated points. (In the point process literature, such processes are referred to as orderly.)

Let (M, N) be a stationary bivariate point process on the real line with differential increments at time t given by $\{dM(t), dN(t)\} = \{M(t, t + dt], N(t, t + dt]\}$. The *mean intensity*, p_M, of the process M is defined by

$$E\{dM(t)\} = p_M \, dt \,. \tag{2.1}$$

Because the points of the process have been assumed to be isolated, expression (2.1) may be interpreted as

$$\text{Prob}\,\{M \text{ point in } (t, t + dt]\}\,.$$

The mean intensity, p_N, of the N process is defined in a similar manner.

The *second-order cross product density* at lag u, $p_{NM}(u)$, is defined by

$$E\{dN(t+u)\,dM(t)\} = p_{NM}(u)\,du\,dt\,, \qquad u \neq 0\,. \tag{2.2}$$

Expression (2.2) may also be interpreted as giving

$$\text{Prob}\,\{N \text{ point in } (t+u, t+u+du] \text{ and } M \text{ point in } (t, t+dt]\}\,.$$

The other second-order product densities, $p_{MM}(u)$, $p_{NN}(u)$ are defined through (2.2) by equating M and N.

These parameters may be used to define the *conditional mean intensity*

$$E\{dN(t+u)\,|\,M\{t\} = 1\} = p_{NM}(u)\,du/p_M\,, \qquad u \neq 0\,, \tag{2.3}$$

which may be interpreted as

$$\text{Prob}\,\{N \text{ point in } (t+u, t+u+du]\,|\,M \text{ event at } t\}\,.$$

As $|u| \to \infty$, the increments $dN(t+u)$ and $dM(t)$ are tending to become independent for many processes. This phenomenon leads to the definition of the *cross-covariance density*

$$q_{NM}(u) = p_{NM}(u) - p_N p_M\,, \qquad u \neq 0 \tag{2.4}$$

which tends to 0 as $|u| \to \infty$. The autocovariance densities, $q_{MM}(u)$, $q_{NN}(u)$ are defined similarly.

Provided M points and N points do not occur simultaneously we can write

$$dC_{NM}(u)\,dt = \text{Cov}\,\{dN(t+u), dM(t)\} = q_{NM}(u)\,du\,dt\,.$$

However in the case of the components themselves we must write

$$dC_{MM}(u)\,dt = \text{Cov}\,\{dM(t+u), dM(t)\} = (\delta(u) + q_{MM}(u))\,du\,dt$$
$$dC_{NN}(u)\,dt = \text{Cov}\,\{dN(t+u), dN(t)\} = (\delta(u) + q_{NN}(u))\,du\,dt\,,$$

where $\delta(\cdot)$ is the Dirac delta function, to take account of the singularity at $u = 0$.

The *cross-spectrum* of the two process at frequency λ, $f_{NM}(\lambda)$, is now defined by

$$\begin{aligned} f_{NM}(\lambda) &= (2\pi)^{-1} \int \exp\{-iu\lambda\}\,dC_{NM}(u) \\ &= (2\pi)^{-1} \int \exp\{-iu\lambda\}\,q_{NM}(u)\,du \end{aligned} \tag{2.5}$$

for $-\infty < \lambda < \infty$, provided the integral exists. The *power spectrum* of the process M, $f_{MM}(\lambda)$, is defined by

$$\begin{aligned} f_{MM}(\lambda) &= (2\pi)^{-1} \int \exp\{-iu\lambda\}\,dC_{MM}(u) \\ &= (2\pi)^{-1}p_M + (2\pi)^{-1} \int \exp\{-iu\lambda\}q_{MM}(u)\,du \end{aligned} \tag{2.6}$$

with a similar definition for $f_{NN}(\lambda)$.

We may continue in the previous manner and define higher-order parameters such as the third-order product density

$$p_{MMM}(u, v) = E\{dM(t+u)\,dM(t+v)\,dM(t)\}/dt\,du\,dv\,, \tag{2.7}$$
$$u \neq v,\ u \neq 0,\ v \neq 0\,,$$

the third order cumulant density

(2.8) $$q_{MMM}(u, v) = \text{cum}\,\{dM(t+u), dM(t+v), dM(t)\}/dt\,du\,dv\,, \qquad u \neq v,\ u \neq 0,\ v \neq 0\,,$$

and even higher-order spectra, see Brillinger (1972).

The parameters defined in (2.1), (2.2), (2.3), (2.7) have the advantage, over corresponding parameters defined in the case of ordinary time series, of possessing a further interpretation as probabilities.

Given a segment $\{M(0, t], N(0, t]\}$, $0 < t \leqq T$, of a realization of an M, N process satisfying some regularity conditions, each of the parameters defined above may be estimated consistently as $T \to \infty$, and the asymptotic distributions of the estimates are known, see Cox and Lewis (1966, 1972) and Brillinger (1972, 1975b). Estimates of third-order densities are given in Brillinger (1975a). In this section, like Bartlett (1963), we have eschewed the mathematical problems about existence of mean intensities, autocovariance density functions, etc. Lewis (1972) contains papers concerned with these issues.

3. System parameters and system identification. Suppose that we are dealing with a time invariant system with input process M and output process N. A key element of the character of such a system is provided by

$$\text{Prob}\,\{N \text{ point in } (t, t+dt] \mid M\} \sim E\{dN(t) \mid M\}\,.$$

In connection with it we suppose

(3.1) $$\lim_{h\downarrow 0} \text{Prob}\,\{N \text{ point in } (t, t+h] \mid M\}/h = \mu_M(t)$$

for given input process $M \in \mathscr{N}$. Let us discuss plausible forms for $\mu_M(t)$ for a succession of input processes.

(i) Suppose we take as input to the system $M(\cdot) \equiv 0$, that is no input events. Then we might be willing to assume that $\mu_M(t)$ exists and is equal to a constant,

(3.2) $$\mu_M(t) = s_0\,.$$

The system is here assumed to be emitting points at rate s_0.

(ii) Next, suppose we take as input to the system, M corresponding to a single event at time σ. Then we might alter (3.2) to

(3.3) $$\mu_M(t) = s_0 + s_1(t-\sigma) = s_0 + \textstyle\int s_1(t-u)\,dM(u)\,.$$

$s_1(t)$ represents the effect, on the output intensity, of inputting a single point at time 0. For example, in a service system with service time density $g(t)$, we have (3.3) with $s_0 = 0$, $s_1(t) = g(t)$.

(iii) Suppose next we take as input to the system, M corresponding to points at times σ_1 and σ_2. Were there no interaction of the two points we might be prepared to write (3.1) as

(3.4) $$\mu_M(t) = s_0 + s_1(t-\sigma_1) + s_1(t-\sigma_2) = s_0 + \textstyle\int s_1(t-u)\,dM(u)\,.$$

For example if the service system has 2 or more servers, then (3.4) holds with $s_0 = 0$, $s_1(t) = g(t)$.

If there were an interaction, then we might write (3.1) as

$$\begin{aligned}(3.5)\qquad \mu_M(t) &= s_0 + s_1(t-\sigma_1) + s_1(t-\sigma_2) + s_2(t-\sigma_1, t-\sigma_2)\\ &= s_0 + \int s_1(t-u)\,dM(u) + \iint_{u\neq v} s_2(t-u, t-v)\,dM(u)\,dM(v)\end{aligned}$$

where the function $s_2(\cdot)$ gives the effect of the interaction. If the service system above has but 1 server, then (3.1) has the form

$$\mu_M(t) = g(t-\sigma_1) + \int_{\sigma_2}^t g(v-\sigma_1)g(t-v)\,dv\,, \qquad \sigma_1 < \sigma_2 < t\,,$$

which is of the form of (3.5).

(iv) It is now evident that we may proceed in a recursive manner building up a succession of models for (3.1) of the form

$$\begin{aligned}(3.6)\qquad \mu_M(t) = s_0 + \sum_{k=1}^K \int \cdots \int_{u_1,\cdots,u_k;\text{distinct}} s_k(t-u_1, \cdots,\\ t-u_k)\,dM(u_1)\cdots dM(u_k)\end{aligned}$$

where the function $s_K(t-\sigma_1, \cdots, t-\sigma_K)$ may be interpreted as the interaction effect at time t when the input process consists of K events at times $\sigma_1, \cdots, \sigma_K$. The expansion of (3.6) is a point process analog of the Volterra expansions considered in Wiener (1958) for Gaussian processes.

We shall say that the system is *linear* when $K = 1$ in (3.6), that is

$$(3.7)\qquad \lim_{h\downarrow 0} \text{Prob}\,\{N \text{ event in } (t, t+h]\,|\,M\}/h = s_0 + \int s_1(t-u)\,dM(u)\,.$$

By analogy with the terminology of the ordinary time series case, we might call $s_1(\cdot)$ in (3.7), the *average impulse response* of the system. We remark that (3.7) is an average property of the system, not a sample path property. We say that the system is *realizable* or *causal*, when $s_1(u) = 0$ for $u < 0$.

EXAMPLE 3. *The $G/G/\infty$ queue.* Suppose that the jth customer of a service facility arrives at time σ_j and experiences service time γ_j, $j = 0, \pm 1, \cdots$. Suppose that the γ_j are random variables with density function $g(u)$. Then, symbolically,

$$dN(t) = (\textstyle\sum_j \delta(t-\sigma_j-\gamma_j))\,dt\,,$$

and so

$$\begin{aligned}E\{dN(t)\,|\,M\} &= (\textstyle\sum_j \int \delta(t-\sigma_j-\gamma)g(\gamma)\,d\gamma)\,dt\\ &= (\textstyle\sum_j g(t-\sigma_j))\,dt\\ &= (\textstyle\int g(t-u)\,dM(u))\,dt\,.\end{aligned}$$

This is of the form of (3.7) with $s_0 = 0$, $s_1(u) = g(u)$ and an example of the random translation of Example 1.

EXAMPLE 4. *A Hawkes' process.* Suppose the system may be described by

$$(3.8)\qquad \mu_M(t) = \mu + \int_{-\infty}^t a(t-u)\,dN(u) + \int_{-\infty}^t b(t-u)\,dM(u)$$

and that it generates a stationary process $N(\cdot)$ when a stationary $M(\cdot)$ is taken as input. Expression (3.8) leads to

$$\mu_M(t) = \nu + \int_{-\infty}^{t} c(t-u)\, dM(u)$$

where $p_N = \mu + A(0)p_N + B(0)p_M = \nu + C(0)p_M$, $C(\lambda) = (1 - A(\lambda))^{-1}B(\lambda)$; $A(\cdot)$, $B(\cdot)$, $C(\cdot)$ being the Fourier transforms of $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ respectively; see Hawkes (1972) for further details and references.

EXAMPLE 5. For some $\alpha, \Delta > 0$

$$\mu_M(t) = \alpha M(t - \Delta, t]$$

The output intensity is here assumed to be proportional to the number of input points in the immediately previous time interval of length Δ. This model has the form of (3.7) with $s_0 = 0$,

$$\begin{aligned} s_1(u) &= \alpha\,, \qquad 0 \leqq u < \Delta \\ &= 0\,, \qquad \text{otherwise.} \end{aligned}$$

We now turn to the problem of identifying the linear system (3.7). Provided the process $N(\cdot)$ is well-defined, the relationship (3.7) leads to the equalities,

$$\begin{aligned} p_N &= s_0 + p_M \int s_1(u)\, du \\ p_{NM}(t) &= s_0 p_M + s_1(t)p_M + \int s_1(t-u)p_{MM}(u)\, du \\ q_{NM}(t) &= s_1(t)p_M + \int s_1(t-u)q_{MM}(u)\, du \\ f_{NM}(\lambda) &= S_1(\lambda)f_{MM}(\lambda) \end{aligned} \tag{3.9}$$

where $S_1(\cdot)$ is the Fourier transform of $s_1(\cdot)$. These relations suggest the estimates

$$\begin{aligned} \hat{S}_1(\lambda) &= \hat{f}_{NM}(\lambda)\hat{f}_{MM}(\lambda)^{-1} \\ \hat{s}_0 &= \hat{p}_N - \hat{p}_M \hat{S}_1(0) \\ \hat{s}_1(t) &= (2\pi)^{-1} \int \hat{S}_1(\lambda) \exp\{it\lambda\}\, d\lambda \end{aligned}$$

where $\hat{p}_M, \hat{p}_N, \hat{f}_{MM}(\lambda), \hat{f}_{NM}(\lambda)$ are estimates of $p_M, p_N, f_{MM}(\lambda), f_{NM}(\lambda)$. Details of this estimation procedure may be found in Brillinger (1974). An example of its use with neurophysiological data is given in Section 7 of this paper.

An alternate identification procedure that may be of use in certain situations is the following. Suppose that it is known that $s_1(u)$ vanishes for $|u| > \Delta$. Suppose that the input points are spaced farther than 2Δ apart. Then the individual terms of

$$\mu_M(t) = s_0 + \int s_1(t-u)\, dM(u) = s_0 + \sum_j s_1(t - \sigma_j)$$

do not interfere. This suggests that $s_0 + s_1(u)$ can be estimated, reasonably, by an expression such as

$$\#\{|\tau_k - \sigma_j - u| < \beta\}/(2\beta M(0, T])$$

for some small β, where the τ_k denote the times of observed output events from the system. This estimate is suggested by first principles. It is also suggested by the second equation of expression (3.9) as $p_{MM}(u) = 0$ for $|u| \leqq 2\Delta$ here.

Even when the model (3.7) is not satisfied, the function $s_1(\cdot)$ satisfying (3.9) is of some interest. It provides the best linear mean-squared error predictor of the process N based on M. The relations (3.9), most especially the third, suggest that the simplest way to identify the system is to take Poisson noise as input to the system, for then $q_{MM}(u) = 0$ identically, and so $s_1(t) = q_{NM}(t)/p_M$. Finally we remark that (3.8) gives an answer to the interesting question of what sort of input behavior is most likely to lead to an output point, say at 0. We see that the increments $dM(t)$ should mimic the shape of $a(-t)$.

The above discussion indicates that, provided one has sufficient data, a linear point process system may be identified fairly directly. Unfortunately things are not so nice in the nonlinear case. Consider the model (3.6) with $K = 2$. It is convenient to set it down in an alternate form. With $M'(u) = M(u) - up_M$, we write it as

$$(3.10)\qquad \mu_M(t) = r_0 + \int r_1(t-u)\,dM'(u) + \int\int_{u\neq v} r_2(t-u, t-v)\,dM'(u)\,dM'(v)\,.$$

Supposing $r_2(u, v) = r_2(v, u)$, expression (3.10) leads to

$$\begin{aligned}
p_N &= r_0 + \int\int r_2(-u, -v)q_{MM}(u-v)\,du\,dv \\
q_{NM}(t) &= r_1(t)p_M + \int r_1(t-u)q_{MM}(u)\,du + 2\int r_2(t, t-v)q_{MM}(v)\,dv \\
&\quad + \int\int r_2(t-u, t-v)q_{MMM}(u, v)\,du\,dv \\
q_{NMM}(s, t) &= r_1(s-t)q_{MM}(0) + r_1(s)q_{MM}(t) + \int r_1(s-u)q_{MMM}(u, t)\,du \\
&\quad + 2r_2(s, s-t)p_M^2 + 2\int r_2(s-u, s-t)q_{MM}(u)\,du \\
&\quad + 2\int r_2(s, s-t-v)q_{MM}(v)\,dv \\
&\quad + 2\int\int r_2(s-u, s-t-v)q_{MM}(u)q_{MM}(v)\,dv\,du \\
&\quad + 2\int r_2(s-u, s)q_{MMM}(u, t)\,du + 2\int r_2(s-u, s-t)q_{MMM}(u, t)\,du \\
&\quad + \int\int r_2(s-u, t-v)q_{MMMM}(u, v, t)\,du\,dv\,.
\end{aligned}$$

It is not at all apparent how we could make direct use of these relationships without making further assumptions. We do note that if $q_{MM}(u)$, $q_{MMM}(u, v)$, $q_{MMMM}(u, v, w)$ are all identically 0, as would be the case for a process with independent increments, such as the Poisson, then the relationships give

$$\begin{aligned}
r_0 &= p_N \\
r_1(u) &= q_{NM}(u)/p_M \\
r_2(u, v) &= q_{NMM}(u, u-v)/(2p_M^2)
\end{aligned}$$

and the functions r_1 and r_2 may be identified directly. The above discussion suggests that we should probe a point process system with Poisson noise whenever possible. Unfortunately in practise this is often not possible because the noise generating device has a "dead time", that is a nonnegligible minimum interval between points. Other procedures for identifying polynomial systems involving ordinary time series are given in Brillinger (1970). It is not presently clear if these may be adapted to the point process case usefully. The Fourier-Hermite orthogonal polynomials discussed there for Gaussian processes (and introduced into the context of system identification by Wiener (1958)) could be replaced by the Poisson–Charlier polynomials (see Hida (1970)) for Poisson noise.

An alternate nonlinear model for the conditional intensity (3.1) is the multiplicative model

$$\mu_M(t) = \beta \prod_j b(t - \sigma_j)$$
$$= \exp\{\alpha + \int a(t - u)\, dM(u)\}$$

with $\alpha = \log \beta$, $a(u) = \log b(u)$. If we expand the exponential, then we see that this corresponds to the model (3.6) with $K = \infty$. In the case that M is Poisson, this model leads to the relationships

$$p_N = \exp\{\alpha + p_M \int [b(u) - 1]\, du\}$$
$$p_{NM}(u) = p_M p_N b(u) .$$

Another nonlinear model of some interest is provided by

$$\mu_M(t) = \alpha \quad \text{if} \quad M(t - \Delta, t] \geqq k$$
$$= 0 \quad \text{otherwise}$$

for some $\Delta > 0$. An output point occurs here only if there are at least k input events in the previous time interval of length Δ.

So far we have only discussed models for the first-order system parameter (3.1). A related second-order parameter is the following,

$$\text{(3.11)} \qquad \mu_M(s, t) = \lim_{h \downarrow 0} \text{Prob}\{N \text{ points in } (s, s + h_1] \text{ and } (t, t + h_2] \mid M\}/(h_1 h_2) .$$

This parameter would be especially useful were input points stimulating pairs of output points. From what has been said already we might consider modelling (3.11) by

$$r_0(s - t) + \int r_1(s - u, t - u)\, dM'(u)$$

where $M'(u) = M(u) - up_M$, $r_0(-u) = r_0(u)$, $r_1(s, t) = r_1(t, s)$. This model leads to the relationships

$$p_N = \lim_{u \to \infty} r_0(u)$$
$$p_{NN}(u) = r_0(u)$$
$$p_{NNM}(s, t) - p_{NN}(s - t)p_M = p_M r_1(s, t) + \int r_1(s - u, t - u) q_{MM}(u)\, du .$$

Denoting the Fourier transform of r_1 by R_1 and of the left hand side of the last expression by P, we see that

$$R_1(\lambda_1, \lambda_2) = P(\lambda_1, \lambda_2)/(2\pi f_{MM}(\lambda_1 + \lambda_2)) .$$

We end this section by mentioning that there is a growing literature concerning a martingale approach to point processes. (See Segall et al. (1975), Segall and Kailath (1975), Boel et al. (1973), Van Schuppen and Wong (1974), Dolvio (1974), for example.) It makes use of the Doob–Meyer decomposition of submartingales and results of Doléans-Dadé and Meyer (1970) among other things. It is concerned with formalizing representations of the form

$$N(0, t] = A(t) + w(t)$$

where $A(t)$, $w(t)$ are respectively a predictable increasing process and a zero mean martingale on some σ-algebras $\mathscr{B}_t \subset \sum_{\{N(s);s\leq t\}}$. The cases where $A(t)$ is differentiable are analagous to our assumption of (3.1). The topics covered in the literature include: detection, control, forecasting, likelihood ratios and the representation of martingales in the basic process.

4. The measurement of association. A problem of some interest is the measurement of the degree of interdependence of two point processes. This involves addressing ourselves to the question of whether the input to a point process system affects the output at all and if it does to what degree?

We begin by noting that

$$\begin{aligned}\operatorname{corr}\{dN(t+u), dM(u)\} &= (p_{NM}(u)\,dt\,du - p_N p_M\,dt\,du)/(p_N\,dt\,p_M\,du)^{\frac{1}{2}} \\ &\propto p_{NM}(u) - p_N p_M\,.\end{aligned}$$

This remark suggests our considering the measure $p_{NM}(u) - p_N p_M$. This particular measure may also be interpreted as

$$\frac{\operatorname{Prob}\{dN(t+u)=1 \text{ and } dM(t)=1\} - \operatorname{Prob}\{dN(t+u)=1\}\operatorname{Prob}\{dM(t)=1\}}{dt\,du}\,.$$

An equivalent measure is

$$\frac{p_{NM}(u)}{p_M} - p_N = (\operatorname{Prob}\{dN(t+u)=1 \mid dM(t)=1\} - \operatorname{Prob}\{dN(t+u)=1\})/dt\,.$$

Both of these measures are 0 in the case of independence.

The problem can also be viewed as one of looking for association in the 2×2 table:

$dN(t+u)$	$dM(t)$		Totals
	0	1	
0	1	$p_M\,dt$	1
1	$p_N\,du$	$p_{NM}(u)\,dt\,du$	$p_N\,du$
Totals	1	$p_M\,dt$	

A variety of measures of association have been suggested for 2×2 tables, see pages 536–540 in Kendall and Stuart (1961). In the present context, these lead to

(i) the *cross-product ratio*

$$\alpha(u) = p_{NM}(u)/[p_N p_M]\,,$$

(ii) *Yule's coefficient of association*

$$Q(u) = [p_{NM}(u) - p_N p_M]/[p_{NM}(u) + p_N p_M],$$

(iii) *Yule's coefficient of colligation*

$$Y(u) = [(p_{NM}(u))^{\frac{1}{2}} - (p_N p_M)^{\frac{1}{2}}]/[(p_{NM}(u))^{\frac{1}{2}} + (p_N p_M)^{\frac{1}{2}}],$$

(iv) *Pearson's* ϕ^2

$$\phi^2(u) = [p_{NM}(u) - p_N p_M]^2/[p_N p_M].$$

The "null" values of these measures occur in the case that $p_{NM}(u) = p_N p_M$.

An alternate manner in which to proceed is to look at the degree of correlation of certain combinations of the values of the process. For example if we set

$$d_M{}^T(\lambda) = \int_0^T \exp\{-i\lambda t\}\, dM(t), \qquad d_N{}^T(\lambda) = \int_0^T \exp\{-i\lambda t\}\, dN(t)$$

then

$$\begin{aligned} \lim_{T\to\infty} |\text{corr}\,\{d_M{}^T(\lambda), d_N{}^T(\lambda)\}|^2 &= \lim_{T\to\infty} \frac{|\text{Cov}\,\{d_M{}^T(\lambda), d_N{}^T(\lambda)\}|^2}{\text{Var}\, d_M{}^T(\lambda)\ \text{Var}\, d_N{}^T(\lambda)} \\ &= |f_{MN}(\lambda)|^2/|f_{MM}(\lambda) f_{NN}(\lambda)|^2 \qquad (4.1) \\ &= |R_{MN}(\lambda)|^2. \end{aligned}$$

This last measure is called the *coherence* of the two processes at frequency λ. Its values lie between 0 and 1, with 0 occurring in the case of independence.

5. Multidimensional systems. So far we have been considering the case in which the system has a single input and a single output. In many interesting situations, the input and output processes are multidimensional. No great difficulties appear in extending the linear system of (3.7) to the multidimensional case. Specifically, we might postulate

$$\lim_{h\downarrow 0} E\{\mathbf{N}(t, t+h]/h \mid \mathbf{M}\} = \mathbf{s}_0 + \int \mathbf{s}_1(t-u)\, d\mathbf{M}(u) \qquad (5.1)$$

with the process $\mathbf{M}$ being r dimensional, the process $\mathbf{N}$ being s dimensional, $\mathbf{s}_0$ being an s vector and $\mathbf{s}_1(\cdot)$ being an $s \times r$ matrix. If $\mathbf{S}_1$ denotes the Fourier transform of $\mathbf{s}_1$, if $\mathbf{f}_{MM}(\lambda)$ denotes the spectral density matrix of the process $\mathbf{M}$ and if $\mathbf{f}_{NM}(\lambda)$ denotes the cross-spectral density matrix of the two processes, then the relation (5.1) leads to the equality $\mathbf{f}_{NM}(\lambda) = \mathbf{S}_1(\lambda)\mathbf{f}_{MM}(\lambda)$, showing that the system may be identified through estimating spectral density matrices.

In the multidimensional case we may be interested in certain partial parameters. Consider a univariate process M and a bivariate process $\mathbf{N}$ with component N_1 and N_2, corresponding to M being input to two systems $\mathscr{S}_1$ and $\mathscr{S}_2$ with outputs N_1 and N_2 respectively. In practise, the outputs N_1 and N_2 may appear to be related. However this association may only be due to the fact that the two systems had the same input M and not due to any further connection. Partial spectra provide a tool for checking into this possibility. Consider the model

$$dN_1(t) = (\mu_1 + \int a_1(t-u)\, dM(u))\, dt + d\varepsilon_1(t)$$
$$dN_2(t) = (\mu_2 + \int a_2(t-u)\, dM(u))\, dt + d\varepsilon_2(t)$$

where ε_1 and ε_2 are processes with stationary increments. This model leads to the relationships

$$\begin{aligned} f_{\varepsilon_j\varepsilon_k}(\lambda) &= f_{N_jN_k\cdot M}(\lambda) \\ &= f_{N_jN_k}(\lambda) - f_{N_jM}(\lambda)f_{MN_k}(\lambda)/f_{MM}(\lambda) \end{aligned}$$

for $j, k = 1, 2$. In the case that the processes ε_1 and ε_2 are uncorrelated the partial cross-spectrum $f_{N_1N_2\cdot M}(\lambda)$ and consequently the partial coherence

$$|R_{N_1N_2\cdot M}(\lambda)|^2 = |R_{\varepsilon_1\varepsilon_2}(\lambda)|^2 = |f_{\varepsilon_1\varepsilon_2}(\lambda)|^2/[f_{\varepsilon_1\varepsilon_1}(\lambda)f_{\varepsilon_2\varepsilon_2}(\lambda)] \tag{5.2}$$

will be identically 0 allowing an examination of the hypothesis through estimates of these functions. An example of the checking of such a hypothesis for some neurophysiological data is given in Section 7.

6. Finite parameter models. On occasion we may find ourselves in a situation where a system of interest is characterized by a finite dimensional parameter θ. Suppose that in such a situation we may derive the form of the spectral density matrix assuming stationary input and output processes and that it is given by

$$\begin{bmatrix} f_{MM}(\lambda; \theta) & f_{MN}(\lambda; \theta) \\ f_{NM}(\lambda; \theta) & f_{NN}(\lambda; \theta) \end{bmatrix}.$$

Suppose further that

$$\lim_{|\lambda|\to\infty} f_{MM}(\lambda; \theta) = \mu_M(\theta), \qquad \lim_{|\lambda|\to\infty} f_{NN}(\lambda; \theta) = \mu_N(\theta).$$

Set

$$\begin{aligned} g_{MM}(\lambda; \theta) &= f_{MM}(\lambda; \theta)/\mu_M(\theta), \qquad g_{NN}(\lambda; \theta) = f_{NN}(\lambda; \theta)/\mu_N(\theta) \\ g_{NM}(\lambda; \theta) &= f_{NM}(\lambda; \theta)/(\mu_N(\theta)\mu_M(\theta))^{\frac{1}{2}}. \end{aligned}$$

Let $\hat{p}_M = M(0, T]/T$ and $\hat{p}_N = N(0, T]/T$, then under regularity conditions (see Brillinger (1975 b)) the variate $\mathbf{h}^T(\lambda) = \{d_M{}^T(\lambda)/(\hat{p}_M)^{\frac{1}{2}}, d_N{}^T(\lambda)/(\hat{p}_N)^{\frac{1}{2}}\}$ is asymptotically bivariate complex normal with mean $\mathbf{0}$ and covariance matrix

$$T\begin{bmatrix} g_{MM}(\lambda; \theta) & g_{MN}(\lambda; \theta) \\ g_{NM}(\lambda; \theta) & g_{NN}(\lambda; \theta) \end{bmatrix} = T\mathbf{g}(\lambda; \theta)$$

for $\lambda \neq 0$. This suggests setting down the following approximate "log likelihood" function

$$-\sum_{s=1}^{S} \left\{\log \operatorname{Det} \mathbf{g}\left(\frac{2\pi s}{T}; \theta\right) + \operatorname{tr}\left(\mathbf{J}^T\left(\frac{2\pi s}{T}\right)\mathbf{g}\left(\frac{2\pi s}{T}; \theta\right)^{-1}\right)\right\} \tag{6.1}$$

where $\mathbf{J}^T(\lambda) = \mathbf{h}^T(\lambda)^{\tau}\overline{\mathbf{h}^T(\lambda)}$, and then estimating θ by $\hat{\theta}$, the value maximizing expression (6.1). This procedure is a point process version of a procedure suggested by Whittle (1953, 1961) for ordinary time series. Under regularity conditions (see Brillinger (1975 b)) it may be shown that the estimate $\hat{\theta}$ is consistent and asymptotically normal with mean θ and covariance matrix $2\pi T^{-1}\mathbf{A}^{-1}(\mathbf{A} + \mathbf{B})\mathbf{A}^{-1}$

where $\mathbf{A}$, $\mathbf{B}$ are matrices with entries

$$A_{jk} = \int_0^{2\pi S/T} \mathrm{tr}\left(\frac{\partial \mathbf{g}(\alpha)}{\partial \theta_j} \mathbf{g}(\alpha)^{-1} \frac{\partial \mathbf{g}(\alpha)}{\partial \theta_k} \mathbf{g}(\alpha)^{-1}\right) d\alpha$$

$$B_{jk} = \int \sum_a \sum_b \sum_c \sum_d C_{abj}(\alpha) C_{cdk}(\beta) g_{abcd}(\alpha, -\alpha, -\beta)\, d\alpha\, d\beta$$

with $C_{abj}(\alpha)$ the entry in row a, column b of the matrix

$$\mathbf{g}(\alpha)^{-1} \frac{\partial \mathbf{g}(\alpha)}{\partial \theta_j} \mathbf{g}(\alpha)^{-1}\,.$$

Estimates constructed in the above manner cannot be expected to be efficient as they are based only upon first and second order parameters and statistics. It would be interesting to construct a procedure involving third order parameters as well.

7. Some examples based on neurophysiological data. The field of neurophysiology is an excellent source of problems and data relating to point process systems. The paper by Bryant et al. (1973) is a good example of recent quantitative work in the field. The data discussed below were provided to this worker by those authors.

When a microelectrode is inserted into a nerve cell, a changing voltage may be recorded. Figure 1 is an example of such records for two neighboring cells, (L10, L3), of the sea slug (*Aplysia californica*). Here, and in many cases, the records are made up of pulses of large amplitude and short duration. Consequently the times of the pulses may reasonably be thought of as realizations of point processes. Figure 2 provides estimates of certain of the parameters mentioned in this paper with M referring to the times at which the cell L10 of a sea slug fired and N referring to the corresponding times at which L3 fired. In all there were 2548 M events and 1532 N events corresponding to mean rates of $\hat{p}_M = 2.21$ and $\hat{p}_N = 1.33$ events/sec. respectively. A and B are estimates of $(p_{MM}(u)/p_M)^{\frac{1}{2}}$, $(p_{NN}(u)/p_N)^{\frac{1}{2}}$, $0 \leqq u \leqq 12.5$ sec. respectively. The construction of such estimates is described in Brillinger (1975 b). The square roots are taken, because the estimates then have stable variance, *ibid.* The graphs have dips near 0 because of the cells' dead times. (There is a refractory period, after a nerve cell has fired, during which it cannot fire again.) The horizontal lines of A and B are at the levels $(\hat{p}_M)^{\frac{1}{2}}$, $(\hat{p}_N)^{\frac{1}{2}}$ respectively corresponding to estimates of the level for processes with orthogonal increments. The L10 cell was here stimulated

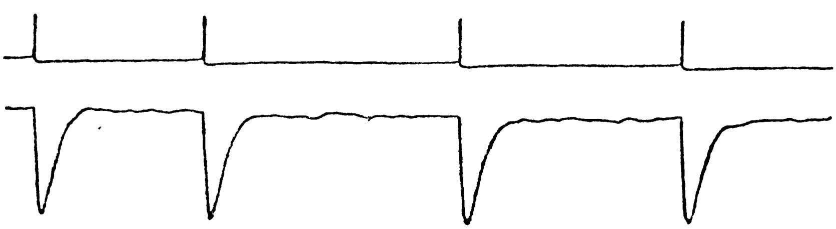

FIG. 1.

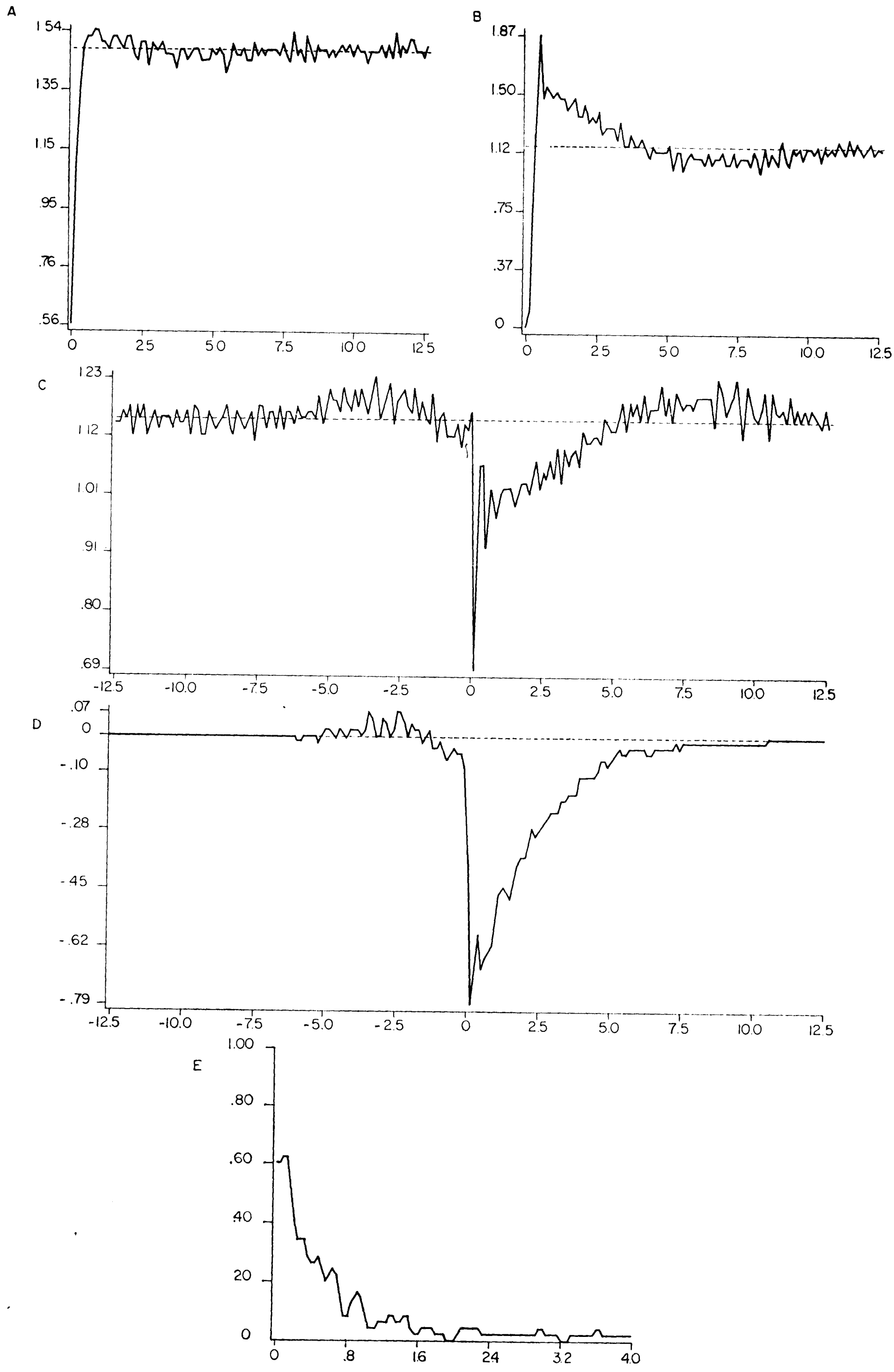
A
1.54
1.35
1.15
.95
.76
.56
0 2.5 5.0 7.5 10.0 12.5
B
1.87
1.50
1.12
.75
.37
0
0 2.5 5.0 7.5 10.0 12.5
C
1.23
1.12
1.01
.91
.80
.69
-12.5 -10.0 -7.5 -5.0 -2.5 0 2.5 5.0 7.5 10.0 12.5
D
.07
0
-.10
-.28
-.45
-.62
-.79
-12.5 -10.0 -7.5 -5.0 -2.5 0 2.5 5.0 7.5 10.0 12.5
E
1.00
.80
.60
.40
.20
0
0 .8 1.6 2.4 3.2 4.0

FIG. 2.

to fire in as Poisson a manner as possible. C is an estimate of $(p_{NM}(u)/p_M)^{\frac{1}{2}}$. The horizontal line is at the level $(\hat{p}_N)^{\frac{1}{2}}$ corresponding to unassociated M and N processes. The graph suggests that there is a drop in the rate of N events for up to 5 sec. after the occurrence of an M event. D provides an estimate of the average impulse response function, $s_1(\cdot)$, of (3.7). The estimate suggests that $s_1(u)$ is near 0 for $u < 0$, in accordance with the neurophysiologists' understanding of the relationship between the cells, and it suggests that the rate of L3 pulses drops for a period after the arrival of an L10 pulse. E is an estimate of the coherence function, $|R_{MN}(\lambda)|^2$, of (4.1). The estimate is significantly different from 0.0, at the 95 per cent level, for 94 of the 100 points plotted. The apparent coherence at low frequencies is surprisingly large, considering that coherence is a measure of degree of linear association and the system is nonlinear here. (Other such coherences may be found in Figure 3.) Graphs C and D are here so similar because the input is near Poisson.

Figure 3 presents some of the results of an analysis of the sort described in Section 5 for a three cell network, (L10, L3, L2), of the sea slug. In the notation of that section, M corresponds to L10, N_1 to L3 and N_2 to L2. Graphs A, B, C are estimates of the coherences $|R_{MN_1}(\lambda)|^2$, $|R_{MN_2}(\lambda)|^2$, $|R_{N_1N_2}(\lambda)|^2$ respectively. The horizontal line corresponds to the 95 per cent point of the null distribution in each case. These graphs suggests that the three cells are intercorrelated. The neurophysiologists suspected, for these particular cells, that L10 was driving both L3 and L2 and that there was no direct path between L3 and L2. Graph

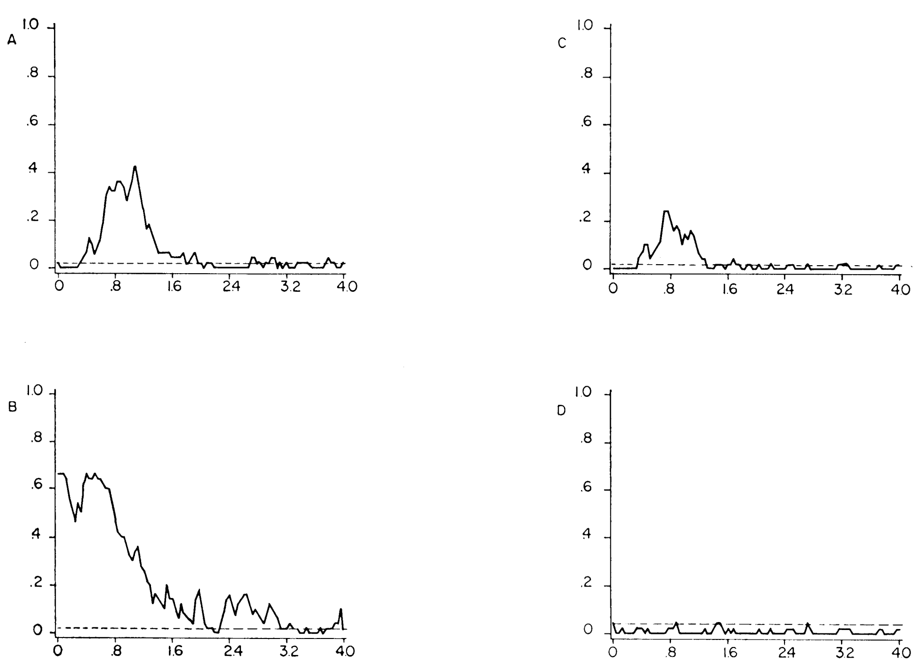

FIG. 3.

D is in accord with the suspicion. It is an estimate of the partial coherence, $|R_{N_1N_2\cdot M}(\lambda)|^2$, of (5.2). The horizontal line corresponds to the 95 per cent point of the null distribution. There is no suggestion that the partial coherence is not 0.0.

8. Acknowledgment. I would like to thank Professor J. P. Segundo and Dr. H. L. Bryant, Jr. of the U.C.L.A. Brain Research Institute for many stimulating conversations on the topics of nerve cell network analysis and point processes and for providing the data analyzed in the previous section. Dr. Daryl Daley made helpful remarks leading to improvement in the exposition of the paper.

REFERENCES

BARTLETT, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25** 264-296.

BOEL, R., VARAIYA, P. and WONG, E. (1973). Martingales on jump processes: I. Representation results, II. Applications. Memoranda ERL-M 407, 409. Electronics Research Laboratory, Berkeley.

BRILLINGER, D. R. (1970). The identification of polynomial systems by means of higher order spectra. *J. Sound Vibration* **12** 301-313.

BRILLINGER, D. R. (1972). The spectral analysis of stationary interval functions. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 483-513, Univ. of California Press.

BRILLINGER, D. R. (1974). Cross-spectral analysis of processes with stationary increments including the stationary $G/G/\infty$ queue. *Ann. Probability* **2** 815-827.

BRILLINGER, D. R. (1975a). Estimation of product densities. *Proc. Computer Science and Statist. Eighth Annual Symp. on the Interface.* Health Sciences Computing Facility, U.C.L.A. 431-438.

BRILLINGER, D. R. (1975b). Statistical inference for stationary point processes. *Stochastic Processes and Related Topics* **1** 55-99 (M. L. Puri, ed.). Academic, New York.

BRYANT, H. L. JR., RUIZ MARCOS, A. and SEGUNDO, J. P. (1973). Correlations of neuronal spike discharges produced by monosynaptic connections and by common inputs. *J. Neurophysiology* **36** 205-225.

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events.* Methuen, London.

COX, D. R. and LEWIS, P. A. W. (1972). Multivariate point processes. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 401-448, Univ. of California Press.

DOLÉANS-DADÉ, C. and MEYER, P. (1970). Integrales stochastiques par rapport aux martingales locales. *Seminaire de Probabilites* IV. Lecture Notes in Mathematics **124** Springer-Verlag, New York, 77-107.

DOLVIO, F. (1974). Counting processes and integrated conditional rates: a martingale approach with application to detection. Ph. D. thesis, Univ. Michigan.

HAWKES, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. *Stochastic Point Processes.* (P. A. W. Lewis, ed.). Wiley, New York.

HIDA, T. (1970). *Stationary Stochastic Processes.* Princeton Univ. Press.

KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics* **2**. Griffin, London.

LEWIS, P. A. W. (1972). *Stochastic Point Processes.* Wiley, New York.

SEGALL, A., DAVIS, M. H. A. and KAILATH, T. (1975). Nonlinear filtering with counting observations. *IEEE Trans. Information Theory* **IT-21** 143-149.

SEGALL, A. and KAILATH, T. (1975). The modelling of randomly modulated jump processes. *IEEE Trans. Information Theory* **IT-21** 135-142.

VAN SCHUPPEN, J. H. and WONG, E. (1974). Transformation of local martingales under a change of law. *Ann. Probability* **2** 879-888.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat. Astron. Fys.* **2** 423–434.

WHITTLE, P. (1961). Gaussian estimation in stationary time series. *Bull. Inst. Internat. Statist.* **33** 105–130.

WIENER, N. (1958). *Nonlinear Problems in Random Theory.* M. I. T. Press, Cambridge.

DEPARTMENT OF STATISTICS,
UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA 94720

DISCUSSION ON PROFESSOR BRILLINGER'S PAPER

D. R. COX (*Imperial College, London*) My comments concern the statistical aspects of Dr. Brillinger's interesting paper. First, when it is required to study the dependence of a process $\{N\}$ on an explanatory process $\{M\}$, there are often strong arguments for arguing conditionally on the observed process $\{m\}$. In particular, assumptions about $\{M\}$ itself are avoided; even its stationarity is not required so long as the interrelations are time-invariant.

Secondly, some qualification seems desirable of Dr. Brillinger's blanket recommendation that $\{M\}$ should, where possible, be chosen to be Poisson. Will not much depend on the constraints on observation and on the nature of the interrelations? For instance, one can envisage situations where it would be more informative to take $\{M\}$ as a regular sequence of widely spread points, supplemented, perhaps, by some pairs of points close together to examine linearity.

Thirdly, an alternative to the study of interrelations is via the modulation of simple models for $\{N\}$ (Cox, 1972). In this the intensity of the $\{N\}$ process is modified by a factor depending on relevant aspects of the $\{M\}$ process. Two advantages of this approach are that in certain cases likelihood functions can be obtained and that simple relations, nonlinear in Dr. Brillinger's special sense, can be accommodated; for example, the backward recurrence time in the $\{M\}$ process may be particularly relevant. An advantage of Dr. Brillinger's approach is that special assumptions about $\{N\}$ are avoided.

REFERENCES

COX, D. R. (1972). The statistical analysis of dependences in point processes. In *Stochastic Point Processes*, P. A. W. Lewis, ed. Wiley, New York.

P. Z. MARMARELIS (*California Institute of Technology*) Professor Brillinger's well-written paper on the identification of point process systems fulfills, among others, a long-standing need for such work in the field of neurophysiological system analysis. I expect that many applications of these techniques on point process systems (certainly on neural systems) will come to fruition following Brillinger's work.