Comparative Aspects of the Study of Ordinary Time Series and of Point Processes†

DAVID R. BRILLINGER

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
BERKELEY, CALIFORNIA
AND
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF AUCKLAND
AUCKLAND, NEW ZEALAND

† This research was partially supported by the J. S. Guggenheim Memorial Foundation and National Science Foundation Grant MCS76-06117.

ISBN 0-12-4266 01-0

1. INTRODUCTION

Both the literature concerning the subject of ordinary time series analysis and that of point process analysis are large and are rapidly growing. The following books may be mentioned: Bartlett (1966), Whittle (1963), Rozanov (1966), Box and Jenkins (1970), Hannan (1970), Anderson (1971), Koopmans (1974), and Brillinger (1975a) in the case of ordinary time series, and Harris (1963), Cox and Lewis (1966), Kerstan *et al.* (1974), Srinivasan (1974), Snyder (1975), and Murthy (1974) in the case of point processes. Generally speaking, the literatures of these two subjects have developed quite independently of each other, although Bartlett (1966) and Brillinger (1972) are exceptions. There are many similarities between the two subjects and it appears that each can benefit from a consideration of the methods of the other. The intention of this chapter is to indicate cases in which the concepts and procedures of ordinary time series (or point processes) have direct analogs in the study of point processes (or ordinary time series). Certain cases in which one subject has unique facets and there are no apparent immediate analogs will also be mentioned. There are gains to be had from adopting a unified approach. Indeed, nowadays data are being collected that are hybrid, part ordinary time series and part point process (e.g., see Bryant and Segundo, 1975), so some unified method of analysis is clearly called for. The structure of this chapter is one of parallel sections indicating corresponding results for ordinary time series and for point processes. The reader is generally referred to the original literature for detailed statements of theorems and most general results. About half of the material was presented in lectures to graduate students in mathematics at the University of Auckland during March to July 1976. Alan J. Lee made helpful comments concerning the manuscript.

1.1. Time Series and Some History of Their Study

The general phenomena studied and recorded by scientists depend on time. In many circumstances this dependence on time may be ignored. The intent of this work is to study phenomena that depend on time in an essential manner. Measurements corresponding to continuous real-valued functions of time are called *time series* and are denoted by $X(t)$, $0 \le t < \infty$, assuming their domain of definition may be thought of as the interval $[0, \infty)$. Simple examples of time series include the current at a particular junction in an electric circuit as it varies in time, the displacement of the needle of a seismometer from its rest position as a function of time, the temperature at a given location on the earth's surface as a function of time, and finally the height of a sea's surface along a given parallel of latitude as a function of longitude. (In this last example the "time" parameter is really a distance parameter.) The books mentioned in Section 1 may be consulted for specific references to a wide variety of interesting time series.

The scientific analysis of time series has a very long history. It may be thought of as having commenced in 1664 when Issac Newton decomposed a light signal (or time series) into frequency components by passing the signal through a glass prism. A multicolored image was cast upon an opposite wall. He called this the spectrum. This analysis corresponds to the bandpass filtering operation discussed in Section 2.6. Newton did not carry out a quantitative analysis of his time series. However, in 1800 W. Herschel did, by using thermometers. He measured the average energy in various frequency bands of the sunlight's spectrum by placing thermometers along that spectrum. Mathematical foundations began to be laid in the mid-1800s for the analysis of time series when Gouy represented white light as a Fourier series. Later Rayleigh replaced the series by an integral. In 1881 S. P. Langley refined Herchel's experiment considerably by measuring the light energy with a spectral bolometer. (A device that he had invented, it makes use of electric current generated in a wire by incident radiation.)

In 1872 Lord Kelvin built a harmonic analyzer and a harmonic synthesizer for use in the analysis and prediction of the series $X(t)$ of the height of the tide at a particular location and time t. His devices were mechanical, based on pulleys. During the same time period a variety of workers (e.g., G. G. Stokes) were carrying out numerical Fourier analyses using computation schedules. In particular, in 1891 S. C. Chandler carried out an analysis of the variation of latitude with time. His analysis led him to suggest that the motion of the earth's pole of rotation was composite, containing components of period 12 and approximately 14 months.

A substantial advance in the analysis of time series corresponding to light signals occurred in 1891 when A. A. Michelson invented the interferometer. This device allowed the measurement of the average value

$$\left\{\int_0^T [X(t) + X(t+u)]^2 \, dt\right\}\Big/T \tag{1.1.1}$$

for large values of T and nominated values of the lag u. In consequence, it allowed the estimation of the autocovariance function (see Section 2.4) of the incident signal. In 1898 Michelson and Stratton described a harmonic analyzer (based on springs) and used it to obtain the Fourier transform of the function (1.1.1). This Fourier transform provided an estimate of the power spectrum of the signal. Michelson envisaged the signal as being a sum of cosines. He saw the estimated spectra as descriptive statistics of the light emitting sources.

In 1894 M. I. Pupin invented the electric wave filter (also a sort of bandpass filter). This device considerably broadened the frequency domain over which time series could be analyzed. The power of an electric signal could now be measured in a range of frequency bands.

In a series of papers written during the years 1894–1898, A. Schuster proposed and discussed the periodogram statistic

$$\left| \sum_{t=1}^{T} \exp\{-i\lambda t\} X(t) \right|^2 \tag{1.1.2}$$

based on an observed stretch of time series $X(t), t = 1, 2, \ldots, T$. His motivation was a search for "hidden periodicities." In the succeeding years periodograms and their equivalents were computed for a variety of phenomena by many workers.

In 1922 Crandall and MacKenzie at Bell Telephone Laboratories used resonance tubes to measure the energy distribution of speech as a function of frequency. Once again, the frequency domain over which phenomena could be studied was usefully broadened. Also during the 1920s time series of turbulent flow were investigated by G. I. Taylor (who may have been the first to define the autocovariance function of a time series). Certain of Taylor's methods were applied to meteorological series by G. Walker. During the 1920s and 1930s the field of quantum mechanics and its related form of spectral analysis underwent considerable development.

In the time period 1930–1950 substantial developments in the area of time series analysis were provided by N. Wiener, H. Cramér, A. N. Kolmogorov, M. S. Bartlett, and J. W. Tukey. Details of their contributions may be found in the books mentioned in Section 1.

The range of problems studied by time series analysts covers smoothing, forecasting, control, seasonal adjustment, detection, parameter estimation, discrimination of series from different populations, checking for association between series, and isolation of more elementary series. A variety of these topics are discussed in this chapter.

1.2. Point Processes and Some History of Their Study

The preceding section was concerned with phenomena evolving continuously in time. Point processes refer to isolated events occurring haphazardly in time. A stochastic point process is a random, nonnegative, integer-valued measure. If I is an interval of the real line and ω is a random element, then the values of this measure may be denoted by $N(\cdot, \omega)$ with $N(I, \omega)$ denoting the number of points in the interval I for the realization corresponding to ω. Here the atoms of the measure correspond to a particular set of points. Throughout this chapter, the notation $N(t) = N((0, t], \omega)$ will be employed. Then $dN(t)$ refers to the number of points in the small interval $(t, t + dt]$.

The variety of data that arise in practice and have point process character is staggering. Subject areas leading directly to the collection of point process data include traffic systems, queues, neuronal electrical activity, microscopic theory of gases, resistance noise, heartbeats, population growth, radioactivity, seismology, accident or failure processes, telephone systems, cosmic rays, and fluctuation of photoelectrons. A wide variety of examples are discussed by Snyder (1975) and in Lewis (1972). A recent paper discussing point processes in seismology is that of Udias and Rice (1975), and one in neurophysiology is by Brillinger *et al.* (1976).

Point process data are typically stored either in terms of actual times of events (locations of points) or in terms of lengths of successive intervals between events. In the case of points of several types, the data may be stored separately for each type or by the actual times of events with a "flag" to indicate the event type. Point process data are processed in both analog and digital manners.

The beginnings of the study of point processes may be found in the history of population mathematics. J. Graunt (1620–1674) constructed a life table. Such a table corresponds to the superposition of many independent point processes, each containing a single point at the time of death of an individual. Other workers in this area were C. Huygens (1629–1695) and E. Haley (1656–1742). Much of their concern was over the mean duration of life, the disadvantages of such a measure, and the calculation of the values of life annuities. Specific functional forms were proposed for the force of mortality by de Moivre, Lambert, Gompertz, and Makeham among others. This early history is described in some detail by Westergaard (1968).

The next area of point process research activity related to the Poisson distribution and process. The distribution is credited to de Moivre in 1718 and Poisson in 1837 (see Haight, 1967, p. 113). The name of von Bortkiewicz (1868–1931) is closely associated with the Poisson distribution, especially because of his use of it to model the frequency of death by horse-kick in the Prussian Army. The Poisson process was introduced over a long period. In 1868 Boltzmann determined the expression $\exp\{-\mu t\}$ for the probability of no points in an interval of length t (Haight, 1967, p. 114), and in 1910 Bateman determined the counting distributions by solving a set of differential equations (Haight, 1967, p. 120). In 1903 F. Lundberg investigated a process of which the Poisson is a particular case (see Cramér, 1976).

In 1909 Erlang applied the Poisson process to traffic studies, proposed the truncated Poisson, and considered the process with intervals made up of right displaced exponentials (Haight, 1967, p. 121–123). Erlang's interest was in building better telephone systems, for example, determining the optimum

number of circuits. He may be said to have initiated the study of queueing systems—involving input and output point processes corresponding to times of arrival and departure of customers. Later workers who made substantial contributions to queueing theory include Molina, Fry, Khinchin, Palm, and Pollaczek (see Bhat, 1969).

Another class of point processes with a long history of study is that of the renewal processes. In these the successive intervals between points are independent nonnegative variates. Lotka (1957) ascribes the first serious investigation of these processes to Herbelot in 1909. Other historically important references may be found in Lotka's paper. A modern reference is Cox (1962).

The work of Bateman, mentioned earlier, was stimulated by problems of particle physics. The late 1930s saw the commencement of substantial developments in the modeling of point processes by physicists. Point processes occur in radioactive decay, in particle bombardment experiments, and in coincidence experiments among other areas of physics. In 1937 Bahba and Heitler and simultaneously Carlson and Oppenheimer applied stochastic methods to the cascade phenomena of cosmic ray showers. The Bhabha–Heitler approach led to the Poisson process. Furry criticized it and proposed a simple birth process instead. This approach was criticized in turn by Scott and Uhlenbeck who proposed a further model. Explicit details of this work may be found in the books by Bharucha-Reid (1960) and Srinivasan (1969).

A class of point correlation functions was introduced by Yvon (1935) to study the dependency properties of certain point processes. Later multidimensional product density functions were introduced by Ramakrishnan (1950) to study the higher order dependencies of point processes. These had appeared earlier in a specific situation in an article by Rice (1945).

The doubly stochastic Poisson process may be said to have been introduced by Quenouille (1949). The name of D. R. Cox is often associated with it also. The early history of branching processes, another form of point process, is described in some detail by Kendall (1975).

A point process may often be characterized by its conditional intensity function $\gamma(t, \omega)$. McFadden (1965) introduced this concept and used it to show that the rate of change of the entropy of a point process may be written

$$E\{\gamma(t, \omega)[1 - \log \gamma(t, \omega)]\}$$

Wang (1968) provides an early example of a $\gamma(t, \omega)$ of nontrivial form.

1.3. Some Notation

It is convenient to collect in this section some of the notation used throughout this chapter. When a concept is introduced its name will be displayed in italics: $\in$ means "is an *element of*." The *semiopen interval* of real numbers t satisfying $a < t \leq b$ is denoted $(a, b]$. If I is the set of real numbers $\{r : r \in I\}$, then $I + t$ represents the set of corresponding *translated numbers* $\{r + t : r \in I\}$. If I is an interval, $|I|$ is its *length*. The set of *all real numbers*, $(-\infty, \infty)$, is denoted by R. If s is a complex number, Re s denotes its *real part*; i is $\sqrt{-1}$. Given two sets, I and J, $I \times J$ denotes their *direct product* made up of pairs (r, s) with $r \in I$ and $s \in J$; log refers to the logarithm base e. $\prod_{j \in J} a_j$ represents the *product* of the a_j with $j \in J$. Given Q_{jk}, with $j = 1, \ldots, J$ and $k = 1, \ldots, K$, $[Q_{jk}]$ represents the *matrix* with the element Q_{jk} in row j and column k. The notation $N^{(K)}$ means $N(N-1)\cdots(N-K+1)$. When a domain of integration is not indicated, it is to be taken as the whole space. $\delta(t)$ denotes the *Dirac delta function* with the property

$$\int \delta(t) f(t)\, dt = f(0)$$

for functions $f(t)$ continuous at 0.

Turning to probabilistic considerations, Ω denotes a *sample space*. $B(\Omega)$ is the smallest *Borel field* generated by certain of the subsets of Ω (e.g., when Ω has a topology, by all open sets). P denotes a *probability measure* with values $P(A)$ for $A \in B(\Omega)$. *Integrals* with respect to P are given either by $\int f(\omega)\, dP(\omega)$ or $\int f(\omega) P(d\omega)$. Given a random variable $Y(\omega)$, its *expected value* is denoted by $EY(\omega)$ (or by EY when there is little chance for confusion). Its *variance* is var Y. The *covariance* of two random variables is written as $\operatorname{cov}\{Y_1, Y_2\}$. The joint *cumulant* of K random variables is $\operatorname{cum}\{Y_1, \ldots, Y_K\}$. A *stochastic time series* is a function-valued random variable $X(t, \omega)$, $t \in R$, $\omega \in \Omega$. It is sometimes denoted by $X(t)$ and sometimes by X, when confusion seems unlikely. A *stochastic point process* is a nonnegative integral step-function-valued random variable $N(t, \omega)$, $t \in R$, $\omega \in \Omega$. On occasion it is written as $N(t)$ or N. Differential notation is often used with $dN(t) = N(t + dt) - N(t)$ for infinitesimal increments dt. When θ is a parameter, $\hat{\theta}$ stands for an *estimate* of θ. T means the length of the time period of observation of a time series or point process. In asymptotic studies, $T \to \infty$.

2. FOUNDATIONS

2.1. Formal Definitions of an Ordinary Time Series

Several approaches are available to the theory of *ordinary time series*. The intent of each approach is to provide a structure within which one can manipulate and deal with real-valued functions $X(t)$ for t in some index set. [In this work the index set is the interval $[0, \infty)$ and it is assumed that $X(t)$ is

a continuous function of t.] The various constructions allow the definition of certain useful parameters.

One means of proceeding is to consider a function of two variables $X(t, \omega)$, with t in $[0, \infty)$ and ω in Ω, where $(\Omega, B(\Omega), P)$ is a probability space, $X(t, \omega)$ is a measurable function of ω for each t, and $X(t, \omega)$ is a continuous function of t for almost all ω. The section $X(\cdot, \omega)$ for fixed ω is called a *realization* or *trajectory* of the stochastic time series X. In this setup $X(\cdot, \omega)$ may be regarded as a random continuous function. An example of proceeding in this manner would be to define

$$X(t, \omega) = \rho \cos(\lambda t + \omega), \qquad 0 \le t < \infty$$

where ρ, λ are fixed and ω is a uniform variate on $(-\pi, \pi)$. A more complicated example would be a diffusion process defined through an integral equation involving Brownian motion.

A second means of proceeding is to imagine a collection of consistent *finite-dimensional distribution functions*

$$F_K(x_1, \ldots, x_K; t_1, \ldots, t_K) \tag{2.1.1}$$

$-\infty < x_1, \ldots, x_K < \infty$; $0 \le t_1, \ldots, t_K < \infty$, with (2.1.1) thought of as being

$$\text{Prob}\{X(t_1) \le x_1, \ldots, X(t_K) \le x_K\} \tag{2.1.2}$$

The Kolmogorov extension theorem now indicates the existence of a probability space $(\Omega, B(\Omega), P)$ and $X(t, \omega)$, $0 \le t < \infty$, $\omega \in \Omega$, such that (2.1.2) equals (2.1.1) for all $t_1, \ldots, t_K$, $K = 1, 2, \ldots$, with $X(t) = X(t, \cdot)$. Further conditions must be set down in order to ensure that $X(\cdot, \omega)$ is almost surely a continuous function. An example of proceeding by this approach is when a Gaussian process is defined by stating that all the finite-dimensional distributions of the process are consistent multivariate normals.

A third means of proceeding is to imagine being given $C[0, \infty)$ as the space of continuous functions on $[0, \infty)$, with an appropriate Borel field of sets for the function space, and then to assume the existence of a probability measure on the space. In terms of the previous notation, $\Omega = C[0, \infty)$. Many useful theorems have been developed for this structure. This approach is considered by Billingsley (1968) and Nelson (1959) for example.

In some circumstances one proceeds through conditional distribution or transition probability functions, for example, by assuming the existence of the functions

$$\text{Prob}\{X(t) \in A \,|\, X(t_1), \ldots, X(t_K); \quad 0 \le t_1 < t_2 < \cdots < t_K < t\} \tag{2.1.3}$$

for A in $(-\infty, \infty)$, $K = 1, 2, \ldots$, and of an initial probability function

$$\text{Prob}\{X(0) \in A\}$$

This is the usual means by which a Markov process on $[0, \infty)$ is defined. It may be used to define an autoregressive scheme of finite order also. Far-reaching generalizations of the notion of studying processes through conditional variates are given by Schwartz (1973) and Knight (1975).

Another means by which a stochastic time series may be characterized is through its *characteristic functional* (or, equivalently, the distribution of a broadly defined linear functional). The characteristic functional is discussed, e.g., by Bochner (1960) and Prohorov (1961). It may be thought of as being given by

$$C[\xi] = E \exp\left\{i \int \xi(t) X(t)\, dt\right\} \tag{2.1.4}$$

Provided a functional $C[\cdot]$ satisfies certain regularity conditions for $\xi(\cdot)$ in a sufficiently large function space, there is a corresponding time series. One thing needed is consistent characteristic functions for all finite-dimensional distributions. Stable processes are usually defined through the characteristic functional (see Bochner, 1960).

As a final means of introducing stochastic time series, consider the class of *harmonizable processes*. This is made up of series of the form

$$X(t, \omega) = \int_{-\infty}^{\infty} \exp\{i\lambda t\}\, dZ(\lambda, \omega) \tag{2.1.5}$$

for $Z(\lambda, \omega)$ a complex-valued random process, with the integral of (2.1.5) defined in a probabilistic manner (e.g., as a limit in the mean of approximating Stieltjes sums). Harmonizable processes are discussed by Loève (1955, p. 474). $X(t, \omega)$ is seen to be a linear function of $Z(\lambda, \omega)$ so its moments can be given directly in terms of those of $Z(\lambda, \omega)$. This manner of introducing a time series proves especially useful in the case of stationary series.

By a *stationary time series* is meant one with the property that the characteristics of finite collections of its values

$$\{X(t + u_1), \ldots, X(t + u_K)\} \tag{2.1.6}$$

do not depend on t for $0 \le t + u_1, \ldots, t + u_K < \infty$ and $K = 1, 2, \ldots$. In particular, the finite-dimensional distribution of the variate (2.1.6) should not depend on t. In the stationary case there are alternative ways of beginning.

One is to suppose that a probability space $(\Omega, B(\Omega), P)$ is given as well as a semigroup of measure-preserving transformations of Ω. Then the series is constructed through setting

$$X(t, \omega) = Y(U_t \omega) \tag{2.1.7}$$

for some random variable $Y(\omega)$, with U_t, $t \in [0, \infty)$, the generator of the semigroup. This approach is discussed, e.g., by Doob (1953, p. 507). It is apparent that (2.1.5) implies, for example,

$$X(t + s, \omega) = Y(U_t U_s \omega) = X(t, U_s \omega), \qquad t, s \geq 0$$

Immediate ways of constructing stationary time series include beginning with a consistent family of finite-dimensional distributions that are invariant under shifts $(u_1, \ldots, u_K) \to (t + u_1, \ldots, t + u_K)$ or beginning with a characteristic functional $C[\cdot]$ unaffected by the transformation $\xi(t) \to \xi(t - u)$, $0 \leq t, t + u$.

An important means of introducing stationary time series is through the spectral representation (2.1.5). One can assume that the characteristic functional of the process $Z(\lambda, \omega)$, namely

$$E \exp\left\{i \int \xi(\lambda)\, dZ(\lambda, \omega)\right\}$$

is invariant under the transformations $\xi(\lambda) \to \xi(\lambda) \exp\{i\lambda u\}$ or that the process is determined by its moments and the moment measures

$$E\{dZ(\lambda_1) \cdots dZ(\lambda_K)\} \tag{2.1.8}$$

are concentrated in the hyperplanes

$$\lambda_1 + \cdots + \lambda_K = 0, \qquad K = 1, 2, \ldots$$

Each of the preceding approaches has been stochastic in character. An observed stretch of time series is to be thought of as a segment of some realization of a stochastic process. Wiener (1930) introduced a structure based on single nonstochastic functions. For a given function $X(t)$, $0 \leq t < \infty$, limits such as

$$\lim_{T \to \infty} \left\{\int_0^T X(t)\, dt\right\} \Big/ T, \qquad \lim_{T \to \infty} \left\{\int_0^T X(t + u) X(t)\, dt\right\} \Big/ T$$

are assumed to exist (and in the case of the latter to be continuous at 0). This approach goes under the name of *generalized harmonic analysis*.

Throughout this section, the basic time domain has been taken to be the semi-infinite interval $[0, \infty)$. In the stationary case, the series may always be extended to be stationary on the whole line $(-\infty, \infty)$. One simply takes for the finite-dimensional distribution at any collection of time points $-\infty < t_1 < t_2 < \cdots < t_K < \infty$, the distribution at $0, t_2 - t_1, \ldots, t_K - t_1$. The Kolmogorov extension theorem then indicates the existence of a process

defined for $-\infty < t < \infty$. The extension of a discrete time series ($t = 0, 1, 2, \ldots$) is considered by Breiman (1968, p. 105). An alternative means of carrying out an extension in the case of continuous time is presented by Furstenberg (1960).

This section concludes by mentioning that certain function spaces provide important representations of time series. To begin, consider a time series $X(t, \omega)$ having finite second-order moments. Then $L_2(X; t)$ is defined to be the closed linear manifold spanned by the values $X(u, \omega)$ with $u \leq t$ and the norm of an element U taken to be $E|U|^2$. That is, it is the smallest Hilbert space which contains all random variables of the form

$$c_1 X(t_1, \omega) + \cdots + c_K X(t_K, \omega)$$

for real numbers $c_1, \ldots, c_K$, for $t_1, \ldots, t_K \leq t$, and $K = 1, 2, \ldots$. The spaces $L_2(X; t)$ are most useful in problems concerning linear prediction. They are all contained in $L_2(X) = L_2(X, \infty)$. Using these spaces Cramér has developed many important properties. Cramér's work may be found in the book by Ephremides and Thomas (1973). A Gaussian process is determined by its first- and second-order moments. Consequently, when studying such a process, it is often enough to consider these spaces $L_2(X; t)$.

In the general case, however, it is often necessary to consider $N_2(X; t)$, the smallest Hilbert space which contains random variables with finite second-order moment and of the form

$$g(X(t_1), \ldots, X(t_K))$$

for $t_1, \ldots, t_K \leq t$ and $K = 1, 2, \ldots$ (see Parzen, 1962). The space $N_2(X; t)$ is contained in $L_2(\Omega) = N_2(X; \infty)$ for each t. It is especially useful in problems concerning nonlinear prediction.

Before indicating the next class of function spaces, it is necessary to mention the existence of *reproducing kernel Hilbert spaces*. Let $R(\psi, \phi)$ be a real nonnegative definite function on the product space $\Phi \times \Phi$, for some index space Φ. Aronszan (1950) has shown the existence of a Hilbert space $H(R)$ of real-valued functions $f(\phi)$, ϕ in Φ, with the following properties: (i) $R(\cdot, \phi)$ belongs to $H(R)$ for each ϕ; (ii) if $\langle\ ,\ \rangle$ denotes the inner product of $H(R)$, then $\langle f(\cdot), R(\cdot, \phi)\rangle = f(\phi)$ for every f in $H(R)$; (iii) $H(R)$ is generated by linear combinations, and limits of these, of the functions $R(\cdot, \phi)$, ϕ in Φ. This space, $H(R)$, is called the reproducing kernel Hilbert space corresponding to the *kernel* R. Some of the reproducing kernel Hilbert spaces important in time series analysis include those corresponding to (a) $\Phi = [0, \infty)$ and $R(\psi, \phi) = E\{X(\psi)X(\phi)\}$ (see Parzen, 1959). The members of this $H(R)$ have the form $E\{UX(\phi)\}$ with U in $L_2(X)$. (b) $\Phi = \{\phi(t)$ for suitable $\phi(\cdot)\}$ and

$R(\psi, \phi) = E \exp\{i \int X(t)[\psi(t) - \phi(t)]\, dt\} = C[\psi - \phi]$ (see Hida, 1970). (c) $\Phi = \{(\phi_0, \phi_1(t_1), \phi_2(t_1, t_2), \phi_3(t_1, t_2, t_3), \ldots\}$ and

$$R(\psi, \phi) = \sum_{J,K} \int \cdots \int \psi_J(s_1, \ldots, s_J)\phi_K(t_1, \ldots, t_K) \times E\{X(s_1) \cdots X(s_J)X(t_1) \cdots X(t_K)\}\, ds_1 \cdots ds_J\, dt_1 \cdots dt_K \quad (2.1.9)$$

In this case and when the series X is determined by its moments, $H(R)$ is congruent to $L_2(\Omega) = N_2(X, \infty)$.

Hida (1960) used the reproducing kernel space of (a) to show that purely nondeterministic series may be represented as

$$X(t) = \sum_{i=1}^{N} \int_{-\infty}^{t} g_i(t, u)\, dz_i(u) \quad (2.1.10)$$

where N ($\leq \infty$) is the multiplicity of $X(t)$, the $dz_i(u)$ are mutually orthogonal random innovation processes, and the $g_i(t, u)$ are nonstochastic. The representation (2.1.10) is useful in that the best linear predictor of the value $X(t + h)$, $h > 0$, based on $\{X(u), u \leq t\}$ may be written

$$\sum_{i=1}^{N} \int_{0}^{t} g_i(t + h, u)\, dz_i(u)$$

2.2. Formal Definitions of Point Processes

There are a number of approaches to the foundations of point processes and certain of these correspond directly with the preceding constructions for ordinary time series. Once again assume that the time domain is $[0, \infty)$. Assume also that the processes have no multiple points.

Initially, consider a probability space $(\Omega, B(\Omega), P)$ and take a (stochastic) *point process* to be a function $N(t, \omega)$, $0 \leq t < \infty$, ω in Ω, with $N(t, \omega)$ a measurable function of ω for each t; $N(0, \omega) = 0$ or 1 (depending on whether or not there is an initial point at the origin); for each ω, $N(t, \omega)$ is finite, nondecreasing, purely discontinuous but right continuous, with unit jumps. This is one way of defining a point process.

Instead of dealing with functions in t, another means is to deal with atomic measures and to consider $N(I, \omega)$, $I \in B[0, \infty)$, $\omega \in \Omega$, with $N(I, \omega)$ a measurable function of ω for each I; and for each ω, $N(\cdot, \omega)$ is a measure with discrete support and mass 1 at each point of the support. The $N(t, \omega)$ of the preceding paragraph may be determined from this $N(\cdot, \omega)$ through the correspondence

$$N(t, \omega) = N([0, t], \omega) \quad (2.2.1)$$

The correspondence (2.2.1) may likewise be used to deduce $N(I, \omega)$ from $N(t, \omega)$.

A procedure analogous to the second one of the preceding section is to assume a consistent family of finite-dimensional probability functions

$$f(n_1, \ldots, n_K; I_1, \ldots, I_K) \quad (2.2.2)$$

for $n_1, \ldots, n_K = 0, 1, 2, \ldots$; bounded $I_1, \ldots, I_K$ in $B[0, \infty)$, $K = 1, 2, \ldots$, to be thought of as giving

$$\text{Prob}\{N(I_1) = n_1, \ldots, N(I_K) = n_K\} \quad (2.2.3)$$

Consistency now includes requirements necessary for the realizations of the process to be a measure such as, for disjoint I_1, I_2,

$$f(n_1, n_2, n_3; I_1, I_2, I_1 \cup I_2) = 1 \quad \text{if} \quad n_1 + n_2 = n_3$$
$$= 0 \quad \text{otherwise}$$

(see Jagers, 1974). The Poisson process is usually defined in this particular manner.

When they exist, the infinitesimal probability functions, or *product densities*,

$$p_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K = \text{Prob}\{dN(t_1) = 1, \ldots, dN(t_K) = 1\} \quad (2.2.4)$$

$t_1, \ldots, t_K$ distinct, corresponding to (2.2.2) and (2.2.3), often prove very useful and even characterize a point process under certain conditions. (However, there are examples of distinct point processes having the same product densities; see Ruelle, 1969, p. 106.) Product densities are discussed, e.g., by Macchi (1975) and Brillinger (1975a). The conditional Poisson of Macchi (1975) is defined via product densities.

A surprising result concerning the theory of point processes is that point processes $N(I, \omega)$, without multiple points, may be characterized by their *zero probability functions*

$$\phi(I) = \text{Prob}\{N(I, \omega) = 0\} \quad (2.2.5)$$

for bounded I in $B[0, \infty)$ (see Kurtz, 1974; Jagers, 1974). Given a function $\phi(I)$ on $B[0, \infty)$ satisfying certain conditions, there is a corresponding point process satisfying (2.2.5). Equivalently, one only needs to specify $f(0, \ldots, 0; I_1, \ldots, I_K)$ of (2.2.2) for finite intervals $I_1, \ldots, I_K$ in order to specify the process. For example, the *Poisson process* with parameter measure $\mu(I)$ may be characterized as the point process with

$$\phi(I) = \exp\{-\mu(I)\}$$

for bounded I in $B[0, \infty)$ (see Renyi, 1967).

The zero probability function and product densities are related through

$$\phi(I) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_I \cdots \int_I p_k(t_1, \ldots, t_k)\, dt_1 \cdots dt_k$$

$$p_1(t)\, dt = 1 - \phi(t, t + dt)$$

$$p_2(t_1, t_2)\, dt_1\, dt_2 = 1 - \phi(t_1, t_1 + dt_1) - \phi(t_2, t_2 + dt_2) + \phi((t_1, t_1 + dt_1) \text{ or } (t_2, t_2 + dt_2))$$

$$p_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K = (-1)^K \Delta_{dt_1} \cdots \Delta_{dt_K} \phi(\varnothing)$$

where $\Delta_{dt}\, \phi(I) = \phi(I \text{ or } (t, t + dt)) - \phi(I)$ and $\varnothing$ denotes the empty set (see Kurtz, 1974).

A point process may also be characterized by giving a *probability generating functional* $G[\xi]$, to be thought of as

$$G[\xi] = E \exp\left\{\int \log \xi(t)\, dN(t)\right\} \tag{2.2.6}$$

satisfying certain regularity conditions (see Daley and Vere-Jones, 1972, Theorem 3.10). This is the way the Gauss–Poisson process is defined, for example (see Milne and Westcott, 1972), as well as the infinitely divisible point process. The probability generating functional may be written out in terms of the product densities, specifically

$$G[\xi] = \sum_k \frac{1}{k!} \int \cdots \int (\xi(t_1) - 1) \cdots (\xi(t_k) - 1) p_k(t_1, \ldots, t_k)\, dt_1 \cdots dt_k \tag{2.2.7}$$

(e.g., see Vere-Jones, 1968). The finite-dimensional distributions of (2.2.3) may be determined from $G[\xi]$ by setting

$$\xi(t) = z_k \quad \text{for} \quad t \text{ in } I_k$$
$$= 1 \quad \text{otherwise}$$

and then determining the coefficient of $z_1^{n_1} \cdots z_K^{n_K}$ in the Taylor series expansion of the functional.

An approach to the foundations of point processes which is related to those just given, but yet has a somewhat different character, involves taking Ω to be the set of all locally finite subsets, ω, of $[0, \infty)$. A stochastic point process is then defined to be a probability measure on Ω or, equivalently, a measurable map from a probability space into $(\Omega, B(\Omega))$. The connection with the previous approach is indicated by setting

$$N(I, \omega) = \text{card}(I \cap \omega)$$

for I in $B[0, \infty)$ and ω in Ω. The individual elements of the subsets are called points of the process.

Because points on the real line are ordered, it might be desired to consider, instead of the space of the preceding paragraph, the space $\mathcal{T} = \{(\tau_0, \tau_1, \ldots, \tau_N)\colon 0 < \tau_0 < \tau_1 < \cdots < \tau_N < \infty;\ N = 1, 2, \ldots\}$, and to consider a stochastic point process to be a measurable map into $\mathcal{T}$ from a probability space $(\Omega, B(\Omega), P)$. A realization of the process can then be denoted $\tau_0(\omega), \tau_1(\omega), \ldots$, with the $\tau_j(\omega)$ the points of the realization. The connection with the previous approach is seen through setting

$$N(I, \omega) = \text{the number of } \tau_j(\omega) \text{ in } I$$

In the reverse direction one has the following connection: $\tau_j(\omega)$ is the t such that $N[0, t) = j < N[0, t]$. If $e\{t\}$ denotes the point measure of mass 1 at t, then an alternative way of writing the realizations of the process is

$$N(\cdot, \omega) = \sum_j e\{\tau_j(\omega)\}$$

Turning to specific descriptions of the probabilistic structure of the process within this last framework, the process may clearly be described by the finite-dimensional distributions of the discrete process $\{\tau_0(\omega), \tau_1(\omega), \tau_2(\omega), \ldots\}$ or by the conditional distributions

$$G_j(t; t_j, j < J) = \text{Prob}\{\tau_J > t \,|\, \tau_j = t_j, j < J\} \tag{2.2.8}$$

Macchi (1971) gives the joint probability density function of the q successive points following t being at $t_1 < t_2 < \cdots < t_q$ as

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{(t, t_q]^k} \cdots \int p_{k+q}(t_1, \ldots, t_q, u_1, \ldots, u_k)\, du_1 \cdots du_k \tag{2.2.9}$$

This expression may be determined directly from the probability generating functional (2.2.7) by using it to evaluate

$$\text{Prob}\{N(t, t_1) = 0, dN(t_1) = 1, N(t_1, t_2) = 0, dN(t_2) = 1, \ldots, dN(t_q) = 1\}$$

A further space to consider is $[0, \infty) \times Y$ where $Y = \{(y_1, \ldots, y_N)\colon 0 < y_1, \ldots, y_N;\ N = 1, 2, \ldots\}$. These y terms are to be thought of as the distances between successive points of a realization of a stochastic point process, $y_j = \tau_j - \tau_{j-1}$, $j = 1, 2, \ldots$. The individual points of the process may be reconstructed from τ_0 and the y_j through the relationship $\tau_j = \tau_0 + y_1 + \cdots + y_j$, $j = 1, 2, \ldots$. Defining a measurable map from a probability space $(\Omega, B(\Omega), P)$ into $[0, \infty) \times Y$ is now seen to be a further means of constructing a stochastic point process. A realization of the process is now denoted $\tau_0(\omega), y_1(\omega), y_2(\omega), \ldots$. This is the means by which a renewal process is generally constructed, or the Markov interval process studied by Vere-Jones (1975a), or the autoregressive model of Gaver and Lewis (1976).

Given a stochastic point process N, certain conditional probabilities, the Palm probabilities, are occasionally important. These are conditional prob-

abilities given that a point of the process occurs at a particular location, and may be denoted by

$$P(A\,|\,t) = \text{Prob}\{A\,|\,t \text{ is a point of the realization}\}$$

for events A in $B(\Omega)$. $P(A\,|\,t)$ may be defined as the Radon–Nikodyn derivative of the relationship

$$\int_A N(I, \omega)\, dP(\omega) = \int_I P(A\,|\,t)\, d\mu(t)$$

for I in $B[0, \infty)$, A in $B(\Omega)$, and where μ is the intensity measure of the process (see Jagers, 1974, Section 6). These probabilities may be chosen such that for all t, one has a corresponding point process which will be denoted by $N(\ \ |t)$. In fact the original process N is determined by the specification of $P(A\,|\,t)$ and μ. For example,

$$\text{Prob}\{N(I) = k\} = \left\{\int_I \text{Prob}\{N(I\,|\,t) = k\}\, d\mu(t)\right\}\Big/k \qquad (2.2.10)$$

for $k = 1, 2, \ldots$. It is clear that the zero probability function of N may be determined using expression (2.2.10).

The product densities of the process $N(\ \ |t)$ may be determined directly from those of the process N, namely

$$p_K(t_1, \ldots, t_K\,|\,t) = p_{K+1}(t, t_1, \ldots, t_K)/p_1(t) \qquad (2.2.11)$$

Its probability generating functional is therefore given by

$$G[\xi\,|\,t] = \sum (1/k!) \int \cdots \int (\xi(t_1) - 1) \cdots (\xi(t_k) - 1)$$
$$\times\, p_{k+1}(t, t_1, \ldots, t_k)\, dt_1 \cdots dt_k/p_1(t) \qquad (2.2.12)$$

It follows from (2.2.9) and (2.2.12) that the joint probability density function of q successive points being at $t_1 < t_2 < \cdots < t_q$ following an immediately preceding point at t is given by

$$\frac{1}{p_1(t)} \sum \frac{(-1)^k}{k!} \int_{(t, t_q)^k} \cdots \int p_{k+q+1}(t, t_1, \ldots, t_q, u_1, \ldots, u_k)\, du_1 \cdots du_k$$

(see Macchi, 1971).

By analogy with the transition probability functions of (2.1.3), one might be led to consider a *conditional intensity function* $\gamma(t, \omega)$ defined by

$$\text{Prob}\{dN(t) = 1\,|\,\mathcal{F}_t\} = \gamma(t, \omega)\, dt \qquad (2.2.13)$$

where $\mathcal{F}_t$ denotes the Borel field generated by the variates $N(u)$, $u \le t$. Such a $\gamma(t, \omega)$ need not always exist (an example of nonexistence is given by Segall and Kailath, 1975); however, when it does exist it often provides a very useful variate. If A is an event of $\mathcal{F}_t$, then from (2.2.13)

$$\text{Prob}\{dN(t) = 1\,|\,A\} = \int_A \gamma(t, \omega)\, dP(\omega)\, dt/\text{Prob}\{A\}$$

and so

$$\text{Prob}\{A\,|\,dN(t) = 1\} = \left[\int_A \gamma(t, \omega)\, dP(\omega)\right]\Big/p_1(t) \qquad (2.2.14)$$

where $p_1(t)$ is the intensity of the process at time t. The left-hand side of (2.2.14) may be interpreted as the *Palm probability* $P(A\,|\,t)$. Expression (2.2.14) indicates that the probability measure $P(A\,|\,t)$ is absolutely continuous with respect to the probability measure P over $\mathcal{F}_t$ and the Radon–Nikodyn derivative is $\gamma(t, \omega)/p_1(t)$. (In the stationary case this result is developed by Papangelou, 1974.)

In the case in which the distributions of (2.2.8) are absolutely continuous

$$\gamma(t, \omega) = -g_{N(t)}(t; \tau_j, j < N(t))/G_{N(t)}(t; \tau_j, j < N(t))$$
$$= -\frac{\partial}{\partial t} \log G_{N(t)}(t; \tau_j, j < N(t)) \qquad (2.2.15)$$

for $\tau_{N(t)-1} \le t < \tau_{N(t)}$ and

$$G_J(t; t_j, j < J) = \exp\left\{-\int_{t_{J-1}}^{t} \gamma(u, \omega)\, du\right\} \qquad (2.2.16)$$

(e.g., Snyder, 1975, p. 245). An important use of the conditional intensity is in calculating the joint probability function for the number of points in an interval $[0, T]$ and their locations. From (2.2.16) the expression is

$$\exp\left\{-\int_0^T \gamma(t, \omega)\, dt + \int_0^T \log \gamma(t, \omega)\, dN(t, \omega)\right\} \qquad (2.2.17)$$

(see Rubin, 1972; Vere-Jones, 1975b).

Defining

$$\Gamma(t, \omega) = E\{N(t)\,|\,\mathcal{F}_t\} = \int_0^t \gamma(u, \omega)\, du$$

the process $N(t, \omega) - \Gamma(t, \omega)$ may be seen to be a martingale. This occurrence leads to many useful results (see Chou and Meyer, 1975; Segall and Kailath, 1975). Among the results is the fact that the process N may be transformed into a Poisson process N^* by the random time change

$$\Gamma^{-1}(u) = \inf\{t\colon \Gamma(t) > u\}$$

Specifically, the process $N^*(u) = N(\Gamma^{-1}(u))$ is Poisson with rate 1 (see Bremaud, 1974; Meyer, 1971; Aalen, 1975; Papangelou, 1974). In fact, the result is strongly suggested by expression (2.2.17) and may be seen from (2.2.15).

The product densities $p_K(\cdot)$ of (2.2.4) may be determined from $\gamma(t, \omega)$ as

$$p_K(t_1, \ldots, t_K) = E\{\gamma(t_1, \omega)\} \prod_{1<k\le K} E\{\gamma(t_k, \omega) | \text{points at } t_1, \ldots, t_{k-1}\}$$

(see Snyder, 1975, p. 275). In particular the intensity or rate function is simply given by $E\gamma(t, \omega)$.

The conditional intensity function, as defined by (2.2.13), was based on $\mathscr{F}_t$, the Borel field generated by $\{N(u), u \le t\}$. On occasion it is convenient to use a larger Borel field $\mathscr{G}_t \supset \mathscr{F}_t$ (see Segall and Kailath, 1975). This allows one to study point processes that depend, for example, on some signal of interest.

It is by now apparent that there are many intertwined yet different manners of introducing stochastic point processes. The particular one adopted in any given situation relates to the peculiarities of the situation.

In the case in which the process is *stationary*, certain simplifications occur. The joint probability functions of (2.2.2) are now invariant under shifts of time:

$$f(n_1, \ldots, n_K; I_1 + t, \ldots, I_K + t) = f(n_1, \ldots, n_K; I_1, \ldots, I_K)$$

for t such that the intervals $I + t$ contain no negative values. The product densities of (2.2.4) depend on one less argument

$$p_K(t_1, \ldots, t_K) = p_K(t_1 - t_K, \ldots, t_{K-1} - t_K)$$

The zero probability function is invariant under shifts

$$\phi(I + t) = \phi(I) \tag{2.2.18}$$

for t such that $I + t$ contains no negative values. The probability generating functional (2.2.6) now has the property

$$G[S^t\xi] = G[\xi] \tag{2.2.19}$$

where $S^t\xi(u) = \xi(t + u)$. Conversely, invariance as in (2.2.18) and (2.2.19) implies stationarity of the corresponding process.

The process has so far been considered to have domain $[0, \infty)$. As in the ordinary time series situation, in the stationary case the process may be extended to be stationary on the whole line $(-\infty, \infty)$. This may be demonstrated by the Kolmogorov extension theorem; alternatively, the zero probability function may be used to construct the extended process.

In the case in which the process is characterized by the location τ_0 of the initial point and the interpoint distances $y_1, y_2, \ldots$, stationarity has a very important consequence. Namely, $y_1, y_2, \ldots$ is a stationary sequence conditional on the location of τ_0, i.e., under the Palm measure $P(\ |\tau_0 = t)$. This may be seen directly from (2.2.12). General proofs may be found in the works by Ryll-Nardzewski (1961) and Neveu (1968). In this stationary case, the Palm measure $P(\ |\tau_0 = t)$ will not depend on t, and on many occasions τ_0 is taken to be at 0.

A stationary N may also be thought of as a random counting measure with the property $N(I + t, \omega) = N(I, U^t\omega)$, for a semigroup of measure preserving U^t of Ω. Using these maps U^t, simple expressions may be set down relating the Palm and unconditional probability measures, namely

$$P(A|0) = \int \left\{ \int_0^1 I_A(U^t\omega)\, dN(t) \right\} dP(\omega)$$

$$P(A) = p_1 \int \left\{ \int_0^{y_1} I_A(U^{-t}\omega)\, dt \right\} dP(\omega|0) \tag{2.2.20}$$

where U^t is a typical map, p_1 the intensity of the process, y_1 the distance to the first positive point, A is in $B(\Omega)$, and I_A denotes the indicator function of the event A (see de Sam Lazaro and Meyer, 1975, Ryll-Nardzewski, 1961).

A deterministic approach to the foundations of point processes in a "stationary" case is also available. Consider an increasing sequence of points $0 \le \tau_0 < \tau_1 < \tau_2 < \cdots$ along the half-line. Let $N(t)$ denote the number of τ_k in the interval $[0, t]$. One now proceeds by assuming that limits such as

$$\lim_{T\to\infty} N(T)/T, \qquad \lim_{T\to\infty} \left\{ \int_0^T [N(y + u) - N(t)]\, dN(t) \right\} \Big/ T \tag{2.2.21}$$

exist (see Brillinger, 1973). The limit in (2.2.21) may also be interpreted as that of

$$\lim_{T\to\infty} \{\text{number of } (j, k) \text{ with } \tau_j - \tau_k \le u\}/T$$

Function spaces analogous to those introduced at the end of Section 2.1 are also of use in the point process case. The Hilbert space $L_2(N; t)$ spanned by linear combinations of the $N(u)$, $u \le t$, is introduced for prediction purposes in the article by Vere-Jones (1974). The space $N_2(N; t)$ may be introduced in a directly analogous manner, and also the reproducing kernels.

The kernel (2.1.7) appears in Ruelle (1969) in a point process version.

Certain ordinary series may be associated with a given point process in a direct manner, for example,

$$X(t) = \int a(t, u)\, dN(u)$$

$$X(t) = N(t, t+1]$$

$$X(t) = \int \exp\{-\mu(t-u)\}\, dN(u)$$

The last series is discussed in some detail for the stationary case in the article by Vere-Jones (1974).

2.3. Other Formal Definitions

The preceding two sections have provided separate discussions of the foundations of ordinary time series analysis and point process analysis. In fact, there are several methods by which one may simultaneously lay foundations for time series analysis and point process analysis.

Both types of processes may be discussed as particular cases of the *random additive set functions* of Bochner (1960). In Bochner's theory one deals with random interval functions $\alpha(I, \omega)$, I an interval of $[0, \infty)$, ω in Ω with $(\Omega, B(\Omega), P)$ a probability space, and

$$\alpha(I \cup J, \omega) = \alpha(I, \omega) + \alpha(J, \omega) \quad \text{with probability 1}$$

for I and J disjoint intervals. The interval representation $N(I, \omega)$ of a point process is immediately seen to be of this character. If the ordinary series $X(t, \omega)$ is such that

$$\int_I E|X(t)|\, dt < \infty \tag{2.3.1}$$

for bounded intervals I, then an additive set function may be associated with X by setting

$$\alpha(I, \omega) = \int_I X(t)\, dt \tag{2.3.2}$$

Bochner (1960) develops an extension theorem, defines a characteristic functional, discusses stationarity, and develops an harmonic analysis for such random interval functions. Brillinger (1972) develops further aspects and considers the particular case of point processes in some detail.

An essentially equivalent development would result from considering time series and point processes as particular kinds of *random signed measures* on $[0, \infty)$. In fact, a point process corresponds directly to a random measure. A time series with nonnegative values corresponds to a random measure through the relationship (2.3.2).

The theories of time series and point processes may be subsumed under the theory of *random generalized functions* (random Schwartz distributions). Let $\mathcal{D}$ denote the space of infinitely differentiable functions $\phi(t)$ of compact support. A continuous linear functional $Y(\phi)$ on $\mathcal{D}$ is called a generalized function (Schwartz distribution). A measurable map from a probability space $(\Omega, B(\Omega), P)$ to the space of generalized functions is called a random generalized function (see Yaglom, 1962; Gelfand and Vilenkin, 1964; Brillinger, 1974). Its image may be denoted $Y(\phi, \omega)$. A time series $X(t)$, satisfying (2.3.1), gives rise to the random generalized function

$$X(\phi, \omega) = \int \phi(t) X(t, \omega)\, dt, \qquad \phi \text{ in } \mathcal{D} \tag{2.3.3}$$

A point process $N(t)$ gives rise to the random generalized function

$$N(\phi, \omega) = \int \phi(t)\, dN(t, \omega), \qquad \phi \text{ in } \mathcal{D} \tag{2.3.4}$$

(see Daley, 1969). The generalized function of (2.3.4) may alternatively be denoted by

$$\sum_j \delta(t - \tau_j) \tag{2.3.5}$$

where $\delta(\cdot)$ is the Dirac delta function (a generalized function!) and the τ_j are the positions of the points of the process. Further developments in the theory of generalized random functions include work by Rozanov (1969), who developed linear prediction theory and defined the space $L_2(Y; t)$ spanned by the $Y(\phi, \omega)$ with ϕ zero to the right of t; Kailath (1971), who considered the reproducing kernel Hilbert space associated with the kernel $R(\psi, \phi) = EY(\psi, \omega)Y(\phi, \omega)$; and Hida (1970), who considered the Hilbert space associated with the kernel $R(\psi, \phi) = E \exp\{iY(\psi - \phi, \omega)\}$.

The point processes considered so far are particular cases of *marked point processes*. These are point processes with auxiliary values (marks) associated with each point. If the marks are constant, a marked point process is equivalent to an ordinary point process. If the marks are real valued, the process may be thought of as being piecewise constant, but jumping (up or down) by the value of the mark at the points of the process. If the mark is M_j at the point τ_j, then the marked point process may be denoted

$$\sum_j M_j\, \delta(t - \tau_j) \tag{2.3.6}$$

a notation extending that of (2.3.5). If the mark M_j is k-dimensional, then the marked point process may be thought of as a standard point process with domain R^{k+1}, i.e., a point process with points at the (τ_j, M_j). A vector-valued point process may be thought of as a marked point process with the mark giving the type of point occurring at a given time. A point process with multiple points may be thought of as a marked point process with the mark giving the multiplicity of the point. A useful generalized process related to the one of (2.3.6) is provided by

$$\sum_j X(\tau_j)\,\delta(t-\tau_j) \tag{2.3.7}$$

where $X(t)$ is an ordinary time series. The process of (2.3.7) corresponds to sampling the series X at stochastic times.

Incidentally, if the moments $EX(t)^k$ of an ordinary time series or the moments $EN(t)^k$ of a point process grow more slowly than some power of t as $t \to \infty$, then Fourier transforms of the moments and indeed of the processes themselves are well defined using the theory of generalized functions. This fact leads to a direct definition of certain useful parameters of the processes, even in the nonstationary case.

The sample paths of both the time series and the point processes considered are members of $D[0, \infty)$, the *Skorokhod space* of real-valued functions on $[0, \infty)$ which are right continuous and have left limits. In consequence, general results, such as central limit theorems, developed for random variables with values in $D[0, \infty)$ will apply directly to the processes considered in this chapter. One general reference is the book by Billingsley (1968).

Further, the processes considered are particular cases of processes with a more general time parameter, for example, of a *spatial series*

$$X(t_1, \ldots, t_p), \qquad t_1, \ldots, t_p \text{ in } [0, \infty)$$

or a *spatial point process*

$$N(I), \qquad I \subset B[0, \infty)^p$$

with $N(I)$ giving the number of points lying in the p-dimensional interval I. A not so obvious correspondence is the one between random hyperplanes in R^p and random points in the hypercylinder $R \times [0, 2\pi)^{p-1}$ (see Krickeberg, 1974). The correspondence is through the characterization of a hyperplane by its angles of orientation and distance from the origin.

Processes of structure simpler than that considered so far in this chapter are often very useful and illustrative. In particular, consider the discrete time series

$$X(th), \qquad t = 0, 1, 2, \ldots$$

for some $h > 0$. As $h \downarrow 0$, the behavior of these series comes close to that of continuous time series. The particular case in which $X(th)$ takes on only the values 0 and 1 is a simple analog of a point process. This analogy is made use of by Ryll-Nardzewski (1961), Breiman (1968), and Lewis (1970) for example.

In the stationary case, both $N(t, \omega)$ and $\int_0^t X(u, \omega)\,du$ are examples of processes with *stationary increments* and *helices*. A variety of useful results have been developed for such processes (e.g., see Doob, 1953, p. 551; Yaglom, 1962; de Sam Lazaro and Meyer, 1975; Masani, 1972, including the definition of Palm measures in the case in which the process is non-decreasing (see Geman and Horowitz, 1973). Further, in the "stationary" case, it is possible to construct a generalized harmonic analysis for individual generalized functions if so desired (see Pfaffelhuber, 1975). Certain limits are assumed to exist in this development. The results may be specialized to obtain results either for individual time series or individual step functions.

2.4. Parameters for Ordinary Time Series

The preceding sections have been concerned with the complete characterization of time series and point processes. In fact, much information is often contained in certain simple parameters related to the processes. In this context parameters will be viewed as quantities providing descriptive features of a probability distribution of interest.

Basic parameters of a time series include the *moment functions*

$$m_K(t_1, \ldots, t_K) = E\{X(t_1) \cdots X(t_K)\} \tag{2.4.1}$$

$K = 1, 2, \ldots$, when these exist. In some circumstances the moment functions taken in totality characterize the distribution of a time series. In terms of the finite-dimensional distribution functions (2.1.1), the moment function (2.4.1) is given by

$$\int \cdots \int x_1 \cdots x_K F_K(dx_1, \ldots, dx_K; t_1, \ldots, t_K)$$

Under regularity conditions (e.g., $E|X(t)|^K < C^K$, for some finite C and K sufficiently large), the characteristic functional (2.1.4) may be expanded in terms of the moment functions:

$$C[\xi] = \sum_{K=0}^{\infty} \frac{i^K}{K!} \int \cdots \int \xi(t_1) \cdots \xi(t_K) m_K(t_1, \ldots, t_K)\,dt_1 \cdots dt_K \tag{2.4.2}$$

and the individual moment functions may be determined by evaluating the Gateaux derivatives of $C[\xi]$. The moment functions are seen to be symmetric in their arguments from the definition (2.4.1).

Of the moment functions, those of first and second order, $m_1(t) = EX(t)$ and $m_2(t_1, t_2) = EX(t_1)X(t_2)$, are used most often. $m_1(t)$ gives the level about which the series fluctuates at time t. Its units are those of X; $m_2(t, t) = E|X(t)|^2$ gives an indication of the expected magnitude of the series at time t. The function $m_2(t_1, t_2)$ is nonnegative definite, i.e.,

$$\sum_j \sum_k \alpha_j \alpha_k m_2(t_j, t_k) \geq 0 \tag{2.4.3}$$

for times t_j and scalars α_j. That (2.4.3) is true is immediate, since the expression on the left is

$$E\left|\sum_k \alpha_k X(t_k)\right|^2$$

The functions $m_1(t)$ and $m_2(t_1, t_2)$ are most useful in the linear analysis of time series and in the anlysis of Gaussian series.

In cases in which the values of a series are becoming less and less dependent (statistically) as they are becoming further separated in time, the moment functions possess certain factorization properties such as

$$\lim_{U \to \infty} m_{J+K}(t_1, \ldots, t_J, t_{J+1} + U, \ldots, t_{J+K} + U) = m_J(t_1, \ldots, t_J) m_K(t_{J+1}, \ldots, t_{J+K}) \tag{2.4.4}$$

Of course, for an m-dependent series (wherein collections of values more than m time units apart are statistically independent) expression (2.4.4) holds without the limit, provided $U > m$.

The degree of dependence of the values of a time series is more directly measured by the *cumulant functions*

$$c_K(t_1, \ldots, t_K) = \operatorname{cum}\{X(t_1), \ldots, X(t_K)\} \tag{2.4.5}$$

when they exist. These are defined, for example, by the functional expansion

$$\log C[\xi] = \sum_{k=1}^{\infty} \frac{i^K}{K!} \int \cdots \int \xi(t_1) \cdots \xi(t_K) c_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K \tag{2.4.6}$$

and have the property that $c_K(t_1, \ldots, t_K) = 0$ whenever any subset of the variates $X(t_1), \ldots, X(t_K)$ is independent of the remaining variates (for a discussion of this and further properties of cumulant functions, see Brillinger, 1975a). The degree of dependence of a time series (or degree of mixing) may be described by specifying the rate at which its cumulant functions fall to 0 as the $|t_k|$ increase.

As in the case of the moment functions, the *mean function* $c_1(t) = EX(t) = m_1(t)$ and the *autocovariance function*

$$c_2(t_1, t_2) = \operatorname{cov}\{X(t_1), X(t_2)\} = m_2(t_1, t_2) - m_1(t_1) m_1(t_2)$$

are the ones most often used in practice. The autocovariance function is also nonnegative definite as

$$\operatorname{var}\left\{\sum_k \alpha_k X(t_k)\right\} = \sum_j \sum_k \alpha_j \alpha_k c_2(t_j, t_k) \geq 0$$

The *variance function*, $\operatorname{var} X(t) = c_2(t, t)$, measures the extent of fluctuations of the series $X(t)$ about its mean level $c_1(t)$ at time t.

It will be seen in Section 2.6 that both the moment and cumulant functions transform in a direct manner when a series is subjected to a linear transformation. Expressions for the moments and cumulants of polynomials in the values of a time series may also be set down, however not so directly as in the linear case.

The moment and cumulant functions are based on expected values. There are other expected values, based on time series, of some interest. Suppose that the finite-dimensional distribution function (2.1.1) is absolutely continuous with density

$$f_K(x_1, \ldots, x_K; t_1, \ldots, t_K) \tag{2.4.7}$$

Then it may be written as

$$E\{\delta(X(t_1) - x_1) \cdots \delta(X(t_K) - x_K)\} \tag{2.4.8}$$

where $\delta(\cdot)$ is the Dirac delta function. Expression (2.4.8) suggests how the density function (2.4.7) might be estimated by using an approximate delta function. This is useful to do in the case in which it is possible that a time series has finite-dimensional distributions of a specific form and it is desired to check this possibility.

Another expected value which has a direct connection with the parameters of point processes is provided by

$$E\left\{\int_I \cdots \int_{I \neq t_k} \delta(X(t_1) - a) \cdots \delta(X(t_K) - a) |X'(t_1)| \cdots |X'(t_K)|\, dt_1 \cdots dt_K\right\} \tag{2.4.9}$$

where $X'(t)$ denotes the derivative of $X(t)$. Expression (2.4.9) represents the Kth factorial moment of the number of crossings of the level a, by the series $X(t)$, in the time interval I. If $g_K(x_1, \ldots, x_K, y_1, \ldots, y_K; t_1, \ldots, t_K)$ denotes the probability density of the variate $\{X(t_1), \ldots, X(t_K), X'(t_1), \ldots, X'(t_K)\}$ then, under regularity conditions, expression (2.4.9) is given by

$$\int_I \cdots \int_I \int_0^{\infty} \cdots \int_0^{\infty} y_1 \cdots y_K g_K(a, \ldots, a, y_1, \ldots, y_K; t_1, \ldots, t_K) \times dy_1 \cdots dy_K\, dt_1 \cdots dt_K \tag{2.4.10}$$

Matters of this sort are investigated by Leadbetter (1972).

Of course, conditional expected values such as

$$E\{X(t+h) \mid X(u), u \le t\}, \qquad h > 0 \tag{2.4.11}$$

are very important in prediction theory. Specifically, the random variable V of (2.4.11) provides the minimum of

$$E\,|X(t+h) - V|^2$$

for measurable functions V of $\{X(u), u \le t\}$. Alternatively, (2.4.11) is the variate, based on the past, having the highest correlation with $X(t+h)$. The value (2.4.11) may be viewed as a limit of expected values based on the conditional distributions (2.1.3). Were the process Markov, these would simplify. In addition in the Markov case, one would be interested in estimating the transition probability elements:

$$\text{Prob}\{x < X(t+h) < x + dx \mid X(t) = x_0\}$$

Other parameters of occasional use in dealing with time series include the fractiles (especially the median) and the fractile ranges.

In a situation wherein one has a number of realizations of a time series available for analysis, one can proceed to estimate the parameters just defined directly. More often, however, one has available only a segment of a single realization. Even in the case in which the series involved is approximately stationary, one can still set down reasonable estimates of parameters of interest.

In the stationary case, the moment functions of (2.4.1) and (2.4.5) depend on one fewer argument. Specifically,

$$m_K(t_1, \ldots, t_K) = m_K(t_1 - t_K, \ldots, t_{K-1} - t_K, 0) \tag{2.4.12}$$

$$c_K(t_1, \ldots, t_K) = c_K(t_1 - t_K, \ldots, t_{K-1} - t_K, 0) \tag{2.4.13}$$

This has the advantage that now a sensible estimate of (2.4.12) may be based on the expression

$$\int_I \cdots \int_I X(t_1 + u) \cdots X(t_K + u)\, du$$

where the domain of integration is over the region of observation. In the case of $K = 1$, the mean level of the series may be estimated by

$$\left[\int_0^T X(u)\, du\right] \Big/ T \tag{2.4.14}$$

given values of the process in the interval $(0, T)$.

Because of the simplification of expressions (2.4.12) and (2.4.13), the "0"

appearing in them will be dropped and, in the stationary case, the definitions

$$m_K(u_1, \ldots, u_{K-1}) = E\{X(t+u_1) \cdots X(t+u_{K-1})X(t)\}$$

$$c_K(u_1, \ldots, u_{K-1}) = \text{cum}\{X(t+u_1), \ldots, X(t+u_{K-1}), X(t)\}$$

will be stated.

The mean function $m_1 = m_1(t) = EX(t) = c_1(t) = c_1$ is constant in the stationary case. The series is seen to be fluctuating about a constant level c_1. The autocovariance function $c_2(u) = \text{cov}\{X(t+u), X(t)\}$ is particularly useful in practice. Its shape provides many indications of the inherent character of the series X. The value $c_2(u)$ measures the degree of linear statistical dependence of values of the series lag u time units apart.

In the case in which $c_2(u)$ dies off sufficiently rapidly as $|u| \to \infty$, the *power spectrum* at frequency λ, $f_2(\lambda)$, may be defined by

$$f_2(\lambda) = (2\pi)^{-1} \int_{-\infty}^{\infty} \exp\{-i\lambda u\} c_2(u)\, du \tag{2.4.15}$$

for $-\infty < \lambda < \infty$, with the accompanying inverse relationship:

$$c_2(u) = \int_{-\infty}^{\infty} \exp\{iu\lambda\} f_2(\lambda)\, d\lambda \tag{2.4.16}$$

In the general case in which $c_2(u)$ is finite and continuous at 0, Khintchine (1934) has shown the existence of a unique finite spectral measure F on the real line satisfying

$$c_2(u) = \int_{-\infty}^{\infty} \exp\{iu\lambda\} F(d\lambda) \tag{2.4.17}$$

This last means, for example, that $f_2(\lambda)$ of (2.4.15) is nonnegative. Interpretations of $f_2(\lambda)$ and $F(d\lambda)$ will be provided in a later section after the notion of a bandpass filter has been introduced. The usefulness of the parameter $f_2(0)$ is suggested by the remark that the large sample variance of the statistic (2.4.14) is $2\pi f_2(0)/T$.

In the case in which

$$\int \cdots \int |c_K(u_1, \ldots, u_{K-1})|\, du_1 \cdots du_{K-1} < \infty \tag{2.4.18}$$

as, for example, when values at a distance of the series are only weakly dependent, the cumulant spectrum of order K of the series may be defined by

$$f_K(\lambda_1, \ldots, \lambda_{K-1}) = (2\pi)^{-K+1} \int_{-\infty}^{\infty} \cdots \int \exp\{-i(\lambda_1 u_1 + \cdots + \lambda_{K-1} u_{K-1})\} \times c_K(u_1, \ldots, u_{K-1})\, du_1 \cdots du_{K-1} \tag{2.4.19}$$

with the inverse relationship

$$c_K(u_1, \ldots, u_{K-1}) = \int_{-\infty}^{\infty} \cdots \int \exp\{i(u_1\lambda_1 + \cdots + u_{K-1}\lambda_{K-1})\} \times f_K(\lambda_1, \ldots, \lambda_{K-1})\, d\lambda_1 \cdots d\lambda_{K-1} \tag{2.4.20}$$

for $K = 2, 3, \ldots$. In general, there is no relationship analogous to (2.4.17) for $K > 2$. However, provided the functions $c_K(u_1, \ldots, u_{K-1})$ are of slow growth, the Fourier transform of (2.4.18) may be defined as a generalized function, with an inverse relationship analogous to (2.4.20).

In fact, a stationary series $X(t)$ has a direct Fourier representation. Under the conditions leading to the representation (2.4.17), Cramér (1942) develops the result

$$X(t) = \int_{-\infty}^{\infty} \exp\{it\lambda\}\, dZ(\lambda) \tag{2.4.21}$$

with the "integral" defined as a limit in mean of Stieltjes sums and with $Z(\lambda)$ a stochastic function having the properties

$$E\, dZ(\lambda) = \delta(\lambda)c_1\, d\lambda$$

$$\operatorname{cov}\{dZ(\lambda_1), dZ(\lambda_2)\} = \delta(\lambda_1 - \lambda_2)F(d\lambda_1)\, d\lambda_2$$

where $F(d\lambda)$ appears in (2.4.17). In the case in which (2.4.18) is satisfied,

$$\operatorname{cum}\{dZ(\lambda_1), \ldots, dZ(\lambda_K)\} = \delta(\lambda_1 + \cdots + \lambda_K) f_K(\lambda_1, \ldots, \lambda_{K-1})\, d\lambda_1 \cdots d\lambda_K \tag{2.4.22}$$

This last expression provides an interpretation for the cumulant spectrum of order K. It is proportional to the joint cumulant of order K of the increments $dZ(\lambda_1), \ldots, dZ(\lambda_K)$ with $\lambda_1 + \cdots + \lambda_K = 0$. In the case $K = 2, f_2(\lambda)$ is seen to be proportional to the variance of $dZ(\lambda)$.

Returning to the *Cramér representation* (2.4.21), the time series $\exp\{it\lambda\}\, dZ(\lambda)$ appearing therein is called the *component of frequency* λ in the series $X(t)$. If the series $X(t)$ is extended to be stationary on the whole real line, then the function $Z(\lambda)$ may be determined as the limit in mean of the variate

$$(2\pi)^{-1} \int_{-T}^{T} X(t) \frac{1 - \exp\{-i\lambda t\}}{-it} dt$$

as $T \to \infty$.

In the case in which the series is not stationary, but its moments are of slow growth, a representation of the form (2.4.21) may still be given, but now

involving random generalized functions. Now, however, the reduction of the values to the hyperplane $\lambda_1 + \cdots + \lambda_K = 0$ occurring in (2.4.22) will no longer take place. In the case in which the second-order moment function of $Z(\lambda)$ is of bounded variation, the series $X(t)$ is called *harmonizable*.

A class of time series that are nonstationary, strictly speaking, but yet that possess many of the properties of stationary series is comprised of those of the form $X(t) = a(t) + \varepsilon(t)$ where $a(t)$ is a nonstochastic, nonconstant function and $\varepsilon(t)$ is a zero mean stationary time series. For this $X(t)$, the mean function is $a(t)$, and so nonconstant; however, all of its cumulant functions of order greater than 1 are those of a stationary series. In the case in which $a(t)$ is slowly varying, $a(\cdot)$ is called the *trend function* of the series. In engineering literature $a(t)$ is sometimes called the *signal* and $\varepsilon(t)$ the *noise*. Commonly used forms for $a(t)$ include $a(t) = \alpha + \beta t$ and $a(t) = \rho \cos(\omega t + \phi)$. The first is called a *linear trend model*, the second a *hidden periodicity model*.

The cumulant spectra of a stationary series may be obtained from the expansion of the logarithm of the characteristic functional, specifically,

$$\begin{aligned}\log C[\xi] &= \log E \exp\left\{i \int \xi(t)X(t)\, dt\right\} \\ &= \log E \exp\left\{i \int \hat{\xi}(\lambda)\, dZ(\lambda)\right\} \\ &= \sum_{k=1}^{\infty} \frac{i^k}{k!} \int \cdots \int \hat{\xi}(\lambda_1) \cdots \hat{\xi}(\lambda_{k-1}) \overline{\hat{\xi}(\lambda_1 + \cdots + \lambda_{k-1})} \\ &\quad \times f_k(\lambda_1, \ldots, \lambda_{k-1})\, d\lambda_1 \cdots d\lambda_{k-1}\end{aligned}$$

where $\hat{\xi}(\lambda) = \int \xi(t) \exp\{i\lambda t\}\, dt$.

2.5. Parameters for Point Processes

Point process parameters directly analogous to the product moment functions (2.4.1) of ordinary time series analysis, are provided by the *product densities* $p_K(t_1, \ldots, t_K)$ of

$$p_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K = E\{dN(t_1) \cdots dN(t_K)\} \tag{2.5.1}$$

$K = 1, 2, \ldots$, when they exist. In practice, because of the step function character of $N(t)$, it is only reasonable to expect the absolute continuity of (2.5.1) to occur when $t_1, \ldots, t_K$ are distinct. It has already been assumed that the points of the process are isolated. Suppose, in fact, that

$$\operatorname{Prob}\{N(I) = n\} < L_\delta |I|^n \tag{2.5.2}$$

for some finite L_δ, whenever I is an interval of length $|I| \leq \delta$. Then, for example,

$$E\{dN(t)\, dN(t)\} = E\, dN(t) + O(dt^2)$$
$$= p_1(t)\, dt + O(dt^2) \tag{2.5.3}$$

with similar reductions occurring in the case of general K whenever some of the t_k are equal (see Brillinger, 1972).

The parameter $p_1(t)$ is called the *mean rate* of the process at time t. A representation for it, as an alternative to (2.5.1), is provided by

$$\int_0^t p_1(u)\, du = EN(t) \tag{2.5.4}$$

It is a parameter with an intimate connection to the appearance of realizations of the process. When $p_1(t)$ is large, there tend to be many points in the neighborhood of t. When $p_1(t)$ is small, points tend to be rare near t. In the case $K = 2$, it follows from (2.5.1) that

$$\int_0^t \int_0^t p_2(u_1, u_2)\, du_1\, du_2 = E\left\{ \iint_{u_1 \neq u_2} dN(u_1)\, dN(u_2) \right\}$$
$$= E\left\{ \left(\int_0^t dN(u) \right)^2 - \int_0^t (dN(u))^2 \right\}$$
$$= E\{N(t)(N(t) - 1)\} \tag{2.5.5}$$

using (2.5.3). That is, the integral of $p_2(t_1, t_2)$ gives the second factorial moment of $N(t)$. In general,

$$\int_I \cdots \int_I p_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K = EN(I)^{(K)} \tag{2.5.6}$$

where $N^{(K)} = N(N-1)\cdots(N-K+1)$, $K = 1, 2, \ldots$. If desired, the ordinary moment $EN(I)^K$ may be deduced from the factorial moments of order less than or equal to K. The *factorial moment generating functional* of the point process N is defined to be

$$E \exp\left\{ \int \log[1 + \xi(t)]\, dN(t) \right\} \tag{2.5.7}$$

If the points of the process are represented by $\tau_0, \tau_1, \ldots$, then (2.5.7) may be written

$$E\left\{ \prod_j [1 + \xi(\tau_j)] \right\} = E\left\{ \sum_k \frac{1}{k!} \prod \xi(\tau_{j_1}) \cdots \xi(\tau_{j_k}) \right\} \tag{2.5.8}$$

where $\tau_{j_1}, \ldots, \tau_{j_k}$ are distinct. It follows that the functional (2.5.7) has the representation

$$\sum_k \frac{1}{k!} \int \cdots \int \xi(t_1) \cdots \xi(t_k) p_k(t_1, \ldots, t_k)\, dt_1 \cdots dt_k \tag{2.5.9}$$

In cases in which increments of the process, well separated in time, are only weakly dependent it is often more convenient to consider functions q_k defined by the expansion of the logarithm of (2.5.7), namely by

$$\log E \exp\left\{ \int \log[1 + \xi(t)]\, dN(t) \right\}$$
$$= \sum_k \frac{1}{k!} \int \cdots \int \xi(t_1) \cdots \xi(t_k) q_k(t_1, \ldots, t_k)\, dt_1 \cdots dt_k \tag{2.5.10}$$

These functions are called the *cumulant densities* and have the interpretations

$$q_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K = \operatorname{cum}\{dN(t_1), \ldots, dN(t_K)\} \tag{2.5.11}$$

when the t_k are distinct. They provide direct measures of the degree of statistical dependence of a process N. Those of order 1 and 2 are given by $q_1(t) = p_1(t)$, $q_2(t_1, t_2) = p_2(t_1, t_2) - p_1(t_1)p_2(t_2)$. The factorial cumulant of order K of $N(I)$ may be written

$$\int_I \cdots \int_I q_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K \tag{2.5.12}$$

Conditions under which expansions such as (2.5.9) and (2.5.10) may be manipulated may be found in an article by Wescott (1972).

Under a condition like (2.5.2), small increments $dN(t)$ generally contain no or one point. It follows then that the product density has the further important interpretation

$$p_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K = \operatorname{Prob}\{dN(t_1) = 1, \ldots, dN(t_K) = 1\} \tag{2.5.13}$$

for distinct t_k and $K = 1, 2, \ldots$. In particular,

$$\operatorname{Prob}\{dN(t) = 0\} = 1 - p_1(t)\, dt$$
$$\operatorname{Prob}\{dN(t) = 1\} = p_1(t)\, dt$$

The first product density, $p_1(t)$, is seen to give the mean intensity with which points occur at time t. Its units are those of points per unit time. It is an important parameter to estimate in practice.

The product density of order 2, $p_2(t_1, t_2)$, provides a measure of the

intensity with which points simultaneously occur near t_1 and near t_2. A related useful function is provided by

$$p_2(t_1, t_2)/p_1(t_1) = \text{Prob}\{dN(t_2) = 1 \,|\, N\{t_1\} = 1\}/dt_2 \qquad (2.5.14)$$

This gives the intensity with which points are occurring near t_2 given that there is a point at t_1. It may be thought of as the mean rate, $p_1(t_2 \,|\, t_1)$, of the Palm process $N(t \,|\, t_1)$. The measure (2.5.14) may also be written in terms of cumulant densities as

$$q_1(t_2) + q_2(t_1, t_2)/q_1(t_1) \qquad (2.5.15)$$

In the case in which well-separated increments of the process are approximately independent, the measure (2.5.14) is seen to be approximately $p_1(t_2) = q_1(t_2)$ since the second term in (2.5.15) is negligible.

The first- and second-order densities are important in constructing a further useful point process parameter, the *index of dispersion*,

$$I(t) = \text{var } N(t)/EN(t)$$
$$= 1 + \int_0^t \int_0^t q_2(u_1, u_2)\, du_1\, du_2 \Big/ \int_0^t q_1(u)\, du \qquad (2.5.16)$$

This index measures, to some extent, the departure of the process N from being a Poisson process (for which the index is identically 1). Also useful is the *variance time curve*

$$V(t) = \int_0^t q_1(u)\, du + \int_0^t \int_0^t q_2(u_1, u_2)\, du_1\, du_2 \qquad (2.5.17)$$

A further parameter, based on expression (2.5.14), is the *renewal function*

$$U(u; t) = E\{N(t, t+u] \,|\, N\{t\} = 1\}$$
$$= \left\{\int_t^{t+u} p_2(v, t)\, dv\right\} \Big/ p_1(t) \qquad (2.5.18)$$

giving the expected number of points (or renewals) within u time units of a point at t. The higher order factorial moments

$$E\{N(t, t+u]^{(K)} \,|\, N\{t\} = 1\}, \qquad K = 1, 2, \ldots$$

find some use in practice as well. They may be determined by integrating the conditional product densities (2.2.11).

Certain other parameters based on the Palm process $N(\cdot \,|\, t)$ may be defined. The *survivor function* (or distribution of lifetime) is given as

$$\text{Prob}\{\text{next point from an event at } t \text{ occurs after } u \text{ time units}\}$$
$$= \text{Prob}\{N(t, t+u] = 0 \,|\, N\{t\} = 1\}$$
$$= 1 - G(u; t) \qquad (2.5.19)$$

where $G(u; t) = \text{Prob}\{N(t, t+u] \geq 1 \,|\, N\{t\} = 1\}$. The *hazard function* or *force of mortality* is given by

$$\mu(u; t) = \text{Prob}\{dN(t+u) = 1 \,|\, N\{t\} = 1 \quad \text{and} \quad N(t, t+u] = 0\}/dt$$
$$= g(u; t)/[1 - G(u; t)] \qquad (2.5.20)$$

where g is the derivative of G. The inverse relation to this last expression is

$$1 - G(u; t) = \exp\left\{-\int_t^{t+u} \mu(v; t)\, dv\right\}$$

These parameters are often used in reliability theory and population mathematics. Many specific functional forms have been proposed for them. The probability (2.5.19) is an example of the general Palm probability

$$\text{Prob}\{N(t, t+u] = n \,|\, N\{t\} = 1\}$$

As suggested by the result (2.2.10), these probabilities are directly related to unconditional probabilities. Further details concerning this are given by Cramér *et al.* (1971) and Daley and Vere-Jones (1972).

The parameters introduced so far have been based directly on the step functions $N(t)$. Certain other parameters are more easily introduced through the sequence τ_j, $j = 0, 1, 2, \ldots$, of locations of points and the sequence $y_j = \tau_{j+1} - \tau_j$, $j = 1, 2, \ldots$, of interpoint distances. The *forward recurrence time* is the variate $\tau_{N(t)} - t$. Its distribution is given by

$$\text{Prob}\{\tau_{N(t)} - t \leq u\} = 1 - \text{Prob}\{N(t, t+u] = 0\}$$
$$= 1 - \phi((t, t+u]) \qquad (2.5.21)$$

This distribution is useful in extrapolating the behavior of the process ahead of t. Expression (2.5.21) may be given in terms of the product densities as

$$1 - \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_t^{t+u} \cdots \int p_k(t_1, \ldots, t_k)\, dt_1 \cdots dt_k$$

In the reverse direction, one has the *backward recurrence time* $t - \tau_{N(t)-1}$. Its distribution is given by

$$\text{Prob}\{t - \tau_{N(t)-1} \leq v\} = 1 - \text{Prob}\{N[t-v, t) = 0\}$$

The joint distribution of the two variates $\tau_{N(t)} - t$ and $t - \tau_{N(t)-1}$ is sometimes of use, as is the distribution of their sum $\tau_{N(t)} - \tau_{N(t)-1}$. The distribution of this last variate should not be confused with that of any of the $\tau_{j+1} - \tau_j$ unconditionally or based on a Palm measure.

Using the individual times of events, an alternative expression may be given for (2.5.14). Let $\sigma_1, \sigma_2, \ldots$ denote the times of successive points follow-

ing t_1 for the Palm process $N(t\,|\,t_1)$. Let $g_k(\sigma; t)$ denote the probability density of σ_k. Then it is clear from first principles that

$$\text{Prob}\{\text{point in } (t, t+dt)\,|\,\text{point at } t_1\} = \sum_k g_k(t; t_1)\,dt$$

and so

$$\frac{p_2(t_1, t)}{p_1(t_1)} = \sum_k g_k(t; t_1) \tag{2.5.22}$$

In the stationary case, the product densities and the cumulant densities depend on one less parameter than in the general case. Specifically,

$$E\{dN(t+u_1)\cdots dN(t+u_{K-1})\,dN(t)\} = p_K(u_1, p_{pK}(u_1, \ldots, u_{K-1})\,du_1\cdots du_{K-1}\,dt \tag{2.5.23}$$

$$\text{cum}\{dN(t+u_1), \ldots, dN(t+u_{K-1}), dN(t)\} = q_K(u_1, \ldots, u_{K-1})\,du_1\cdots du_{K-1}\,dt \tag{2.5.24}$$

for $u_1, \ldots, u_{K-1}, 0$ distinct and $K = 1, 2, \ldots$. This reduction has the important implication that plausible estimates of the parameters can now be based on single realizations of the process. For example, the mean rate may be estimated by $\hat{p}_1 = N(T)/T$ given the data stretch $N(t)$, $0 < t \le T$. As in the case of ordinary time series, the first- and second-order parameters seem to be the most important ones in practice. It is convenient to describe these by

$$\begin{aligned} E\,dN(t) &= p_1\,dt = q_1\,dt \\ E\{dN(t+u)\,dN(t)\} &= \{\delta(u)p_1 + p_2(u)\}\,du\,dt \\ \text{cov}\{dN(t+u), dN(t)\} &= \{\delta(u)q_1 + q_2(u)\}\,du\,dt \end{aligned} \tag{2.5.25}$$

In the case in which

$$\int_{-\infty}^{\infty} |q_2(u)|\,du < \infty \tag{2.5.26}$$

the *power spectrum* of the point process N may be defined to be

$$\begin{aligned} f_2(\lambda) &= (2\pi)^{-1}\left[\int_{-\infty}^{\infty} \exp\{-i\lambda u\}\,\text{cov}\{dN(t+u), dN(t)\}\right]/dt \\ &= (2\pi)^{-1}q_1 + (2\pi)^{-1}\int_{-\infty}^{\infty} \exp\{-i\lambda u\}q_2(u)\,du \end{aligned} \tag{2.5.27}$$

for $-\infty < \lambda < \infty$. The parameter λ here is called the *frequency*. One important manner in which the power spectrum of a process satisfying (2.5.26) differs from that of an ordinary series satisfying (2.4.18) is that

$$\lim_{\lambda\to\infty} f_2(\lambda) = q_1/2\pi \neq 0 \tag{2.5.28}$$

The power spectrum of a point process is similar to that of an ordinary time series in that it is symmetric, $f_2(-\lambda) = f_2(\lambda)$, and nonnegative. Inverse relations to the definition (2.5.27) are provided by (2.5.27) and

$$q_2(u) = \int_{-\infty}^{\infty} \exp\{iu\lambda\}[f_2(\lambda) - q_1/2\pi]\,d\lambda$$

In the higher order case one sets down the assumption

$$\int\cdots\int |q_K(u_1, \ldots, u_{K-1})|\,du_1\cdots du_{K-1} < \infty \tag{2.5.29}$$

and the definition of the *cumulant spectrum*

$$f_K(\lambda_1, \ldots, \lambda_{K-1}) = (2\pi)^{-K+1}\left[\int\cdots\int \exp\{-i(\lambda_1 u_1 + \cdots + \lambda_{K-1}u_{K-1})\} \times \text{cum}\{dN(t+u_1), \ldots, dN(t+u_{K-1}), dN(t)\}\right]/dt \tag{2.5.30}$$

$-\infty < \lambda_K < \infty$, $K = 2, 3, \ldots$. The definition is completed by setting $f_1 = p_1$. From (2.5.10) it is clear that the cumulant spectra may also be defined as "coefficients" in the expansion

$$\log E\exp\left\{\int \xi(t)\,dN(t)\right\} = \sum_{k=0}^{\infty}\frac{1}{k+1!}\int\cdots\int \hat{\xi}(\lambda_1)\cdots\hat{\xi}(\lambda_k)\hat{\xi}(-\lambda_1 - \cdots - \lambda_k) \times f_{k+1}(\lambda_1, \ldots, \lambda_k)\,d\lambda_1\cdots d\lambda_k$$

where $\hat{\xi}(\lambda) = \int \exp\{i\lambda t\}\xi(t)\,dt$.

In the general case in which $N(t)$ is continuous in mean square, Kolmogorov (1940) has developed the representations

$$\text{cov}\{N(t+u), N(t)\} = \int_{-\infty}^{\infty} \frac{\exp\{iu\lambda\} - 1}{i\lambda}\,F(d\lambda) \tag{2.5.31}$$

and

$$N(t) = \int_{-\infty}^{\infty} \frac{\exp\{it\lambda\} - 1}{i\lambda}\,dZ(\lambda) \tag{2.5.32}$$

where $F(d\lambda)$ is a measure satisfying $\int (1+\lambda^2)^{-1}F(d\lambda) < \infty$ and $Z(\lambda)$ is a random process with the properties

$$\begin{aligned} E\,dZ(\lambda) &= \delta(\lambda)p_1\,d\lambda \\ \text{cov}\{dZ(\lambda_1), dZ(\lambda_2)\} &= \delta(\lambda_1 - \lambda_2)F(d\lambda_1)\,d\lambda_2 \end{aligned}$$

In the case in which (2.5.29) holds, $Z(\lambda)$ has the further properties

$$\text{cum}\{dZ(\lambda_1), \ldots, dZ(\lambda_K)\} = \delta(\lambda_1 + \cdots + \lambda_K)f_K(\lambda_1, \ldots, \lambda_{K-1})\,d\lambda_1\cdots d\lambda_K \tag{2.5.33}$$

$K = 1, 2, \ldots$. This relationship is identical to the corresponding result (2.4.22) for ordinary time series. Expression (2.5.33) provides one possible means of interpreting the cumulant spectra of a point process.

The spectral representation (2.5.32) gives a means of interpreting certain ordinary series associated with point processes. Namely, if

$$X(t) = \int a(t-u)\, dN(u)$$

then

$$X(t) = \int \exp\{it\lambda\} A(\lambda)\, dZ(\lambda)$$

where $A(\lambda) = \int \exp\{-i\lambda t\} a(t)\, dt$. Using (2.5.33) it is now clear that the Kth-order cumulant spectrum of the series X is given by

$$A(\lambda_1) \cdots A(\lambda_{K-1}) A(-\lambda_1 - \cdots - \lambda_{K-1}) f_K(\lambda_1, \ldots, \lambda_{K-1}) \quad (2.5.34)$$

It is seen that if $A(\cdot)$ does not vanish, then estimates of the cumulant spectra of N may be determined directly from estimates of the cumulant spectra of X.

In the case of a stationary point process, a whole new family of cumulant functions and spectra may be associated with the process through the Palm measure. It was seen in Section 2.2 that the random sequence $y_j = \tau_{j+1} - \tau_j$ of interpoint distances is a stationary sequence for the Palm measure $P(\cdot|0)$. From (2.2.12) or (2.2.20) the mean level of this sequence may be seen to be $Ey_j = 1/p_1$. As would be expected a high rate for the point process corresponds to a small average between-point distance.

The second-order product density and the power spectrum of the process N may be determined from the characteristics of the time series y_j, $j = 1, 2, \ldots$, via the relationship (2.5.22). The autocovariance function and the power spectrum of the series y_j, $j = 1, 2, \ldots$, may be determined from the characteristics of the process N via the relationship (2.2.12). These matters are discussed by Cox and Lewis (1966, Section 4.6). The relationships are most useful in the case of particular processes, for example renewal processes. Data benefit from an analysis both in terms of the step function $N(t)$ and the interpoint distances y_j. It must be remembered, however, that the latter is stationary for the measure $P(\cdot|0)$, and in practice it is the process $N(t)$ which is observed, not $N(\cdot|0)$. However, the two processes lead to the same asymptotic results, so in situations in which the data set is large, the sequence y_j, $j = 1, 2, \ldots$, may typically be treated as stationary.

2.6. Operations on Time Series

A variety of physical operations are applied to time series before and after their collection. Sometimes these operations are applied deliberately,

but sometimes not. In any case, it is important that the effect of the operations be understood to the greatest degree possible.

One very common operation is that of forming some linear combination

$$Y = \int \xi(t) X(t)\, dt$$

for a suitable function $\xi(\cdot)$. The variate Y has already appeared in the definition of the characteristic functional (2.1.4). It is apparent that Y is well defined when, for example, $\int |\xi(t)| E\{|X(t)|\}\, dt < \infty$. The moments and cumulants of Y are given by

$$EY^K = \int \cdots \int \xi(t_1) \cdots \xi(t_K) m_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K$$

$$\operatorname{cum}_K Y = \int \cdots \int \xi(t_1) \cdots \xi(t_K) c_K(t_1, \ldots, t_K)\, dt_1 \cdots dt_K \quad (2.6.1)$$

when they exist.

The preceding operation formed a real-valued variate from a given time series. By far the most important operations form time series from given time series. Consider, for example, the linear operation

$$Y(t) = \int a(t, u) X(u)\, du \quad (2.6.2)$$

for a suitable function $a(t, u)$. The moment and cumulant functions of this new series are given by

$$\int \cdots \int a(t_1, u_1) \cdots a(t_K, u_K) m_K(u_1, \ldots, u_K)\, du_1 \cdots du_K$$

$$\int \cdots \int a(t_1, u_1) \cdots a(t_K, u_K) c_K(u_1, \ldots, u_K)\, du_1 \cdots du_K \quad (2.6.3)$$

while the characteristic functional is given by

$$E \exp\left\{ i \int Y(t)\, d\theta(t) \right\} = C\left[\int^{\tau} \left\{ \int a(u, v)\, d\theta(u) \right\} dv \right] \quad (2.6.4)$$

in terms of the characteristic functional of the series $X(t)$. The characteristics of the two series are seen to be directly connected in this case.

A related linear operation is described by requiring that the series $Y(t)$ satisfies

$$\int b(t, u) Y(u)\, du = X(t) \quad (2.6.5)$$

for some function $b(t, u)$, given the series $X(t)$. For example, the series $Y(t)$ might be defined to be the solution of the differential equation

$$\sum_{k=0}^{K} \beta_k \frac{d^k Y(t)}{dt^k} = X(t)$$

in the case in which $Y(t)$ is K times differentiable. This is of the form (2.6.5) with

$$b(t, u) = \sum_{k=0}^{K} \beta_k \frac{d^k \delta(t-u)}{dt^k}$$

In the case in which there exists $a(v, t)$ such that

$$\int a(v, t) b(t, u)\, dt = \delta(v-u)$$

the solution of (2.6.5) may be written as (2.6.2) and the relations (2.6.3) and (2.6.4) are seen to apply to this series $Y(t)$ also.

On other occasions, the series $Y(t)$ may be determined from the series $X(t)$ through a polynomial relationship of the form

$$Y(t) = \sum_{j=0}^{J} \int \cdots \int a_j(t, u_1, \ldots, u_j) X(u_1) \cdots X(u_j)\, du_1 \cdots du_j \quad (2.6.6)$$

(see Wiener, 1958). Relations analogous to (2.6.3) may be written down for the series $Y(t)$ in the case in which the series $X(t)$ is Gaussian; however, matters are complicated in the non-Gaussian case (see Shiryaev, 1963).

The operations introduced so far have been smooth in character. On occasion one wishes to apply a strongly nonlinear operator, for example, $Y(t) = G[X(t)]$, with $G[\cdot]$ discontinuous. An important example is the *hard limiter*

$$\begin{aligned} Y(t) &= \quad 1 \quad \text{if} \quad X(t) > 0 \\ &= -1 \quad \text{if} \quad X(t) < 0 \end{aligned}$$

or the *discretizer*

$$Y(t) = j \quad \text{if} \quad jh \le X(t) < (j+1)h$$

given $h > 0$, for $j = 0, \pm 1, \ldots$. Such operations must be investigated separately of any general theory.

The functions $a(\cdot)$ of (2.6.2), $b(\cdot)$ of (2.6.5), and $a_j(\cdot)$ of (2.6.6) have been assumed nonstochastic. On occasion it is appropriate to consider them to be stochastic and to consider series $Y(t)$ defined by expressions of the form

$$Y(t) = \int a(t, u, \omega) X(u, \omega)\, du$$

Expressions for the characteristics of the series $Y(t)$ in terms of those of the series $X(t)$ now appear to exist only for very specific series and operations.

An operation of alternate character from those considered so far is that of time substitution, specifically

$$Y(t) = X(\Gamma(t))$$

where Γ is a nondecreasing function with domain $[0, \infty)$. In the case in which Γ is linear, the series $Y(t)$ behaves essentially like the series $X(t)$. Nonlinear Γ leads to quite drastically altered characteristics on many occasions, especially when Γ is a random nonlinear function.

So far the discussion has been concerned with operations on a single real-valued series. In practice one often has to consider operations on vector-valued series. As in the case of a single series, linear operations may be described quite succinctly. For example, suppose the characteristic functional is given for the r vector-valued series $\{X_1(t), \ldots, X_r(t)\}$, namely

$$C[\theta_1, \ldots, \theta_r] = E \exp\left\{ i \int X_1(t)\, d\theta_1(t) + \cdots + i \int X_r(t)\, d\theta_r(t) \right\} \quad (2.6.7)$$

Then the generalization of the relationship (2.6.4) to a series such as

$$Y(t) = \int a_1(t, u) X_1(t)\, dt + \cdots + \int a_r(t, u) X_r(t)\, dt$$

is immediately apparent. In particular, the characteristic functional of the superposed series

$$Y(t) = X_1(t) + \cdots + X_r(t)$$

is $C[\theta, \ldots, \theta]$. In the case in which the series are independent and identically distributed it is $C[\theta]^r$. In many circumstances, this latter result may be used to prove that the superposed series is asymptotically Gaussian as $r \to \infty$.

An important class of operations is now introduced. Consider an operation $\mathscr{A}$ carrying real-valued functions $X(t)$, $-\infty < t < \infty$, over into real-valued functions $Y(t)$, $-\infty < t < \infty$, with the properties of (i) linearity

$$\mathscr{A}[\alpha_1 X_1 + \alpha_2 X_2](t) = \alpha_1 \mathscr{A}[X_1](t) + \alpha_2 \mathscr{A}[X_2](t)$$

and (ii) time invariance

$$\mathscr{A}[S^u X](t) = \mathscr{A}[X](t+u)$$

where S^u denotes the shift operator $[S^u X(t) = X(t+u)]$. Then $\mathscr{A}$ has the property of carrying complex exponentials, $\exp\{i\lambda t\}$, over into complex exponentials, specifically $\mathscr{A}[e](t) = A(\lambda) e(t)$, $-\infty < t < \infty$, for $e(t) = \exp\{i\lambda t\}$ (e.g., see Brillinger, 1975a, Lemma 2.7.1). An operation with properties (i) and (ii) is called a (linear) *filter*. The function $A(\lambda)$, $-\infty < \lambda < \infty$, is called

the *transfer function* of the filter, and it is generally complex-valued. Its amplitude, $|A(\lambda)|$, is called its *gain*, and its argument, $\arg A(\lambda)$, is called its *phase*. Many of the operations applied to a time series, by an analyst or nature, seem to be well approximated by filters.

An important class of filters of ordinary time series takes the form

$$Y(t) = \int X(t-u)\, da(u) \tag{2.6.8}$$

for some definition of the integral. The transfer function of this filter is

$$A(\lambda) = \int \exp\{-i\lambda u\}\, da(u) \tag{2.6.9}$$

If $a(u)$ is differentiable, its derivative is called the *impulse response* of the filter, since it is the output series when $X(t) = \delta(t)$ is the input. Important filters include the *bandpass filter* at frequency v with bandwidth Δ and gain A where

$$\begin{aligned} A(\lambda) &= A \qquad \text{for} \quad |\lambda \pm v| < \Delta/2 \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{2.6.10}$$

and the *Hilbert transform*, with

$$\begin{aligned} A(\lambda) &= -i, \qquad \lambda > 0 \\ &= 0, \qquad \lambda = 0 \\ &= i, \qquad \lambda < 0 \end{aligned}$$

If the series $X(t)$ has a spectral representation

$$X(t) = \int \exp\{it\lambda\}\, dZ(\lambda) \tag{2.6.11}$$

for some process $Z(\lambda)$, then the output of the filter with transfer function $A(\lambda)$ has representation

$$Y(t) = \int \exp\{it\lambda\} A(\lambda)\, dZ(\lambda)$$

The output of the bandpass filter (2.6.10) with small Δ is seen to be approximately $A[\exp\{itv\}\, \Delta Z(v) + \exp\{-itv\}\, \Delta Z(-v)]$.

In the case of a stationary series $X(t)$, it is apparent that the series $Y(t)$ is also stationary with mean $A(0)EX(t)$ and cumulant spectra

$$A(\lambda_1) \cdots A(\lambda_{K-1}) A(-\lambda_1 - \cdots - \lambda_{K-1}) f_K(\lambda_1, \ldots, \lambda_{K-1}) \tag{2.6.12}$$

In particular, the power spectrum of the series $Y(t)$ is $|A(\lambda)|^2 f_2(\lambda)$. If

$$Y_k(t) = \int \exp\{it\lambda\} A_k(\lambda)\, dZ(\lambda)$$

$k = 1, \ldots, K$, then

$$\begin{aligned} \operatorname{cum}\{Y_1(t), \ldots, Y_K(t)\} = \int \cdots \int & A_1(\alpha_1) \cdots A_{K-1}(\alpha_{K-1}) A_K(-\alpha_1 - \cdots - \alpha_{K-1}) \\ & \times f_K(\alpha_1, \ldots, \alpha_{K-1})\, d\alpha_1 \cdots d\alpha_{K-1} \end{aligned} \tag{2.6.13}$$

This last result provides an interpretation of the Kth-order cumulant spectrum f_K. Let $A_k(\cdot)$ be the transfer function of a bandpass filter at frequency λ_k, with small bandwidth Δ, $k = 1, \ldots, K$, and suppose $\lambda_1 + \cdots + \lambda_K = 0$. Then the cumulant of (2.6.13) is seen to be proportional to $\operatorname{Re}\{f_K(\lambda_1, \ldots, \lambda_{K-1})\}$. In words, the real part of the Kth-order cumulant spectrum at $(\lambda_1, \ldots, \lambda_{K-1})$ may be interpreted as the scaled Kth-order joint cumulant of the output of a bank of narrow bandpass filters at frequencies $\lambda_1, \ldots, \lambda_K$ where $\lambda_1 + \cdots + \lambda_K = 0$. In particular, the power spectrum at frequency λ may be interpreted as the scaled variance of the output of a narrow bandpass filter at frequency λ. The imaginary part of f_K may be obtained by taking the transfer function of one of the filters to be the product of the Hilbert transform and the previous narrow bandpass filter and once again considering the joint cumulant (2.6.13).

An operation somewhat out of the ordinary may be defined by choosing $Y(t)$ to be the value that minimizes

$$\int_{t-U}^{t+U} |X(v) - Y(t)|\, dv$$

for some U. The series $Y(t)$ could be called the *running median* of the series $X(t)$.

2.7. Operations on Point Processes

In general, it appears that the class of interesting operations is much larger in the case of point processes than in the case of time series. Another distinction is that the typical operation appears to have stochastic character. The description of a specific operation is sometimes best given in terms of the step functions $N(t)$, sometimes best in terms of the sequence $\tau_0, \tau_1, \ldots$ of locations of points, sometimes in terms of the series $y_1, y_2, \ldots$ of intervals between events, and sometimes in terms of $\gamma(t, \omega)$, the conditional intensity function. The concern in this section is with operations carrying point processes over into point processes.

. The analog of the linear operation (2.6.2) seems to be one carrying a process with points at $\tau_0, \tau_1, \ldots$ over into one with points at $\tau_j + u_i(\tau_j)$, $i = 1, 2, \ldots, I, j = 0, 1, \ldots$, for given functions $u_i(t)$. A representation for the new process is

$$\sum_j \sum_i \delta(t - \tau_j - u_i(\tau_j)) \tag{2.7.1}$$

the mean rate for which is

$$\sum_i p_1(t - u_i(t))$$

and the higher order product densities are given by

$$\sum_{i_1} \cdots \sum_{i_K} p_K(t_1 - u_{i_1}(t_1), \ldots, t_K - u_{i_K}(t_K))$$

In the case in which $I = 1$, this operation is called a *displacement*. In the case in which the $u_i(t)$ do not depend on t, the image series will be stationary when the domain series is.

This particular operation does not seem to have a great deal of use in practice; however, its stochastic analog, with the $u_i(t)$ random, is very important. For example, consider the case in which the point τ_j is replaced by the points $\tau_j + u_{ji}$, $i = 1, \ldots, I_j, j = 0, 1, \ldots$, with the u_{ji} and the I_j random. In the case in which $I_j = 1$, the operation might be thought of as corresponding to the action of a service system, with no waiting time, in which the τ_j are the arrival times of customers, u_j the service time of the jth customer, and $\tau_j + u_j$ his departure time. The process of points $\{\tau_j + u_{ji}, i = 1, \ldots, I_j, j = 0, 1, \ldots\}$ is called a *cluster process*. If $N^j(A)$ denotes the number of u_{ji} in the set A for $i = 1, \ldots, I_j$, then $N^j(\cdot)$ is seen to be a point process and the cluster process is seen to have the representation

$$\sum_j N^j(A - \tau_j) \tag{2.7.2}$$

Supposing that the process $N^j(\cdot)$ has intensity function $p_1^*(t)$ for each j and is independent of the process N, the process (2.7.2) is seen to have intensity function

$$\int p_1(u) p_1^*(t - u)\, du$$

This process is well defined if this integral is finite. In the case in which the process $N^j(\cdot)$ has probability generating functional $G^*[\xi]$ for each j, and where $N(\cdot), N^0(\cdot), N^1(\cdot), \ldots$ is a sequence of independent processes, the process (2.7.2) is seen to have probability generating functional

$$G[G^*[S^t\xi]] \tag{2.7.3}$$

where S^t is the shift operator. Expression (2.7.3) shows directly that the process (2.7.2) is stationary when the process N is.

A useful operation in the study of point processes is that of time substitution involving the replacement of the process $N(t)$ by the process $N(\Gamma(t))$, for $\Gamma(t)$ a nondecreasing function, which may either be fixed or stochastic. If the process $N(t)$ jumps at the points τ_j, then the image process jumps at the points

$$\sigma_j = \inf\{\sigma: \Gamma(\sigma) = \tau_j\} = \Gamma^{-1}(\tau_j)$$

In Section 2.2, it was seen that a random time substitution could, in certain circumstances, transform a given process into a Poisson process of rate 1. The intervals between the points of the two processes are related by

$$\tau_{j+1} - \tau_j = \Gamma(\sigma_{j+1}) - \Gamma(\sigma_j)$$

In the case in which the rate of the transformed process is high, this last relationship may be written approximately

$$\tau_{j+1} - \tau_j = \gamma(\sigma_j)(\sigma_{j+1} - \sigma_j) \tag{2.7.4}$$

In the case in which $\Gamma(\cdot)$ is deterministic, the product densities of the image process are given by

$$p_K(\Gamma(t_1), \ldots, \Gamma(t_K))\gamma(t_1) \cdots \gamma(t_K) \tag{2.7.5}$$

where $\gamma(t)$ is the derivative of $\Gamma(t)$. The probability generating functional of the image process is given by $G[\xi \circ \Gamma^{-1}]$. It is clear that, in this case, unless $\Gamma(\cdot)$ is linear, a stationary process will not be transformed into a stationary process.

A very useful means of generating further point processes from given point processes is to employ a stochastic $\Gamma(\cdot)$, independent of $N(\cdot)$. This particular procedure leads to the doubly stochastic Poisson when N is Poisson. From expression (2.7.5), the product densities of the transformed process are seen to be given by

$$E\{p_K(\Gamma(t_1), \ldots, \Gamma(t_K))\gamma(t_1) \cdots \gamma(t_K)\} \tag{2.7.6}$$

The time substitution of Section 2.2 depended on the process N in a very direct manner, however, and an expression such as (2.7.6) does not apply. An important use of time substitutions is in the transformation of a given nonstationary point process into one that is (approximately) stationary.

Moving on to a different class of transformations, suppose that a process $N(\cdot)$ has the representation

$$\sum_j \delta(t - \tau_j) \tag{2.7.7}$$

Consider the process with representation

$$\sum_j I(\tau_j)\,\delta(t-\tau_j) \tag{2.7.8}$$

where $I(\tau_j)$ is a random indicator variable taking on the value 0 or 1. The operation of forming the process (2.7.8) from the process (2.7.7) is called *thinning, skipping,* or *random deletion.* If $E\{I(\tau_j)\,|\,\tau_j\} = \pi(\tau_j)$, then the intensity function of the process (2.7.8) is given by

$$p_1(t)\pi(t) \tag{2.7.9}$$

If, in addition, $I(\tau_0), I(\tau_1), \ldots$ given N is a sequence of independent random variables, then the probability generating functional of the process (2.7.8) is given by

$$E\prod\{\xi(\tau_j)^{I(\tau_j)}\} = \underset{N}{E}\prod\{\xi(\tau_j)\pi(\tau_j)+1-\pi(\tau_j)\}$$
$$= G[\xi\pi+1-\pi] \tag{2.7.10}$$

Defining the factorial cumulant generating functional of the process by

$$H[\xi] = \log G[1+\xi] = \sum_{k>0}\frac{1}{k!}\int\cdots\int \xi(t_1)\cdots\xi(t_k)q_k(t_1,\ldots,t_k)\,dt_1\cdots dt_k$$

the effect of the operation may be described more directly by

$$H[\xi] \to H[\pi\xi] \tag{2.7.11}$$

The cumulant densities are given, in terms of those of the process N, by

$$\pi(t_1)\cdots\pi(t_K)q_K(t_1,\ldots,t_K) \tag{2.7.12}$$

A related operation is that of *censoring* in which, for example, only every other point is retained. In terms of the function $I(\cdot)$, this operation might correspond to requiring $I(\tau_{j+1}) = 1 - I(\tau_j)$.

An important transformation of a point process occurs when one is fed into a physical counter meant to record the times of its events. Many counters have the unpleasant but not unexpected property of becoming inoperative for a brief period (called the *dead time*) following each registration of an event. This property is analogous to the one for recording apparatus for ordinary time series that do not respond if the series is changing too rapidly. Two important classes of counters are usually distinguished. *Type I* has the property of no output events occurring during the dead time whatever. *Type II* is such that events occurring during the dead time make the dead time begin again (the counter is paralyzable). Applying the operation of the type

II counter is one means used to decluster point process realizations (for a seismological example, see Udias and Rice, 1975). In analytic terms, a type I counter may be described by

$$\text{Prob}\{dN(t)=1\,|\,N(u), M(u); u\le t\} = \gamma(t,\omega)\,dt, \qquad \tau_{N(t)-1}+\Delta \le t$$
$$= 0 \qquad \text{otherwise} \tag{2.7.13}$$

where M denotes the input series, N the recorded series, Δ the dead time, and $\gamma(t,\omega)$ the conditional intensity of M. The operation corresponds to modulation by a 0–1 function. Clearly $p_2(u) = 0$ for $|u| \le \Delta$ here. A type II counter may be described by

$$\text{Prob}\{dN(t)=1\,|\,N(u), M(u); u\le t\} = \gamma(t,\omega), \qquad \sigma_{M(t)-1}+\Delta \le t$$
$$= 0 \qquad \text{otherwise}$$

with $\sigma_0, \sigma_1, \ldots$ denoting the times of input events. It is clear that the rate of the output of a type II counter may be determined from

$$\text{Prob}\{dN(t)=1\} = \text{Prob}\{M(t-\Delta, t] = 0 \text{ and } dM(t) = 1\}$$

as

$$p_1^*(t) = \sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\int_{t-\Delta}^{t}\cdots\int p_{k+1}(u_1,\ldots,u_k,t)\,du_1\cdots du_k$$

using the expansion (2.2.7). The higher order product densities may be determined in a similar manner.

An interesting aspect of the operations just given is that the sample path of the transformed process is in each case absolutely continuous with respect to the path of the original process.

A further form of transformation occurs when a point process is passed through a service system, i.e., the points of the original process are envisaged as the times at which customers arrive at a queueing system with, say, k servers. If all the servers are busy, a customer must wait for a free server. When a server becomes available, the customer first in line experiences a random service time. The image process of the transformation is taken to be the point process corresponding to the exit times of the customers. In the case in which there are an infinite number of servers, the output process is the cluster process mentioned earlier, with one member per cluster. There is an extensive literature concerning queueing theory of particular situations, e.g., Cox and Smith (1961), Benes (1963), and Prahbu (1965).

If τ_j denotes the arrival time of the jth customer, ω_j his waiting time, and σ_j his service time, then the transformed process may be represented by

$$\sum_j \delta(t - \tau_j - \omega_j - \sigma_j)$$

Generally the ω_j have quite complicated stochastic structure and this is what leads to difficulties in the analysis of queues. If $M(t)$ denotes the arrival process and $N(t)$ the departure, then the number of customers in the system at time t is $M(t) - N(t)$. The number being served at time t is $\min\{M(t) - N(t), k\}$. The times at which the number of customers in the system falls to 0 are important in the analysis. They are called *points of regeneration*, since the system starts anew at these times (in a certain sense).

Suppose next that a point process M has conditional intensity function defined by

$$\gamma(t, \omega)\, dt = \text{Prob}\{dM(t) = 1 \,|\, M(u), u \le t\}$$

$$= \gamma(t; \sigma_0, \ldots, \sigma_{M(t)-1})\, dt$$

where $\sigma_0, \sigma_1, \ldots$ are the locations of its points. Suppose $\beta(t, \omega) = \beta(t; \sigma_0, \ldots, \sigma_{M(t)-1})$ is a nonnegative function of the past of the process. The operation of forming a point process N with conditional intensity function

$$\beta(t, \omega)\gamma(t, \omega) \tag{2.7.14}$$

provided $E\{\beta(t, \omega)\gamma(t, \omega)\} < \infty$, is called *modulation*, discussed, e.g., by Cox (1972) and Varaiya (1975). From (2.2.16) the conditional distribution functions of the points $\tau_0, \tau_1, \ldots$ of the process N are given by

$$\text{Prob}\{\tau_J > t \,|\, \tau_j = t_j, j < J\}$$

$$= \exp\left\{-\int_{t_{J-1}}^{t} \beta(u; t_j, j < J)\gamma(u; t_j, j < J)\, du\right. \tag{2.7.15}$$

Equation (2.7.15) indicates a means by which a modulation process might be simulated given $\beta(t, \omega)$ and $\gamma(t, \omega)$. The standard Poisson process has $\gamma(t, \omega) \equiv 1$. Expression (2.7.12) shows that a point process with general conditional intensity $\beta(t, \omega)$, say, is a modulated version of the standard Poisson process. Cox (1972) considers the particular case of nonstochastic $\beta(t)$ in some detail. The joint probability distribution for the number of points of the process N in the interval $[0, T]$ and their locations may be determined from (2.7.15) and (2.2.17). Varaiya (1975) shows that, if the probability measures corresponding to the processes M and N are mutually absolutely continuous, then there exists a $\beta(t, \omega)$ of this character, in many circumstances.

He also shows that the likelihood ratio of the process N with respect to the process M, on the interval $[0, T]$, may be written

$$\exp\left\{\int_0^T \log \beta(t, \omega)\, dN(t, \omega) - \int_0^T [\beta(t, \omega) - 1]\gamma(t, \omega)\, dt\right\} \tag{2.7.16}$$

The operations considered so far have been defined on univariate point processes. In the case of vector-valued processes, $N_1(t), \ldots, N_r(t)$, referring to points of r different types, an important operation is that of *pooling* or *superposition*. Here one forms the process

$$N_1(t) + \cdots + N_r(t)$$

If the probability generating functional of the original process is

$$G[\xi_1, \ldots, \xi_r] = E \exp\left\{\int \log \xi_1(t)\, dN_1(t) + \cdots + \int \log \xi_r(t)\, dN_r(t)\right\} \tag{2.7.17}$$

then that of the superposed process is clearly $G[\xi, \ldots, \xi]$. In the case in which the component series are independent realizations of a process with probability generating functional $G[\xi]$, the superposed process has probability generating functional $G[\xi]^r$. In a variety of circumstances, this last result may be used to demonstrate that the limit, as $r \to \infty$, of superposed and re-time-scaled point processes is a Poisson process (e.g., see Vere-Jones, 1968). The conditional intensity is here the sum of the individual intensities.

Finally, it is clear that many of the operations of discrete time series analysis may be applied to the interval series $\{y_1, y_2, \ldots\}$ in order to obtain a further interval series and in consequence a new point process.

2.8. The Identification of Time Series Systems

By a *time series system* is meant the collection of a space of input series, a space of output series, and an operation carrying an input series over into an output series. A common form of a time series system is provided by the specification

$$Y(t) = \mathscr{A}[X](t) + \varepsilon(t) \tag{2.8.1}$$

where $\mathscr{A}[\cdot]$ is a deterministic operator of the kind discussed in Section 2.6, X denotes an input series, Y the output series, and ε an unobservable stochastic error series. The problem of system identification is that of determining $\mathscr{A}$, or essential properties of $\mathscr{A}$, from a stretch of input and corresponding output data $\{X(t), Y(t)\}$, $0 < t < T$. Part of the motivation for studying this problem is the desire to be able to indicate properties of output series

corresponding to specific input series. A system is called *time invariant* if the bivariate series $\{X(t), Y(t)\}$ is stationary when the series $X(t)$ is stationary. A system is called *causal* when $\mathscr{A}[X](t)$ depends only on $X(u)$, $u \le t$, and the future error $\varepsilon(v)$, $v > t$, is independent of the past $\{X(u), Y(u)\}$, $u \le t$.

In a few circumstances the identification problem has a fairly direct solution. For example, consider the (regression) system

$$Y(t) = \mu + \int X(t-u)\, da(u) + \varepsilon(t) \tag{2.8.2}$$

where $\varepsilon(t)$ is a zero mean stationary series independent of the stationary series $X(t)$. Then (2.8.2) leads to the relationship

$$f_{11}(\lambda) = A(-\lambda) f_{20}(\lambda) \tag{2.8.3}$$

where $A(\lambda) = \int \exp\{-i\lambda u\}\, da(u)$, $f_{20}(\lambda)$ is the power spectrum of the series X, and

$$f_{11}(\lambda) = (2\pi)^{-1} \int \exp\{-i\lambda u\} \operatorname{cov}\{X(t+u), Y(t)\}\, du \tag{2.8.4}$$

is the *cross spectrum* of the series X with the series Y. Expression (2.8.3) indicates that the transfer function $A(\lambda)$ of the linear filter of (2.8.2) may be estimated once estimates of the second-order spectra $f_{20}(\lambda)$ and $f_{11}(\lambda)$ are available. (The construction of such estimates is described by Brillinger, 1975.) This method of system identification was proposed by Wiener (1949).

The relationship (2.8.3) has a direct extension to a nonlinear system in one case. Suppose the system is determined by

$$Y(t) = \mu + \int X(t-u)\, da(u) + \iint X(t-u)X(t-v)\, db(u, v) + \varepsilon(t) \tag{2.8.5}$$

where $\varepsilon(t)$ is a zero mean stationary series independent of the stationary Gaussian series $X(t)$. The relationship (2.8.3) still holds for this system. In addition, one has

$$f_{21}(\lambda, \nu) = 2B(-\lambda, -\nu) f_{20}(\lambda) f_{20}(\nu) \tag{2.8.6}$$

where $B(\lambda, \nu) = \iint \exp\{-i(\lambda u + \nu v)\}\, db(u, v)$ and

$$f_{21}(\lambda, \nu) = (2\pi)^{-1} \iint \exp\{-i(\lambda u + \nu v)\} \operatorname{cum}\{X(t+u), X(t+v), Y(t)\}\, du\, dv \tag{2.8.7}$$

is a third-order joint cumulant spectrum of the bivariate series $\{X(t), Y(t)\}$. The functions $A(\lambda)$, $B(\lambda, \nu)$ may therefore be estimated once estimates of the second- and third-order spectra of the series $\{X(t), Y(t)\}$ are available.

Consider, next, the system defined by the equations

$$U(t) = \int X(t-u)\, da(u)$$

$$V(t) = G[U(t)]$$

$$Y(t) = \mu + \int V(t-u)\, db(u) + \varepsilon(t)$$

where $\varepsilon(t)$ is again a zero mean stationary series independent of the stationary Gaussian series $X(t)$, and where $G[\cdot]$ is a function from reals to reals. The system (2.8.2) corresponds to $G[\cdot]$ and $b(\cdot)$ identities. When $G[u]$ is a quadratic in u, this system is a particular case of (2.8.5). Now, for jointly normal variates U, V, W, and $G[\cdot]$ such that

$$E|G[U][1 + |U| + |V| + |W| + |VW| + |U|^2]| < \infty,$$

$$\operatorname{cov}\{V, G[U]\} = \operatorname{cov}\{V, U\} \operatorname{cov}\{U, G[U]\}/\operatorname{var} U$$

$$\operatorname{cum}\{W, V, G[U]\} = [\operatorname{cov}\{V, U\} \operatorname{cov}\{W, U\} \operatorname{cum}\{U, U, G[U]\}]/\operatorname{var}^2 U$$

Hence for the previous system, assuming the series $X(t)$ Gaussian

$$f_{11}(\lambda) = L_1 A(-\lambda) B(-\lambda) f_{20}(\lambda)$$

$$f_{21}(\lambda, \nu) = L_2 A(-\lambda) A(-\nu) B(-\lambda - \nu) f_{20}(\lambda) f_{02}(\nu)$$

where $L_1 = \operatorname{cov}\{U(0), V(0)\}/\operatorname{var} U(0)$ and

$$L_2 = \operatorname{cum}\{U(0), U(0), V(0)\}/\operatorname{var}^2 U(0).$$

When either of the filters $a(\cdot)$, $b(\cdot)$ is the identity, the transfer function of the other may therefore be determined, up to a constant of proportionality, by $f_{11}(-\lambda) f_{20}(\lambda)^{-1}$. In the general case, the previous relations lead to

$$|A(\lambda)| \propto \{|f_{XXY}(\lambda, -\lambda)|\}^{1/2} / f_{XX}(\lambda)$$

on setting $\nu = -\lambda$. If $\phi(\lambda) = \arg A(\lambda)$, $\psi(\lambda, \nu) = \arg\{f_{XXY}(\lambda, \nu)/f_{XY}(\lambda + \nu)\}$, then

$$\phi(\lambda) = \left\{2 \int_0^{\lambda} \phi(\alpha)\, d\alpha + \int_0^{\lambda} \psi(\alpha, \lambda - \alpha)\, d\alpha\right\} \Big/ \lambda$$

provides a recursive procedure for obtaining $\phi(\lambda)$ given $\phi(\alpha)$, $0 \le \alpha < \lambda$. Korenberg (1973) considered this system and developed related expressions for the case of $X(\cdot)$ Gaussian white noise and $G[u]$ a polynomial in u.

A classical procedure that has been used to estimate the transfer function $A(\lambda)$ of the system (2.8.2) is to take for input series $X(t) = \exp\{i\lambda t\}$ (i.e.,

separately input the series $\cos \lambda t$ and $\sin \lambda t$). Then the output series is

$$Y(t) \doteq \mu + A(\lambda) \exp\{i\lambda t\}$$

and $A(\lambda)$, $\lambda \neq 0$, may be estimated by

$$\left[\int_0^T [Y(t) - \hat{m}] \exp\{-i\lambda t\}\, dt\right] \Big/ T$$

for example, where $\hat{m}$ is the mean of the Y terms. This procedure has the disadvantage of requiring the use of a whole family of input series, $\exp\{i\lambda t\}$, covering the λ domain.

A variant of this procedure exists for the identification of a system such as (2.8.5) in which one takes the series $\exp\{i\lambda t\} \pm \exp\{ivt\}$ as input (see Brillinger, 1970, and the references cited therein).

It is worth noting that the filter with transfer function $A(\lambda)$ determined by (2.8.3) has an alternative interpretation. Given input and output series $X(t)$, $Y(t)$ of a system, consider the problem of determining the best linear approximant of the system, i.e., finding the $a(u)$ that minimizes

$$E\left[Y(t) - \mu - \int X(t-u)\, da(u)\right]^2 \tag{2.8.8}$$

In the case in which $\{X(t), Y(t)\}$ is a stationary series, it may be seen that one wants to choose μ and $A(\lambda)$ to satisfy (2.8.3) and $EY(t) = \mu + A(0)EX(t)$. When dealing with nonlinear filters, it is often useful to consider their approximate linear effect.

A system is called a *linear dynamic system* if it has the representation

$$\begin{aligned} dS(t) &= F(t)S(t)\, dt + G(t)X(t)\, dt + K(t)\, dW_1(t) \\ dY(t) &= H(t)S(t)\, dt + J(t)X(t)\, dt + L(t)\, dW_2(t) \end{aligned} \tag{2.8.9}$$

$0 \leq t < \infty$, where $F(t), \ldots, L(t)$ are fixed (matrix-valued) functions, $W_1(t)$, $W_2(t)$ independent noise series, $S(t)$ an unobservable (vector-valued) state series, $X(t)$ the input series, and $Y(t)$ the corresponding output series. A considerable literature exists concerning the theory of such systems (e.g., see the December, 1974, number of the *IEEE Transactions on Automatic Control*, Vol. 19). With the series and functions appearing vector-valued, the model is exceedingly general. The functions $F(t), \ldots, L(t)$ are parameters of the system. In some cases they are fixed functions of a common set of parameters. The problem of identification then is concerned with the estimation of the parameters of some canonical form of the system, given stretches of input and corresponding output. The description (2.8.9) is often called the *state space* or *state variable* model. The state process $S(t)$ is meant to represent the totality of information from the past and present input to be transmitted to the future output. In practice, one seeks minimal realizations of the system, wherein the dimension of $S(t)$ is as small as possible. An appealing aspect of the description (2.8.9) is its clear indication of the dynamic development of the system.

Much of the literature of state space models is concerned with obtaining expressions for the series

$$\hat{S}(t) = E\{S(t) \mid X(u),\ Y(u);\ u \leq t\}$$

In the case in which (i) $W_1(t)$, $W_2(t)$ are independent Wiener processes, (ii) $S(0)$ is a normal variate independent of the process $\{W_1, W_2\}$, (iii) the series $X(t)$ is fixed, the results of Kailath (1970) and Balakrishnan (1973) show that the likelihood ratio of the series $Y(t)$ relative to the process $L(t)W_2(t)$ is given by

$$\exp\left\{\int_0^T L(t)^{-2}(H(t)\hat{S}(t) + J(t)X(t))\, dt - \tfrac{1}{2}\int_0^T (H(t)\hat{S}(t) + J(t)X(t))^2 L(t)^{-2}\, dt\right\} \tag{2.8.10}$$

This variate is useful in problems of estimation and detection. An important role is played in the analysis of the system (2.8.9) by the *innovations process* defined as

$$dv(t) = dY(t) - H(t)\hat{S}(t)\, dt - J(t)X(t)\, dt$$

The identification of systems like (2.8.9) has generally proceeded by parametrizing them in some canonical manner, and then maximizing the likelihood or some other criterion function involving the parameters.

The system (2.8.9) is linear. In the case in which the $F(t), \ldots, L(t)$ are constant and the noise series W has stationary increments, the system is time invariant. When the noise processes are absent, the system corresponds to a linear filter with matrix-valued transfer function $J + H(i\lambda I - F)^{-1}G$. The various entries of this matrix are rational functions of λ.

There is one class of nonlinear systems that may be identified quite readily, when the experimenter has the freedom to use any input series. It is represented by

$$Y(t) = \sum_{K=0}^{\infty} \int \cdots \int_{u_k \text{ distinct}} a_K(u_1, \ldots, u_K) X(t-u_1) \cdots X(t-u_K)\, du_1 \cdots du_K + \varepsilon(t) \tag{2.8.11}$$

with $a_K(u_1, \ldots, u_K)$ symmetric. Suppose that the experimenter takes as input

series the (generalized) Gaussian series with mean 0 and $\text{cov}\{X(t+u), X(t)\} = \delta(u)$. Then as Wiener (1958) shows

$$a_K(u_1, \ldots, u_K) = E\{X(t-u_1)\cdots X(t-u_K)Y(t)\}/K!$$

for the u_k distinct. A number of the papers in the proceedings edited by McCann and Marmarelis (1975) are concerned with the details of (approximately) carrying out this identification procedure in practice.

Consider next a situation involving a trivariate time series $\{X_1(t), X_2(t), X_3(t)\}$. Suppose that there is an apparent association between the series X_1 and X_2, but it is felt that this association may simply be due to their individual associations with the series X_3. A system that allows an investigation of such a possibility is provided by

$$X_1(t) = \mu_1 + \int a_1(t-u)X_3(u)\,du + \varepsilon_1(t)$$
$$X_2(t) = \mu_2 + \int a_2(t-u)X_3(u)\,du + \varepsilon_2(t) \qquad (2.8.12)$$

where $\{\varepsilon_1, \varepsilon_2, X_3\}$ is a stationary series with the series X_3 independent of the series ε_1 and ε_2. In the case in which the series ε_1 and ε_2 were independent of each other, one could say that the association of the series X_1 and X_2 is simply due to the common influence of the series X_3.

Now, one measure of the degree of association of two stationary time series is provided by the *coherence* function defined as

$$|g_{12}(\lambda)|^2/[g_{20}(\lambda)g_{02}(\lambda)] \qquad (2.8.13)$$

where g_{20} and g_{02} are the power spectra of the series and g_{11} their cross spectrum. The values of the coherence function (2.8.13) may be shown to lie in the interval [0, 1], with 0 occurring if the series are independent and 1 occurring if the series are connected in a linear time invariant manner (see Brillinger, 1975a). For the system (2.8.12) it may be shown that the second-order spectra of the series $\varepsilon_1, \varepsilon_2$ are given by

$$g_{jk}(\lambda) = \{f_{jk0}(\lambda) - [f_{j01}(\lambda)f_{0k1}(\lambda)]\}/f_{002}(\lambda) \qquad (2.8.14)$$

$j + k = 2$, in terms of the second-order spectra of the process $\{X_1, X_2, X_3\}$. With the substitution (2.8.14), the expression (2.8.13) is called the *partial coherence* at frequency λ of the series X_1 and X_2 with the linear effects of the series X_3 removed. This function may be estimated once estimates of the second-order spectra are available.

2.9. The Identification of Point Process Systems

By a *point process system* is meant a collection of a space of input step functions $M(t)$, a space of output step functions $N(t)$, and an operation

carrying input functions over into output functions. A common form of the point process system is given by

$$N(I) = \mathscr{A}[M](I) + E(I) \qquad (2.9.1)$$

I in $B[0, \infty)$, where $\mathscr{A}$ is a point process operation of the kind discussed in Section 2.7 and $E(t)$ is a further point process. In contrast to the situation for ordinary time series systems, the operation $\mathscr{A}$ is generally stochastic [and it is sometimes convenient to consider the process $E(t)$ as part of it]. The problem of point process system identification is that of determining essential properties of the operator $\mathscr{A}$ from a stretch of input and output data $\{M(t), N(t)\}$, $0 < t \le T$. A system is called *time invariant* if the bivariate process $\{M(t), N(t)\}$, $0 \le t < \infty$, is stationary whenever the input process $M(t)$, $0 \le t < \infty$, is. An operator $\mathscr{A}$ is called *causal* when $\mathscr{A}[M_j](t)$ is the same for two realizations $M_j(\cdot)$ satisfying $M_1(v, \omega) = M_2(v, \omega)$ for $v > t$. The system (2.9.1) is called causal when the operator $\mathscr{A}$ is causal and the future increments $E(v) - E(t)$, $v > t$, are independent of the past $\{M(u), N(u)\}$, $u \le t$. A point process system will be said to have a *refractory period* if there exists a time interval immediately following an output event, during which there can be no further output events.

As an example of a point process system, consider the noisy random displacement. Suppose that input points occur at σ_j, $j = 0, 1, \ldots$. Suppose that σ_j is displaced, randomly, to $\sigma_j + u_j$, $j = 0, 1, \ldots$. Suppose that γ_j, $j = 0, 1, \ldots$, denote the points of a further point process $E(t)$. Let the points of the output process be the union of the $\sigma_j + u_j$ and the γ_j. The output process may be represented by

$$\sum_j \delta(t - \sigma_j - u_j) + \sum_j \delta(t - \gamma_j) \qquad (2.9.2)$$

In the case in which the process $E(t)$ is stationary with mean rate μ and where u_j is independent of σ_j with density function $a(u)$, $j = 0, 1, \ldots$, it follows from (2.9.2) that

$$E\{dN(t)\,|\,M\} = \left[\mu + \int a(t-u)\,dM(u)\right]dt \qquad (2.9.3)$$

This expression is seen to be analogous to the regression model (2.8.2) of ordinary time series analysis.

In the case in which the process M is stationary with power spectrum $f_{20}(\lambda)$ and where

$$f_{11}(\lambda) = (2\pi)^{-1}\left[\int_{-\infty}^{\infty} \exp\{-i\lambda u\}\,\text{cov}\{dM(t+u), dN(t)\}\right]\Big/dt \qquad (2.9.4)$$

is the *cross spectrum* of the process M with the process N, it follows from (2.9.3) that

$$f_{11}(\lambda) = A(-\lambda) f_{20}(\lambda) \tag{2.9.5}$$

Hence the system with output (2.9.2) may be identified, in the sense that $A(\lambda)$, and hence $a(u)$, may be estimated once estimates of $f_{11}(\lambda), f_{20}(\lambda)$ are available. Incidentally, in terms of the random displacement model,

$$A(-\lambda) = \int \exp\{i\lambda u\} a(u)\, du = E \exp\{i\lambda u\}$$

is the characteristic function of the distribution of translations.

In the case in which there are a number of displacements u_{jk}, $k = 1, \ldots, K_j$, applied to each input point, the relationship (2.9.5) continues to hold with

$$A(\lambda) = E \sum_k \exp\{-i\lambda u_{jk}\}$$

provided this last function does not depend on j.

It should be remarked that the specification (2.9.3) does not characterize the distribution of the process completely, in the manner that (2.8.2) characterized that of Y. Were it given that

$$\text{Prob}\{dN(t) = 1 \,|\, N(u), u \le t, M\} = \left[\mu + \int a(t-u)\, dM(u)\right] dt \tag{2.9.6}$$

for the system, then the distribution of N could be characterized as time inhomogeneous Poisson given M.

A point process system analogous to those of Hawkes (1972) and suggested by (2.9.6) is the one defined by

$$\text{Prob}\{dN(t) = 1 \,|\, N(u), u \le t, M\} = \left[\mu + \int a(t-u)\, dM(u) + \int^t b(t-u)\, dN(u)\right] dt \tag{2.9.7}$$

For this system, the cross spectrum between the input and output processes is given by

$$f_{11}(-\lambda) = (1 - B(\lambda))^{-1} A(\lambda) f_{20}(\lambda)$$

where $B(\lambda) = \int_0^\infty \exp\{-i\lambda u\} b(u)\, du$. This expression generalizes (2.9.5). Rice (1975) discusses some aspects of the problem of identification of the system (2.9.7).

The modulation model of Section 2.7 is sometimes useful in describing certain point process systems. Suppose that $E(t)$ is a point process with

conditional intensity function $\gamma(t, \omega)$. Suppose that $M(t)$ is a deterministic step function. Suppose that the output process $N(t)$ is to have conditional intensity function given by

$$\text{Prob}\{dN(t) = 1 \,|\, N(u), u \le t, M\} = \beta(t, M, \omega)\gamma(t, \omega)\, dt \tag{2.9.8}$$

The likelihood ratio of the process N relative to the process E is here

$$\exp\left\{\int_0^T \log \beta(t, M, \omega)\, dN(t) - \int_0^T [\beta(t, M, \omega) - 1]\gamma(t, \omega)\, dt\right\} \tag{2.9.9}$$

In this situation, $\gamma(t, \omega)$ may be thought of as specifying the output process when there is no input M, and $\beta(t, M, \omega)$ as indicating the effect that input has upon the output. In the case of the model (2.9.6), $\gamma(t, \omega) = 1$ corresponding to a unit Poisson process, and

$$\beta(t, M, \omega) = \beta(t, M) = \mu + \int a(t-u)\, dM(u) \tag{2.9.10}$$

In the case of the model (2.9.7), $\gamma(t, \omega) = 1$ again, and

$$\beta(t, M, \omega) = \mu + \int a(t-u)\, dM(u) + \int^t b(t-u)\, dN(u) \tag{2.9.11}$$

In a situation in which the functions $a(\cdot)$ and $b(\cdot)$ of these last two cases depend on a parameter θ, the likelihood ratio (2.9.9) with the substitution (2.9.11) may be used to estimate θ and hence identify the system, in certain circumstances.

Consider next a system with the property that for each pair (σ_j, σ_k) of input points σ_j, σ_k, an output point appears at $\sigma_j + u_{jk}$ where u_{jk} is a random variable with density function $h(u, \sigma_k - \sigma_j)$. Then the output process may be represented by

$$\sum_{j<k} \delta(t - \sigma_j - u_{jk})$$

and it is the case that

$$\frac{1}{dt} E\{dN(t) \,|\, M\} = \sum_{j<k} h(t - \sigma_j, \sigma_k - \sigma_j) = \sum_{j<k} b(t - \sigma_j, t - \sigma_k) = \iint_{u<v} b(t-u, t-v)\, dM(u)\, dM(v)$$

where $b(u, v) = h(u, u - v)$. The system in this case is seen to be "quadratic." Continuing in this manner a system may be constructed such that

$$\frac{1}{dt} E\{dN(t) \,|\, M\} = \sum_{K=0} \int_{u_k \text{distinct}} \cdots \int a_K(u_1, \ldots, u_K)\, dM(t-u_1) \cdots dM(t-u_K) \tag{2.9.12}$$

to parallel the ordinary time series system (2.8.11). Suppose that one is able to employ unit Poisson noise as input process M. Then from (2.9.12)

$$E\{dM'(t-u_1)\cdots dM'(t-u_K)\,dN(t)\} = a_K(u_1,\ldots,u_K)\,du_1\cdots du_K\,dt\;K! \tag{2.9.13}$$

in the case in which a_K is symmetric and where $M'(t) = M(t) - p_1 t$. Hence the system of (2.9.12) may be identified once one can estimate joint product densities of order K of bivariate point processes. This estimation problem is considered by Brillinger (1975b).

One general remark that may be made concerning point process systems of the character just considered is that a useful parameter for use in the modeling of systems is $\mu_M(t)$ defined by

$$E\{dN(t)\,|\,M\} = \mu_M(t)\,dt$$

The identification problem then comes down to, in part, the estimation of $\mu_M(t)$ given a stretch of input and output data of the system.

The linear dynamic system (2.8.9) may be paralleled to some extent through the point process system specified by

$$\begin{aligned} dS(t) &= F(t)S(t)\,dt + G(t)\,dM(t) + K(t)\,dW(t) \\ \text{Prob}\{dN(t) = 1\,|\,N(u), u \le t, S, M, W\} &= [\mu + H(t)S(t)]\,dt \ge 0 \end{aligned} \tag{2.9.14}$$

where $M(t)$ is the input process, $S(t)$ a (vector-valued) state process, and $W(t)$ a noise process. In this system, following expression (2.2.17), the likelihood function is given by

$$\exp\left\{\int_0^T \log[\mu + H(t)\hat{S}(t)]\,dN(t) - \int_0^T [\mu + H(t)\hat{S}(t)]\,dt\right\} \tag{2.9.15}$$

assuming the input process fixed. Here the variate $\hat{S}(t)$ is defined by

$$\hat{S}(t) = E\{S(t)\,|\,N(u), u \le t\}$$

and may be determined by Kalman–Bucy type of recursive equations. Some details concerning this sort of model are given by Snyder (1975). In the stationary case with $F(t), \ldots, J(t)$ constant and the process W uncorrelated with the process M, the second-order spectra of the input and output are related by

$$f_{11}(-\lambda) = H(i\lambda I - F)^{-1}Gf_{20}(\lambda) \tag{2.9.16}$$

In the case in which the $F, \ldots, J$ have been parametrized to make the system identifiable, the relationship (2.9.16) or the criterion (2.9.15) may be used on occasion to estimate the parameters.

Consider next the system

$$N_1(t) = M_1(t) + M_3(t), \qquad N_2(t) = M_2(t) + M_4(t) \tag{2.9.17}$$

where the processes M_1 and M_2 depend on a process N_3, but the bivariate process $\{M_3, M_4\}$ is independent of the process $\{M_1, M_2, N_3\}$. In particular, suppose that

$$\text{Prob}\{dM_j(t) = 1\,|\,N_3\} = \left[\mu_j + \int a_j(t-u)\,dN_3(u)\right]dt \tag{2.9.18}$$

for $j = 1, 2$. One may be interested in the question of whether an association observed between the processes N_1 and N_2 indicates a proper connection, or whether it is simply due to their common association with the process N_3. The model (2.9.17), (2.9.18) is one means of examining this question. In the case in which the processes M_3 and M_4 are independent, the association would be apparent, not real. Now the degree of dependence of the processes M_3 and M_4 may be measured by their coherence function

$$|g_{11}(\lambda)|^2/[g_{20}(\lambda)g_{02}(\lambda)] \tag{2.9.19}$$

with g_{jk} given in terms of the second-order spectra of the process N_1, N_2, N_3 as

$$g_{jk}(\lambda) = \{f_{jk0}(\lambda) - [f_{j0k}(\lambda)f_{0k1}(\lambda)]\}/f_{002}(\lambda)$$

$j + k = 2$. The function (2.9.19) is here called the partial coherence at frequency λ of the processes N_1 and N_2 with the linear effects of the process N_3 removed. For networks of three nerve cells some estimates of partial coherences are presented by Brillinger *et al.* (1976). The parameter proved useful in investigating the connections between the cells, specifically whether each pair of cells had a direct link.

Numerous models of point process systems are provided by the various models that have been proposed for the operation of a nerve cell, driven by a spike train to emit a further spike train. The effect of the input train may be excitatory or inhibitory, corresponding to an increase or decrease of the instantaneous output rate. As an example, consider the following model of an excitatory system, of the kind discussed by Moore *et al.* (1966). Let the input spike train be denoted by $M(t)$. Let the output spike train be denoted by $N(t)$ and suppose its spikes occur at the times $\tau_0, \tau_1, \ldots$. Define the time series

$$X(t) = \alpha \int_{\tau_{N(t)-1}}^{t} \exp\{-\alpha(t-u)\}\,dM(u) \tag{2.9.20}$$

for $\tau_{N(t)-1} < t \le \tau_{N(t)}$ and $\alpha > 0$. Then $\tau_{N(t)}$ is defined to be

$$\tau_{N(t)} = \inf\{t: t > \tau_{N(t)-1}, X(t) \ge \theta\}$$

The value θ is assumed to be positive and is called the *threshold* value. $X(t)$ is called the *postsynaptic potential* of the cell at time t. To be realistic a nerve cell system model must also have a refractory period. Other models of nerve cell activity are described by Feinberg (1974).

2.10. Some Particular Time Series

Many specific properties are known concerning certain types of ordinary time series. The class that has been subjected to the highest level of development is undoubtedly that of *Gaussian time series*. A series $X(t)$ is called Gaussian when all of its finite-dimensional distributions are multivariate normal. In the Gaussian case, the transition probability functions (2.1.2) will also be normal. The characteristic functional is given by

$$C[\xi] = \exp\left\{i \int \xi(t)c_1(t)\,dt - \tfrac{1}{2} \iint \xi(t_1)\xi(t_2)c_2(t_1, t_2)\,dt_1\,dt_2\right\} \tag{2.10.1}$$

with $c_1(t) = EX(t)$ and $c_2(t_1, t_2) = \operatorname{cov}\{X(t_1), X(t_2)\}$. It follows from expression (2.10.1) that a Gaussian series has moments and cumulants of all orders and that the cumulant functions of order greater than 2 are identically 0. A Gaussian series is determined by its first- and second-order moment functions, and given any function $c_1(t)$ and continuous nonnegative definite $c_2(t_1, t_2)$, there exists a Gaussian series with these parameters. Linear operations on Gaussian series produce Gaussian series, and conditional distributions based on linear combinations of its values remain Gaussian. The Gaussian series appears as a limit when independent series satisfying finite second-order moment conditions are added. The importance of the use of Gaussian series in the identification of nonlinear systems was indicated in Section 2.7.

One important Gaussian series is the *Wiener process* $W(t)$ satisfying $W(0) = 0$, $c_1(t) = 0$, $c_2(t_1, t_2) = \min\{t_1, t_2\}$. The increments of the Wiener process are stationary and independent. The generalized process $dW(t)/dt$ is called *Gaussian white noise*. Its covariance function may be represented by $c_2(t_1, t_2) = \delta(t_1 - t_2)$. It provides a continuous time analog of a sequence of independent standardized normal variates.

A variety of conditions have been set down to ensure that the sample paths of a Gaussian series are continuous. A useful sufficient condition that ensures this (for any series) is

$$E|X(t_2) - X(t_1)|^a \le b|t_2 - t_1|^{1+c} \tag{2.10.2}$$

for some $a, b, c > 0$ and any t_1, t_2 (e.g., see Cramér and Leadbetter, 1967). This criterion shows that the Wiener process has continuous paths.

Another important class of time series is made up of the *diffusion processes*. These are Markov processes with continuous sample paths. In particular,

$$\text{Prob}\{X(v) \le y \mid X(u),\ u \le t\} = \text{Prob}\{X(v) \le y \mid X(t)\}$$

for all $v > t$. Suppose the transition density function is given by

$$p(t, x; v, y) = \frac{d}{dy}\text{Prob}\{X(v) \le y \mid X(t) = x\} \tag{2.10.3}$$

Then it satisfies

$$p(t, x; v, y) = \int p(t, x; u, z)p(u, z; v, y)\,dz$$

for $t < u < v$ and also $p(t, x; t, y) = \delta(x - y)$. Suppose

$$E\{dX(v) \mid X(t) = x\} = \mu(t, x)\,dv \tag{2.10.4}$$

and

$$E\{(dX(v))^2 \mid X(t) = x\} = \sigma^2(t, x)\,dv \tag{2.10.5}$$

The parameter $\mu(t, x)$ is called the *local mean* or *drift*. The parameter $\sigma^2(t, x)$ is called the *local variance*, and its reciprocal is called the *speed*. The transition density function satisfies certain differential equations:

$$-\frac{\partial f(t, x; v, y)}{\partial t} = \tfrac{1}{2}\sigma^2(t, x)\frac{\partial^2 f(t, x; v, y)}{\partial x^2} + \mu(t, x)\frac{\partial f(t, x; v, y)}{\partial x}$$

called the *backward Kolmogorov equation*, and

$$\frac{\partial f(t, x; v, y)}{\partial t} = \frac{1}{2}\frac{\partial^2[\sigma^2(v, y)f(t, x; v, y)]}{\partial y^2} - \frac{\partial[\mu(v, y)f(t, x; v, y)]}{\partial y}$$

the *forward Kolmogorov equation*. This diffusion process may be approximated by the following discrete model: Suppose $X(t) = x$. Then $X(t + dt) = x + dx$ with probability

$$\frac{1}{2} + \frac{\mu(t, x)}{2\sigma(t, x)}\sqrt{dt} \tag{2.10.6}$$

and $X(t + dt) = x - dx$ with probability $1 - (2.10.6)$, $dx = \sigma(t, x)\sqrt{dt}$ independently of its behavior before t (see Prohorov and Rozanov, 1969, p. 263).

Particular diffusion processes include the Wiener process with $\mu(t, x) = 0$ and $\sigma(t, x) = 1$ and the *Ornstein–Uhlenbeck process* with $\mu(t, x) = -1$ and $\sigma^2(t, x) = 1$. The latter process is the most general Gaussian Markov stationary process.

The Wiener process may be used to derive a large class of diffusion processes. Ito (1951) suggests solving the equation

$$X(t) - X(0) = \int_0^t \mu(s, X(s))\, ds + \int_0^t \sigma(s, X(s))\, dW(s) \qquad (2.10.7)$$

where $W(t)$ is the Wiener process. Under appropriate conditions, the equation has a solution that is a diffusion process. Equation (2.10.7) is called a *stochastic Ito differential equation*. A considerable literature exists concerning first passage time distributions for diffusion processes. Differential equations may be set down for certain characteristic functions and expected values based on a diffusion process.

Certain transformations carry diffusion processes over into diffusion processes. Let g_x and g_t denote the partial derivatives of $g(x, t)$. Suppose $g_x(x, t) > 0$. Let $X(t)$ be the diffusion process just discussed and $Y(t) = g(X(t), t)$. Then $Y(t)$ is a diffusion process with parameters

$$\mu_Y(t, x) = \mu(t, x)g_x(x, t) + \tfrac{1}{2}\sigma^2(t, x)g_{xx}(x, t) + g_t(x, t)$$
$$\sigma_Y(t, x) = \sigma(x, t)g_x(x, t)$$

(see Gihman and Skorohod, 1972). Alternatively, consider a random time transformation defined by

$$\frac{d\tau(t, \omega)}{dt} = \frac{1}{V(t, X(\tau))} > 0$$

and the process $Y(t) = X(\tau(t, \omega))$. Then $Y(t)$ is also a diffusion with the parameters

$$\mu_Y(t, x) = \mu(t, x)/V(t, x), \qquad \sigma_Y(t, x) = \sigma(t, x)/\{V(t, x)\}^{1/2}$$

The transformations described here may be used to carry a given process over into one with a simpler description.

Likelihood ratios may be determined for diffusion processes in certain cases. For example, suppose that $\sigma(t, x) = 1$; then under conditions including $X(0) = 0$, the likelihood ratio of the process of (2.10.7) relative to the process W is given by

$$\exp\left\{\int_0^T \mu(t, X(t))\, dX(t) - \tfrac{1}{2}\int_0^T \mu^2(t, X(t))\, dt\right\} \qquad (2.10.8)$$

(see Gihman and Skorohod, 1972, p. 90). This ratio is used to determine maximum likelihood estimates of the parameters $\theta_1, \ldots, \theta_K$ in the case in which

$$\mu(t, x) = \sum_{k=1}^{K} \theta_k \phi_k(t, x)$$

with the ϕ_k known (see Taraskin, 1974).

A generalization of the Gaussian Markov process is provided by the *N-tuple Markov Gaussian* of Hida (1960). It is defined as a Gaussian process with the property that the variates

$$E\{X(t_k) \mid X(u),\ u \le t_0\}, \qquad k = 1, \ldots, K$$

$t_0 \le t_1 < \cdots < t_K$ are linearly independent for $K = N$, but linearly dependent for $K = N + 1$. The process may be shown to have a representation

$$X(t) = \int_0^t \sum_{n=1}^{N} f_n(t)g_n(u)\, dW(u) \qquad (2.10.9)$$

where $W(t)$ is a Wiener process. In the stationary case the kernel of (2.10.9) may be shown to be a linear combination of the functions $t^k u^{n-k} \exp\{-(\rho + i\mu)(t - u)\}$ (see Hida, 1960). The power spectrum is a rational function of λ.

As a final general class of time series consider the *linear processes* defined by $X(t) = \int a(t, u)\, dV(u)$, where $V(t)$ is a process with independent increments. Examples of independent increment processes include the Wiener and Poisson processes. The general increment process is completely characterized by the first-order distribution of $V(t)$ and the increment distribution of $V(t) - V(u)$. The process $V(t)$ may be shown to have the representation

$$V(t) = V_c(t) + \int_{|x| \le 1} x[N([0, t] \times dx) - \mu([0, t] \times dx)] + \int_{|x| > 1} xN([0, t]\, dx) \qquad (2.10.10)$$

where $V_c(t)$ is a continuous Gaussian process with independent increments and $N(t, x)$ a Poisson process in the plane with parameter $\mu(t, x)$ independent of the process V_c (see Prohorov and Rozanov, 1969). Supposing $E\, dV_c(t) = \alpha(t)\, dt$, $\operatorname{cov}\{dV_c(t_1), dV_c(t_2)\} = \beta(t_1)\, \delta(t_1 - t_2)\, dt_1\, dt_2$,

$$\mu((t, t + dt] \times dx) = m(t, x)\, dt\, dx,$$

the cumulant functions of the process V are given by

$$E\{dV(t)\} = \left[\alpha(t) + \int_{|x| > 1} xm(t, x)\, dx\right] dt$$

$$\operatorname{cov}\{dV(t_1), dV(t_2)\} = \left[\beta(t_1) + \int_{|x| > 1} x^2 m(t_1, x)\, dx\right] \times \delta(t_1 - t_2)\, dt_1\, dt_2$$

$$\operatorname{cum}\{dV(t_1), \ldots, dV(t_K)\} = \int x^K m(t_1, x)\, dx\, \delta(t_1 - t_2) \cdots \times \delta(t_1 - t_K)\, dt_1 \cdots dt_K$$

$K > 2$. The cumulant functions of the linear process may be derived directly from these expressions. In the case in which the process $V(t)$ has stationary increments and $X(t) = \int a(t-u)\,dV(u)$, $A(\lambda) = \int \exp\{-i\lambda u\}a(u)\,du$, the cumulant spectrum of order K of $X(t)$ will be proportional to

$$A(\lambda_1)\cdots A(\lambda_{K-1})A(-\lambda_1 - \cdots - \lambda_{K-1})$$

Certain additional aspects concerning linear processes may be found in the article by Westcott (1970).

2.11. Some Particular Point Processes

Foremost among the point processes is the Poisson process. The *Poisson process* with intensity function $p(t)$ is defined by the requirement that for $I_1, \ldots, I_K$ disjoint intervals, the variates $N(I_1), \ldots, N(I_K)$ are independent Poisson random variables with means $P(I_1), \ldots, P(I_K)$ where

$$P(I) = \int_I p(t)\,dt \tag{2.11.1}$$

The zero probability function also characterizes the Poisson. It is given by

$$\phi(I) = \exp\{-P(I)\} \tag{2.11.2}$$

for I in $B[0, \infty)$. The probability generating functional is given by

$$G[\xi] = \exp\left\{\int [\xi(t) - 1]p(t)\,dt\right\} \tag{2.11.3}$$

It follows from (2.11.3), or directly from the definition of the process, that the product densities of the Poisson are given by

$$p_K(t_1, \ldots, t_K) = p(t_1)\cdots p(t_K) \tag{2.11.4}$$

and the cumulant densities by $q_1(t) = p(t)$ and $q_K(t_1, \ldots, t_K) = 0$, $K = 2, 3, \ldots$. The conditional intensity function is constant in ω and given by

$$\gamma(t, \omega) = p(t) \tag{2.11.5}$$

If the points observed in the interval $[0, T]$ are $0 \le \tau_0 < \tau_1 < \cdots$, then the corresponding likelihood function is

$$\exp\left\{\int_0^T \log p(t)\,dN(t) - \int_0^T p(t)\,dt\right\} \tag{2.11.6}$$

and the likelihood ratio relative to the unit Poisson process is

$$\exp\left\{\int_0^T \log p(t)\,dN(t) - \int_0^T (p(t) - 1)\,dt\right\} \tag{2.11.7}$$

From expression (2.2.9), the joint density of the q successive intervals after τ_0 is given by

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^\infty \int\cdots\int_{(0, t+y_1+\cdots+y_q)^k} p_{k+q+1}(t, t+y_1, \ldots, t+y_1+\cdots + y_q, u_1, \ldots, u_k)\,du_1\cdots du_k\,dt \tag{2.11.8}$$

In the Poisson case this reduces to

$$\int_0^\infty p(t)p(t+y_1)\cdots p(t+y_1+\cdots+y_q)\ \exp\left\{-\int_0^{t+y_1+\cdots+y_q} p(u)\,du\right\}dt \tag{2.11.9}$$

In the stationary case, $p(t) = p$, and (2.11.9) becomes

$$p^q \exp\{-p(y_1 + \cdots + y_q)\} \tag{2.11.10}$$

and the successive intervals are seen to be independent exponentials with mean $1/p$. Expression (2.2.9) may also be used to show that τ_0 is independent of $\{y_1, \ldots, y_q\}$ and exponential with mean $1/p$.

From (2.2.12), the probability generating functional of the Palm process is given by

$$G[\xi\,|\,t] = \exp\int [\xi(t) - 1]p(t)\,dt$$

showing that in the Poisson case the distribution of the Palm process is the same as that of the original process. Among other things, this implies that the survivor function is given by

$$\text{Prob}\{N(t, t+u] = 0\,|\,N\{t\} = 1\} = \exp\left\{-\int_t^{t+u} p(v)\,dv\right\}$$

and that the hazard function is given by

$$\text{Prob}\{dN(t+u) = 1\,|\,N\{t\} = 1,\ N(t, t+u] = 0\}/du = p(t+u)$$

The forward recurrence time distribution is given by

$$\text{Prob}\{\tau_{N(t)} - t \le u\} = 1 - \exp\left\{-\int_t^{t+u} p(v)\,dv\right\}$$

The *unit Poisson* process has intensity $p(t)$ identically 1. If N is a Poisson process with intensity $p(t)$ and one sets

$$P(t) = \int_0^t p(u)\,du$$

then $N(P^{-1}(t))$ is a unit Poisson process. Conversely, if $M(t)$ is a unit Poisson, then $M(P(t))$ is Poisson with mean rate $p(t)$.

The general Poisson process is clearly a Markov process with state space 0, 1, 2, ... and transition distributions given by

$$\text{Prob}\{N(t+dt) = j+1 \,|\, N(t) = j\} = p(t)\,dt$$
$$\text{Prob}\{N(t+dt) = j \,|\, N(t) = j\} = 1 - p(t)\,dt$$

$j = 0, 1, \ldots$. It is the simplest discontinuous Markov process.

When the Poisson process is thinned, as indicated in the discussion following (2.7.8), its probability generating functional becomes

$$G[\xi\pi + 1 - \pi] = \exp\left\{\int [\xi(t) - 1]\pi(t)p(t)\,dt\right\}$$

That is, it becomes a Poisson process with intensity $\pi(t)p(t)$.

When r independent Poissons, with intensities $p_1(t), \ldots, p_r(t)$, are superposed, as indicated by the discussion following (2.7.17), the probability generating functional of the superposed process is

$$\exp\left\{\int [\xi(t) - 1]p_1(t)\,dt + \cdots + \int [\xi(t) - 1]p_r(t)\,dt\right\}$$

i.e., the superposed process is Poisson with intensity $p_1(t) + \cdots + p_r(t)$.

If $\tau_0, \tau_1, \ldots$ are the points of a Poisson process with rate $p(t)$ and if independent random displacements U_j, $j = 0, 1, \ldots$, with cumulative distribution function (c.d.f.) $F(u)$ are applied to these points, then the process with points $\tau_j + U_j$, $j = 0, 1, \ldots$, has, from expression (2.7.3), probability generating functional (p.g.f.) given by

$$G\left[\int \xi(\cdot + u)\,dF(u)\right] = \exp\int\left[\int \xi(u+t)\,dF(u) - 1\right]p(t)\,dt$$
$$= \exp\left\{\int [\xi(v) - 1]\left(\int p(v-u)\,dF(u)\right)dv\right\}$$

i.e., the displaced process is also Poisson. It has mean rate $\int p(v-u)\,dF(u)$. This result is due to Mirasol (1963).

Next, consider a stationary Poisson process N with intensity p. Because of the independence properties of the increments of N one has

$$\text{cum}\{dN(t_1), \ldots, dN(t_K)\} = p\,\delta(t_1 - t_K)\cdots\delta(t_{K-1} - t_K)\,dt_1\cdots dt_K$$

$K = 2, 3, \ldots$, and so from (2.5.30), the cumulant spectra of N are given by

$$f_K(\lambda_1, \ldots, \lambda_{K-1}) = (2\pi)^{-K+1}p \qquad (2.11.11)$$

for $K = 2, 3, \ldots$. Being stationary, the process has a spectral representation

$$N(t) = \int \frac{\exp\{it\lambda\} - 1}{i\lambda}\,dZ(\lambda)$$

In view of (2.11.11) and (2.5.33), the process $Z(\lambda)$ has the interesting property

$$\text{cum}\{dZ(\lambda_1), \ldots, dZ(\lambda_k)\} = (2\pi)^{-K+1}p\,\delta(\lambda_1 + \cdots + \lambda_k) \qquad (2.11.12)$$

When a type I counter with dead time Δ is applied to a stationary Poisson process, the interpoint distribution simply becomes the original exponential translated by Δ. To see this, note from (2.7.11) that the conditional intensity function of the output process is given by

$$\gamma(t, \omega) = p \qquad \text{for} \quad \tau_{N(t)-1} + \Delta \le t$$
$$= 0 \qquad \text{otherwise}$$

Therefore, from (2.2.16)

$$\text{Prob}\{\tau_J > t \,|\, \tau_j = t_j, j < J\} = 1 \qquad \text{for} \quad t < t_{J-1} + \Delta$$
$$= \exp\left\{-\int_{t_{J-1}+\Delta}^{t} p\,du\right\} \qquad \text{for} \quad t \ge t_{J-1} + \Delta$$

giving the result.

Consider now a Poisson process $N_\theta(t)$ with intensity function $\theta p(t)$. Consider the characteristic functional of the process

$$X_\theta(t) = (N_\theta(t) - \theta p(t))/\sigma$$

where $\sigma^2 = \theta p(t)$. From (2.11.3) the functional is

$$E\exp\left\{i\int X_\theta(t)\xi(t)\,dt\right\} = \exp\left\{-i\int \sigma\xi(t)\,dt\right\}$$
$$\times \exp\left\{\int (\exp\{i\xi(t)/\sigma\} - 1)\,\theta p(t)\,dt\right\}$$
$$\to \exp\left\{-\int \xi(t)^2\,dt\right\}$$

as $\theta \to \infty$. This is the characteristic functional of the Wiener process. This result suggests that a Poisson process with high rate may be approximated by a simple function of a Wiener process.

Suppose now that $H[\xi]$ is the factorial cumulant generating functional of a stationary point process with finite intensity p. Let

$$N_n(t) = M_1(t/n) + \cdots + M_n(t/n)$$

where the M_j are independent realizations of M. Then the factorial cumulant generating functional of N_n is

$$nH[\xi(nt)] = n\left[\int \xi(nt)p\,dt + o\left(\frac{1}{n}\right)\right] \to p\int \xi(t)\,dt$$

as $n \to \infty$. This is the factorial cumulant generating functional of a Poisson process with intensity p. Hence the limit in distribution of rescaled superpositions of independent replicas of a point process is Poisson. This form of argument appears in the article by Vere-Jones (1968). More general results on the Poisson limit occurring when point processes are superposed appear in the chapter by Cinlar (1972).

Consider a process M with intensity p and factorial cumulant generating functional $H[\xi]$. Consider the effect of applying the thinning operation represented by (2.7.8) with $EI(t) = \pi$ and letting $\pi \to 0$. (This would be the case if the same thinning operation with $\pi < 1$ were applied repeatedly.) Let N_π denote the thinned process. Consider the rescaled process $N_\pi(t/\pi)$. From (2.7.10) its factorial cumulant generating functional (f.c.g.fl.) is given by

$$\begin{aligned} H[\pi\xi(t\pi)] &= \log E\exp\left\{\int \log[1+\pi\xi(\pi t)]\,dM(t)\right\} \\ &\sim \log E\exp\left\{\int \pi\xi(\pi t)\,dM(t)\right\} \\ &\to \log E\exp\left\{\left(\lim_{\pi\to 0}\pi M(1/\pi)\right)\int \xi(u)\,du\right\} \end{aligned} \tag{2.11.13}$$

using a Wiener type of formula. When the process M is ergodic

$$\lim_{\pi\to 0}\pi M(1/\pi) = p$$

almost surely and (2.11.13) becomes $p\int \xi(u)\,du$, the factorial cumulant generating functional of the Poisson process with intensity p. This result is suggestive of why the Poisson process is sometimes said to correspond to the law of rare events. It may also be developed via the zero probability function. More formal developments of thinning results may be found in the article by Kallenberg (1975).

As an alternative procedure leading to a Poisson limit consider the following result of Volkonskii and Rozanov (1959). Let $\mathscr{F}_{a,b}$ denote the Borel field generated by events of the form

$$\{N(v_1) - N(u_1) \le n_1, \ldots, N(v_K) - N(u_K) \le n_K\}$$

$a < u_k < v_k \le b$, n_k a nonnegative integer, $K = 1, 2, \ldots$. The process N is called *strong mixing* with mixing coefficient $\alpha(\tau)$ when

$$\alpha(\tau) = \sup\{|P(AB) - P(A)P(B)| : A \text{ in } \mathscr{F}_{-\infty,t},\ B \text{ in } \mathscr{F}_{t+\tau,\infty}\} \to 0$$

as $\tau \to \infty$. Volkonskii and Rozanov consider a sequence of stationary processes N^T, $T = 1, 2, \ldots$, with N^T (i) having intensity of the form $p\varepsilon_T$, $\varepsilon_T \to 0$, as $T \to \infty$; (ii) having mixing coefficient $\alpha^T(\tau) \to 0$ uniformly as $\tau \to \infty$, and such that $\alpha^T(\tau) \approx \alpha(\tau)$; and (iii) $E\{N^T(t/\varepsilon_T)[N^T(t/\varepsilon_T) - 1]\} = o(t)$ as $t \to 0$, $T \to \infty$. They show that under these conditions $N^T(t/\varepsilon_T)$ tends in distribution to a Poisson variate with mean pt.

The point process obtained by retaining every kth point of a stationary Poisson process is called an *Erlang process*. Its interpoint distributions will be independent gammas, in view of the exponential distributions of the Poisson. An Erlang process may be viewed as the output of a counter that remains paralyzed for $k - 1$ points. It is a renewal process. Useful properties may be determined as particular cases of the renewal process properties to be set down shortly.

Consider an ordinary time series with nonnegative sample paths $p(t, \omega')$, ω' in Ω', $0 \le t < \infty$. Having obtained a particular realization of this series, generate a Poisson process with intensity $p(t, \omega')$. The point process N obtained in this manner is called a *doubly stochastic Poisson*. The product densities of N are given by

$$\begin{aligned} p_K(t_1, \ldots, t_K) &= E_{\omega'}\{p(t_1, \omega')\cdots p(t_K, \omega')\} \\ &= m_K(t_1, \ldots, t_K) \end{aligned} \tag{2.11.14}$$

the moments of the original series. In the case in which $p(t, \omega') = p(\omega')$ and the latter variate has c.d.f. $F(p)$, the product densities are constant in time and given by

$$p_K(t_1, \ldots, t_K) = \int p^K\,dF(p) \tag{2.11.15}$$

The cumulant densities of the process are given by

$$\begin{aligned} q_K(t_1, \ldots, t_K) &= \operatorname{cum}\{p(t_1, \omega'), \ldots, p(t_K, \omega')\} \\ &= c_K(t_1, \ldots, t_K) \end{aligned} \tag{2.11.16}$$

i.e., they are those of the original process. The zero probability function is

$$\phi(I) = E_{\omega'}\exp\left\{-\int_I p(t, \omega')\,dt\right\} \tag{2.11.17}$$

The probability generating functional is

$$G[\xi] = E_{\omega'}\exp\left\{\int [\xi(t) - 1]p(t, \omega')\,dt\right\} = C[-i[\xi(t) - 1]] \tag{2.11.18}$$

in terms of the characteristic functional of the original process.

The conditional intensity function is

$$\gamma(t, \omega) = \underset{\omega'}{E}\{p(t, \omega') \mid N(u), u \le t\} \tag{2.11.19}$$

It is seen to be the minimum mean squared error estimate of the level of the intensity process at time t, given the history of the point process. One may write

$$G_J(t; t_j, j < J) = \text{Prob}\{\tau_J > t \mid \tau_j = t_j, j < J\}$$
$$= \frac{E_{\omega'}\{p(t_0, \omega') \cdots p(t_{J-1}, \omega') \exp\{-\int_0^t p(u, \omega')\, du\}\}}{E_{\omega'}\{p(t_0, \omega') \cdots p(t_{J-1}, \omega') \exp\{-\int_0^{t_J} p(u, \omega')\, du\}\}}$$

Using this representation and (2.2.15)

$$\gamma(t, \omega) = \frac{E_{\omega'}\{p(t, \omega') \exp\{\int_0^t \log p(u, \omega')\, dN(u) - \int_0^t p(u, \omega')\, du\}\}}{E_{\omega'}\{\exp\{\int_0^t \log p(u, \omega')\, dN(u) - \int_0^t p(u, \omega')\, du\}\}} \tag{2.11.20}$$

The likelihood function may be written in two distinct ways: as

$$\underset{\omega'}{E} \exp\left\{\int_0^T \log p(t, \omega')\, dN(t) - \int_0^T p(t, \omega')\, dt\right\} \tag{2.11.21}$$

or, from (2.2.17), as

$$\exp\left\{\int_0^T \log \gamma(t, \omega)\, dN(t) - \int_0^T \gamma(t, \omega)\, dt\right\}$$

with $\gamma(t, \omega)$ given by (2.11.19).

When the doubly stochastic Poisson is thinned, as at (2.7.8), its probability generating functional becomes

$$G[\xi\pi + 1 - \pi] = \underset{\omega'}{E} \exp\{[\xi(t) - 1]\pi(t)p(t, \omega')\, dt\}$$

and so the thinned process is doubly stochastic Poisson as well.

If $\{p_1(t, \omega'), \ldots, p_r(t, \omega')\}$ is an r vector-valued ordinary time series with nonnegative components, then suppose that independent Poissons are generated with intensities $p_1(t, \omega'), \ldots, p_r(t, \omega')$, respectively. Suppose that these Poissons are superposed. Then the probability generating functional of the superposed process is

$$\underset{\omega}{E} \exp\left\{\int [\xi(t) - 1][p_1(t, \omega') + \cdots + p_r(t, \omega')]\, dt\right\}$$

and the superposed process is seen to be doubly stochastic Poisson.

Suppose next that $\tau_0, \tau_1, \ldots$ are the points of a realization of a doubly stochastic Poisson. Suppose the points are subjected to independent random displacements $U_0, U_1, \ldots$, respectively, with c.d.f. $F(u)$. Then from (2.7.3) and (2.11.18) the probability generating functional of the displaced process is

$$G\left[\int (\cdot + u)\, dF(u)\right] = \underset{\omega'}{E} \exp\left\{\int [\xi(v) - 1]\left(\int p(v - u, \omega')\, dF(u)\right) dv\right\}$$

and the displaced process is seen to be doubly stochastic Poisson.

Let us return to the discussion of thinning a general point process given preceding expression (2.11.13). Suppose now that

$$\lim_{\varepsilon \to 0} \varepsilon M(1/\varepsilon) = p(\omega)$$

with the limit depending on the particular realization. Then the limit of the factorial cumulant generating functional is

$$\log E \exp\left\{p(\omega) \int \xi(u)\, du\right\}$$

and the limit is seen to be a doubly stochastic point process in this case.

If the process $p(t, \omega')$ is stationary, then the form of the probability generating functional of N shows that N is stationary as well. Its intensity is m_1, the mean level of the process $p(t, \omega')$. From (2.11.16) its autointensity function is

$$p_2(u) = c_2(u) \tag{2.11.22}$$

and hence from (2.5.27), its power spectrum is given by

$$f_2(\lambda) = (2\pi)^{-1} m_1 + f_2'(\lambda) \tag{2.11.23}$$

where $f_2'(\lambda)$ is the power spectrum of the process $p(t, \omega')$. Expression (2.11.23) indicates a defect that doubly stochastic processes have regarding their use in the general modeling of point processes. Their power spectra are necessarily bounded below by $(2\pi)^{-1}$ times their intensity. This is a reflection of the fact that the doubly stochastic Poisson is more disorderly than a Poisson process with the same intensity. Expression (2.11.22) indicates that the autointensity function may be any function that can be the autocovariance function of a time series with nonnegative sample paths. If the higher order cumulant spectra of the series $p(t, \omega')$ are denoted by $f_K'(\lambda_1, \ldots, \lambda_{K-1})$, then using (2.11.16) the Kth-order cumulant spectrum of the process N is given by

$$f_K(\lambda_1, \ldots, \lambda_{K-1}) = \sum_{k=1}^{K} (2\pi)^{k-K} f_k'\left(\sum_{j \in \nu_1} \lambda_j, \ldots, \sum_{j \in \nu_{k-1}} \lambda_j\right) \tag{2.11.24}$$

where the summation extends over all partitions $(\nu_1, \ldots, \nu_k)$ of the set $(1, \ldots, K)$, $\lambda_K = -\lambda_1 - \cdots - \lambda_{K-1}$ and $f_1' = m_1$ (see Brillinger, 1972, Theorem 3.3). Expression (2.11.23) is the case $K = 2$.

Macchi (1975) discusses the fact that the product densities of a point process may have the form (2.11.14) without the process $p(t, \omega')$ being non-negative. Asymptotic independence and mixing properties of the process $p(t, \omega')$ generally carry over to a doubly stochastic process N. See Westcott (1972) for the case of mixing.

A *cluster process* is a point process of point processes. It has two components, a process M' of cluster centers and a process $M''(\cdot\,|\,t)$ of cluster members (centered at t). Each point of M is assumed to initiate an independent process of cluster members. The cluster process itself then consists of the superposition of all the various cluster members. Suppose that the probability generating functional of M' is $G'[\xi]$ and that of $M''(\cdot\,|\,t)$ is $G''[\xi\,|\,t]$. Then the probability generating functional of the cluster process N is

$$G[\xi] = E\prod_j G''[\xi\,|\,\tau_j] = G'[G''[\xi\,|\,t]] \tag{2.11.25}$$

This relationship is discussed in the article by Moyal (1962). The densities of the process N may be determined in terms of those of the processes M', M'' using Faa de Bruno formulas. The cases of $K = 1, 2$ are

$$q_1(t_1) = \int q_1'(t) q_1''(t_1\,|\,t)\,dt \tag{2.11.26}$$

$$q_2(t_1, t_2) = \int q_1'(t) q_2''(t_1, t_2\,|\,t)\,dt + \iint q_2'(t, u) q_1''(t_1\,|\,t) q_1''(t_2\,|\,u)\,dt\,du \tag{2.11.27}$$

If $M''(\cdot\,|\,\tau_j)$, $j = 0, 1, \ldots$, denote the successive clusters, then the process may be represented by

$$N(I) = \sum_j M''(I\,|\,\tau_j) = \int M''(I\,|\,t)\,dM'(t)$$

In many cases the cluster distribution is the same for all clusters with $M''(I\,|\,t) = M''(I - t)$ and

$$G''[\xi\,|\,t] = E\exp\left\{\int \log \xi(u + t)\,dM''(u)\right\}$$

If the clusters each have a single member with density $a(u)$, then

$$G''[\xi\,|\,t] = E\xi(u + t) = \int \xi(u + t) a(u)\,du$$

If M'' is Poisson with intensity $p''(u)$, then

$$G''[\xi\,|\,t] = \exp\left\{\int [\xi(u + t) - 1] p''(u)\,du\right\}$$

In the case in which the process M' is stationary as well, the process N is stationary. The relationships (2.11.25)–(2.11.27) then become

$$G[\xi] = G'[G''[\xi(t + \cdot)]]$$

$$q_1 = q_1' \int q_1''(v)\,dv = q_1' E\{M''(-\infty, \infty)\}$$

$$q_2(v) = q_1' \int q_2''(v - t, -t)\,dt + \iint q_2'(t - u) q_1''(v - t) q_1''(-u)\,dt\,du \tag{2.11.28}$$

Taking the Fourier transform of (2.11.28), the power spectrum of the process N may be seen to be given by

$$(2\pi)^{-1} p_1' \operatorname{var}\left\{\int \exp\{-i\lambda u\}\,dM''(u)\right\} + f_2'(\lambda)\left|E\int \exp\{-i\lambda u\}\,dM''(u)\right|^2 \tag{2.11.29}$$

where

$$E\int \exp\{-i\lambda u\}\,dM''(u) = \int \exp\{-i\lambda u\} q_1''(u)\,du$$

$$\operatorname{var}\left\{\int \exp\{-i\lambda u\}\,dM''(u)\right\} = \int q_1''(u)\,du + \iint \exp\{-i\lambda u + i\lambda v\} q_2''(u, v)\,du\,dv$$

The expression (2.11.29) may be found in the chapter by Daley and Vere-Jones (1972). Westcott (1971) shows that a stationary cluster process is mixing if the cluster center process is mixing.

In the *Poisson cluster process*, the process M' of cluster centers is Poisson. Suppose the intensity of the Poisson is $p'(t)$. Then from (2.11.25) N has probability generating functional

$$\exp\left\{\int [G''[\xi\,|\,t] - 1] p'(t)\,dt\right\} \tag{2.11.30}$$

The cumulant densities may be determined as follows:

$$\log G[\xi + 1] = \int \left(\sum \frac{1}{k!} \int \cdots \int q_k''(t_1, \ldots, t_k)\ \xi(t_1) \cdots \xi(t_k)\,dt_1 \cdots dt_k\right) p'(t)\,dt$$

and so

$$q_K(t_1, \ldots, t_K) = \int q_K''(t_1, \ldots, t_K\,|\,t) p'(t)\,dt \tag{2.11.31}$$

Some particular Poisson cluster processes may be distinguished. In the *Neyman–Scott process*, the cluster members are independent random variables with density $a_t(u)$ and the probability that the cluster is of size n is $\pi_t(n)$. If $g_t(z) = \sum_n \pi_t(n)z^n$, then

$$G''[\xi\,|\,t] = \underset{n}{E}\,\underset{u}{E}\prod_{j=1}^{n}\xi(t+u_j) = \underset{n}{E}\left[\int \xi(t+u)a_t(u)\,du\right]^n$$
$$= g_t\left[\int \xi(t+u)a_t(u)\,du\right] \tag{2.11.32}$$

Suppose that the variate n has factorial moments $\mu_t^{\langle k\rangle}$ given by

$$g_t(z+1) = \sum_{k=0}^{\infty}\frac{1}{k!}\mu_t^{\langle k\rangle}z^k$$

The cumulant densities $q_K''(\cdot\,|\,t)$ may then be determined as

$$G''[\xi+1\,|\,t] = g_t\left[\int \xi(t+u)a_t(u)\,du + 1\right]$$
$$= \sum_k \frac{1}{k!}\mu_t^{\langle k\rangle}\left[\int \xi(v)a_t(v-t)\,dv\right]^k$$

and so

$$q_K''(t_1,\ldots,t_K\,|\,t) = \mu_t^{\langle K\rangle}a_t(t_1-t)\cdots a_t(t_K-t) \tag{2.11.33}$$

From (2.11.31), the cumulant densities of the Neyman–Scott process are given by

$$\int \mu_t^{\langle K\rangle}a_t(t_1-t)\cdots a_t(t_K-t)p'(t)\,dt \tag{2.11.34}$$

In the stationary case $p'(t) = p'$, $\mu_t^{\langle K\rangle} = \mu^{\langle K\rangle}$, $a_t(u) = a(u)$, and, in particular, one has

$$q_1 = p'E\{n\}$$

$$q_2(u) = p'E\{n(n-1)\}\int a(u-t)a(t)\,dt$$

If the characteristic function of the density is given by $\phi(\lambda) = \int \exp\{i\lambda u\}a(u)\,du$, then

$$f_2(\lambda) = (2\pi)^{-1}p'E\{n\} + (2\pi)^{-1}p'E\{n(n-1)\}\,|\phi(\lambda)|^2$$

and

$$f_K(\lambda_1,\ldots,\lambda_{K-1}) = (2\pi)^{-K+1}p'\sum_{k=1}^{K}\phi\left(\sum_{j\in\nu_1}\lambda_j\right)\cdots$$
$$\times\,\phi\left(\sum_{j\in\nu_{k-1}}\lambda_j\right)\phi\left(\sum_{j\in\nu_k}\lambda_j\right)\mu^{\langle k\rangle}$$

where the summation extends over all partitions $(\nu_1,\ldots,\nu_k)$ of the set $(1,\ldots,K)$ and $\lambda_K = -\lambda_1 - \cdots - \lambda_{K-1}$.

The *Bartlett–Lewis process* is a Poisson cluster process with $p'(t) = p'$ and the individual cluster processes renewal processes (i.e., the points are located at $0 = u_0, u_1, u_2, \ldots$, with the $u_{j+1} - u_j$, $j = 0, 1, \ldots$, independent and identically distributed). The probability that the cluster is of size n is $\pi(n)$. The cumulant densities for this process may be determined from expression (2.11.31) once expressions have been determined for the cumulant densities of a renewal process. This is done later in this section.

One important problem concerning the analysis of cluster processes is that of identifying the points of the cluster center process M' given the points of the overall process N.

Consider next a Poisson process N_R on the plane $(-\infty,\infty)^2$, having parameter measure R satisfying $R(I\times(-\infty,\infty)), R((-\infty,\infty)\times I) < \infty$ for compact I in $B(-\infty,\infty)$. The probability generating functional of N_R is given by

$$E\exp\left\{\iint \log \mathcal{N}(u,v)N_R(du,dv)\right\} = \exp\left\{\iint [\mathcal{N}(u,v)-1]R(du,dv)\right\} \tag{2.11.35}$$

Consider the point process $N_R{}^*$ on the line determined by superposing the two marginals of N_R, namely

$$N_R{}^*(t) = N_R((-\infty,t]\times(-\infty,\infty)) + N_R((-\infty,\infty)\times(-\infty,t])$$

Its probability generating functional is determined from (2.11.35) by setting $\mathcal{N}(u,v) = \xi(u)\xi(v)$ and may be written

$$G[\xi] = \exp\left\{\int [\xi(u)-1][R(du\times(-\infty,\infty)) + R((-\infty,\infty)\times du]\right.$$
$$\left. + \iint [\xi(u)-1][\xi(v)-1]R(du,dv)\right\} \tag{2.11.36}$$

In the case in which R is absolutely continuous with density $r(u,v)$ and

$$r_1(u) = \int r(u,v)\,dv < \infty, \qquad r_2(v) = \int r(u,v)\,du < \infty$$

expression (2.11.36) shows that $N_R{}^*$ has intensity function $r_1(t) + r_2(t)$ and autocovariance density $2r(t_1,t_2)$. This construction shows that any nonnegative function $q_2(t_1,t_2)$ which is integrable in either variable, may be the autocovariance density of a point process. In the stationary case $q_2(t_1,t_2) = q_2(t_1-t_2)$, and it is seen that any nonnegative integrable function can be the autocovariance density of a stationary point process. The rate of the process, as constructed earlier, will be $q_1 = \int q_2(u)\,du$.

The *Gauss–Poisson process* (introduced by Newman, 1970, and discussed by Milne and Westcott, 1972) has probability generating functional

$$\exp\left\{\int [\xi(u)-1]Q_1(du) + \tfrac{1}{2}\iint [\xi(u)-1][\xi(v)-1]Q_2(du, dv)\right\} \qquad (2.11.37)$$

with $Q_2(du, dv) = Q_2(dv, du)$. It is seen that the Gauss–Poisson process may be represented as $N + N^*_{Q_2/2}$ where N is a Poisson process independent of N with intensity measure $Q_1(dt) - Q_2(dt \times (-\infty, \infty))$. It is clear that this procedure may be extended to construct point processes with given (integrable in $K-1$ arguments) nonnegative cumulant density of order K. From expression (2.5.16) the index of dispersion of the process here will be

$$I(t) = 1 + \frac{\{\int_0^t \int_0^t q_2(t_1 - t_2)\, dt_1\, dt_2\}}{\{\int_0^t \int q_2(t_1 - t_2)\, dt_1\, dt_2\}}$$

and so $I(\infty) = 2$. Also, $f_2(0) = q_1/\pi$ and $|f_2(\lambda) - q_1/2\pi| \le |f_2(0) - q_1/2\pi|$, indicating that this process too is overdispersed compared to a Poisson with the same intensity.

Hawkes (1972) introduced the class of *self-exciting point processes*. These are defined on the interval $(-\infty, \infty)$ and have conditional intensity of the form

$$\gamma(t, \omega) = \mu + \int_{-\infty}^{t} a(t-u)\, dN(u)$$
$$= \mu + \sum_{\tau_j \le t} a(t - \tau_j) \qquad (2.11.38)$$

where $\mu, a(u) \ge 0$, and $\int_0^\infty a(u)\, du < 1$. (The first condition here ensures that the conditional intensity is nonnegative; the second ensures that the process has finite intensity.) The parameters of such a process satisfy the following relationships:

$$p_1 = \mu + p_1 \int_0^\infty a(u)\, du$$

$$q_2(u) = \int_0^\infty a(v) q_2(u - v)\, dv$$

$$f_2(\lambda) = p_1/[2\pi |1 - A(\lambda)|^2]$$

$u > 0$, where $A(\lambda) = \int_0^\infty \exp\{-i\lambda u\} a(u)\, du$. The process is stationary, and has the following interpretation as a cluster process: (i) Immigrants arrive in accordance with a Poisson process of intensity μ. (ii) The immigrant who arrived at time τ_j generates descendants in accordance with a Poisson process of rate $a(t - \tau_j)$. (iii) The descendants in turn generate descendants, and so on. (This representation is discussed by Hawkes and Oakes, 1974.)

The process may be modulated by $\beta(t, \omega) = \beta$ to produce a further self-exciting process, provided $\beta \int_0^\infty a(u)\, du < 1$. By arranging for $A(\lambda)$ to be a rational function of λ, the power spectrum may be a rational function and the process will be analogous to the stationary N-tuple Markov Gaussian series.

An explicit expression is not known for the probability generating functional of a self-exciting process; however, it may be represented as

$$G[\xi] = \exp\left\{-\mu \int_{-\infty}^{\infty} (1 - H[\xi(\cdot + u)])\, du\right\}$$

where $H[\xi]$ satisfies the integral equation

$$H[\xi] = \xi(0) \exp\left\{-\int (1 - H[\xi(\cdot + u)]) a(u)\, du\right\}$$

$H[\xi]$ is here the probability generating functional of the cluster generated by an immigrant arriving at time 0 (see Hawkes and Oakes, 1974).

Using expressions (2.2.17) and (2.11.38) the likelihood function may be written down directly in terms of μ, $a(\cdot)$, and $N(t)$, $0 \le t \le T$. Also, from expression (2.2.16) one has, for example,

$$\text{Prob}\{\tau_J > t \,|\, \tau_j, j < J\}$$
$$= \exp\left\{-\mu(t - \tau_{J-1}) - \int_{-\infty}^{t} \int_{\max(v, \tau_{J-1})}^{t} a(u - v)\, du\, dN(v)\right\}$$

Rice (1975) has considered the problem of determining $A(\lambda)$ and $a(u)$ from p_1 and $f_2(\lambda)$. The particular case of $a(u) = \alpha \exp\{-\beta u\}$ is investigated in some detail by Oakes (1975).

Suppose that the members of the sequence $\tau_0, y_1, y_2, \ldots$ are statistically independent variates. Then the corresponding point process is called a *renewal process*. The Poisson and Erlang processes are examples of renewal processes. Generally the variates $y_1, y_2, \ldots$ are assumed to be identically distributed. Suppose their c.d.f. is $A(y)$, their density $a(y)$, and their hazard function $h(y) = (d/dy) \log[1 - A(y)]$. Suppose $A_0(y)$, $a_0(y)$, $h_0(y)$ are the corresponding parameters for τ_0. Then the conditional intensity function is here given by

$$\gamma(t, \omega) = h_0(t) \qquad \text{for} \quad N(t) = 0$$
$$= h(t - \tau_{N(t)-1}) \qquad \text{for} \quad N(t) \ge 1 \qquad (2.11.39)$$

By first principles, the intensity function of the process is given by

$$p_1(t) = a_0(t) + \int a(t-u) a_0(u)\, du$$
$$+ \iint a(t-u) a(u-v) a_0(v)\, du\, dv + \cdots \qquad (2.11.40)$$

Suppose that the Laplace transform of $p_1(t)$ is denoted, for Re $s > 0$, by $\hat{p}_1(t) = \int_0^\infty \exp\{-st\}p_1(t)\,dt$, with a similar definition for $\hat{a}_0(s)$, $\hat{a}(s)$. Taking the Laplace transform of each side of Eq. (2.11.40), we obtain

$$\begin{aligned}\hat{p}_1(s) &= \hat{a}_0(s) + \hat{a}_0(s)\hat{a}(s) + \hat{a}_0(s)\hat{a}(s)^2 + \cdots \\ &= \hat{a}_0(s)/[1 - \hat{a}(s)] \qquad (2.11.41)\end{aligned}$$

If the process is to be stationary, then $p_1(t) = p_1$ and Eq. (2.11.41) reads

$$p_1/s = \hat{a}_0(s)/[1 - \hat{a}(s)]$$

and therefore the initial distribution should have Laplace transform $\hat{a}_0(s) = p_1[1 - a(s)]/s$. It may be verified that this is the Laplace transform of $a_0(y) = p_1[1 - A(y)]$, and it may be further verified that for the density to integrate to 1, it is necessary that $p_1 = 1/E\{y_j\}$. The inverse character of the relationship between the expected interval length and the intensity should have been anticipated. In Section 2.5 it was seen that the relationship applied to general stationary point processes. The general theory of Palm distributions may be applied to show that the point process is strictly stationary with the choice $a_0(y) = [1 - A(y)]/E\{y_j\}$.

The product density of order 2 of the process is given by

$$p_2(t_1, t_2) = \text{Prob}\{dN(t_2) = 1 \,|\, N\{t_1\} = 1\}p(t_1)/dt_2$$

Now for $t_2 > t_1$

$$\begin{aligned}&\text{Prob}\{dN(t) = 1 \,|\, N\{0\} = 1\}/dt_2 \\ &\quad = a(t) + \int a(t-u)a(u)\,du + \iint a(t-u)a(u-v)a(v)\,du\,dv + \cdots \\ &\quad = b(t)\end{aligned}$$

The Laplace transform of $b(t)$ is here given by

$$\begin{aligned}\hat{b}(s) &= \int_0^\infty \exp\{-st\}b(t)\,dt = \hat{a}(s) + \hat{a}(s)^2 + \cdots \\ &= \hat{a}(s)/[1 - \hat{a}(s)] \qquad (2.11.42)\end{aligned}$$

for Re $s > 0$. In particular cases this relationship may be inverted to obtain $b(t)$, and then one has the product density of order 2 as

$$p_2(t_1, t_2) = p_1(t_1)b(t_2 - t_1)$$

for $t_2 > t_1$. In the stationary case $p_2(u) = p_1 b(|u|)$. This relationship may be used to determine the power spectrum. Except at $\lambda = 0$

$$\begin{aligned}f_2(\lambda) &= (2\pi)^{-1}p_1 + (2\pi)^{-1}\int \exp\{-i\lambda u\}p_2(u)\,du \\ &= (2\pi)^{-1}p_1 + (2\pi)^{-1}\int_0^\infty \exp\{-i\lambda u\}p_2(u)\,du \\ &\quad + (2\pi)^{-1}\int_0^\infty \exp\{i\lambda u\}p_2(u)\,du \\ &= (2\pi)^{-1}p_1 + (2\pi)^{-1}p_1\hat{b}(i\lambda) + (2\pi)^{-1}p_1\hat{b}(-i\lambda) \qquad (2.11.43)\end{aligned}$$

(by continuity, the same formula applies at $\lambda = 0$). This, together with (2.11.42), gives the required expression. A similar expression may be set down for the general cumulant spectrum. First set

$$g_k(\lambda_1, \ldots, \lambda_k) = p_1 \sum_P \hat{b}(i\lambda_{P1})\hat{b}(i\lambda_{P1} + i\lambda_{P2}) \cdots \hat{b}(\lambda_{P1} + \cdots + i\lambda_{Pk-1})$$

where the summation is over all permutations P of the set $(1, \ldots, k)$. Then

$$f_K(\lambda_1, \ldots, \lambda_{K-1}) = (2\pi)^{-K+1} \sum_{k=1}^{K} g_k\left(\sum_{j\in\nu_1}\lambda_j, \ldots, \sum_{j\in\nu_k}\lambda_j\right) \qquad (2.11.44)$$

where the summation is over all partitions $(\nu_1, \ldots, \nu_k)$ of $(1, \ldots, K)$ and it is understood that $\lambda_K = -\lambda_1 - \cdots - \lambda_{K-1}$. The corresponding product density is given by

$$p_K(t_1, \ldots, t_K) = p_1 b(t_2 - t_1) \cdots b(t_K - t_{K-1}) \qquad (2.11.45)$$

for $t_1 < t_2 < \cdots < t_K$.

The likelihood function of a given set of data on $[0, T]$ may be written

$$\begin{aligned}&h_0(\tau_0)h(y_1) \cdots h(y_J)\exp\left\{-\int_0^{\tau_0} h_0(u)\,du - \int_0^{y_1} h(u)\,du - \cdots\right. \\ &\qquad \left. - \int_0^{y_J} h(u)\,du - \int_0^{T-\tau_J} h(u)\,du\right\} \qquad (2.11.46)\end{aligned}$$

with $J = N(T) - 1$. Cox (1972) develops some properties of the modulated renewal process having conditional intensity function

$$\exp\{\beta_1 z_1(t) + \cdots + \beta_p z_p(t)\}\gamma(t, \omega)$$

where $\gamma(t, \omega)$ is the conditional intensity function of a renewal process, $\beta_1, \ldots, \beta_p$ are unknown parameters, and $z_1(t), \ldots, z_p(t)$ are known functions. For example, Cox suggests consideration of the likelihood conditional on the order statistics of $y_1, \ldots, y_{N(t)-1}$.

A variety of generalizations of the renewal process are possible. Expression (2.11.39) could be generalized to the case of nonidentical interpoint distributions by writing

$$\gamma(t, \omega) = h(t - \tau_{N(t)-1}, N(t))$$

where $h(u, j)$, $j = 0, 1, 2, \ldots$, are successive hazard functions. The renewal process may be generalized to the *random walk point process* by allowing the random variables $\tau_0, y_1, y_2, \ldots$ to take on negative values and then reordering the points obtained. Daley and Oakes (1974) develop a variety of properties of this process. Another form of extension allows a number of successive intervals to be dependent. Wold (1948) introduced the class of *Markov-dependent interval processes*. Vere-Jones (1975a) developed certain results for these processes.

P. A. W. Lewis and co-workers (see Jacobs and Lewis, 1976) have introduced classes of point processes in which the individual intervals have exponential distributions (as in the Poisson case), but the intervals are dependent. Their EARMA (1, 1) model takes the following form: (i) $\{\varepsilon_j\}$ is a sequence of independent exponential variates; (ii) $\{U_j\}$, $\{V_j\}$ are independent sequences of independent variates taking the values $\{0, 1\}$ with $\text{Prob}\{U_j = 0\} = \beta$, $\text{Prob}\{V_j = 0\} = \rho$; (iii) $A_j = \rho A_{j-1} + V_j \varepsilon_j$, $j = 1, 2, \ldots$, and $A_0 = \varepsilon_0$; (iv) the interval sequence of the process is given by

$$y_j = \beta \varepsilon_j + U_j A_{j-1}$$

$j = 1, 2, \ldots$; (v) $\tau_0 = 0$. This process is stationary. In the case $\rho = 0$, an "exponential moving average model" is obtained with power spectrum the ratio of fourth-order polynomials in λ.

Consider next a stationary Markov process $X(t)$, $0 \le t < \infty$, with state space $\{0, 1, 2, \ldots\}$ and transition intensities given by

$$\begin{aligned} \text{Prob}\{X(t + dt) = j \mid X(t) = k\} &= q_k Q_{kj}\, dt, \qquad j \neq k \\ &= 1 - q_k\, dt, \qquad j = k \end{aligned}$$

with $Q_{kk} = 0$, $\sum_j Q_{kj} = 1$. Rudemo (1973) considers the point process whose points correspond to the transition times of the process X. This process may be represented as follows: let $U(j)$, $j = 0, 1, \ldots$, be a Markov chain with transition matrix $Q = [Q_{kj}]$. Suppose the process began at its stationary distribution. Let $\{e(j)\}$, $j = 0, 1, \ldots$, be a sequence of independent exponentials with mean 1, and independent of the sequence $\{U(j)\}$. Then the points of the process may be represented as $\tau_0 = 0$,

$$\tau_{j+1} = \tau_j + \varepsilon(j)/q_{U(j)}$$

When the point process itself is Markov, it is referred to as a *pure birth process*. Its conditional intensity function is now given by

$$\gamma(t, \omega) = \beta(t, N(t)) \tag{2.11.47}$$

and initial conditions, such as $N(0) = 1$, are assumed. When $\beta(t, N) = \beta(t)N$, $\beta(t)$ is referred to as the birthrate at time t. Expression (2.11.47) shows that the process corresponds to modulating a unit Poisson process by the function $\beta(t, N(t))$. The likelihood function may be written

$$\exp\left\{-\int_0^T \beta(t, N(t))\, dt + \int_0^T \log \beta(t, N(t))\, dN(t)\right\}$$

In the case in which $\beta(t, N) = \beta N$, $N(0) = 1$, the stochastic process is really $N(t) - 1$ and the likelihood given data on $[0, T]$ is

$$\exp\left\{-\beta \int_0^T N(t)\, dt\right\} \times \beta^{N(T)-1} \times 2 \times 3 \times \cdots \times (N(T) - 1) \tag{2.11.48}$$

Details of this birth process have been developed by Keiding (1974).

A *stationary Markov point process* is provided by a doubly stochastic Poisson with time-independent intensity process $p(\omega')$. From expression (2.11.20), the conditional intensity is given by

$$\gamma(t, \omega) = \frac{E_{\omega'}\{p(\omega')^{N(t)+1} \exp\{-tp(\omega')\}\}}{E_{\omega'}\{p(\omega')^{N(t)} \exp\{-tp(\omega')\}\}}$$

showing that the process is in fact Markov. The intensity of the process is $\mu = Ep(\omega')$, the autocovariance density $\sigma^2 = \text{var}\, p(\omega')$, and the power spectrum is

$$\frac{\mu}{2\pi} + \frac{\sigma^2}{2\pi}\delta(\lambda)$$

The delta function at the origin is indicative of the nonmixing character of the process. The product densities of this process are all constant.

A direct counterpart of the pure birth process is provided by the *pure death process*, wherein the points of the process correspond to times of death. One way of setting up such a process is to consider elements such that

$$\text{Prob}\{\text{failure in } (t, t + dt)\} = \exp\left\{-\int_0^t h(u)\, du\right\} h(t)$$

The conditional intensity function of the process N_j corresponding to the life of a single element is

$$\gamma_j(t, \omega) = h(t)[1 - N_j(t^-)]$$

If $N(t) = \sum_j N_j(t)$ corresponds to the times of failure of a population of J independent elements, then its conditional intensity function is

$$h(t)[J - N(t^-)]$$

A great deal of literature exists concerning *birth and death processes* (e.g., Bharucha-Reid, 1960). These may be considered bivariate point processes $\{N_1(t), N_2(t)\}$ with the process N_1 referring to the times of births and the process N_2 referring to the times of deaths. The number living at time t is then $L(t) = N_1(t) - N_2(t)$ (assuming that there have been no immigrants). If births and deaths cannot occur simultaneously, then the processes N_1 and N_2 may be determined directly from the process L. An equivalent notation could apply to a queueing system with N_1 referring to the arrival times of customers, N_2 to departure times, and L to the number of customers in the system.

An extensive literature also exists concerning *branching processes*. Here $N(t)$ may refer to the number of individuals alive at time t. The process evolves through the property that each individual has a probability of producing further individuals for the population. The individuals may be classified as to age or generation. A general reference to the theory of branching processes is provided by Athreya and Ney (1972).

It is perhaps sensible to conclude this section with a warning to the effect that a number of the processes discussed can lead to identical functional forms for certain of the parameters considered. Daley and Oakes (1974) mention that a particular power spectrum that is a ratio of quadratics in λ can arise as (i) a self-exciting process with exponential kernel, (ii) a Neyman–Scott process with exponential displacements and a Poisson number of cluster members, (iii) a random walk point process with the generating distribution having exponential tails. In a similar vein, a flat spectrum occurs for either a stationary Poisson process or a doubly stochastic Poisson with white noise random intensity $p(t, \omega')$.

3. INFERENCE

3.1. Linear Statistics for Ordinary Time Series

Suppose that the data $\{X(t), 0 < t < T\}$ are available for analysis. In many circumstances it is of interest to compute a linear statistic of the form

$$\int_0^T \phi(t)X(t)\,dt \tag{3.1.1}$$

for a given function ϕ. For example, suppose that the series X is given by

$$X(t) = 0\phi(t) + \varepsilon(t) \tag{3.1.2}$$

$-\infty < t < \infty$, ϕ known, 0 unknown, and ε a zero mean series. Then the least squares estimate of 0 is

$$\hat{0} = \left\{\int_0^T \phi(t)X(t)\,dt\right\} \Big/ \left\{\int_0^T |\phi(t)|^2\,dt\right\} \tag{3.1.3}$$

a statistic based on (3.1.1). A theorem is presently given indicating the large sample behavior of such statistics. First, an assumption is formulated concerning the coefficient functions ϕ.

Assumption 3.1. Given the (possibly complex-valued) functions $\phi_j{}^T(t)$, $0 \le t < \infty$, $j = 1, \ldots, J$, suppose that as $T \to \infty$

$$\text{(i)} \qquad \left\{\int_0^T \phi_j{}^T(t+u)\overline{\phi_k{}^T(t)}\,dt\right\} \Big/ T \to b_{jk}(u)$$

for $0 \le u < \infty$ with $b_{jj}(u)$ finite and continuous at 0;

$$\text{(ii)} \qquad \{\sup_{0 \le t \le T} |\phi_j{}^T(t)|\}/\sqrt{T} \to 0$$

for $j, k = 1, \ldots, J$.

The conditions of this assumption are similar to those of Grenander (1954) and Hannan (1970, p. 215). Examples of functions satisfying the conditions are given shortly. Under the assumption, there exist functions of bounded variation $G_{jk}(\lambda)$, $-\infty < \lambda < \infty$, such that

$$b_{jk}(u) = \int_{-\infty}^{\infty} \exp\{iu\lambda\}\,dG_{jk}(\lambda)$$

for $0 \le u < \infty$ and $j, k = 1, \ldots, J$. The $dG_{jk}(\lambda)$ may be viewed as the limit of

$$dG_{jk}^T(\lambda) = (2\pi T)^{-1}\left[\int_0^T \exp\{-i\lambda t\}\phi_j{}^T(t)\,dt\right]\left[\int_0^T \exp\{i\lambda t\}\overline{\phi_k{}^T(t)}\,dt\right]d\lambda$$

in the sense of weak convergence. Now one can state the following theorem:

Theorem. Let $X(t)$, $-\infty < t < \infty$, be a series of the form $X(t) = \mu(t) + \varepsilon(t)$, with $\mu(t)$ nonstochastic and $\varepsilon(t)$ a zero mean stationary series satisfying condition (2.4.18) and having power spectrum $f_{\varepsilon\varepsilon}(\lambda)$. Suppose $\phi_j{}^T, j = 1, \ldots, J$, are functions satisfying Assumption 3.1. Let

$$U_j{}^T = \int_0^T \phi_j{}^T(t)X(t)\,dt$$

Then, as $T \to \infty$, the variate $\{U_1^T, \ldots, U_J^T\}$ is asymptotically normal with

$$EU_j^T = \int_0^T \phi_j^T(t)\mu(t)\,dt$$

$$\operatorname{cov}\{U_j^T, U_k^T\} \sim 2\pi T \int_{-\infty}^{\infty} f_{\varepsilon\varepsilon}(\lambda)\,dG_{jk}(-\lambda) \tag{3.1.4}$$

for $j, k = 1, \ldots, J$.

This theorem is proved quite simply by evaluating joint cumulants and showing that, when standardized, those of order greater than 2 tend to 0 as $T \to \infty$. The case of discrete time and of mixing conditions other than (2.4.18) has been considered by Hannan (1973a).

Example 1. Suppose $J = 1$, $\phi^T(t) = 1$. Then $b(u) = 1$, $dG(\lambda) = \delta(\lambda)\,d\lambda$, and

$$\int_0^T X(t)\,dt$$

is seen to be asymptotically normal with mean $\int_0^T \mu(t)\,dt$ and variance $2\pi T f_{\varepsilon\varepsilon}(0)$.

Example 2. Suppose $J = 1$, $\phi^T(t) = \exp\{-i\lambda_0 t\}$. Then $b(u) = \exp\{-i\lambda_0 u\}$, $dG(\lambda) = \delta(\lambda + \lambda_0)\,d\lambda$, and

$$\int_0^T \exp\{-i\lambda_0 t\}X(t)\,dt$$

is asymptotically (complex) normal with variance $2\pi T f_{\varepsilon\varepsilon}(\lambda_0)$.

Example 3. Suppose $J = 1$, $\phi^T(t) = (t/T)^n$. Then $b(u) = 1/(2n + 1)$, $dG(\lambda) = b(0)\,\delta(\lambda)\,d\lambda$, and

$$\left\{\int_0^T t^n X(t)\,dt\right\} \Big/ T^n$$

is asymptotically normal with variance $2\pi T f_{\varepsilon\varepsilon}(0)/(2n + 1)$.

Example 4. Suppose $J = 1$, $\phi^T(t) = \rho^{t/T}$, with $0 < \rho \leq 1$. Then $b(u) = (1 - \rho^2)/[2 \ln 1/\rho]$, $dG(\lambda) = b(0)\,\delta(\lambda)\,d\lambda$, and

$$\int_0^T \rho^{t/T} X(t)\,dt$$

is asymptotically normal with variance $2\pi T f_{\varepsilon\varepsilon}(0)b(0)$.

Example 5. Suppose the series is given by (3.1.2) with ϕ satisfying Assumption 3.1. Then $\hat{\theta}$ of (3.1.3) is asymptotically normal with mean θ and variance

$$2\pi T^{-1}\left[\int_{-\infty}^{\infty} f_{\varepsilon\varepsilon}(\lambda)\,dG(\lambda)\right] \Big/ b(0)^2$$

Example 6. Suppose the series X is given by (3.1.2) with ϕ satisfying Assumption 3.1. Consider choosing θ to minimize

$$E\left\{\int_0^T \int_0^T w(s - t)[X(s) - \theta\phi(s)][X(t) - \theta\phi(t)]\,ds\,dt\right\}$$

with $w(t) = \int \exp\{it\lambda\}W(\lambda)\,d\lambda$, $W(\lambda)$ being symmetric, nonnegative, bounded, and integrable. The extreme value of θ is given by

$$\left\{\int_0^T \int_0^T w(s - t)\overline{\phi(s)}\,X(t)\,ds\,dt\right\} \Big/ \left\{\int_0^T \int_0^T w(s - t)\overline{\phi(s)}\,\phi(t)\,ds\,dt\right\} \tag{3.1.5}$$

From the theorem, this statistic is asymptotically normal with mean θ and variance

$$2\pi T^{-1}\left[\int f_{\varepsilon\varepsilon}(\lambda)|W(\lambda)|^2\,dG(\lambda)\right] \Big/ \left|\int W(\lambda)\,dG(\lambda)\right|^2$$

By Schwarz's inequality

$$\int f_{\varepsilon\varepsilon}(\lambda)|W(\lambda)|^2\,dG(\lambda) \int f_{\varepsilon\varepsilon}(\lambda)^{-1}\,dG(\lambda) \geq \left|\int W(\lambda)\,dG(\lambda)\right|^2$$

and so the best choice of $W(\lambda)$ is $f_{\varepsilon\varepsilon}(\lambda)^{-1}$.

Results similar to those of Examples 5 and 6 may be developed for the case in which

$$X(t) = \theta_1 \phi_1(t) + \cdots + \theta_P \phi_P(t) + \varepsilon(t)$$

Related references include Kholevo (1969) and Rozanov (1969). Central limit theorems may also be developed directly for the case of a series X such that $\mathscr{A}[X]$ is a series satisfying (2.4.18) for some linear filter $\mathscr{A}$.

3.2. Linear Statistics for Point Processes

In the case of a point process N, given the events of a realization on $(0, T]$, one may compute linear statistics

$$\int_0^T \phi(t)\,dN(t) = \phi(\tau_0) + \cdots + \phi(\tau_{N(T)-1}) \tag{3.2.1}$$

for given functions ϕ. For example, suppose that N is a process with rate function of the form $p_1(t) = \theta\phi(t)$, ϕ known and θ unknown. Then

$$\hat{\theta} = \left\{\int_0^T \phi(t)\, dN(t)\right\} \Big/ \left\{\int_0^T |\phi(t)|^2\, dt\right\}$$

is a linear statistic providing an unbiased estimate of θ. In particular, the choices $\phi(t) = \cos \lambda t$ and $\phi(t) = \sin \lambda t$ lead to the computation of the finite Fourier transform

$$\int_0^T \exp\{-i\lambda t\}\, dN(t) \tag{3.2.2}$$

Alternatively, taking $\phi(t)$ to be approximately the Dirac delta $\delta(t - t_0)$ leads to a statistic whose expected value is approximately $p_1(t_0)$, the rate of the process at time t_0. In the case of a Poisson process with rate function $p(t)$, from expression (2.11.6) the log likelihood function is

$$\int_0^T \log p(t)\, dN(t) - \int_0^T p(t)\, dt$$

and again a linear statistic appears in an important fashion. Grandell (1972) and Clevenson and Zidek (1975) consider the use of linear statistics in the estimation of the rate function in the case of doubly stochastic and time inhomogeneous Poisson processes, respectively.

Turning to the consideration of large sample properties of linear statistics based on point processes, suppose that for the process N of interest

$$\operatorname{cum}\{dN(t_1), \ldots, dN(t_K)\} = C_K(dt_1, \ldots, dt_K)$$

$K = 1, 2, \ldots$, with the C_K of bounded variation in finite intervals, in the manner of Brillinger (1972). Suppose

$$U_j{}^T = \int_0^T \phi_j{}^T(t)\, dN(t), \qquad j = 1, \ldots, J$$

Then one has the following theorem:

Theorem. Suppose that the

$$\int_0^T \int_0^T \phi_j{}^T(t_1)\phi_j{}^T(t_2) C_2(dt_1, dt_2)$$

$j = 1, \ldots, J$, are each of order of magnitude $\sigma_T{}^2$ as $T \to \infty$. Suppose in addition that

$$\int_0^T \cdots \int_0^T \phi_{j_1}^T(t_1) \cdots \phi_{j_K}^T(t_K) C_K(dt_1, \ldots, dt_K) = o(\sigma_T{}^K)$$

for $j_1, \ldots, j_K = 1, \ldots, J$ and $K = 3, 4, \ldots$. Then the variate $\{U_1{}^T, \ldots, U_J{}^T\}$ is asymptotically normal.

The proof of this theorem is immediate, as the standardized joint cumulants of order greater than 2 tend to 0 under the indicated assumptions.

In the case in which the process N is stationary, satisfying the mixing condition (2.5.29), and the $\phi_j{}^T$ satisfy Assumption 3.1, the conditions of the theorem are satisfied. The variate $\{U_1{}^T, \ldots, U_J{}^T\}$ is asymptotically normal with the covariance of $U_j{}^T$ and $U_k{}^T$ given by $2\pi T \int f_2(\lambda)\, dG_{jk}(-\lambda)$. In particular, the finite Fourier transform (3.2.2) is asymptotically (complex) normal with variance $2\pi T f_2(\lambda)$.

The discussion of this section has concentrated on the counts of the process. Clearly linear statistics may be formed in the intervals. The interval sequence is an ordinary discrete time series and results of the character of those of Section 3.1 may be set down directly.

3.3. Quadratic Statistics for Ordinary Time Series

A variety of the quantities computed in the analysis of ordinary time series are quadratic in the observations. Suppose that the data $\{X(t), 0 < t < T\}$ are available. Some examples are as follows:

(i) $$c_2{}^T(u) = T^{-1} \int_0^{T-|u|} [X(t+u) - c_1{}^T][X(t) - c_1{}^T]\, dt$$

with $c_1{}^T = T^{-1} \int_0^T X(t)\, dt$, computed as an estimate of the autocovariance function $c_2(u)$;

(ii) the periodogram

$$I_2{}^T(\lambda) = (2\pi T)^{-1} \left| \int_0^T \exp\{-i\lambda t\} X(t)\, dt \right|^2$$

(iii) The empirical spectral measure

$$F_2{}^T(\lambda) = \frac{2\pi}{T} \sum_{0 < 2\pi s/T \le \lambda} I_2{}^T\left(\frac{2\pi s}{T}\right)$$

(iv) $$f_2{}^T(\lambda) = \frac{2\pi}{T} \sum_s W^T\left(\lambda - \frac{2\pi s}{T}\right) I_2{}^T\left(\frac{2\pi s}{T}\right) \tag{3.3.1}$$

with W^T concentrated near 0 and integrating to 1, or

$$f_2{}^T(\lambda) = L^{-1} \sum_{j=1}^{L} I_2{}^T\left(\frac{2\pi s_j}{T}\right) \tag{3.3.2}$$

with $s_1, \ldots, s_J$ distinct positive integers near $T\lambda/2\pi$, as estimates of the power spectrum $f_2(\lambda)$

$$\text{(v)} \qquad \frac{(2\pi)^2}{T} \sum_s I_{ee}^T\left(\frac{2\pi s}{T}\right)\left|W\left(\frac{2\pi s}{T}\right)\right|^2 I_{\phi\phi}^T\left(\frac{2\pi s}{T}\right)$$

with $e(t)$ the residual series $X(t) - \hat{\theta}\phi(t)$, $\hat{\theta}$ given by (3.1.5), as an estimate of the numerator of the variance (3.1.6) and finally

$$\text{(vi)} \qquad \sum_{s\neq 0}\left[\log f\left(\frac{2\pi s}{T}\right)\Big/ g\left(\frac{2\pi s}{T}\right) + I_2{}^T\left(\frac{2\pi s}{T}\right)\left\{\left[1/f\left(\frac{2\pi s}{T}\right)\right] - \left[1/g\left(\frac{2\pi s}{T}\right)\right]\right\}\right]$$

computed for various functions f and g in the Gaussian estimation of time series parameters (see Dzhaparidze and Yaglom, 1973).

These examples lead to the consideration of quadratic expressions of the form

$$V_j{}^T = \frac{2\pi}{T}\sum_{s\neq 0} A_j{}^T\left(\frac{2\pi s}{T}\right) I_2{}^T\left(\frac{2\pi s}{T}\right) = \iint a_j{}^T(s-t)X(s)X(t)\,ds\,dt \tag{3.3.3}$$

where

$$a_j{}^T(t) = \frac{2\pi}{T}\sum_{s\neq 0}\exp\left\{-i\frac{2\pi st}{T}\right\} A_j{}^T\left(\frac{2\pi s}{T}\right)$$

$j = 1, \ldots, J$. In connection with these variates, one has the following theorem:

Theorem. Suppose that the series X satisfies condition (2.4.18). Suppose $A_j{}^T(\alpha)$ of (3.3.3) $= A_j(\alpha)$ with A_j continuous, bounded, and absolutely integrable. Then the variate $\{V_1{}^T, \ldots, V_J{}^T\}$ is asymptotically normal with

$$EV_j{}^T \sim \int A_j(\alpha) f_2(\alpha)\,d\alpha \tag{3.3.4}$$

$$\operatorname{cov}\{V_j{}^T, V_k{}^T\} \sim \frac{2\pi}{T}\int A_j(\alpha)\overline{A_k(\alpha)} f_2(\alpha)^2\,d\alpha + \frac{2\pi}{T}\int A_j(\alpha)A_k(\alpha) f_2(\alpha)^2\,d\alpha + \frac{2\pi}{T}\iint A_j(\alpha)\overline{A_k(\beta)} f_4(\alpha, -\alpha, -\beta)\,d\alpha\,d\beta \tag{3.3.5}$$

as $T \to \infty$. Suppose $A_j{}^T(\alpha) = B_T^{-1} W(B_T^{-1}[\alpha - \lambda_j])$ in the manner of (3.3.1), with W continuous, bounded, absolutely integrable, and integrating to 1.

Then the variate $\{V_1{}^T, \ldots, V_J{}^T\}$ is asymptotically normal with

$$EV_j{}^T \sim f_2(\lambda_j)$$
$$\operatorname{cov}\{V_j{}^T, V_k{}^T\} \sim 2\pi\int W(\alpha)^2\,d\alpha\,(B_T T)^{-1} f_2(\lambda_j)^2 \quad \text{if} \quad \lambda_j = \pm\lambda_k \neq 0$$
$$\sim 0 \quad \text{if} \quad \lambda_j \neq \lambda_k$$

Suppose $J = 1$ and A^T is such as to give the statistic (3.3.2). Then the variate (3.3.2) is asymptotically $f_2(\lambda)\chi^2_{2L}/(2L)$.

The large sample distributions of the statistics at the beginning of this section may be obtained by choice of appropriate $A_j{}^T$.

Hannan (1973b) develops certain large sample results for quadratic statistics in discrete time series.

3.4. Quadratic Statistics for Point Processes

Important second-order parameters for stationary point processes include the *product density* $p_2(u)$ defined by

$$p_2(u)\,dt\,du = \text{Prob}\{dN(t+u) = 1 \text{ and } dN(t) = 1\} \tag{3.4.1}$$

$u \neq 0$; the *covariance density* $q_2(u)$ of

$$q_2(u)\,dt\,du = \operatorname{cov}\{dN(t+u), dN(t)\} \tag{3.4.2}$$

$u \neq 0$, so that $q_2(u) = p_2(u) - p_1{}^2$; the *power spectrum*

$$f_2(\lambda) = (2\pi)^{-1}\left[p_1 + \int \exp\{-i\lambda u\} q_2(u)\,du\right] \tag{3.4.3}$$

$-\infty < \lambda < \infty$; the *spectral measure*

$$F_2(\lambda) = \int_0^\lambda f_2(\alpha)\,d\alpha \tag{3.4.4}$$

the *variance time curve*

$$V(t) = \operatorname{var} N(t) = tp_1 + 2\int_0^t (t-u) q_2(u)\,du \tag{3.4.5}$$

$$= \int\left(\frac{\sin t\alpha/2}{\alpha/2}\right)^2 f_2(\alpha)\,d\alpha \tag{3.4.6}$$

and finally

$$\operatorname{var}\left\{\int \phi(t)\,dN(t)\right\} = p_1\int \phi(t)^2\,dt + \iint \phi(s)\phi(t) q_2(s-t)\,ds\,dt = \int |\Phi(\alpha)|^2 f_2(\alpha)\,d\alpha \tag{3.4.7}$$

where $\Phi(\alpha) = \int \exp\{i\alpha t\}\phi(t)\,dt$.

Given the stretch of data $\{N(t),\ 0 < t \le T\}$, with events at points $\tau_0 < \tau_1 < \cdots < \tau_{N(T)-1}$, estimates of each of these parameters may be constructed. To begin, suppose that β_T is a nonnegative scale factor tending to 0 as $T \to \infty$. As an estimate of $p_2(u)$ now consider

$$p_2{}^T(u) = \text{card}\{(j, k) \text{ such that } u - \beta_T < \tau_j - \tau_k < u + \beta_T\}/(2\beta_T T) \tag{3.4.8}$$

and

$$p_2{}^T(u) = \sum_{j,k} \frac{w([u - \tau_j + \tau_k]/\beta_T)}{(\beta_T T \int w(t)\, dt)} \tag{3.4.9}$$

with $w(t)$ nonnegative, as estimates of $p_2(u)$. Suppose the process N satisfies condition (2.5.29). Then when $\beta_T = L/T$, L fixed, the estimate (3.4.8) is distributed asymptotically as $\mathscr{P}/2L$ where $\mathscr{P}$ is Poisson with mean $2Lp_2(u)$. On the other hand, when $\beta_T \to 0$, but $\beta_T T \to \infty$, the estimate (3.4.9) is asymptotically normal with variance

$$\left[p_2(u) \int w(t)^2\, dt\right] \Big/ \left(\beta_T T \left|\int w(t)\, dt\right|^2\right)$$

These results may be found in the chapter by Brillinger (1975b). The covariance density may be estimated by

$$q_2{}^T(u) = p_2{}^T(u) - (N(T)/T)^2$$

involving either of the statistics (3.4.8) or (3.4.9).

The periodogram of the data under discussion is the quadratic statistic given by

$$I_2{}^T(\lambda) = (2\pi T)^{-1} \left|\int_0^T \exp\{-i\lambda t\}\, dN(t)\right|^2$$

Estimates of the power spectrum may be based on it through the formula (3.3.1) or (3.3.2). As was the case with ordinary time series, the first estimate will be asymptotically normal, the second distributed asymptotically as a multiple of a chi-squared.

A further situation in which a quadratic statistic based on point process data occurs is with the approximate log likelihood function

$$-\sum_{s=1}^{S} \left[\log f_2\left(\frac{2\pi s}{T}\right) + I_2{}^T\left(\frac{2\pi s}{T}\right) \Big/ f_2\left(\frac{2\pi s}{T}\right)\right] \tag{3.4.10}$$

derived by thinking of the periodogram ordinates as independent exponential variates (e.g., Hawkes and Adamopoulos, 1973).

The expressions (3.4.4), (3.4.6), (3.4.7), and (3.4.10) each suggest consideration of quadratic statistics of the form

$$V_j{}^T = \frac{2\pi}{T} \sum_{s=1}^{S} A_j\left(\frac{2\pi s}{T}\right) I_2{}^T\left(\frac{2\pi s}{T}\right)$$

$j = 1, \ldots, J$. Provided the $A_j(\alpha)$ are continuous, bounded, and absolutely integrable, the variate $\{V_1{}^T, \ldots, V_J{}^T\}$ is asymptotically normal with first- and second-order moments given by (3.3.4) and (3.3.5).

The construction of estimates of certain higher order "polynomial" parameters of time series and point processes by polynomial functions of the observations is discussed by Brillinger (1972, 1975b, c).

3.5. General Parameter Estimation for Ordinary Time Series

In a variety of circumstances, the probability distribution of an ordinary time series X depends on a finite-dimensional parameter θ. For example, the series might be defined by

$$X(t) = \alpha_0 \varepsilon(t) + \alpha_1 \varepsilon^{(1)}(t) + \cdots + \alpha_p \varepsilon^{(p)}(t) \tag{3.5.1}$$

or

$$\beta_0 X(t) + \beta_1 X^{(1)}(t) + \cdots + \beta_q X^{(q)}(t) = \varepsilon(t) \tag{3.5.2}$$

where ε is a white noise series of variance σ^2, and one might be interested in estimating the parameter $\theta = (\alpha_0, \ldots, \alpha_p, \sigma^2)$ in case (3.5.1) and $\theta = (\beta_0, \ldots, \beta_q, \sigma^2)$ in case (3.5.2). In other situations the series might have the form corresponding to linear regression on $0 < t < T$

$$X(t) = \theta_1 \phi_1{}^T(t) + \cdots + \theta_P \phi_P{}^T(t) + \varepsilon(t) \tag{3.5.3}$$

with ε a stationary error series, or the form corresponding to nonlinear regression

$$X(t) = \psi^T(t; \theta) + \varepsilon(t) \tag{3.5.4}$$

where the functional form of ψ^T is known, but not the actual value of θ.

A number of general estimation procedures take the following form: a loss function $Q^T(\theta; X)$ is given, and an estimate $\hat{\theta}$ taken as the value of θ giving the minimum of $Q^T(\theta; X)$. Examples of such procedures include *linear least squares* where $\theta = (\theta_1, \ldots, \theta_P)$ of (3.5.3) is estimated by minimizing

$$Q^T(\theta; X) = T^{-1} \int_0^T |X(t) - \theta_1 \phi_1{}^T(t) - \cdots - \theta_P \phi_P{}^T(t)|^2\, dt \tag{3.5.5}$$

in the case of (3.5.3); *nonlinear least squares* where

$$Q^T(\theta; X) = T^{-1} \int_0^T |X(t) - \psi^T(t; \theta)|^2\, dt \tag{3.5.6}$$

in the case of (3.5.4); *Gaussian estimation* where, as in the case of (3.5.1) and (3.5.2), the functional form $f_2(\lambda; \theta)$ of the power spectrum is known and

$$Q^T(\theta; X) = \frac{2\pi}{T} \sum_{s=1}^{S} \left[\log f_2\left(\frac{2\pi s}{T}; \theta\right) \Big/ g\left(\frac{2\pi s}{T}\right) + I_2{}^T\left(\frac{2\pi s}{T}\right) \left| 1 \Big/ f_2\left(\frac{2\pi s}{T}; \theta\right) - 1 \Big/ g\left(\frac{2\pi s}{T}\right) \right| \right] \tag{3.5.7}$$

for some given function $g(\lambda)$; *maximum likelihood* where $Q^T(\theta; X)$ is taken as the negative of the likelihood ratio relative to some fixed measure, of the given data.

A lemma indicating conditions under which a $\hat{\theta}$ constructed in this manner is consistent, is presented by Brillinger (1975b). Among the conditions are the existence of $Q(\theta)$ with $Q(\theta) > Q(\theta_0)$ for $\theta \neq \theta_0$, the true parameter value, and $Q^T(\theta; X) \geq Q(\theta) + o_p(1)$, $Q^T(\theta_0; X) = Q(\theta_0) + o_p(1)$. Under further regularity conditions

$$\frac{\partial Q^T(\theta_0; X)}{\partial \theta_0}$$

is asymptotically $N_P(0, \Sigma_T)$,

$$\frac{\partial^2 Q^T(\theta_0; X)}{\partial \theta_0 \, \partial \theta_0^*}$$

tends to Ψ in probability and $\hat{\theta}$ is asymptotically $N_p(\theta_0, \Psi^{-1}\Sigma_T\Psi^{-1})$.

In the case of (3.5.3) and (3.5.5) and where the $\phi_j{}^T$ satisfy Assumption 3.1, one has

$$Q(\theta) = (\theta - \theta_0)[b_{jk}(0)](\theta - \theta_0)^* + \int f_{zz}(\alpha)\, d\alpha$$

$$\Psi = 2[b_{jk}(0)]$$

$$\Sigma_T = T^{-1} 8\pi \left[\int f_{zz}(\alpha)\, dG_{jk}(-\alpha) \right]$$

In the case of (3.5.7),

$$Q(\theta) = \int \left[\frac{\log f_2(\alpha; \theta)}{g(\alpha)} + f_2(\alpha; \theta_0) \left\{ \frac{1}{f_2(\alpha; \theta)} - \frac{1}{g(\alpha)} \right\} \right] d\alpha$$

$$\Psi = \left[\int \frac{\partial \log f_2(\alpha; \theta_0)}{\partial \theta_j} \frac{\partial \log f_2(\alpha; \theta_0)}{\partial \theta_k} d\alpha \right]$$

$$\Sigma_T = T^{-1}\Psi + T^{-1} \left[2\pi \iint \frac{\partial \log f_2(\alpha; \theta_0)}{\partial \theta_j} \frac{\partial \log f_2(\beta; \theta_0)}{\partial \theta_k} \times \frac{f_4(\alpha, -\alpha, -\beta; \theta_0)}{f_2(\alpha; \theta_0) f_2(\beta; \theta_0)} d\alpha\, d\beta \right]$$

As an example of (3.5.4) and (3.5.6), consider the model of a series as a sum of decaying cosines:

$$X(t) = \sum_{k=1}^{K} \alpha_k \exp\left\{ -\frac{\phi_k t}{T} \right\} \cos\{\gamma_k t + \delta_k\} + \varepsilon(t)$$

for $0 < t < T$. The matrices Ψ, Σ_T may be evaluated and the estimates found to be asymptotically normal with

$$\overrightarrow{\text{var}}\, \hat{\alpha} \sim T^{-1} 4\pi f_{zz}(\gamma) I_2(\phi) J(\phi)^{-1}$$

$$\overrightarrow{\text{var}}\, \hat{\phi} \sim T^{-1} 4\pi f_{zz}(\gamma) I_0(\phi) J(\phi)^{-1} \alpha^{-2}$$

$$\overrightarrow{\text{var}}\, \hat{\gamma} \sim T^{-3} 4\pi f_{zz}(\gamma) I_0(\phi) J(\phi)^{-1} \alpha^{-2}$$

$$\overrightarrow{\text{var}}\, \hat{\delta} \sim T^{-1} 4\pi f_{zz}(\gamma) I_2(\phi) J(\phi)^{-1} \alpha^{-2}$$

where $I_j(\phi) = \int_0^1 u^j \exp\{-2\phi u\}\, du$ and $J(\phi) = I_0(\phi) I_2(\phi) - I_1(\phi)^2$.

As an example of the use of the method of maximum likelihood, consider the diffusion process determined by

$$dX(t) = \mu(t, X(t))\, dt + dW(t)$$

where W is the Wiener process, $\mu(t, x) = \sum \theta_k \phi_k(t, x)$ with $\phi_k(t, x)$ known, $\theta_0 = 1$, and $X(0) = 0$. From (2.10.8), the logarithm of its likelihood ratio relative to W is given by

$$\sum_{k=0}^{K} \theta_k \int_0^T \phi_k(t, X(t))\, dX(t) - \tfrac{1}{2} \sum_{j,k=0}^{K} \theta_j \theta_k \int_0^T \phi_j(t, X(t)) \phi_k(t, X(t))\, dt$$

The maximum likelihood estimate of $\theta = (\theta_1, \ldots, \theta_K)$ is the solution of the system of equations

$$\sum_{k=1}^{K} \hat{\theta}_k \int_0^T \phi_j(t, X(t)) \phi_k(t, X(t))\, dt = \int_0^T \phi_j(t, X(t))[dX(t) - \phi_0(t, X(t))\, dt]$$

$j = 1, \ldots, K$. Under regularity conditions, including

$$p\text{-}\lim_{T \to \infty} T^{-1} \int_0^T \phi_j(t, X(t)) \phi_k(t, X(t))\, dt = I_{jk}(\theta)$$

Taraskin (1974) shows that $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)$ is asymptotically normal with mean θ and covariance matrix $T^{-1}[I_{jk}(\theta)]^{-1}$ as $T \to \infty$.

In the case in which $\phi_k(t, x) = \phi_k(x)$, the process X may be stationary. Suppose that it has stationary distribution $G_\theta(x)$; then Taraskin (1974) shows that

$$I_{jk}(\theta) = \int \phi_j(x) \phi_k(x)\, dG_\theta(x)$$

$j, k = 1, \ldots, K$. The stationary case is also considered by Brown and Hewitt (1975) and Kulinich (1975). When a diffusion process has known local var-

iance $\sigma^2(t, x)$, it may be transformed so that $\sigma(t, x) = 1$ as earlier. Brown and Hewitt (1975) remark that in the case in which $\sigma(t, x) = \sigma(x)$, the latter value may be determined almost surely, as

$$\lim_{n\to\infty} \sum_{j=1}^{2^n} [X(jt2^{-n}) - X((j-1)t2^{-n})]^2 = \int_0^t \sigma^2(X(u))\, du$$

almost surely.

As a further use of the method of maximum likelihood consider the linear dynamic system (2.8.9) wherein $F, \ldots, L$ depend on an unknown parameter θ, but not on t, i.e., consider the system

$$\begin{aligned} dS(t) &= F_\theta S(t)\, dt + G_\theta X(t)\, dt + K_\theta\, dW_1(t) \\ dY(t) &= H_\theta S(t)\, dt + J_\theta X(t)\, dt + L_\theta\, dW_2(t) \end{aligned} \tag{3.5.8}$$

where $W_1(t)$, $W_2(t)$ are independent Wiener processes. Suppose that the dependence of $F_\theta, \ldots, L_\theta$ on θ is such that the system (3.5.8) is identifiable. (Conditions under which this occurs may be found, e.g., in Balakrishnan, 1973; Glover and Williams, 1974.) Suppose that the data $\{X(t), Y(t), 0 < t < T\}$ are available and that it is desired to estimate θ. If $L_\theta = L$, then the likelihood ratio of the process Y relative to the process LW_2 would be given by

$$\exp\left\{L^{-2}\int_0^T (H_\theta \hat{S}_\theta(t) + J_\theta X(t))\, dY(t) - L^{-2}\int_0^T (H_\theta \hat{S}_\theta(t) + J_\theta X(t))^2\, dt/2\right\} \tag{3.5.9}$$

where

$$\hat{S}_\theta(t) = F_\theta \int_0^t \hat{S}_\theta(u)\, du + G_\theta \int_0^t X(u)\, du + L^{-2}\int_0^t P_\theta(u) H_\theta^*\, dv_\theta(u)$$

$$v_\theta(t) = Y(t) - H_\theta \int_0^t \hat{S}_\theta(u)\, du - J_\theta \int_0^t X(u)\, du$$

and

$$\frac{dP_\theta(t)}{dt} = F_\theta P_\theta(t) + P_\theta(t) F_\theta^* + K_\theta K_\theta^* - L^{-2} P_\theta(t) H_\theta^* H_\theta P_\theta(t)$$

with $P_\theta(0) = 0$. As an estimate of θ, consider the value $\hat{\theta}$ that maximizes expression (3.5.9) for some L. Balakrishnan (1973) shows that under regularity conditions $\hat{\theta} \to \theta$ in probability provided $L_\theta = L$. Bagchi (1975) shows that this convergence continues to occur for other values of L, e.g., $L = 1$. Gupta and Mehra (1974) consider computational aspects of the problem.

Linear regression analysis of time series is discussed, for example, by Grenander (1954) and Hannan (1970). Nonlinear regression is discussed by Hannan (1971) and Robinson (1972). The particular case of ψ^T a sum of sinusoids is investigated by Hannan (1973a). Gaussian estimation is considered by Munk and MacDonald (1960), Whittle (1962), Walker (1964), and Dzhaparidze and Yaglom (1973). Extension of the Gaussian procedure to the estimation of parameters of linear systems is carried out by Akisik (1975). Davies (1973) describes the construction of asymptotically optimal tests and estimators in the case of Gaussian time series. The use of the likelihood function in the estimation for Markov processes is indicated, for example, by Billingsley (1961) and Roussas (1972).

3.6. General Parameter Estimation for Point Processes

Consider now a point process N whose probability distribution depends on a P-dimensional parameter θ. Suppose that the data $\{N(t), 0 < t \le T\}$ of points located at $\tau_0 < \tau_1 < \cdots < \tau_{N(T)-1}$ are available for analysis. Let the conditional intensity function of the process be given by $\gamma(t; \theta)$. Following expression (2.2.17), the log likelihood function of the data may be written

$$-\int_0^T \gamma(t; \theta)\, dt + \int_0^T \log \gamma(t; \theta)\, dN(t) = -\int_0^T \gamma(t; \theta)\, dt + \sum_{j=0}^{N(T)-1} \log \gamma(\tau_j; \theta)$$

Define

$$Q^T(\theta; N) = T^{-1}\int_0^T \gamma(t; \theta)\, dt - T^{-1}\int_0^T \log \gamma(t; \theta)\, dN(t)$$

in the manner of the preceding section. The maximum likelihood estimate $\hat{\theta}$ is obtained by minimizing $Q^T(\theta; N)$. Under regularity conditions θ is the solution of the system of equations $\partial Q^T/\partial\hat{\theta} = 0$, i.e.,

$$\int_0^T \frac{\partial\gamma(t; \hat{\theta})}{\partial\hat{\theta}_k}\, dt - \sum_{j=0}^{N(T)-1} \frac{\log \gamma(\tau_j; \hat{\theta})}{\partial\hat{\theta}_k} = 0 \tag{3.6.1}$$

As in the preceding section, $\hat{\theta}$ will be consistent and asymptotically normal under regularity conditions. Here

$$EQ^T(\theta; N) = T^{-1}\int_0^T E\{\gamma(t; \theta)\}\, dt - T^{-1}\int_0^T E\{\gamma(t; \theta_0) \log \gamma(t; \theta)\}\, dt \tag{3.6.2}$$

for $\theta = \theta_0$, the entropy of McFadden (1965). In the stationary mixing case, expression (3.6.2) will tend to $E\{\gamma(t; \theta) - \gamma(t; \theta_0) \log \gamma(t; \theta)\}$. Let θ_0 denote the true value of the parameter. As in the preceding section, with

$$\frac{\partial Q^T(\theta_0; N)}{\partial\theta_0}$$

asymptotically $N_P(0, \Sigma_T)$ and

$$\frac{\partial^2 Q^T(\theta_0; N)}{\partial\theta_0\, \partial\theta_0^*} \sim \Psi_T$$

in probability, $\hat{\theta}$ will be asymptotically $N_P(\theta, \Psi_T^{-1}\Sigma_T\Psi_T^{-1})$. Here

$$\Psi_T = T^{-1}\int_0^T E\left\{\gamma(t;\theta)^{-1}\frac{\partial\gamma(t;\theta)}{\partial\theta}\frac{\partial\gamma(t;\theta)^*}{\partial\theta}\right\}dt$$

and

$$\Sigma_T = T^{-2}\int_0^T E\left\{\gamma(t;\theta)^{-1}\frac{\partial\gamma(t;\theta)}{\partial\theta}\frac{\partial\gamma(t;\theta)^*}{\partial\theta}\right\}dt$$

See Vere-Jones (1975b).

As a particular case of these results, consider the time inhomogeneous Poisson process with rate function $p^T(t;\theta)$ on $0 < t < T$. Suppose

$$\lim_{T\to\infty} T^{-1}\int_0^T [p^T(t;\theta) - p^T(t;\theta_0)\log p^T(t;\theta)]\,dt = Q(\theta)$$

where $Q(\theta) > Q(\theta_0)$ for $\theta \neq \theta_0$. Suppose also

$$\lim_{T\to\infty} T^{-1}\int_0^T p^T(t;\theta_0)\frac{\partial\log p^T(t;\theta_0)}{\partial\theta_0}\frac{\partial\log p^T(t;\theta_0)^*}{\partial\theta_0}\,dt = \Psi$$

Then the large sample distribution may be approximated by $N_P(\theta, T^{-1}\Psi)$. The particular case of

$$p^T(t;\theta) = \exp\left\{\sum_1^P \theta_k\phi_k(t)\right\}$$

seems important in practice, with the ϕ_k known functions (see Lewis, 1970; Cox, 1972; Keiding, 1974).

Keiding (1974) considers the case of the pure birth process where $\gamma(t;\theta) = \theta N(t)$ and $N(0) = 1$. The maximum likelihood estimate is here

$$\hat{\theta} = (N(T) - 1)\Big/\left\{\int_0^T N(t)\,dt\right\}$$

and $\{[\int_0^T N(t)\,dt]/\theta\}^{1/2}(\hat{\theta} - \theta)$ is asymptotically $N(0, 1)$.

Snyder (1975, p. 251) considers the case of a Poisson process of rate θ incident on a type I counter with dead time Δ. The conditional intensity function is here

$$\gamma(t;\theta) = \theta \quad \text{for} \quad \tau_{N(t)-1} + \Delta \le t$$
$$= 0 \quad \text{otherwise}$$

from (2.7.13). Snyder shows that the maximum likelihood estimate of θ is

$$\hat{\theta} = \hat{p}\left[\frac{\tau_{N(t)-1} + \Delta}{T} - \hat{p}\Delta\right]^{-1} \quad \text{if} \quad \tau_{N(T)-1} \le T < \tau_{N(t)-1} + \Delta$$
$$= \hat{p}[1 - \hat{p}\Delta]^{-1} \quad \text{if} \quad \tau_{N(T)-1} + \Delta \le T$$

where $\hat{p} = N(T)/T$.

Vere-Jones (1975b) discusses a variety of aspects of maximum likelihood estimation for point processes. Aalen (1975) considers the case of $\gamma(t;\theta) = \theta(t)\phi(t,\omega)$ with ϕ given and possibly depending on the past of N. Aalen investigates the maximum likelihood estimation of θ in the case of constant $\theta(t)$, and considers the estimation of

$$\Theta(t) = \int_0^t \theta(u)J(u)\,du \quad \text{by} \quad \hat{\Theta}(t) = \int_0^t \phi(u)^{-1}J(u)\,dN(u)$$

where $J(t) = \lim_{h\downarrow 0} I\{\phi(t-h) > 0\}$, I here being the indicator function. He obtains a central limit theorem by assuming he is dealing with a sequence of such processes.

An analog of the Gaussian estimation procedure of the preceding section may be set down in this case of point process data. Suppose the process has mean rate $p(\theta)$ and power spectrum $f(\lambda;\theta)$. Set $g(\lambda;\theta) = f(\lambda;\theta)/p(\theta)$, $\hat{p} = N(\tau)/T$. As an estimate of θ, take the value $\hat{\theta}$ minimizing

$$Q^T(\theta; N) = \frac{2\pi}{T}\sum_{s=1}^{S}\left[\log g\left(\frac{2\pi s}{T};\theta\right) + I_2{}^T\left(\frac{2\pi s}{T}\right)\Big/\left\{\hat{p}g\left(\frac{2\pi s}{T};\theta\right)\right\} - 1\right]$$

In Brillinger (1975a), conditions are set down under which these estimates are consistent and asymptotically $N_P(\theta_0, \Psi^{-1}\Sigma_T\Psi^{-1})$ where

$$\Psi = \left[\int \frac{\partial\log g(\alpha;\theta)}{\partial\theta_j}\frac{\partial\log g(\alpha;\theta)}{\partial\theta_k}\,d\alpha\right]_{\theta=\theta_0}$$

$$\Sigma_T = T^{-1}\Psi + T^{-1}\left[2\pi\iint \frac{\partial\log g(\alpha;\theta)}{\partial\theta_j}\frac{\partial\log g(\beta;\theta)}{\partial\theta_k} \times \frac{f_4(\alpha,-\alpha,-\beta;\theta)}{g(\alpha;\theta)g(\beta;\theta)}\,d\alpha\,d\beta\right]_{\theta=\theta_0}$$

Akisik (1975) extends this procedure to certain "linear" point process systems.

Davies (1977) investigates the problem of developing optimum tests of the hypothesis that a given point process is Poisson. Snyder (1975) discusses tests of hypotheses relevant to certain point processes.

REFERENCES

1. Aalen, O. O. (1975). Statistical Inference for a Family of Counting Processes. PhD. Thesis, Univ. of California, Berkeley, California.
2. Akisik, V. A. (1975). On the Estimation of Parametric Transfer Functions. PhD. Thesis, Univ. of California, Berkeley, California.
3. Anderson, T. W. (1971). "Statistical Analysis of Time Series." Wiley, New York.
4. Aronszajn, N. (1950). Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68**, 337–404.
5. Athreya, K. B., and Ney, P. E. (1972). "Branching Processes." Springer-Verlag, Berlin.
6. Bagchi, A. (1975). Continuous time systems identification with unknown noise covariance, *Automatica* **11**, 533–536.
7. Balakrishnan, A. V. (1973). Stochastic Differential Systems I, "Lecture Notes in Economics and Mathematical Systems," No. 84. Springer-Verlag, Berlin.
8. Bartlett, M. S. (1966). "Stochastic Processes." Cambridge Univ. Press, London and New York.
9. Benes, V. E. (1963). "General Stochastic Processes in the Theory of Queues." Addison-Wesley, Reading, Massachusetts.
10. Bharucha-Reid, A. T. (1960). "Elements of the Theory of Markov Processes and Their Applications." McGraw-Hill, New York.
11. Bhat, U. N. (1969). Sixty years of queueing theory, *Management Sci.* **15**, 280–294.
12. Billingsley, P. (1961). "Statistical Inference for Markov Processes." Univ. of Chicago Press, Chicago, Illinois.
13. Billingsley, P. (1968). "Convergence of Probability Measures." Wiley, New York.
14. Bochner, S. (1960). "Harmonic Analysis and the Theory of Probability." Univ. of California Press, Berkeley, California.
15. Box, G. E. P., and Jenkins, G. M. (1970). "Time Series Analysis, Forecasting and Control." Holden Day, San Francisco, California.
16. Breiman, L. (1968). "Probability." Addison-Wesley, Reading, Massachusetts.
17. Bremaud, P. M. (1974). The martingale theory of point processes over the real half line, "Lecture Notes in Economics and Mathematics Systems." Springer-Verlag, Berlin.
18. Brillinger, D. R. (1970). The identification of polynomial systems by means of higher order spectra, *J. Sound Vibrat.* **12**, 301–313.
19. Brillinger, D. R. (1972). The spectral analysis of stationary interval functions. 483–513 in *Proc. Berkeley Symp. 6th* (L. M. Le Cam, J. Neyman, and E. L. Scott, eds.), Vol. 1. Univ. of California Press, Berkeley, California.
20. Brillinger, D. R. (1973). Estimation of the mean of a stationary time series by sampling. *J. Appl. Probability* **10**, 419–431.
21. Brillinger, D. R. (1974). Fourier analysis of stationary processes. *Proc. IEEE* **62**, 1628–1643.
22. Brillinger, D. R. (1975a). "Time Series: Data Analysis and Theory." Holt, New York.
23. Brillinger, D. R. (1975b). Statistical inference for stationary point processes, *in* "Stochastic Processes and Related Topics" (M. L. Puri, ed.), Vol. 1, pp. 55–99. Academic Press, New York.
24. Brillinger, D. R. (1975c). Estimation of product densities, *in Comput. Sci. Statist., Annu. Symp. Interface, 8th UCLA, Los Angeles*, pp. 431–438.
25. Brillinger, D. R., Bryant, H. L., Jr., and Segundo, J. P. (1976). Identification of synaptic interactions, *Biol. Cybernet.* **22**, 213–228.
26. Brown, B. M., and Hewitt, J. I. (1975). Asymptotic likelihood theory for diffusion processes, *J. Appl. Probability* **12**, 228–238.

27. Bryant, H., and Segundo, J. P. (1975). How does the neuronal spike trigger read Gaussian white noise transmembrane current? *in Proc. Symp. Testing Identification Nonlinear Syst., 1st* (G. D. McCann and P. Z. Marmarelis, eds.), pp. 236–247. California Inst. of Technology, Pasadena, California.
28. Chou, C-S., and Meyer, P. A. (1975). Sur la representation des martingales comme integrales stochastiques dans les processus pontuels, "Lecture Notes in Mathematics," No. 465, pp. 226–236. Springer-Verlag, Berlin.
29. Cinlar, E. (1972). Superposition of point processes, *In* "Stochastic Point Processes" (P. A. W. Lewis, ed.), pp. 549–606. Wiley, New York.
30. Clevenson, M. L., and Zidek, J. V. (1975). Bayes linear estimates of the intensity function of the nonstationary Poisson process, *J. Amer. Stat. Assoc.* **72**, 112–120.
31. Cox, D. R. (1962). "Renewal Theory." Methuen, London.
32. Cox, D. R. (1972). The statistical analysis of dependencies in point processes, *in* "Stochastic Point Processes" (P. A. W. Lewis, ed.), pp. 55–66. Wiley, New York.
33. Cox, D. R., and Lewis, P. A. W. (1966). "The Statistical Analysis of Series of Events." Methuen, London.
34. Cox, D. R., and Smith, W. L. (1961). "Queues." Methuen, London.
35. Cramér, H. (1942). On harmonic analysis in certain functional spaces, *Ark. Math. Astr. Fysik* **28**, 1–7.
36. Cramér, H. (1976). Half a century with probability theory: some personal reflections, *Ann. Probability* **4**, 509–546.
37. Cramér, H., and Leadbetter, M. R. (1967). "Stationary and Related Stochastic Processes." Wiley, New York.
38. Cramér, H., Leadbetter, M. R., and Serfling, R. J. (1971). On distribution function–moment relationships in a stationary point process, *Z. Wahrschein.* **18**, 1–8.
39. Daley, D. J. (1969). Spectral properties of weakly stationary point processes, *Bull. Inst. Internat. Statist.* **37**, No. 2, 344–346.
40. Daley, D. J., and Oakes, D. (1974). Random walk point processes, *Z. Wahrschein.* **30**, 1–16.
41. Daley, D. J., and Vere-Jones, D. (1972). A summary of the theory of point processes, *in* "Stochastic Point Processes" (P. A. W. Lewis, ed.), pp. 299–383. Wiley, New York.
42. Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time series, *Advances Appl. Probability* **5**, 469–497.
43. Davies, R. B. (1977). Testing the hypothesis that a point process is Poisson *Adv. Appl. Prob.* (to appear).
44. Doob, J. (1953). "Stochastic Processes." Wiley, New York.
45. Dzhaparidze, K. O., and Yaglom, A. M. (1973). Asymptotically efficient estimates of the spectrum parameters of stationary stochastic processes, *Proc. Prague Symp. Asympt. Statistics*, pp. 55–106. Karlova Univ., Prague.
46. Ephremides, A., and Thomas, J. B. (1973). "Random Processes." Dowden, Hutchinson and Ross, Stroudsburg, Pennsylvania.
47. Feinberg, S. (1974). Stochastic models for single neuron firing trains: a Survey, *Biometrics* **30**, 399–427.
48. Furstenberg, H. (1960). "Stationary Processes and Prediction Theory." Princeton Univ. Press, Princeton, New Jersey.
49. Gaver, D. P., and Lewis, P. A. W. (1977). First order autoregressive Gamma sequences and point processes (to appear).
50. Gelfand, I. M., and Vilenkin, N. Ya. (1964). "Generalized Functions," p. 4. Academic Press, New York.

51. Geman, D., and Horowitz, J. (1973). Remarks on Palm measures, *Ann. Inst. Henri Poincare B* **9**, 215–232.
52. Gihman, I. I., and Skorohod, A. V. (1972). "Stochastic Differential Equations." Springer-Verlag, Berlin.
53. Glover, K., and Williams, J. C. (1974). Parameterization of linear dynamical systems: canonical forms and identifiability, *IEEE Trans. Automatic Control* **AC-19**, 640–646.
54. Grandell, J. (1972). Statistical inference for doubly stochastic Poisson processes, *in* "Stochastic Point Processes" (P. A. W. Lewis, ed.), pp. 90–121. Wiley, New York.
55. Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance, *Ann. Math. Statist.* **25**, 252–272.
56. Gupta, N. K., and Mehra, R. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations, *IEEE Trans. Automatic Control* **AC-19**, December.
57. Haight, F. A. (1967). "Handbook of the Poisson Distribution." Wiley, New York.
58. Hannan, E. J. (1970). "Multiple Time Series." Wiley, New York.
59. Hannan, E. J. (1971). Nonlinear time series regression, *J. Appl. Probability* **8**, 767–780.
60. Hannan, E. J. (1973a). The estimation of frequency, *J. Appl. Probability* **10**, 510–519.
61. Hannan, E. J. (1973b). Multivariate time series analysis, *J. Multivariate Anal.* **3**, 395–407.
62. Hannan, E. J. (1973c). Central limit theorems for time series regression. *Z. Wahrschein.* **26**, 157–170.
63. Harris, T. E. (1963). "The Theory of Branching Processes." Springer-Verlag, Berlin.
64. Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables, *in* "Stochastic Point Processes" (P. A. W. Lewis, ed.), pp. 261–271. Wiley, New York.
65. Hawkes, A. G., and Adamopoulos, L. (1973). Cluster models for earthquakes—regional comparisons, *Bull. Inst. Internat. Statist.* **45(3)**, 454–461.
66. Hawkes, A. G., and Oakes, D. (1974). A cluster process representation of a self-exciting process, *J. Appl. Probability* **11**, 493–504.
67. Hida, T. (1960). Canonical representations of Gaussian processes and their applications, *Mem. Coll. Sci. Kyoto A* **33**, 110–155.
68. Hida, T. (1970). "Stationary Stochastic Processes." Princeton Univ. Press, Princeton, New Jersey.
69. Ito, K. (1951). On stochastic differential equations, *Mem. Amer. Math. Soc.* **4**, 51.
70. Jacobs, P. A., and Lewis, P. A. W. (1977). A mixed autoregressive-moving average exponential sequence and point process *Adv. Appl. Prob.* **9**, 87–104.
71. Jacod, J. (1975). Multivariate point processes: predictable projection, Radon Nikodym derivatives, representation of martingales, *Z. Wahrschein.* **31**, 235–253.
72. Jagers, P. (1974). Aspects of random measures and point processes, *Advan. Probability* **3**, 179–239.
73. Jowett, J. H., and Vere-Jones, D. (1972). The prediction of stationary point processes, *in* "Stochastic Point Processes" (P. A. W. Lewis, ed.), pp. 405–435. Wiley, New York.
74. Kailath, T. (1970). The innovations approach to detection and estimation theory, *Proc. IEEE* **58**, 680–695.
75. Kailath, T. (1971). RKHS approach to detection and estimation problems—Part I: deterministic signals in Gaussian noise, *IEEE Trans. Informat. Theory* **IT-17**, 530–549.
76. Kallenberg, O. (1975). Limits of compound and thinned point processes, *J. Appl. Probability* **12**, 269–278.
77. Keiding, N. (1974). Estimation in the birth process, *Biometrika* **61**, 71–80.
78. Kendall, D. G. (1975). The genealogy of genealogy branching processes before (and after) 1873, *Bull. London Math. Soc.* **7**, 225–253.

79. Kerstan, J., Matthes, K., and Mecke, J. (1974). "Unbegrenzt teilbare Punktprozesse." Akademie-Verlag, Berlin.
80. Khintchine, A. Ya. (1934). Korrelationstheories der stationaren Prozesse, *Math. Ann.* **109**, 604–615.
81. Kholevo, A. S. (1969). On estimation of regression coefficients, *Theory Probability Appl.* **14**, 79–104.
82. Knight, F. B. (1975). A predictive view of continuous time processes, *Ann. Probability* **3**, 573–596.
83. Kolmogorov, A. N. (1940). Curves in Hilbert space invariant with regard to a one parameter group of motions, *Dokl. Akad. Nauk. SSSR* **26**, 6–9.
84. Koopmans, L. H. (1974). "The Spectral Analysis of Time Series." Academic Press, New York.
85. Korenberg, M. J. (1973). Cross-correlation analysis of neural cascades, *Proc. Ann. Rocky Mountain Bioeng. Symp., 10th* 47–51.
86. Krickeberg, K. (1974). Moments of point processes, *in* "Stochastic Geometry" (E. F. Harding and D. G. Kendall, eds.), pp. 89–113. Wiley, New York.
87. Kulinich, G. L. (1975). On the estimation of the drift parameter of a stochastic diffusion equation, *Theory Probability Appl.* **20**, 384–387.
88. Kurtz, T. G. (1974). Point processes and completely monotone set functions, *Z. Wahrschein.* **31**, 51–67.
89. Leadbetter, M. R. (1972). Point processes generated by level crossings, *in* "Stochastic Point Processes" (P. A. W. Lewis, ed.). Wiley, New York.
90. Lewis, P. A. W. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events, *J. Sound Vibrat.* **12**, 353–375.
90a. Lewis, P. A. W. (ed.) (1972). "Stochastic Point Processes." Wiley, New York.
90b. Loève, M. (1955). "Probability Theory." Van Nostrand-Reinhold, Princeton, New Jersey.
90c. Lotka, A. J. (1957). "Elements of Mathematical Biology." Dover, New York.
91. McCann, G. D., and Marmarelis, P. Z. (1975). *Proc. Symp. Testing and Identificat. of Nonlinear Syst., 1st*, California Inst. of Technol., Pasadena, California.
92. McFadden, J. A. (1965). The entropy of a point process, *SIAM J. Appl. Math.* **13**, 988–994.
93. Macchi, O. (1971). Stochastic point processes and multicoincidences, *Proc. IEEE Trans. Informat. Theory* **IT-17**, 2–7.
94. Macchi, O. (1975). The coincidence approach to stochastic point processes, *Advan. Appl. Probability* **7**, 83–122.
95. Masani, P. (1972). On helixes in Hilbert space. I, *Theory Probability Appl.* **17**, 1–19.
96. Meyer, P. A. (1971). Demonstration simplifiee d'un theoreme de Knight, "Lecture Notes in Math," No. 191, pp. 191–195. Springer-Verlag, Berlin.
97. Milne, R. K., and Westcott, M. (1972). Further results for Gauss-Poisson processes, *Advan. Appl. Probability* **4**, 151–176.
98. Mirasol, N. M. (1963). The output of an M/G/∞ queueing system is Poisson, *Operations Res.* **11**, 282–284.
99. Moore, G. P., Perkel, D. H., and Segundo, J. P. (1966). Statistical analysis and functional interpretation of neuronal spike data, *Ann. Rev. Physiol.* **28**, 493–522.
100. Moyal, J. E. (1962). The general theory of stochastic population processes, *Acta Math.* **108**, 1–31.
101. Munk, W. H., and MacDonald, G. J. F. (1960). "The Rotation of the Earth." Cambridge Univ. Press, London and New York.
102. Murthy, V. K. (1974). "The General Point Process." Addison-Wesley, Reading, Massachusetts.

103. Nelson, E. (1959). Regular probability measures on function space, *Ann. Math.* **69**, 630–643.
104. Neveu, J. (1968). Sur la structure des processus ponctuels stationnaires, *C. R. Acad. Sci. Paris* **267**, 561–564.
105. Newman, D. S. (1970). A new family of point processes which are characterized by their second moment properties, *J. Appl. Probability* **7**, 338–358.
106. Oakes, D. (1975). The Markovian self-exciting process, *J. Appl. Probability* **12**, 69–77.
107. Papangelou, F. (1974). On the Palm probabilities of processes of points and processes of lines, *in* "Stochastic Geometry" (E. F. Harding and D. G. Kendall, eds.), pp. 114–147. Wiley, New York.
108. Parzen, E. (1959). Statistical inference on time series by Hilbert space methods, I, *in* "Time Series Analysis Papers" (1967, E. Parzen, ed.), pp. 251–382. Holden-Day, San Francisco, California.
109. Parzen, E. (1962). Extraction and detection problems and reproducing kernel Hilbert spaces, *SIAM J. Control* **1**, 35–62.
110. Pfaffelhuber, E. (1975). Generalized harmonic analysis for distributions, *IEEE Trans. Informat. Theory* **IT-21**, 605–611.
111. Prabhu, N. U. (1965). "Queues and Inventories." Wiley, New York.
112. Prohorov, Yu V. (1961). The method of characteristic functionals, *in Proc. Berkeley Symp. Math. Statist. Probability, 4th* (J. Neyman, ed.), Vol. 2, pp. 403–419. Univ. of California Press, Berkeley, California.
113. Prohorov, Yu V., and Rozanov, Yu A. (1969). "Probability Theory." Springer-Verlag, Berlin.
114. Quenouille, M. H. (1949). Problems in plane sampling, *Ann. Math. Statist.* **20**, 355–375.
115. Ramakrishnan, A. (1950). A stochastic process relating to particles distributed in a continuous infinity of states, *Proc. Cambridge Phil. Soc.* **46**, 596–602.
116. Renyi, A. (1967). Remarks on the Poisson process, *Studia Sci. Math. Hungar.* **2**, 119–123.
117. Rice, J. (1975). Estimated factorisation of the spectral density of a stationary point process, *Advan. Appl. Probability* **7**, 801–817.
118. Rice, S. O. (1945). Mathematical analysis of random noise, *Bell Syst. Tech. J.* **24**, 46–156.
119. Robinson, P. M. (1972). Non-linear regression for multiple time series, *J. Appl. Probability* **9**, 758–768.
119a. Roussas, G. G. (1972). "Contiguous Probability Measures." Cambridge Univ. Press, London.
120. Rozanov, Yu A. (1966). "Stationary Random Processes." Holden-Day, San Francisco, California.
121. Rozanov, Yu. A. (1969). On a new class of estimates, *in* "Multivariate Analysis" (P. R. Krishnaiah, ed.), Vol. II, pp. 437–441. Academic Press, New York.
122. Rubin, I. (1972). Regular point processes and their detection, *IEEE Trans. Informat. Theory* **IT-18**, 547–557.
123. Rudemo, M. (1973). Point processes generated by transitions of Markov chains, *Advan. Appl. Probability* **5**, 262–286.
124. Ruelle, D. (1969). "Statistical Mechanics." Benjamin, New York.
125. Ryll-Nardzewski, C. (1961). Remarks on the process of calls, *in Proc. Berkeley Symp., 4th* (J. Neyman, ed.), Vol. 2, pp. 455–465. Univ. of California Press, Berkeley, California.
126. de Sam Lazaro, J., and Meyer, P. A. (1975). Questions de theorie des flots, *in* "Lecture Notes in Mathematics," No. 465, pp. 2–96. Springer-Verlag, Berlin.
127. Schwartz, L. (1973). Surmartingales regulieres a valeurs mesures et desintegrations regulieres d'une mesure, *J. Anal. Math.* **26**, 1–168.

128. Segall, A., and Kajlath, T. (1975). The modelling of randomly modulated jump processes, *IEEE Trans. Informat. Theory* **IT-21**, 135–143.
129. Shiryaev, A. N. (1963). Some problems in the spectral theory of higher-order moments, I, *Theory Probability Appl.* **5**, 265–284.
130. Snyder, D. L. (1975). "Random Point Processes." Wiley, New York.
131. Srinivasan, S. K. (1969). "Stochastic Theory and Cascade Processes." Amer. Elsevier, New York.
132. Srinivasan, S. K. (1974). "Stochastic Point Processes." Hafner, New York.
133. Taraskin, A. F. (1974). On the asymptotic normality of vector-valued stochastic integrals and estimates of drift parameters of a multidimensional diffusion process, *Theory Probability Math. Stat.* **2**, 209–224.
134. Udias, A., and Rice, J. (1975). Statistical analysis near San Andreas Observatory, Hollister, California, *Bull. Seismol. Soc. Amer.* **65**, 809–827.
135. Varayia, P. (1975). The martingale theory of jump processes, *IEEE Trans. Automatic Control* **AC-20**, 1.
136. Vere-Jones, D. (1968). Some applications of probability generating functionals to the study of input-output streams, *J. Roy. Statist. Soc. B* **30**, 321–333.
137. Vere-Jones, D. (1974). An elementary approach to the spectral theory of stationary random measures, *in* "Stochastic Geometry" (E. F. Harding and D. G. Kendall, eds.), pp. 307–321. Wiley, New York.
138. Vere-Jones, D. (1975a). A renewal equation for point processes with Markov dependent intervals, *Math. Nach.* **68**, 133–139.
139. Vere-Jones, D. (1975b). On updating algorithms for inference for stochastic processes, *in* "Perspectives in Probability and Statistics" (J. Gani, ed.), pp. 239–259. Academic Press, New York.
139a. Valkonskii, V. A., and Rozanov, Yu. A. (1959). Some limit theorems for random functions, I. *Theory Probability Appl.* **4**, 178–197.
140. Walker, A. M. (1964). Asymptotic properties of least squares estimates of parameters of the spectrum of stationary non-deterministic time series, *J. Austral. Math. Soc.* **4**, 363–384.
141. Wang, G. L. (1968). Contagion in stochastic models for epidemics, *Ann. Math. Statist.* **39**, 1863–1889.
142. Westcott, M. (1970). Identifiability in linear processes, *Z. Wahrschein.* **16**, 39–46.
143. Westcott, M. (1971). On existence and mixing results for cluster point processes, *J. Roy. Statist. Soc. B* **33**, 290–300.
144. Westcott, M. (1972). The probability generating functional, *J. Austral. Math. Soc.* **14**, 448–466.
145. Westergaard, H. (1968). "Contributions to the History of Statistics." Agathon, New York.
146. Whittle, P. (1962). Gaussian estimation in stationary time series, *Bull. Inst. Internat. Statist.* **39**, 105–129.
147. Whittle, P. (1963). "Prediction and Regulation." English Univ. Press, London.
148. Wiener, N. (1930). Generalized harmonic analysis, *Acta Math.* **55**, 117–258.
149. Wiener, N. (1949). "Time Series." MIT Press, Cambridge, Massachusetts.
150. Wiener, N. (1958). "Nonlinear Problems in Random Theory." MIT Press, Cambridge, Massachusetts.
151. Wold, H. (1948). On stationary point processes and Markov chains, *Skand. Akt.* **31**, 299–340.
152. Yaglom, A. M. (1962). "An Introduction to the Theory of Stationary Random Functions." Prentice-Hall, Englewood Cliffs, New Jersey.
153. Yvon, J. (1935). "La Theories Statistique des Fluids et l'Equation d'Etat." Herman, Paris.