The Spectral Analysis of Point Processes

Author(s): M. S. Bartlett
Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 25, No. 2 (1963), pp. 264-296
Published by: Wiley for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/2984295
Accessed: 11-03-2019 13:57 UTC

The Spectral Analysis of Point Processes

By M. S. BARTLETT
University College, London

[Read at a RESEARCH METHODS MEETING of the SOCIETY, May 1st, 1963, Professor D. R. Cox in the Chair]

SUMMARY

The spectral analysis of stationary point processes in one dimension is developed in some detail as a statistical method of analysis. The asymptotic sampling theory previously established by the author for a class of doubly stochastic Poisson processes is shown to apply also for a class of clustering processes, the spectra of which are contrasted with those of renewal processes. The analysis is given for two illustrative examples, one an artificial Poisson process, the other of some traffic data. In addition to testing the fit of a clustering model to the latter example, the analysis of these two examples is used where possible to check the validity of the sampling theory.

1. DEFINITIONS

THE statistical estimation of the spectra of stationary stochastic processes has been intensively studied in recent years (see, for example, Bartlett, 1951, 1955, 1963; Grenander and Rosenblatt, 1957), it being well known that some statistical smoothing device is required in the case of estimating spectral density functions. At the end of my 1963 paper, I considered, though rather briefly, the analogous problem for stationary *point* processes. My purpose here will be to discuss the spectral analysis of point processes rather more comprehensively, largely with a view to accumulating further information on its practical value as a method of studying different *kinds* of departures from the completely random or Poisson point process. It is relevant to note that precise statistical methods of *detecting* such departures have been discussed in various contexts (see, for example, Greenwood, 1946; Cox, 1955; Thompson, 1955; cf. also Epstein, 1960), but further analysis and interpretation of a non-Poisson process, while also previously discussed—for example, by Cox and Thompson—is more difficult and arbitrary. For this reason, spectral analysis, while it is not claimed to be more than one among various possible methods (one or two other possibilities were mentioned in my contribution to the discussion on Cox's (1955) paper; cf. also Appendix), is worth study as a systematic procedure of some value and importance. I shall discuss here only the one-dimensional case, although it has obvious extensions to two or three dimensions; these I hope to discuss in a later paper. I will merely note now the relevance of the two-dimensional extension to the problem Thompson (1955) was concerned with, of two-dimensional point processes in plant ecology. There is also an obvious connection with the probing by X-rays of geometrical configurations of molecules or atoms in a gas or liquid (cf. Green, 1952, Ch. III). My own interest in the spectra of point processes is mainly, however, in relation to statistical methods of analysis.

To recapitulate first on one or two definitions, point processes are stochastic processes specified in relation to events or individuals each labelled with the random value of a continuous parameter, which may be the time parameter t itself. This last case is considered here, so that the process can be specified by the cumulative number $N(t)$ of events or individuals up to time t. Under certain conditions, such as prohibiting the occurrence of two or more events at the same time (note that McFadden (1962) does not so restrict point processes), the process $N(t)$ has well-defined "product density" or "factorial moment density" functions (see Bartlett, 1955, Section 3.42), e.g. if E denotes expectation,

$$\frac{E\{dN(t)\}}{dt} = \lambda(t), \quad \frac{E\{dN(t)\,dN(t+\tau)\}}{(dt)^2} = \mu'(t, t+\tau),$$

where $dN(t) = N(t+dt)-N(t)$, dt is the usual notation for the differential of t, and $\tau > 0$. For *stationary* point processes, $\lambda(t)$ is constant, λ say, and $\mu'(t, t+\tau)$ is a function of τ only. A "covariance density" function may then be defined by

$$\mu(\tau) = \frac{E\{dN(t)\,dN(t+\tau)\}}{(dt)^2} - \lambda^2 \quad (\tau > 0). \tag{1}$$

For $\tau < 0$, $\mu(\tau) = \mu(-\tau)$, but for $\tau = 0$ it is important to note that

$$E\{[dN(t)]^2\} = E\{dN(t)\}.$$

It is sometimes convenient to write, for all real τ, the complete covariance density

$$\mu_c(\tau) = \lambda\delta(\tau) + \mu(\tau), \tag{2}$$

where $\delta(\tau)$ is the Dirac delta-function, and $\mu(\tau)$ is continuous at $\tau = 0$.

The complete spectral density function for $N(t)$ is defined by

$$f_c(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\omega}\, \mu_c(\tau)\, d\tau, \tag{3}$$

and it follows that

$$g(\omega) = 2\pi f_c(\omega) = \lambda + \int_{-\infty}^{\infty} e^{-i\tau\omega}\, \mu(\tau)\, d\tau. \tag{4}$$

For events occurring purely at random (the so-called Poisson process), $\mu(\tau) = 0$, and the expression $g(\omega)$ in (4) reduces to λ (for all real ω). As in the case of ordinary stationary processes, it is convenient further to define a spectral density for non-negative ω by

$$g_+(\omega) = 2g(\omega) = 2\lambda + 2g_\mu(\omega), \tag{5}$$

where $g_\mu(\omega)$ is the second term on the right of equation (4).

2. CLUSTERING MODELS

An interesting device, first introduced by Cox (1955) in the study of point processes, is to postulate that $N(t)$ arises from the special case of a Poisson process by the (average) rate of occurrence $\lambda(t)$ at time t being replaced by the value of a stochastic process $\Lambda(t)$. In the case of stationary processes, $N(t)$, or rather $dN(t)$, must be stationary with $E\{\Lambda(t)\} = \lambda$. It is then easy to show that the covariance density function $\mu(\tau)$ is identical with the autocovariance function of $\Lambda(t)$; in fact, the more complete relation holds between the characteristic functionals of $N(t)$ and $\Lambda(t)$ (see my contribution to the discussion following Cox's paper),

$$E\left\{\exp\int_0^T i\theta(t)\,dN(t)\right\} = E_\Lambda\left\{\exp\int_0^T \Lambda(t)\,[e^{i\theta(t)}-1]\,dt\right\}, \tag{6}$$

where E_Λ denotes averaging with respect to $\Lambda(t)$.

Relation (6) follows, of course, on the supposition that $\Lambda(t)$ exists. Although by appropriate choice of $\Lambda(t)$ a wide range of processes can be represented (thus, if $\Lambda(t)$ contains a periodic component, the spectrum of $\Lambda(t)$, and hence of $N(t)$, has a discrete component), in the case of "anomalous" point processes with infinite covariance density this relation does not hold. For example, (6) implies, from the identity of $\mu(\tau)$ with the autocovariance function of $\Lambda(t)$, that $g_\mu(\omega)$ is positive, as distinct from $g(\omega)$. Suppose $N(t)$ is a Poisson process, except that every time an event occurs at t, a second occurs at $t+a$, i.e. events occur in pairs separated by the interval a. Then clearly, if λ_p is the density of occurrence of pairs, and $\tau>0$,

$$\mu'(t,t+\tau) = \lambda_p\,\delta(\tau-a)+\lambda^2$$

and, with $2\lambda_p = \lambda$,

$$2g_\mu(\omega) = 2\lambda\cos a\omega.$$

Although

$$g_+(\omega) = 2\lambda+2\lambda\cos a\omega$$

remains positive, this is not true for $g_\mu(\omega)$ alone.

The above process would become more regular if the interval between pairs had a distribution with a density $f(\alpha)$. More generally, define a "Poisson clustering process" (in one dimension) by postulating that a "parent" or "nucleus" in a Poisson process is followed at intervals $\alpha_1, \alpha_1+\alpha_2, \ldots, \alpha_1+\alpha_2+\ldots\alpha_r$, by r "offspring" or "satellites". The distribution of r is $p(r)$ $(r = 0, 1, \ldots)$. The total intervals $\alpha_1+\alpha_2, \ldots, \alpha_1+\alpha_2+\ldots\alpha_r$ could in general be a set with arbitrary simultaneous density function, but for simplicity (cf. also, however, Section 5) we suppose that the successive quantities $\alpha_1, \alpha_2 \ldots$ are independent and identically distributed, so that the distribution of $\alpha_1+\alpha_2+\ldots\alpha_r$ is the r-fold convolution of the distribution of α_1, say, with density function $f(\alpha)$ and corresponding characteristic function

$$M(\theta) = \int_0^\infty e^{i\theta\alpha} f(\alpha)\,d\alpha.$$

The characteristic function of $\alpha_1+\alpha_2+\ldots+\alpha_r$ is then well known to be $M^r(\theta)$, with corresponding density function $f_r(\alpha)$, say. For simplicity, no inhibition effect between clusters is supposed, though for some applications it would be useful to prohibit cluster overlap.

Contributions to $\mu(\tau)$ arise only within the same cluster (or "family"). Let λ_c be the density of clusters. The first element $dN(t)$ in $dN(t)\,dN(t+\tau)$ may be any one of the cluster. Evaluation of all the possible terms gives, for $\tau>0$,

$$\mu(\tau) = \lambda_c[f(\tau)\{p(1)+2p(2)+3p(3)+\ldots\}+f_2(\tau)\{p(2)+2p(3)+\ldots\} \\ +f_3(\tau)\{p(3)+\ldots\}+\ldots], \tag{7}$$

whence

$$g_\mu(\omega) = \lambda_c[\{M(\omega)+M(-\omega)\}E(r)+\{M^2(\omega)+M^2(-\omega)\}E'(r-1) \\ +\{M^3(\omega)+M^3(-\omega)\}E'(r-2)+\ldots], \tag{8}$$

where the prime denotes finding the contribution to the expectation for *positive* values of the variable. Obviously also

$$\lambda = \lambda_c E(r+1), \tag{9}$$

whence

$$g_+(\omega) = 2\lambda\left[1+\{M(\omega)+M(-\omega)\}\frac{E(r)}{E(r+1)}+\{M^2(\omega)+M^2(-\omega)\}\frac{E'(r-1)}{E(r+1)}+\ldots\right]. \tag{10}$$

For large ω, $M(\omega)$ and $M(-\omega)$ tend to zero (as $f(\alpha)$ is a density function), and $g_+(\omega)\to 2\lambda$. For small ω, $M(\omega)$ is continuous and tends to one. Hence, as $\omega\to 0$,

$$g_+(\omega)\to 2\lambda[1+E\{r(r+1)\}/E\{r+1\}]. \tag{11}$$

Notice that in the previous special case, $p(1) = 1$, $g_+(\omega)\to 4\lambda$, consistently with the further extreme case $f(\alpha) = \delta(\alpha-a)$, $\frac{1}{2}\{M(\omega)+M(-\omega)\} = \cos a\omega$. Also, from (11), the ratio $g_+(0)/g_+(\infty)$ is $1+E\{r(r+1)\}/E\{r+1\}$, providing some information on the average size of clusters for such a process, though not separate from the dispersion of cluster size; in fact, if m is the mean $E\{r+1\}$, and σ^2 the variance of r,

$$1+\frac{E\{r(r+1)\}}{E\{r+1\}} = m+\frac{\sigma^2}{m}. \tag{12}$$

More generally, the ratio $g_+(0)/g_+(\infty)$ may be shown to give the asymptotic ratio of the variance of $N(T)$ to its mean.

3. Statistical Technique

Analogously to the periodogram analysis of a stationary process $X(t)$, we define for a sample length T from a stationary point process $dN(t)$ the quantity

$$J(\omega) = \sqrt{\left(\frac{2}{T}\right)}\sum_{s=1}^{n} e^{iT_s\omega} = \sqrt{\left(\frac{2}{T}\right)}\int_0^T e^{it\omega}\,dN(t) \tag{13}$$

where T_s $(s = 1,\ldots,n)$ are the observed occurrence times for the process; and also define

$$I(\omega) = J(\omega)J^*(\omega). \tag{14}$$

Then it has been shown (Bartlett, 1963) that under suitable conditions $I(\omega)$ has similar sampling properties to those for an $I(\omega)$ derived from a process $X(t)$. In fact, in relation to the equivalence of $N(t)$ to a Poisson process with random occurrence rate $\Lambda(t)$, the simple asymptotic formula holds (for $\omega\neq 0$),

$$K(\theta_1,\theta_2)\sim K_\Lambda(\theta_1,\theta_2)+2\lambda\theta_1\theta_2, \tag{15}$$

where $K(\theta_1,\theta_2)$ is the joint cumulant-generating function of $J(\omega)$ and $J^*(\omega)$, and $K_\Lambda(\theta_1,\theta_2)$ similarly for $J_\Lambda(\omega)$ and $J^*_\Lambda(\omega)$, where

$$J_\Lambda(\omega) = \sqrt{\left(\frac{2}{T}\right)}\int_0^T e^{it\omega}\Lambda(t)\,dt. \tag{16}$$

Note that the extra component $2\lambda\theta_1\theta_2$ in (15) is quadratic in θ_1 and θ_2, implying the joint normality of $J(\omega)$ and $J^*(\omega)$ (for $J_\Lambda(\omega)$, $J^*_\Lambda(\omega)$ normal). We mean more completely by this, for the complex quantities

$$J(\omega) = A(\omega)+iB(\omega),\quad J^*(\omega) = A(\omega)-iB(\omega),$$

that $A(\omega)$ and $B(\omega)$ are (asymptotically) normally distributed with zero means and correlation, and equal variances λ.

The extended result for two frequencies ω, ω' $(\omega \neq \omega' \neq 0)$ is

$$K(\theta_1, \theta_2, \theta_1', \theta_2') \sim K_\Lambda(\theta_1, \theta_2, \theta_1', \theta_2') + 2\lambda\theta_1\theta_2 + 2\lambda\theta_1'\theta_2', \tag{17}$$

and similarly for any finite number of frequencies $\omega, \omega', \omega'', \ldots$. These results imply that the smoothing devices available for estimating the spectral density function of a stationary process $X(t)$ are also available for the stationary point process $dN(t)$, at least if an appropriate $\Lambda(t)$ exists. It is usual to assume $\Lambda(t)$ *linear* (in the sense used by Bartlett (1955), Section 9.2), though as the periodogram would not necessarily be standardized in relation to the total sum of squares in the present context, the assumption that $\Lambda(t)$ is *normal* would be preferable. As $\Lambda(t)$ should be always positive, the last assumption cannot be strictly true, but may be approximately so. Further discussion of the sampling theory of $J(\omega)$ is given in Sections 5 and 6.

There is a slight difficulty about the range of ω, as this is strictly infinite for continuous time. In the case of a *discrete* sequence X_r, the orthogonal transformation from X_r to $J(\omega_p)$ can be made exact, leading (with suitable definitions at the end of the frequency range) to the identity

$$\sum_{r=1}^{n} X_r^2 = \sum_{p=1}^{\frac{1}{2}n} I(\omega_p). \tag{18}$$

In the case of a continuous record, there would be an effective or actual rounding-off error in measurement of t, leading to a maximum value of n of order T/δ, where T is the length of record and δ the minimum interval for t. The corresponding maximum frequency would be π/δ. If this complete range were used, the identity (18) would still be available as a check, though if the values of $I(\omega_p)$ for large ω_p were small the frequencies would no doubt be cut off earlier. For a *point* process $dN(t)$ the X_r are replaced by nearly all zeros, with $N(T)$ ones, so that

$$\sum_{r=1}^{n} X_r^2 = N(T).$$

The identity (18) then merely says that the total number of occurrences $N = N(T)$ tends to be λT as T increases; the cut-off analogous to the $X(t)$ case would have to be judged by the drop of $I(\omega_p)$ to an average level of 2λ, but this is now less obvious because the constant theoretical component is also subject to the same sampling fluctuation as the variable theoretical component. In suitable problems further information might be available that would delimit the useful range of ω. For example, if a clustering model such as (10) were envisaged, with a typical clustering length of order t_c, then the corresponding frequency could be restricted to a maximum value $A\pi/t_c$, where A should be greater than one, but need not perhaps be greater than, say, 4 or 5.

There is also the problem which arises with any sampling theory of continuous time-series, and is particularly relevant for point processes, of the effective number of degrees of freedom. It seems reasonable that if only $N(T)$ events have occurred in the interval T, that in some sense only N degrees of freedom are available, though superficially the number, even for appropriate discrete spacing of ω, could appear much larger if the range of ω is taken large enough. This point is important in assessing the goodness of fit of hypothetical spectra, and is referred to again in Section 7. Unfortunately, it seems difficult to examine theoretically in any adequate manner; for if there are N events and the number of ω points taken is, say, of the order $2N$, the dependence between the spectral values might not be detectable until at least N values were considered *simultaneously*, and even then not perhaps very easily.

4. THE CLASS OF DOUBLY STOCHASTIC POISSON PROCESSES

In the sampling theory of point processes already referred to, it was assumed that the relation (6) held. Such processes, where $N(t)$ is a Poisson process for given $\Lambda(t)$, but $\Lambda(t)$ is itself a stochastic process, might be termed *doubly stochastic* (d.s.) *Poisson processes*. We have seen that anomalous point processes may be specified where events occur either together, or in clusters with *fixed* separation intervals; and these processes would not be classifiable as d.s. Poisson processes. An important question arises as to how far other point processes are so classifiable, for otherwise the sampling theory so far established might be inapplicable.

First of all, it may be remarked that the relation (6) can be used *formally* to try to define a stochastic process $\Lambda(t)$. For the relation (6) represents a set of identities between the density functions

$$f_s(t_1, t_2, \dots t_s) = \frac{E\{dN(t_1)\, dN(t_2) \dots dN(t_s)\}}{dt_1\, dt_2 \dots dt_s}$$

$(t_1 \neq t_2 \dots \neq t_s)$ of $N(t)$ and the respective moments $E\{\Lambda(t_1)\Lambda(t_2)\dots\Lambda(t_s)\}$ of $\Lambda(t)$, of which the first two are:

$$f_1(t_1) = E\{\Lambda(t_1)\}, \quad f_2(t_1, t_2) = E\{\Lambda(t_1)\,\Lambda(t_2)\}.$$

Such identities may be used to define, at least formally, the required process $\Lambda(t)$. The remaining question is whether the process $\Lambda(t)$ so defined is a valid and self-consistent one. Pending a more constructive answer, we may say that it is if the moments so defined constitute a valid and mutually consistent set.

To demonstrate that they do *not* always do so, consider a simplified point process of the clustering type, but with pairs of points ($r = 1$ certainly) separated by an interval with distribution $f(\alpha)\, d\alpha$. Then

$$g(\omega) = \lambda\{2 + M(\omega) + M(-\omega)\}. \tag{19}$$

Suppose further that

$$f(\alpha)\, d\alpha = \alpha\, e^{-\alpha/\gamma}\, d\alpha/\gamma^2, \tag{20}$$

so that

$$M(\omega) = (1 - i\gamma\omega)^{-2}. \tag{21}$$

Then

$$M(\omega) + M(-\omega) = \frac{2(1-\gamma^2\omega^2)}{(1+\gamma^2\omega^2)^2},$$

and

$$g_+(\omega) = 2\lambda\left\{1 + \frac{1-\gamma^2\omega^2}{(1+\gamma^2\omega^2)^2}\right\}. \tag{22}$$

While $g_+(\omega)$ is of course always positive, it will be noted that the second term, representing $g_\mu(\omega)$, does not always remain so (even although there is nothing particularly "anomalous" about the process above defined, i.e. no infinite covariance densities). This, in turn, implies that $E\{\Lambda(t_1)\,\Lambda(t_2)\}$ is not "positive definite", so that $\Lambda(t)$ is not a valid *real* process, let alone a valid *positive* process.

5. SAMPLING THEORY FOR CLUSTERING PROCESSES

The sampling problem arises, therefore, for such processes, as there is no guarantee that the asymptotic results will still apply. However, we may note that the asymptotic result (15) (and extensions of it such as (17)) are not dependent on the nature of $\Lambda(t)$, so that if the sampling properties of $\Lambda(t)$, even if $\Lambda(t)$ only exists in abstract terms, are appropriate, those of $J(\omega)$ or $I(\omega)$ will be also. As far as the dispersion of $I_\Lambda(\omega)$ is concerned, this is affected by the fourth-order cumulant properties, which can be related in the present problem to the properties of the fourth-order moment density properties of the point process $dN(t)$. (A complete distributional theory would be desirable, but it is perhaps relevant to recall that it does not exist even for continuous stationary processes $X(t)$ which fall outside some particular assumed class.)

To investigate the effect of fourth-order moment properties in the case of the clustering model defined in Section 2, we may consider

$$\begin{aligned}\kappa_{1111} = E\{\delta\Lambda(t_1)\,\delta\Lambda(t_2)\,\delta\Lambda(t_3)\,\delta\Lambda(t_4)\} \\ -E\{\delta\Lambda(t_1)\,\delta\Lambda(t_2)\}\,E\{\delta\Lambda(t_3)\,\delta\Lambda(t_4)\} \\ -E\{\delta\Lambda(t_1)\,\delta\Lambda(t_3)\}\,E\{\delta\Lambda(t_2)\,\delta\Lambda(t_4)\} \\ -E\{\delta\Lambda(t_1)\,\delta\Lambda(t_4)\}\,E\{\delta\Lambda(t_2)\,\delta\Lambda(t_3)\},\end{aligned}$$

where $\delta\Lambda(t) = \Lambda(t)-\lambda$. This from the equivalence required may also be written

$$\kappa_{1111} = \frac{E[\{dN(t_1)-\lambda dt_1\}\{dN(t_2)-\lambda dt_2\}\{dN(t_3)-\lambda dt_3\}\{dN(t_4)-\lambda dt_4\}]}{dt_1\,dt_2\,dt_3\,dt_4}$$
$$-\mu(t_1-t_2)\,\mu(t_3-t_4)-\mu(t_1-t_3)\,\mu(t_2-t_4)-\mu(t_1-t_4)\,\mu(t_2-t_3), \tag{23}$$

where $\mu(t-t')$ is the covariance density function for $dN(t)$ defined in equation (1). From (23) we see that contributions to κ_{1111} $(t_1 \neq t_2 \neq t_3 \neq t_4)$ can only arise for four-point dependence in $dN(t)$.

However, in the case of the class of Poisson clustering processes defined in Section 2 we can obtain a more precise result, even when clusters of size four or more may occur. Let us consider a sample of "parent" points at times τ_r $(r = 1, 2, \ldots, n)$ in the interval $(0, T)$. The quantity $J(\omega)$, including the contribution from "offspring", may be denoted by

$$J(\omega) = \sqrt{\left(\frac{2}{T}\right)} \sum_{r=1}^{n} e^{i\tau_r\omega}\,\phi_r(\omega), \tag{24}$$

where $\phi_r(\omega)$ is random but independent of τ_r. The distributional properties of $\phi_r(\omega)$ will be similar for each r, except for an end effect for r near n which can, for clusters or families of finite length, be neglected for large enough T. We consider $J(\omega)$ in (24) and see that it is the sum of n independent (complex) quantities Z_r with the same distribution and hence will tend to normality. Each quantity Z_r will (for ω not too small) have a mean effectively zero from the factor $e^{i\tau_r\omega}$ as $E\{e^{i\tau_r\omega}\} = \{\sin(\omega T)\}/(\omega T)$ and thus $E\{Z_r\} = O(T^{-\frac{1}{2}})$. Similarly, if we write $\phi_r = |\phi_r|\,e^{i\alpha_r}$, the phase angle α_r merely shifts the origin of τ_r, so that the square of Z_r will also have effectively zero mean, implying that it has real and imaginary components with equal variance. This indicates that $J(\omega)\,J^*(\omega)$ is asymptotically of the form

$$J_c(\omega)\,J^*(\omega)\,\overline{|\phi|^2},$$

where the bar denotes averaging over the $|\phi_r|^2$ in the sample, and $J_c(\omega)$ refers to the parent points only; or, for large n,

$$J(\omega)J^*(\omega) \sim J_c(\omega)J_c^*(\omega)E\{|\phi|^2\}. \tag{25}$$

The result (25) implies that the asymptotic sampling results previously established for d.s. Poisson processes apply also for Poisson clustering processes, with the further result on averaging (25) over the τ_r that

$$g_+(\omega) = 2\lambda_c E\{|\phi|^2\}. \tag{26}$$

This result (26) is more general than equation (8), as (apart from minor regularity restrictions on ϕ) any type of cluster may be assumed.

6. The Spectra of Renewal Processes

As an example suppose the cluster or "family" consists of a finite renewal process with n renewals (n fixed) after each parent point. Then it is easy to show that

$$\begin{aligned} E\{|\phi|^2\} &= n\left[1+\sum_{u=1}^{n}\left(1-\frac{u}{n}\right)\{M^u(\omega)+M^u(-\omega)\}\right] \\ &\sim n\left\{1+\frac{M(\omega)}{1-M(\omega)}+\frac{M(-\omega)}{1-M(-\omega)}\right\} \end{aligned} \tag{27}$$

for n large and $|M(\omega)|<1$, where $M(\omega)$ is defined as in Section 2. With $(1+n)\lambda_c = \lambda$, this gives, as n increases, and λ remains finite,

$$g_+(\omega) = 2\lambda\left\{1+\frac{M(\omega)}{1-M(\omega)}+\frac{M(-\omega)}{1-M(-\omega)}\right\}. \tag{28}$$

This is independent of n, and in fact provides the limiting spectrum for a single infinite renewal process with mean rate of occurrence λ. Some care is necessary with limiting operations, especially in this context (see also the remarks below), but an independent derivation of (28) is not difficult by means of the standard renewal formulae connecting renewal times and occurrence times (cf. Bartlett (1955), Section 6.12 for an example of the derivation of $\mu(\tau)$ for a particular renewal process). Thus, let $L(s)$ be the Laplace transform of the renewal distribution. Then the renewal density $r(t)$, say, at $t>0$ (given a renewal point at 0) is well known to have Laplace transform

$$R(s) = L(s)/\{1-L(s)\}.$$

From the formulae (1), (4) and (5), and the identifications

$$\begin{aligned} M(\omega) &= L(-i\omega), \\ \mu(\tau) &= \lambda r(\tau)-\lambda^2 \quad (\tau>0), \\ S(s) &= \lambda R(s)-\lambda^2/s, \end{aligned}$$

where $S(s)$ is the Laplace transform of $\mu(\tau)$, and

$$g(\omega) = \lambda+S(i\omega)+S(-i\omega),$$

result (28) follows.

To particularize (28) further for an important class of renewal process, choose

$$M(\omega) = (1-i\omega/k\lambda)^{-k} \tag{29}$$

corresponding to a gamma distribution for the renewal time. Then (28) becomes

$$g_+(\omega) = 2\lambda\left\{1+\frac{1}{(1-i\omega/k\lambda)^k-1}+\frac{1}{(1+i\omega/k)\lambda^k-1}\right\}. \tag{30}$$

For $k = 1$, this reduces of course to 2λ. For $k = 2$,

$$g_+(\omega) = 2\lambda-\frac{\lambda}{1+\omega^2/16\lambda^2}. \tag{31}$$

For large k, (30) becomes approximately

$$g_+(\omega) \sim 2\lambda\left\{1-\frac{1}{1+(e^{\omega^2/k\lambda^2}-1)/2\{1-\cos(\omega/\lambda)\}}\right\}, \tag{32}$$

which represents a cosine curve starting near zero and rising to a maximum at 2λ, with the successive minima gradually increasing to 2λ.

It should be noticed that as $k\to\infty$ the renewal process becomes one with occurrences at more precise periodic intervals, but it is inadvisable to put $k = \infty$, as there is an absence of long-term order in the renewal process for $k<\infty$ which would imply a discontinuity in behaviour with the case $k = \infty$. The case of strict periodicity is best treated separately, and for slightly greater generality we may suppose a random variation of time of occurrence independently for each event, but with periodic mean values at $a, 2a, 3a, \ldots$ (cf. Lewis (1961), who was, however, more interested in the distribution of intervals for such processes). Then, if we may neglect overlap "effects" for large T,

$$J(\omega) = \sqrt{\left(\frac{2}{T}\right)}(e^{i\tau\omega})\,(e^{i\omega\epsilon_1}+e^{i\omega(a+\epsilon_2)}+e^{i\omega(2a+\epsilon_3)}+\ldots+e^{i\omega([n-1]a+\epsilon_n)}), \tag{33}$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are the independent errors for the n occurrences, say, in the interval $(0, T)$, and τ is the initial random "phase angle" for the first occurrence in relation to the start of the second. In the special case of no errors $\{\epsilon_r\}$, notice that $J(\omega)J^*(\omega)$ contains no random components, in complete contrast with the sampling theory established in the case of doubly stochastic Poisson processes and clustering Poisson processes. Because of the random phase angle, $E\{J(\omega)\} = 0$ from (33), and

$$E\{J(\omega)\,J^*(\omega)\} = \frac{2n}{T}\{1-M(\omega)\,M(-\omega)\}+\frac{2}{T}M(\omega)\,M(-\omega)\frac{1-\cos(na\omega)}{1-\cos(a\omega)}, \tag{34}$$

where $M(\omega)$ is the characteristic function $E\{e^{i\omega\epsilon}\}$. Whether the errors $\{\epsilon_r\}$ are present or not, the second term has very large peaks at the frequencies $2\pi s/a$, of order $n^2/T = \lambda n$, so that as T and hence n increase the spectrum becomes discrete with ordinates at $2\pi s/a$. (The factor $M(\omega)\,M^*(\omega)$ provides, in the case of non-zero errors, a modulating factor on these ordinates plus a spectral density component.)

Returning for a moment to the results (30), (31) and (32), we see that renewal distributions between events in an infinite renewal process will usually depress the spectrum for low frequencies (unless the renewal distribution has a greater coefficient of variation than the exponential distribution), compared with clustering processes, which amplify it. For example, the second term in the expression (31) is an inverted Cauchy distribution, reducing the value of $g_+(\omega)$ by 50 per cent. at $\omega = 0$. We may note the possible complication of both effects together in one process, as the inhibiting effects between parents and clustering effects for families could tend to cancel out.

7. TWO EXAMPLES

The data analysed by this technique have already been referred to in my *Sankhyā* paper (1963), but a somewhat more detailed account is given here. Example 1 consists of times at which 129 successive vehicles passed a point on a road, with a total time

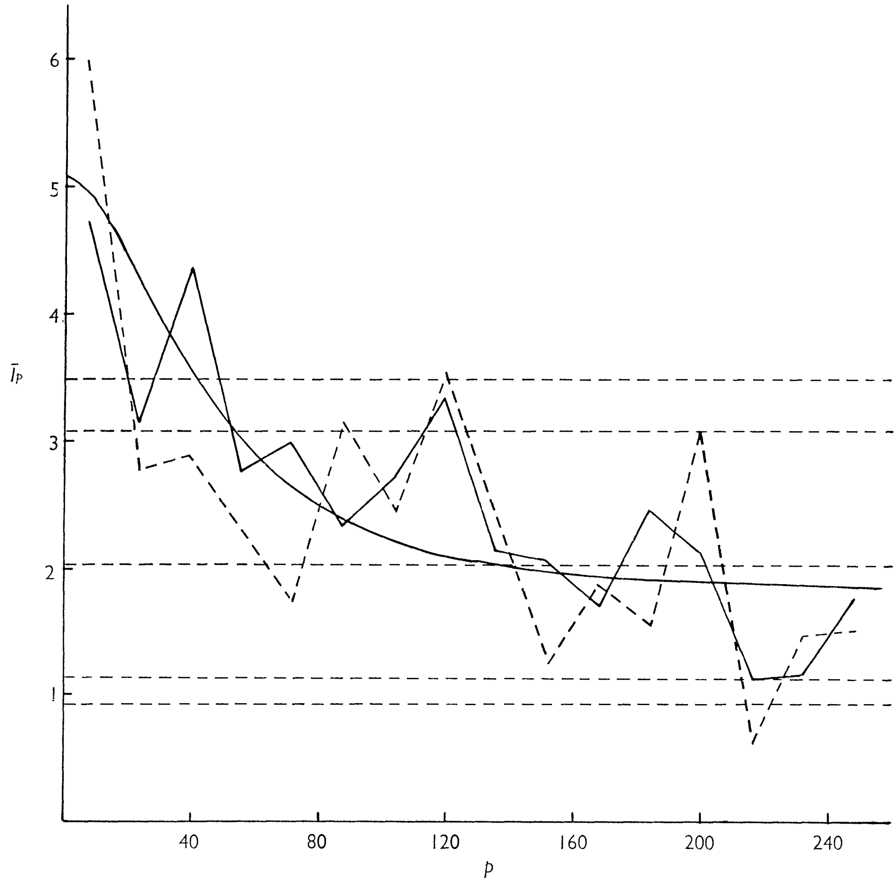

FIG. 1. Smoothed spectra for Example 1 (uniform weighting—continuous line; quadratic weighting—dotted line). The $P = 0{\cdot}05$ and $P = 0{\cdot}01$ significance bands (for individual values in the uniform weighting case) are shown, also the theoretical spectrum for a fitted clustering model.

interval (from the first vehicle) $T = 2023{\cdot}5$ seconds. The average interval of 15·81 seconds was reduced to the order of unity by taking 16 seconds as a unit.† The

† *Further note added at meeting*. In numerical calculations it is better to standardize by putting $T_1 = n$, $T'_s = nT_s/T$. This preserves the orthogonality relations, and the bias

$$2T\lambda^2 \frac{\sin^2(\frac{1}{2}T\omega_p)}{(\frac{1}{2}T\omega_p)^2}$$

is strictly zero. In the examples as calculated, the bias is negligible for Example 1; but for Example 2 is of order 25 per cent. for small ω, owing to the value of T being by chance rather different from its expected value.

values of ω taken were $\omega_p = 2\pi p/n$ with $n = 128$, $p = 1, 2, \ldots, 256$, and values of A_p, B_p, I_p tabulated by an electronic computer, where

$$J(\omega_p) = A_p + iB_p, \quad I_p = A_p^2 + B_p^2.$$

Example 2 consisted of 128 independent and exponentially distributed successive time intervals, with mean unity. The values of p and n were as in Example 1. (The original data for Example 1 have been used as a class example at University College since 1960, and are given in this guise in the Appendix. The data for Example 2 are the first 128 values from Clark and Holz (1960).)

The complete sets of I_p values are not reproduced, but smoothed spectral values are shown in Figs. 1 and 2 for 16-point averaging, both with uniform weighting and

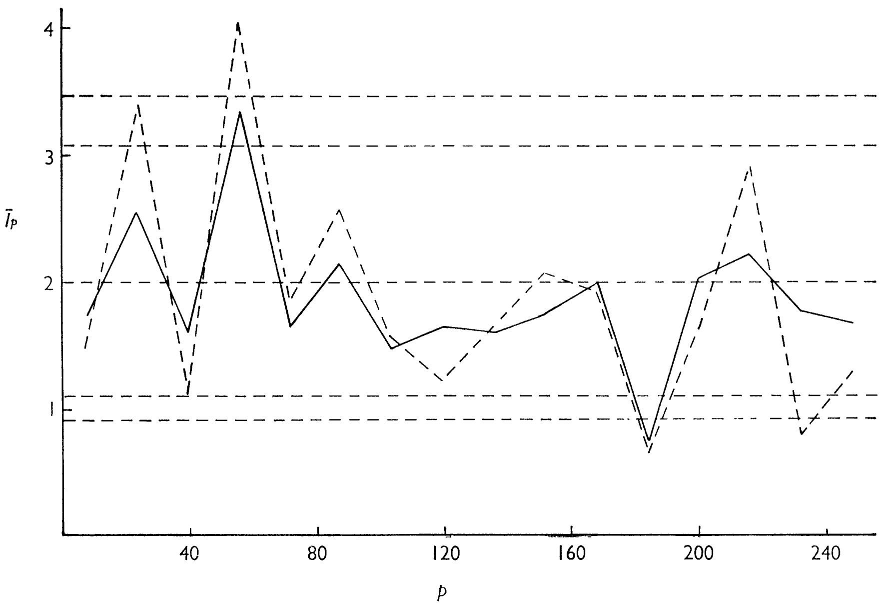

FIG. 2. Smoothed spectra for Example 2 (uniform weighting—continuous line; quadratic weighting—dotted line). The $P = 0{\cdot}05$ and $P = 0{\cdot}01$ significance bands (for individual values in the uniform weighting case) are shown.

quadratic weighting (for the numerical values corresponding to the figures and an explanation of quadratic weighting, see Bartlett (1963)). The $P = 0{\cdot}05$ and $0{\cdot}01$ significance levels are also shown in the figures. These are based on fluctuations from 2 in Example 2, and $2N(T)/T = 2{\cdot}024$ in Example 1, proportional to χ^2 quantities with 32 degrees of freedom, and are appropriate for the uniform weighting estimates. The significance levels do not allow for the possible search for maximum departures from a uniform spectrum, but it is recalled that with $N(T) = 128$, the order of magnitude of the number of independent $\{I_p\}$ may be limited to 64, i.e. to the first four χ^2 quantities. With these χ^2 quantities mainly in mind if clustering is suspected, the significance level need only be dropped by a factor of about four; and this leaves the high values of I_p for small p still obviously significant in Example 1, in contrast with the values for Example 2.

Subsequent discussion of the analysis may conveniently be divided into two parts, the first associated with the rough fit of a clustering model to Example 1, the second with the evidence provided by the numerical analysis of the validity of the asymptotic sampling theory assumed. It is evident from formula (12), even without further fitting, that the mean size of cluster cannot exceed about $2\frac{1}{2}$–3, though of course this mean includes clusters of one. It is also, of course, obvious that a good graduation with a clustering model could not show more than compatibility of the model with the data, and for this reason no exhaustive search for the best fit, or a more suitable model, was made. The model (8) or (10) was used, with a modified geometric distribution for cluster size,

$$p(r) = \begin{cases} 1-c \quad (r=0), \\ c\alpha^{r-1}(1-\alpha) \quad (r=1,2,\ldots), \end{cases}$$

and $M(\omega)$ given by (29) with $k=2$. We then have

$$g_\mu(\omega) = \frac{\lambda_c c}{1-\alpha}\sum_{s=1}^{\infty}\left\{\frac{\alpha^{s-1}}{(1-\frac{1}{2}i\omega/\mu)^{2s}}+\frac{\alpha^{s-1}}{(1+\frac{1}{2}i\omega/\mu)^{2s}}\right\},$$

$$= \frac{2\lambda_c c}{1-\alpha}\left\{\frac{(1-\frac{1}{4}\omega^2/\mu^2-\alpha)}{(1+\frac{1}{4}\omega^2/\mu^2)^2-2\alpha(1-\frac{1}{4}\omega^2/\mu^2)+\alpha^2}\right\}.$$

With†

$$c=\tfrac{1}{9}, \quad \alpha=\tfrac{2}{3}, \quad \lambda_c=\tfrac{3}{4}\lambda,$$

and

$$g_\mu(\omega) = \frac{\frac{1}{2}\lambda(\frac{1}{3}-\frac{1}{4}\omega^2/\mu^2)}{\frac{1}{9}+\frac{5}{6}\omega^2/\mu^2+\frac{1}{16}\omega^4/\mu^4},$$

we obtain

$$g_+(\omega) = 2\lambda\left\{\frac{\frac{5}{18}+17\omega^2/24\mu^2+\omega^4/16\mu^4}{\frac{1}{9}+5\omega^2/6\mu^2+\omega^4/16\mu^4}\right\}.$$

The scale for ω is reasonably fitted if $\mu = 15\pi/8$. With $2\lambda = 2{\cdot}024$, this gives

$$g_+(\omega) = 2{\cdot}024\left\{\frac{\frac{5}{2}+34\phi^2/3+16\phi^4/9}{1+40\phi^2/3+16\phi^4/9}\right\}, \tag{35}$$

where $\phi = 0{\cdot}4\omega/\pi$.

The plot of (35) is shown in Fig. 1. As a check on the adequacy of the fit, the values of I_p for this example were rescaled by dividing by $g_+(\omega)$. The cumulative sums

$$S_p = \sum_{q=1}^{p} I_q$$

are shown in Figs. 3, 4 and 5 for the original I_p of Example 1, for the rescaled values, and for Example 2, respectively. By joining the origin to the final sum S_{256} in each case, and drawing the $P=0{\cdot}05$ and $0{\cdot}01$ boundaries (for the theory of this test see Bartlett, 1955, Section 9.21), we may check whether the spectrum is or has been made uniform. The width of the boundary depends on the effective number of degrees of freedom for the $\{I_p\}$, and this is, on the more cautious assessment, nearer 64 than the

† I am indebted to Mr P. A. W. Lewis for pointing out a slip in my value of c, now corrected.

nominal number 256 (cf. the discussion in Section 3 of this paper) used to calculate the boundaries. However, it is clear that the rescaled I_p for Example 1 are satisfactory, and while the significance of the unscaled I_p in this example would now become more

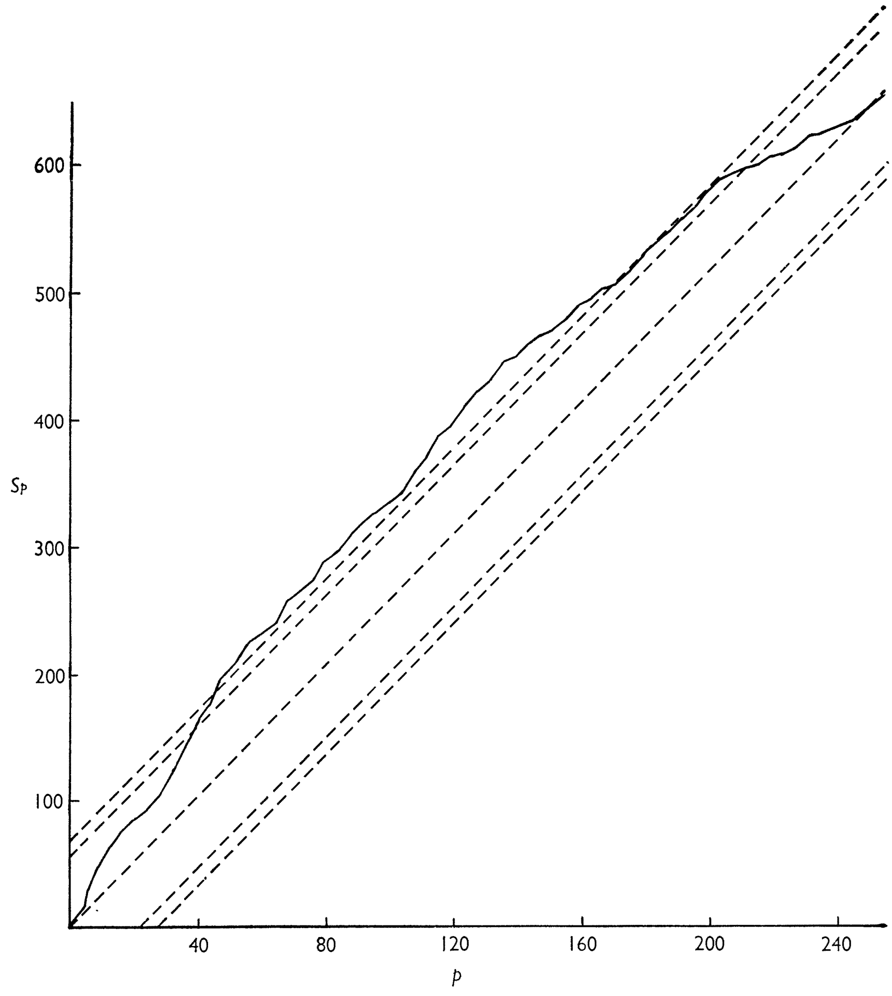

FIG. 3. Cumulative spectral function for Example 1, with $P = 0{\cdot}05$ and $P = 0{\cdot}01$ significance bands (based on 256 degrees of freedom).

borderline by this test (the $P = 0{\cdot}05$ distance of 55·6 and $P = 0{\cdot}01$ distance of 66·6 would be doubled and no longer exceeded), it should be recalled that the inadequacy of a uniform spectrum has already been established by a previous test.

With regard to the adequacy of the sampling theory, this was partially checked in the case of the Poisson process by the construction of a bivariate frequency table for I_p and I_{p+1}, for which the marginal distribution should be exponential, and the correlation zero. In the case of Example 1, this can only be tested on the scaled values.

The results are summarized in Tables 1 and 2 and are seen to be satisfactory. It is, as noted in Section 3, difficult to check the suspected multiple dependence of the I_p, but for Example 2 the mean values taken 16 at a time were tested for homogeneity on

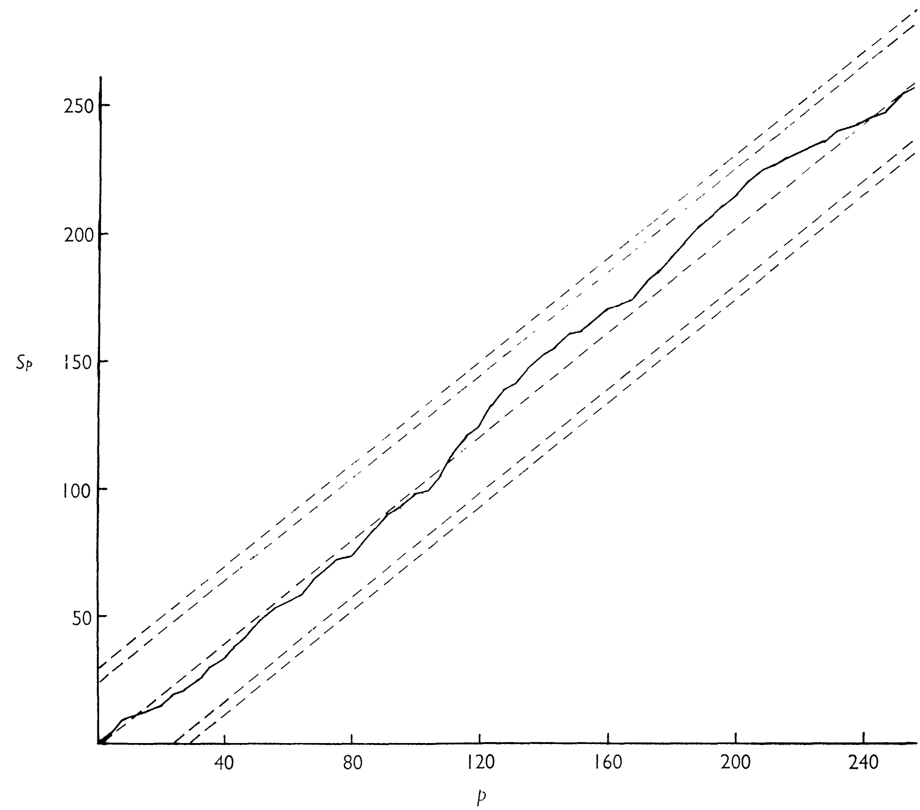

FIG. 4. Cumulative spectral function for Example 1 after scaling by clustering model spectrum, with $P = 0{\cdot}05$ and $P = 0{\cdot}01$ significance bands (based on 256 degrees of freedom).

the null hypothesis that they were proportional to χ^2's with 32 degrees of freedom. While this hypothesis was reasonably used for the first four such quantities in Example 1, it seemed possible that for all 16 values an indication of some dependence might be obtained, though it is by no means clear whether a too large or too small probability level of significance would be expected. In fact we obtain, by the approximate "homogeneity of variances" test, χ^2 (corrected) $= 22{\cdot}04$ (15 d.f.), which does not reach the $P = 0{\cdot}10$ significance level and provides no evidence of dependence between the I_p items. An examination of the distribution of mean values was more convenient on means of 8 (proportional to χ^2's with 16 d.f., the percentile points for which could be read off from standard tables), and is summarized in Table 3. Again there seems no evidence of failure to conform. It is, of course, still doubtful whether we should expect any indication of a departure from the asymptotic theory for the tests used above. The previous comment (Section 3) about degrees of freedom is not obviously the most relevant, as if the events are regarded as scores 0 and 1 on an approximating discrete time basis, the 0's would legitimately add to the degrees of freedom, at least provided their proportion was not excessive; on this basis the nominal degrees of freedom would be the more appropriate. Test criteria with correct

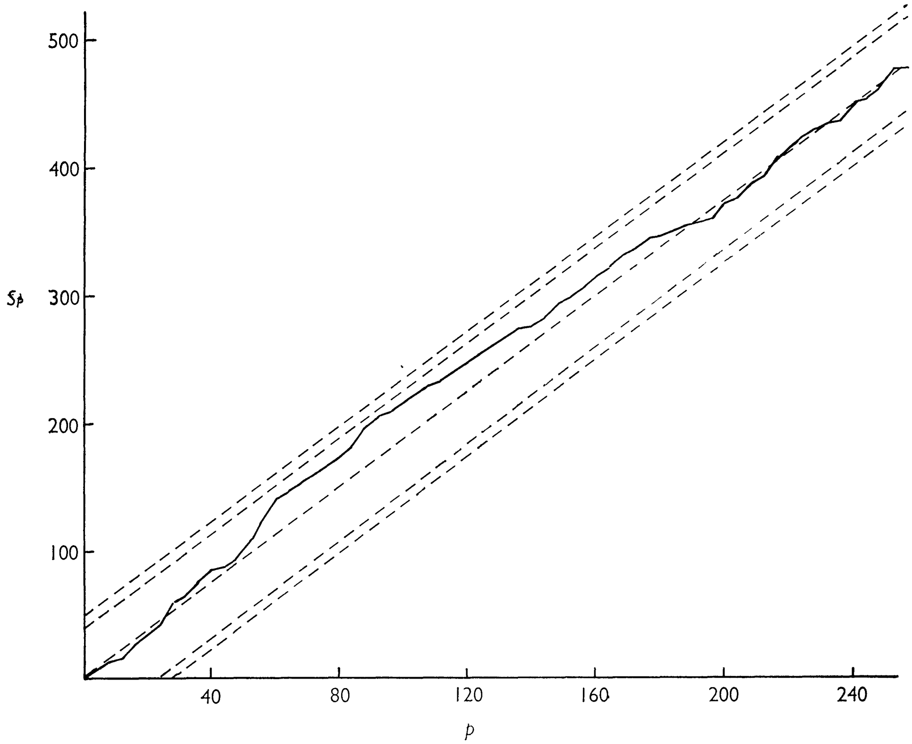

FIG. 5. Cumulative spectral function for Example 2 with $P=0{\cdot}05$ and $P=0{\cdot}01$ significance bands (based on 256 degrees of freedom).

asymptotic variances will be the most "robust"; and on this basis the nominal degrees of freedom might reasonably be allocated to the test criteria used here, involving comparisons and ratios of sample periodogram intensities (cf. Bartlett, 1955, Section 9.21). A more precise conclusion on this point is, however, among the various questions still to be answered and, in the meantime, some degree of caution is advisable. It would, I think, be premature at this stage, without more extensive experience with this technique, to attempt any final assessment of its value compared with possible alternatives. There is always the temptation to subject comparatively simple sets of data to unnecessarily cumbersome analysis. Nevertheless, the technique developed has one important advantage over some of the alternative methods of analysis—that is, the availability of the *same* asymptotic sampling theory for a wide range of non-null hypotheses as for the null case of the Poisson process. There is, of course, the computational problem, but, as with ordinary spectral analysis, this is not so alarming when there is access to an electronic computer.

TABLE 1

Example 1. *Two-way frequency table to* I_p *and* I_{p+1} $(I_{257} = I_1)$ *after scaling by clustering model spectrum*

I_{p+1} \ I_p	0–½–	½–1–	1–1½–	1½–2–	2–2½–	2½–3–	3–3½–	3½–4–	4–	*Total*
0–½–	36	24	12	8	9	3	1	1	1	95
½–1–	22	13	9	6	4	5	2	2	—	63
1–1½–	14	10	1	5	—	1	2	—	—	33
1½–2–	9	2	3	3	4	1	2	—	—	24
2–2½–	7	7	3	1	—	—	—	—	—	18
2½–3–	2	3	3	1	—	—	—	—	1	10
3–3½–	4	2	2	—	—	—	—	—	—	8
3½–4–	1	2	—	—	—	—	—	—	—	3
4½–	—	—	—	—	1	—	1	—	—	2
Total observed	95	63	33	24	18	10	8	3	2	256
(*Expected*	97·2	60·3	37·4	23·2	14·4	8·9	5·6	3·4	5·6	256)

Mean 1·047. *Standard deviation* 0·917. *Correlation coefficient* −0·003.

TABLE 2

Example 2. *Two-way frequency table to* I_p *and* I_{p+1} $(I_{257} = I_1)$

I_{p+1} \ I_p	0–1–	1–2–	2–3–	3–4–	4–5–	5–6–	6–7–	7–8–	8–9–	9–10–	10–	*Total*
0–1–	48	20	18	10	12	5	—	1	1	—	—	115
1–2–	31	7	6	1	4	2	—	—	1	—	1	53
2–3–	13	11	4	3	—	4	—	—	—	1	—	36
3–4–	9	4	3	2	—	—	—	—	—	—	—	18
4–5–	9	5	1	2	—	—	—	—	—	—	—	17
5–6–	3	2	4	—	1	—	—	—	—	1	—	11
6–7–	—	—	—	—	—	—	—	—	—	—	—	0
7–8–	—	1	—	—	—	—	—	—	—	—	—	1
8–9–	—	2	—	—	—	—	—	—	—	—	—	2
9–10–	1	1	—	—	—	—	—	—	—	—	—	2
Total observed	115	53	36	18	17	11	0	1	2	2	1	256
(*Expected*	105·6	62·1	36·5	21·3	12·6	7·4	4·3	2·6	1·5	0·8	1·3	256)

Mean 1·879. *Standard deviation* 1·813. *Correlation coefficient* −0·070.

TABLE 3

Distribution of means of eight I_p (16 d.f.) *compared with the appropriately scaled* χ^2 *distribution*

Boundary values:	0·995	1·151	1·394	1·578	1·918	2·303	2·559	2·943	3·288	
Observed	2	2	2	5	10	6	1	2	0	2
Expected	1·6	1·6	3·2	3·2	6·4	3·2	3·2	3·2	1·6	1·6

ACKNOWLEDGEMENTS

As noted in my *Sankhyā* paper, I am indebted to Dr A. J. Miller for supplying the traffic data for Example 1, and to Mr D. Walley for the programming and further arrangements for the spectral analysis of both examples (also of another interesting set of traffic data from Dr Miller that Dr Miller will refer to in the discussion). I am also indebted to Miss Patricia Jackson and Miss Katharine Solomon for some of the further calculations for Example 1.

REFERENCES

BARTLETT, M. S. (1951), "Periodogram analysis and continuous spectra", *Biometrika*, **37**, 1–16.

—— (1955), *An Introduction to Stochastic Processes.* Cambridge University Press.

—— (1963), "Statistical estimation of density functions", *Sankhyā* (to be published).

CLARK, C. E. and HOLZ, B. W. (1960), *Exponentially Distributed Random Numbers.* Baltimore: Johns Hopkins.

COX, D. R. (1955), "Some statistical methods connected with series of events", *J. R. statist. Soc.* B, **17**, 129–164.

EPSTEIN, B. (1960), "Tests for the validity of the assumptions that the underlying distribution of life is exponential", *Technometrics*, **2**, Part I, 83–101; Part II, 167–183.

GREEN, H. S. (1952), *Molecular Theory of Fluids.* Amsterdam: North-Holland.

GREENWOOD, M. (1946), "The statistical study of infectious diseases", *J. R. statist. Soc.* A, **109**, 85–109.

GRENANDER, U. and ROSENBLATT, M. (1957), *Statistical Analysis of Stationary Time Series.* New York: Wiley.

LEWIS, T. (1961), "The intervals between regular events displaced in time by independent random deviations of large dispersion", *J. R. statist. Soc.* B, **23**, 476–483.

MCFADDEN, J. A. (1962), "On the length of intervals in a stationary point process", *J. R. statist. Soc.* B, **24**, 364–382.

THOMPSON, H. R. (1955), "Spatial point processes, with applications to ecology", *Biometrika*, **42**, 102–115.

APPENDIX

(Example used for lecture course on stochastic processes, Statistics Department, University College, London)

Times to $\frac{1}{10}$th sec. at which consecutive vehicles passed a point

6067	6095	6129	6143	6288	6307	6335	6358
6511	6529	6624	6649	6743	6754	7640	7656
7675	7690	8027	8053	8182	8344	8363	8566
8934	9335	10040	10060	10140	10161	10193	10210
10775	11012	11036	11250	11301	11380	11581	11730
11786	12303	13174	13186	13213	13223	13238	13251
13498	14224	15422	15434	15503	15542	15558	15588
15606	16054	16104	16143	17396	17624	17643	17802
17862	18068	18197	18236	18366	18435	18460	18583
18640	18753	18778	18794	18870	18893	18954	18975
19322	19476	19522	20079	20101	20161	20179	20198
20216	20636	20729	21646	21670	21976	21988	22076
22142	22640	23221	23240	23269	23274	23286	23596
23715	23723	23735	23743	23790	23873	23946	24034
24052	24083	24091	24432	24462	24488	24525	24938
25235	25411	25430	25568	25970	26071	26190	26300
26302							

Analyse the above data with a view to examining:

(i) whether the times of passing constitute a Poisson process;

(ii) if not, whether some form of "bunching" or "clustering" seems to be present.

Possible analyses include:

(a) testing the homogeneity of the consecutive random time-intervals, by means of a partitioning of the degrees of freedom for the total (approximate) χ^2;

(b) testing the homogeneity of counts in consecutive fixed time-intervals, choosing an appropriate interval, and partitioning the degrees of freedom corresponding to the total dispersion by means of an analysis of variance;

(c) testing the correlation between the consecutive random time-intervals;

(d) examining the overall distribution of counts in fixed time-intervals;

(e) examining the overall distribution of the consecutive random time-intervals.

You should undertake at least sufficient of these to answer the questions asked.

Discussion on Professor Bartlett's Paper

Professor P. Whittle: That I should be proposing the vote of thanks to Professor Bartlett is an event whose probability of recurrence is not large, and so, at the risk of seeming irrelevant, I should like to express the great personal pleasure it gives me to do so. When I first entered statistics some fifteen years ago, and soon encountered Professor Bartlett's name, I would have given almost no credence at all to the proposition that I should one day be speaking under these circumstances; even less, that I should be doing so as successor in the Manchester chair which Professor Bartlett endowed with his own distinction.

On the other side of the world, fifteen years ago, "statistics" meant largely "English statistics"; this not being entirely due to blind loyalty, as people here are apt to believe, but to a perfectly practical and proper discernment of the situation as it was then. However, little discernment was required to recognize Professor Bartlett's status; from a 12,000 mile perspective he emerged as one of the necessarily few people who had something real to say, which he said, on a wide variety of topics, in a mathematics which went to the limit in economy and pregnancy, and in a language which sometimes went rather beyond.

The point processes with which Professor Bartlett is dealing, viewed as stationary processes, are of a type which is extreme rather than anomalous. If to each event one associates a pulse of positive duration h and amplitude $1/h$, then the resulting process is quite regular, e.g. the autocovariance is finite and continuous at the origin, and the classic spectral representations of autocovariance and variate are both valid. As $h \downarrow 0$, however, the spectral representations fail unless given special interpretation. Nevertheless, it is clear from the paper that a Fourier analysis of the sample is still useful.

For a conventional process it is known that the attractive sampling properties of the periodogram ordinate $|J|^2$ in the normal case do not hold asymptotically in general, although there are certain types ("linear" processes) for which they do. A principal result of this paper would seem to be that these sampling properties of $|J|^2$ hold also for certain classes of point process, e.g. the clustering process. In general, to demonstrate that k periodogram ordinates tend to independent exponential variates one will have to demonstrate that the corresponding $2k$ sample Fourier transforms tend to joint normality; this demands a certain amount of "independence" of the process.

There are a number of places in the paper where I would have welcomed a more detailed proof or enunciation. For example, does equation (15) imply a stronger result than that J, J^*, J_Λ and J_Λ^* are all asymptotically normally distributed, with the relations between distributions implicit in the equation?

Relation (6) can be readily extended to the case of a "doubly stochastic clustering process". Suppose that one defines a "cluster characteristic functional" by the conditional expectation

$$\xi(t) = E\left\{\exp\left(i\int_t^T \theta(s)\,dM(s)\right)\,\middle|\,\text{cluster begins at } t\right\},$$

where $M(s)$ is the number by time s in the cluster which was initiated at time t. Then, if the probability intensity of cluster-initiation at time t is $\Lambda(t)$, one finds that

$$E\left\{\exp\left(i\int_0^T \theta(t)\,dN(t)\right)\right\} = E_\Lambda\left\{\exp\left(\int_0^T \Lambda(t)\{\xi(t)-1\}\,dt\right)\right\}.$$

Note that there is no need to restrict oneself to single occurrences; indeed, the increment $dN(t)$, when it is not zero, need not even be positive or integral, let alone unity. One might characterize a rather general form of point process by demanding simply that $dN(t)$ be zero with probability $1-O(dt)$.

A last question is that of the stochastic model: Professor Bartlett considers largely models in which independent effects are superimposed. It would, of course, be of great interest if one could treat models in which there was true interaction, such as the "inhibition" mentioned. Such interaction will probably be even commoner in spatial than in temporal models; effects such as competition, attraction and repulsion will often be present. A rather novel source of point processes is a model containing a piecewise linear response term. Consider, for example, the model

$$L\left(\frac{d}{dt}\right)x_t = u(x_t)$$

where $L(d/dt)$ is a linear differential operator, and

$$u(x) = \begin{cases} 1 & (x \geqslant 0), \\ 0 & (x < 0). \end{cases}$$

Then an attempt at Fourier analysis of x_t leads to the formal equation

$$\left|\int x_t\, e^{i\omega t}\,dt\right|^2 = \frac{1}{|L(i\omega)|^2}\left|\sum_r \frac{e^{i\omega t_{2r}} - e^{i\omega t_{2r-1}}}{i\omega}\right|^2,$$

where the t_{2r} and the t_{2r-1} are the instants at which the x-axis is crossed from above and below respectively. Thus, one is led to a Fourier analysis of the point processes $\{t_{2r}\}$, $\{t_{2r-1}\}$. There is no statistical element in this particular model, but this could be supplied.

I have great pleasure in moving that the speaker be accorded a vote of thanks.

Professor D. R. Cox: It is a particular pleasure to congratulate Professor Bartlett on his paper. He has made a most important contribution to statistical technique, illustrated by cogent and interesting numerical examples.

First, I want to comment briefly on what Professor Bartlett calls the doubly stochastic Poisson process. This was first introduced in connection with assemblies of textile fibres in which the point events are the leading ends of fibres. It is found empirically that such processes always show more dispersion than the Poisson process, and the doubly stochastic Poisson process is a very convenient way of representing this. However, it is easily shown, and is obvious from analysis of variance ideas, that one can only represent in this way over-dispersion relative to the Poisson process. This is the reason why Professor Bartlett's example in (19)–(22) is not covered.

Now properties of point processes can be expressed equivalently in terms of numbers of events, or in terms of intervals between events. Suppose, however, that we restrict ourselves to second-degree properties, such as the mutually equivalent functions

(a) the variance (as a function of "time");
(b) the autocovariance;
(c) the spectrum.

Then the second-degree properties of intervals are, of course, not equivalent to the second-degree properties of numbers of events. That is, although there are elegant general formulae that relate the correlation function of intervals to the distributional properties of numbers, and the covariance density of numbers to the distributional properties of intervals, the two correlation functions are mathematically independent. The correlation properties of intervals have been ignored in the paper and I wonder how much information is lost by this. It will rarely be possible to write down the full likelihood for a model of a point process. Therefore I suggest that a searching analysis of a series of events, assumed stationary, requires, in general, simultaneous consideration at least of some second-degree properties of numbers, of the distribution of intervals between successive events, and of some second-degree properties of intervals. In particular, if the main analysis is based on the spectrum of numbers, it seems desirable to examine also perhaps the first serial correlation coefficient for intervals between successive events and the frequency distribution of the intervals. Of course, formal internal tests of significance based on such combinations of statistics may be difficult to justify, but this seems a relatively minor matter at the present stage of the subject.

For example, consider the traffic data. A model alternative to the clustering renewal process is to have two distributions with p.d.f.'s $f_1(x)$, $f_2(x)$. Intervals between successive events are generated by sampling these p.d.f.'s in accordance with a two-state Markov chain. This is a simple form of semi-Markov process and the spectrum can be calculated by a simple modification of renewal theory arguments. In particular, if the Markov chain degenerates to a random sequence, the process becomes a simple renewal process. We interpret $f_1(x)$ as referring to intervals in "free" traffic, and assume it to be an exponential distribution with large mean. We interpret $f_2(x)$ as referring to "bunched" traffic and assume it to be a gamma distribution with a small mean. I hope that later in the discussion Mr P. A. W. Lewis will describe some comparisons he has made with Professor Bartlett's analysis. Even in the renewal case, it does not seem possible to distinguish the models merely from the spectrum of numbers.

I would like to comment briefly on the possible advantages of the spectrum over the equivalent functions (a) and (b) mentioned above, the variance–time function and the autocovariance. In standard time series theory, there seem three main advantages to the spectrum:

(i) the spectral decomposition of the random function itself may have a physical meaning;
(ii) the result of a linear operation is neatly represented by a transfer function;
(iii) the sampling theory of the spectrum is easier than that for the other functions.

I have been unable to see an analogue of (i) for point processes, except possibly for the doubly stochastic Poisson process. As for (ii), one interpretation of a linear operation on a point process is that of translating the points independently by random amounts. An analogue of the transfer function can then be defined. Professor Bartlett's result (34) for the unpunctuality process is a special case. Finally, it is clear from Professor Bartlett's work that the elegant sampling properties of the spectrum carry over and this fact seems to be the main advantage of the spectrum over the autocovariance for the general analysis of point processes. The uses of the variance–time function, which is a double integral of the autocovariance, are somewhat more special.

I am very glad indeed to second the vote of thanks to Professor Bartlett.

The vote of thanks was put to the meeting and carried unanimously.

Dr D. J. Bartholomew: The hypothesis of clustering arises in many fields, and a new method for detecting its presence and determining its nature is to be welcomed. I was present recently at a discussion on the analysis of the statistics of strikes. It was suggested that the occurrence of one strike triggers off other strikes. The "Poisson clustering process" may very well provide a suitable model for the study of this phenomenon. Unlike the application to traffic data, it is quite possible for the clusters to overlap if strikes in several industries are considered simultaneously. Similar problems arise in the investigation of absences and accidents.

Professor Cox has already pointed out that the spectral density and the variance–time curve are mathematically equivalent. The following remarks may help to clarify the choice between them in practice. Cox (1955) showed that $I(T)$ ($= \operatorname{var} N(T)/EN(T)$) may be expressed in the form

$$I(T) = 1+\frac{2}{\lambda}\int_0^T (1-\tau/T)\,\mu(\tau)\,d\tau.$$

The relationship between this method of analysis and the use of the spectral density is brought out by writing $\omega = T^{-1}$ and $G(\omega) = 2\lambda I(1/\omega)$, which gives

$$G(\omega) = 2\lambda+4\int_0^\infty \langle 1-\omega\tau\rangle \mu(\tau)\,d\tau. \tag{1}$$

This has to be compared with the spectral density

$$g_+(\omega) = 2\lambda+4\int_0^\infty \cos(\omega\tau)\,\mu(\tau)\,d\tau. \tag{2}$$

The equivalence of the two methods when $\omega = 0$ is implied by Bartlett's remark at the end of Section 2. We now have to decide whether the linear weighting function of (1) or the harmonic function of (2) is more appropriate. In the case of the simple clustering model mentioned at the beginning of Section 2, where the clusters consist of equally spaced pairs, the spectral method has an obvious advantage. The oscillation of the spectral density is more distinctive than the *J*-shaped curve given by (1). In general we should expect the same conclusion to hold whenever there is any appreciable degree of regularity in the process. On the other hand, if we consider infinite renewal processes or relatively complicated clustering processes, like the one discussed by Bartlett, the advantages of spectral analysis are less obvious. The ease with which $I(T)$ can be estimated may then weigh heavily in its favour.

Both methods suffer from the same practical drawback. The amount of data available for analysis is restricted by the length of time for which the process can be assumed stationary. Unless a very long series is available it appears that we can obtain no more than a general idea of the shape of the curves given by either (1) or (2). The possibility of obtaining precise information on the three probability distributions in Bartlett's model is therefore remote. A fundamental objection to complex models in these circumstances is that simpler models may describe the data equally well. For example, I suspect that an infinite renewal process could be found to provide a good fit to the data of Example 1 if the renewal distribution was more skew than the exponential. This is not a criticism of either method; it is just one of the facts of life.

Mr A. G. Hawkes: Many tests have been proposed to detect departure from a purely Poisson process although little is known about their power against various alternatives. One of these, proposed by Professor Durbin (*Biometrika*, **48** (1961), 41), is a modification of the Kolmogorov random walk test. The Kolmogorov test can be applied directly to the problem of vehicles passing a point on a road by plotting T_n against n, where T_n is the time of arrival of the nth vehicle. If this is done for the data given in this paper the function oscillates about the central line and fails to be significant at the 5 per cent. level. Durbin shows that by reordering the inter-arrival times into ascending order and then using another transformation we get a random walk which can be tested by a one-sided Kolmogorov test. This is more powerful than the original test against alternatives which exhibit a greater degree of bunching than the Poisson process. When applied to the data considered here it is now significant at the 0·1 per cent. level, a very great increase in power. It is possible that the Kolmogorov tests used in Professor Bartlett's paper could be similarly modified, though the rationale for so doing is not obvious in this context.

Mr J. F. C. KINGMAN: I would like to make two comments, the first concerning doubly stochastic Poisson processes. Professor Bartlett has raised the problem of deciding when a given point process can be represented as a doubly stochastic Poisson process, and it may be of interest to give the answer to this problem in the particular case of a renewal process. Thus the following result can be proved. An equilibrium renewal process with "lifetime" distribution function F is a doubly stochastic Poisson process if and only if there exists a strictly positive λ and a non-decreasing function K such that, for all $\theta > 0$,

$$\int_0^\infty e^{-\theta x}\, dF(x) = \lambda\left\{\lambda+\theta+\int_0^\infty (1-e^{-\theta x})\, dK(x)\right\}^{-1}.$$

The corresponding process $\Lambda(t)$ has the property that it takes only the values λ and 0. If $0 < k = K(\infty) - K(0) < \infty$, then the time axis can be broken up into intervals whose lengths are independent random variables and on which $\Lambda(t)$ is constant, the distribution of the lengths of those intervals on which $\Lambda(t) = \lambda$ being exponential with mean $1/k$, and that of those on which $\Lambda(t) = 0$ having distribution function K/k. A more precise description, valid for both finite and infinite k, can be given in terms of the theory of regenerative events (for which see *Bull. Amer. Math. Soc.*, **69** (1963), 268–272).

My second point concerns more generally the specification of point processes. One way of describing the statistical structure of a point process Π would be in terms of the counts of the process in successive intervals of equal length. This, however, is an incomplete specification, due essentially to the phenomenon of "aliasing". One might expect to avoid this by considering the counts in intervals of unequal length, and this could be done by taking intervals of random length. Thus let Π' be a renewal process independent of Π and having lifetime distribution function G, and let N_n be the number of jumps of Π in the nth interval of Π'. Then each N_n is a randomized statistic of Π. In the particular case when Π' is a Poisson process, the joint distributions of the N_n can often be calculated, and these form a specification of Π which is convenient in some contexts. In particular, it arises very naturally in the theory of queues. Moreover, it can be shown that a knowledge of the joint distributions of the N_n suffices to determine those of Π. This is not true for a general choice of G, and in particular it is not true when G is degenerate. A necessary and sufficient condition of G for the joint distributions of the N_n to determine those of Π is that the characteristic function $\phi(\theta)$ of G should take no value more than once as θ ranges over all real values.

Dr A. J. MILLER: I am glad to have the opportunity of taking part in the discussion of this paper. As mentioned in the paper, I have performed spectral analyses upon several more sets of traffic data. All but one of these other sets of data was from the same source as that used by Professor Bartlett. That is, they were recorded by the Swedish State Roads Institute on the same two-lane rural road in Sweden.

The remaining set of data was collected in Birmingham at a point about 300 ft. downstream from traffic signals operating a fixed cycle of 83 seconds. The smoothed periodogram (Fig. 1) shows, as would be hoped, a very good peak at this cycle time. The weighting function used in smoothing the periodogram is shown in Fig. 2. This weighting was achieved by taking three equally weighted moving averages. The interesting feature of this periodogram is that it levels off to its ultimate value (one) very much more quickly than the periodograms of the rural data. To see why this should be, I have examined the product densities, or at least closely related functions, for these data and the set used in the paper. For the data used in the paper it seems reasonable to assume that the first-order density $\lambda(t)$ was a constant. The conditional density $\lambda(t+\tau|t)$ was then estimated by finding the proportion of times that one vehicle was followed by another one (possibly with other vehicles between) after a time lag between $(\tau-\frac{1}{2})$ and $(\tau+\frac{1}{2})$ seconds. If the events (the passing of vehicles in this case) had been independent, then we would find

$$\lambda(t+\tau \mid t) = \lambda(t+\tau).$$

The ratio $\phi(\tau)$ of estimates of conditional and unconditional densities is shown plotted in Fig. 3 for the Swedish data.

For the Birmingham data, the estimated first-order density is shown in Fig. 4. The ratio of conditional to unconditional densities was found to be only weakly dependent upon the

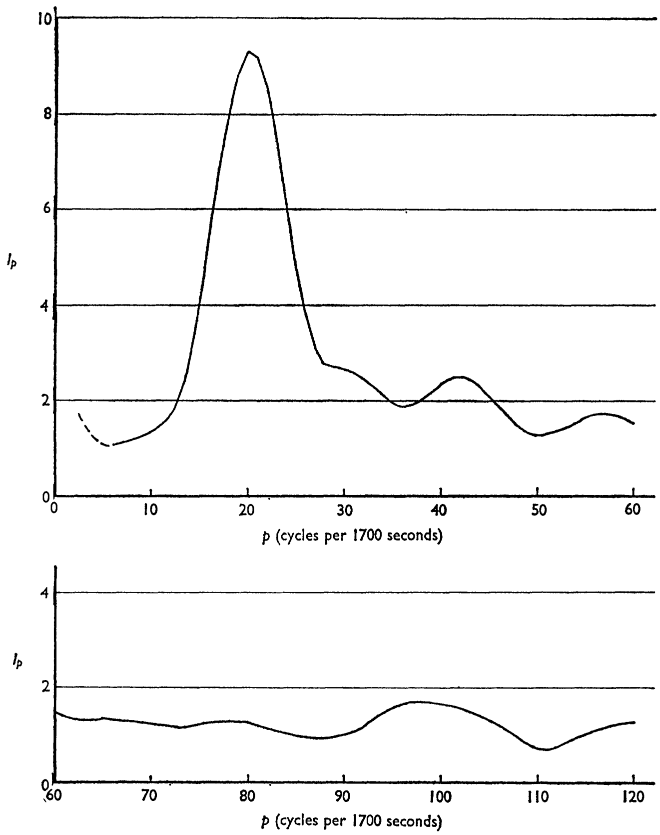

FIG. 1. Smoothed periodogram of traffic data collected downstream from traffic signals with fixed cycle of 83 seconds.

amount of the cycle which had elapsed, and has been averaged over t. $\phi(\tau)$ for these data is shown in Fig. 5. For lags greater than 2 seconds, $\phi(\tau)$ is virtually one, and its highest value is only 1·4. This means that the flow of traffic downstream from signals could be reasonably well simulated by generating vehicles independently of each other, except for a minimum headway, and using a varying generation rate.

I would like to comment briefly on the two distribution models for headways (gaps) between vehicles suggested by Professor Cox. Such models have been proposed by a number of people. In particular I would mention the book *Poisson and Traffic* by Gerlough and Schuhl, and the paper "Vehicle headway distributions" by Buckley which was read at the First Biennial Conference of the Australian Roads Board in Canberra, 1962. If the two distributions considered are the exponential and gamma then the only difference between Professor Cox's model and Professor Bartlett's appears to be that in Professor Bartlett's model the length of a bunch or cluster is subtracted from the next inter-bunch headway and there is a small probability of bunches overlapping. In both models, the distribution of bunch sizes is geometric with the first probability modified. In practice, the modified geometric distribution fits observed distributions of bunch sizes only moderately well.

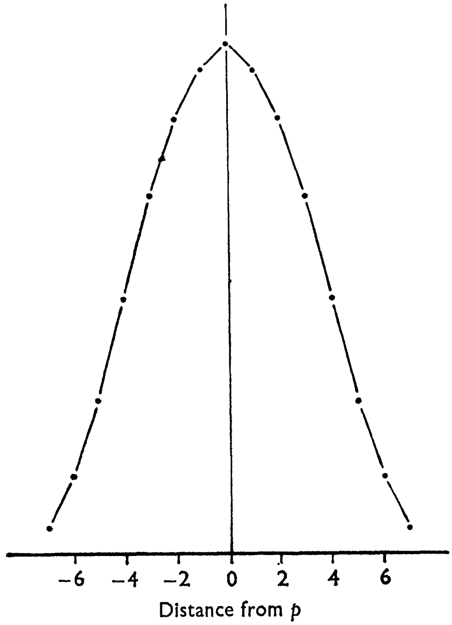

FIG. 2. Weighting function for Fig. 1.

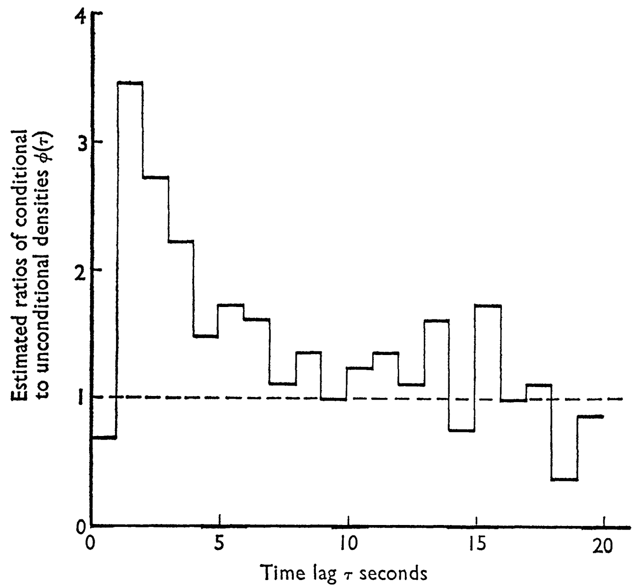

FIG. 3. Ratios of conditional to unconditional densities against time lag for the data used in Professor Bartlett's paper.

Professor Cox has spoken of a choice of two alternative methods of handling point processes, namely by means of counts of events and by means of intervals between events. Product densities provide the bridge between the two methods. In his book on stochastic

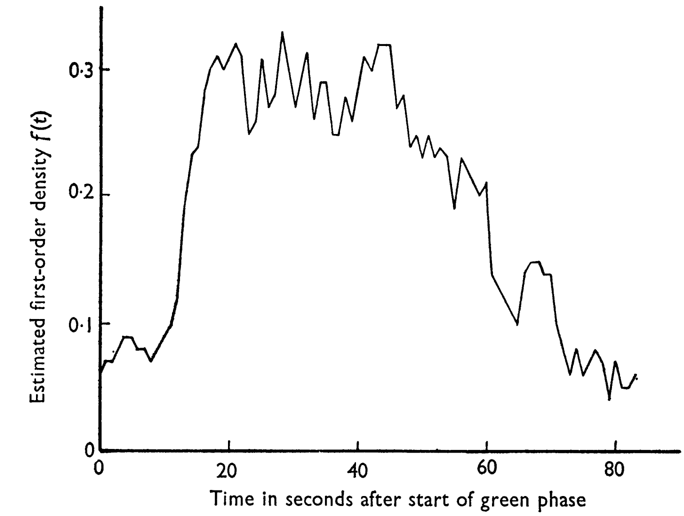

FIG. 4.

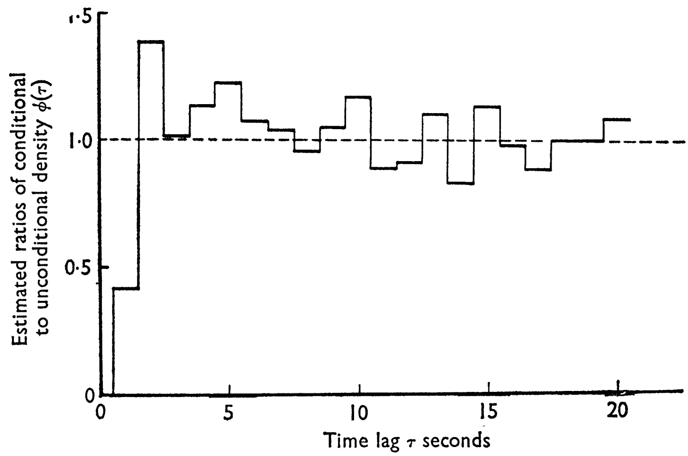

FIG. 5. Flow of vehicles 300 ft. downstream from signals.

processes, Professor Bartlett has shown how product densities may be obtained from the distribution of (independent) intervals, and how the moments of count distributions can be obtained from product densities.

Dr M. B. PRIESTLEY: Professor Bartlett has illustrated the relationship between the spectral analysis of stationary point processes and that of real stationary processes by considering the case of a doubly stochastic Poisson process. I should like to mention an alternative approach to this problem which does not require any specific assumptions regarding the structure of the stationary point process.

Consider a linear transformation (or "filter") of $dN(t)$, of the form,

$$Y(t) = \int_{-\infty}^{\infty} w(t-\tau)\, dN(\tau)$$

$$\equiv \sum_{s=-\infty}^{\infty} w(t-T_s),$$

where the weight function $w(t)$satisfies

$$w(t) = 0, \quad t<0, \quad \int_{-\infty}^{\infty} w(t)\, dt = 1, \quad \int_{-\infty}^{\infty} w^2(t)\, dt < \infty.$$

($Y(t)$ may be regarded as a "smoothed form" of the improper process $dN(t)/dt$.) It is easy to show that, if $dN(t)$ is a second-order stationary point process, $Y(t)$ is a second-order stationary real process, with an autocovariance function $\mu_Y(\tau)$, which is finite for all τ and continuous at $\tau = 0$, and an absolutely continuous spectral density function $f_Y(\omega)$. Further, it may be shown that

$$f_Y(\omega) = |\, W(\omega)\,|^2 f_c(\omega),$$

where

$$W(\omega) = \int_0^{\infty} w(t)\, e^{-i\omega t}\, dt,$$

and also, for large T,

$$J_Y(\omega) \sim V(\omega)\, J(\omega),$$

so that

$$I_Y(\omega) \sim |\, V(\omega)\,|^2\, I(\omega), \tag{*}$$

where

$$J_Y(\omega) = \sum_{t=1}^{N} Y_t\, e^{i\omega t},$$

$$I_Y(\omega) = J_Y(\omega)\, J_Y^*(\omega),$$

and

$$V(\omega) = \sum_{t=0}^{\infty} w(t)\, e^{-i\omega t}.$$

Thus, asymptotically, $I(\omega)$ has similar sampling properties to the periodogram of a real process, and this result would appear to be independent of any assumptions regarding the structure of $dN(t)$. Possible uses of this type of "filter" are as follows.

(a) By a suitable choice of $W(\omega)$, we may introduce a "cut-off" in the spectrum, thereby removing the difficulty of working with an unlimited frequency range. (As the difference between the spectrum of a doubly stochastic Poisson process and that of a clustering process is concentrated in the low frequency end, the use of a "low-pass" filter would not obscure such differences.)

(b) It provides an alternative method of estimating the spectral density function (cf. the "filter method" for real processes), with $|\, V(\omega)\,|^2$ corresponding to the periodogram weight function.

If the bandwidth of $W(\omega)$ is chosen sufficiently large, then we may preserve most of the spectral properties of $dN(t)$. In practice, given a sequence of events $T_1, T_2, \ldots, T_N$, then if we choose the bandwidth of $W(\omega)$ to be less than

$$\min_r\, (T_r - T_{r-1}),$$

the process $Y(t)$ would appear to contain all the second-order moment properties of $dN(t)$.

The problem of choosing the weight function to be used for smoothing the periodogram ordinates is common to all branches of spectral analysis, and one which, so far, has not been completely solved. The quadratic weight function referred to by Professor Bartlett may be regarded as an approximation to the optimum form (on the mean-square error criterion), but only with the restricted class of exponentially damped weight functions (except for the very rare case when the spectral density function is exactly a quadratic function of ω, so that the above restriction may be removed). There are, however, many weight functions which belong to the class of algebraically damped functions, but which nevertheless are very useful in spectral analysis. Consequently, one would be reluctant to ascribe "optimum" properties to the quadratic weight function.

Dr M. B. Priestley added in writing:

Professor Bartlett has pointed out that the sampling properties of a periodogram of a real process are known only in the case of *linear* processes, and consequently the usefulness of equation (*) is limited. Now $Y(t)$ may be represented as a linear process provided

$$\int_{-\infty}^{\infty} \frac{|\log f(\omega)|}{1+\omega^2} d\omega < \infty.$$

This condition, in turn, implies restrictions on the choice of weight function $w(t)$, and on the spectrum $f_c(\omega)$ of the point process. However, even allowing for these restrictions, I feel that the class of point processes for which equation (*) is useful is much wider than the restricted class of doubly stochastic Poisson processes for which the process $\Lambda(t)$ *itself* has to be assumed linear.

Dr G. M. Jenkins: Professor Bartlett has referred to an extension of his spectral technique to multidimensional processes. However, an area where there is an immediate application is to multivariate point processes. Some years ago I was asked to analyse data consisting of records of faults in two output channels from a certain electronic device. It was suspected that the faults in the two channels were not independent. At the time, fairly *ad hoc* methods of analysis suggested themselves, but after reading Professor Bartlett's paper the method of approach seems obvious. Denote the point processes by $dN_i(t)$ $(i = 1, 2)$, and

$$J_i(\omega) = \sum_{s=1}^{n_i} \exp(i\omega T_s^{(i)}), \tag{1}$$

where the $T_s^{(i)}$ are the times of occurrence of faults in the ith channel. The cross periodogram may then be defined by

$$I_{12}(\omega) = J_1(\omega) J_2^*(\omega), \tag{2}$$

where a star denotes a complex conjugate, and this may be split up into in-phase and out-of-phase periodograms in the form

$$I_{12}(\omega) = C_{12}(\omega) - iQ_{12}(\omega). \tag{3}$$

These may now be smoothed to give spectral estimates $c_{12}(\omega)$, $q_{12}(\omega)$, from which may be obtained the cross-amplitude spectrum

$$R_{12}(\omega) = \sqrt{\{c_{12}^2(\omega) + q_{12}^2(\omega)\}}, \tag{4}$$

and the phase spectrum

$$\phi(\omega) = \tan^{-1}\left\{\frac{q_{12}(\omega)}{c_{12}(\omega)}\right\}. \tag{5}$$

If faults in one channel tend to occur near faults in the other, then the cross-periodogram will have maximum power at zero frequency. If there are gaps between the faults in the two channels, then the power is a maximum at higher frequencies, whereas if the faults occur mutually at random, the cross-periodogram would be flat. The phase spectrum will yield useful information concerning delays, for example, if the second process lagged the first process by a fixed time T, the phase spectrum would have a marked peak at frequency $\omega = 2\pi/T$. It should, in theory, be possible to set up stochastic differential equations describing such interactions and to interpret these in the light of $R_{12}(\omega)$ and $\phi(\omega)$.

In addition to applications where the two point processes occur on an equal footing, the method could be applied to situations where they are the input and output to some system which could be approximated linearly by

$$dN_2(t) = \int_0^\infty w(u)\, dN_1(t-u) + n(t), \tag{6}$$

where $n(t)$ is an error. Possible applications of this technique would be:

(a) to investigate the effect of a modification such as a roundabout or road widening on traffic flow by making simultaneous observations on both sides of the modification;

(b) to investigate the service pattern of a queue when this is not regular, e.g. the near approach to random service in airport control procedure when the traffic is exceptionally heavy.

The extra information to the phase required to characterize the transfer function of the system (6) would then be given by the gain spectrum

$$G(\omega) = R_{12}(\omega)/g_{11}(\omega), \tag{7}$$

where $g_{11}(\omega)$ is the auto-spectrum of the input. It may be that the non-linearity of such systems limits the usefulness of such information but this would be a matter for trial.

Recently I was asked to analyse two records, one of which was a point process, and the other a continuous time-series $X(t)$. The above approach could be applied directly in this situation to investigate the effect of the point process on the continuous record.

Professor Bartlett's paper opens up many avenues, but I would like to support Professor Cox's contention that this approach should be coupled with a similar analysis on the intervals. In this context, it would probably be more useful to work with the logarithms of the intervals since these would have a distribution much closer to normal and hence yield a more informative spectrum.

Mr T. Lewis: As it is so late, I shall confine myself to two brief comments. The first relates to equation (33), where $J(\omega)$ is given for a sample length with n events from a periodic process disturbed by independent random time-deviations. Professor Bartlett takes the times of occurrence of the events in this sample to be $(r-1)\,a+\epsilon_r$ with $r = 1, 2, \ldots, n$; but in fact if the dispersion of the ϵ's is at all substantial the values of r for the n events, $r_1, r_2, \ldots, r_n, = \mathbf{r}$, say, will in general fail to coincide with $1, 2, \ldots, n$ near the beginning and end of the sample. (For example, in a typical artificial sample with $n = 128$, $\sigma_\epsilon = 3{\cdot}5a$, the three points $r = 5$, 125 and 127 fell outside the sample range and were replaced by $r = 0$, 129 and 139.) On evaluating $E\{J(\omega)\,J^*(\omega)\}$ conditional on $\mathbf{r}$, the first term $2n(1-MM^*)/T$ on the right-hand side of (34) remains unchanged; the second term is modified, but still becomes of order n^2/T at the frequencies $2\pi s/a$, as is evident by inspecting

$$\sum_{r=r_1}^{r_n} \exp(i\omega ra) \sum_{r=r_1}^{r_n} \exp(-i\omega ra).$$

While this confirms that the general shape of $I(\omega)$ is essentially unmodified by the end effects, it also raises a problem, because it suggests that the spectrum for the type of process above described is equally characteristic of a rather more general process in which the spacing of the mean times of occurrence is not uniform. There is perhaps something to be looked into here from the point of view of sampling theory.

Taking up now a point made by Dr Bartholomew, it certainly does seem, at any rate in the above case of disturbed periodic processes, that the spectrum may not necessarily be the best process characteristic to use if one is concerned with estimating process parameters. With this type of process there are advantages in working in terms of counts of events in successive equal intervals, since it turns out that the autocorrelation function of these counts is essentially the same as the convolution probability density of the difference of two independent ϵ's; the count correlogram can thus be used in a direct way for estimating the parameters of the ϵ-distribution. One has here a slightly curious set-up—the autocorrelation function of counts, which is the Fourier transform of the corresponding (count) spectral density, is also a probability density. The general problem of estimating a probability density function, and its relation to the estimation of a spectral density, have been discussed by Professor Bartlett in other writings of his. In the particular problem I have been mentioning I think (for reasons which I hope to give fully elsewhere) that there are advantages in using a Mellin-type transform, and estimating parameters of the ϵ-distribution in terms of

$$f(\alpha) = \sum_s r_s/s^{\alpha},$$

where the r_s are the observed serial correlations of counts.

Mr A. M. Walker: I should like to make a brief comment in the form of a question to Professor Bartlett. Can he say whether it would be possible to obtain results similar to those given in his paper for processes of a more general type, in which a random variable is associated with each "event", for example the process $\{Y(t)\}$ defined by

$$Y(t) = \int_0^t X(u)\,dN(u), \tag{1}$$

where $\{X(t)\}$ is a stationary process which is independent of $\{N(t)\}$? (A point process could be regarded as a particular case of (1) with $X(t) \equiv 1$.) One might also wish to consider a process where the random variables associated with the different "events" were identically distributed, which might be thought of as a limiting form of (1) obtained by letting the autocorrelation function of $\{X(t)\}$ decrease more and more rapidly as its argument increases; when the events form a Poisson process, this gives us a familiar type of additive process which has arisen in various applications, for example in storage theory (compare P. A. P. Moran, 1959, *Theory of Storage*, p. 68; London: Methuen).

The following written contribution was read by the Honorary Secretary.

Mr P. A. W. Lewis: I have quite recently derived Professor Bartlett's Poisson clustering process as a model for the failure patterns of complex systems such as electronic computers. A simple model for these failure patterns can be derived by assuming that the sequence of failures in each component position in the system constitutes a renewal process. The failure pattern of the computer—the superposition of these renewal processes—should then form a Poisson process, but this has been found to be not true. The reason seems to be that the repair of a computer failure is not always successful, the failure recurring at times $Y_1, Y_1+Y_2, \ldots, Y_1+\ldots Y_r$ after the initial occurrence. The delay between successive recurrences of the failure is due to the fact that only a small proportion of the components in a computer are in use and needed at any given time, or in any given period of time, for the correct operation of the system. In addition many component failures are intermittent rather than permanent, as is assumed in the renewal model.

I think it is clear that the resultant failure model is the same as Professor Bartlett's clustering model. I have examined in some detail the case where the Y_i's are independent and identically distributed and r is a modified geometric distribution, and in what follows I will refer only to this case. The survivor function of the intervals in the stationary process (one minus the distribution function) is given by

$$R(t) = \frac{(1-\alpha+cR_Y(t))}{(1-\alpha+c)}\exp\left\{-\lambda_c t - \frac{c\lambda_c}{(1-\alpha)}\int_0^t R_Y(u)\,du\right\},$$

where $R_Y(t)$ is the survivor function for Y. The variance of the number of events in a period of length t is

$$V(t) = \lambda_c t + \frac{\lambda_c c(3-\alpha)}{(1-\alpha)^2}t - \frac{2\lambda_c cE(Y)}{(1-\alpha)^3} + n(t),$$

the first two terms corresponding to equation (12) in the paper, and $n(t)$ being the transient term which is proportional to the double integral of the autocovariance.

I have used these and other results to check the compatibility of the model with the data on the basis of the properties of the intervals, and the variance–time curve, rather than on the basis of the spectrum, as Professor Bartlett has done. With the parameter values $\lambda = 1{\cdot}012$, $\alpha = \frac{2}{3}$ and $c = \frac{1}{9}$, the maximum deviation of $R(t)$ from the empirical survivor function is 0·08, and the predicted slope of the variance–time curve is 2·5 compared with an estimate of 1·8. The predicted and estimated coefficients of variation are 1·25 and 1·49 respectively, and most importantly the predicted correlation coefficient of lag one for the intervals is $-0{\cdot}173$ compared with an estimate of $+0{\cdot}087$. Now when Y has a gamma distribution with index k greater than one, values of the parameters consistent with clustering always give relatively large and negative correlation coefficients, and it does not seem possible to adjust the parameters to obtain a reasonable consistency with the estimated survivor function and other properties of the process which were referred to previously.

At Professor Cox's suggestion I have investigated the "fit" to the data of the semi-Markov model he has referred to. The Laplace transform of the "renewal density" for the stationary semi-Markov process is

$$h^*(s) = \frac{\pi_1 f_1^*(s) + \pi_2 f_2^*(s) + (1-\alpha_1-\alpha_2) f_1^*(s) f_2^*(s)}{1-\alpha_1 f_1(s) - \alpha_2 f_2(s) - f_1(s) f_2(s)(1-\alpha_1-\alpha_2)},$$

where α_1 is the probability of transition from state 1 to state 1, α_2 is the probability of transition from state 2 to state 2, and

$$\pi_1 = 1-\pi_2 = (1-\alpha_2)/(2-\alpha_1-\alpha_2).$$

The spectrum is then

$$g_+(\omega) = 2\lambda\{1 + h^*(i\omega) + h^*(-i\omega)\}.$$

Now for the special case of the clustering process we are considering, equation (10) in the paper may be written in closed form as

$$g_+(\omega) = 2\lambda\left[1 + \frac{cM(i\omega)}{(1-\alpha+c)\{1-\alpha M(i\omega)\}} + \frac{cM(-i\omega)}{(1-\alpha+c)\{1-\alpha M(-i\omega)\}}\right].$$

It can then be shown that when $f_2(x)$ is a gamma density with parameter k, and Y has the same type of density, the spectra for the two processes have the same functional form, namely the ratio of two polynomials in ω, each of order $2k$, with coefficients of odd powers of ω being zero. It is therefore possible when $k = 2$ to choose the parameters in the semi-Markov process to give a spectrum approximately the same as the spectrum for the clustering process.

It is apparent, however, from the rather odd shape of the logarithm of the empirical survivor function, that by taking $f_2(x)$ to be a gamma density with index $k = 3$, one can get a very good fit to the empirical survivor function. Estimating the four parameters $\alpha_1, \alpha_2, \mu_1, \mu_2$ from the first three sample moments and the sample correlation coefficient gives $\mu_1 = 27{\cdot}15$ sec, $\mu_2 = 2{\cdot}91$ sec, $\alpha_1 = 0{\cdot}652$, $\alpha_2 = 0{\cdot}607$, and a maximum deviation between fitted and empirical survivor functions of 0·035. The slope of the variance–time curve is 2·573, and the spectrum

$$g_+(\omega) = 2{\cdot}024\frac{0{\cdot}000630\varphi^6 + 0{\cdot}834\varphi^4 + 4{\cdot}721\varphi^2 + 2{\cdot}573}{0{\cdot}000630\varphi^6 + 0{\cdot}833\varphi^4 + 8{\cdot}126\varphi^2 + 1}.$$

This curve is almost identical with the curve given in Fig. 1 of the paper up to $p = 128$, drops to 1·652 at $p = 160$ and has a minimum of 1·494 at $p = 256$. Judging by eye, it would seem to be a good fit to the estimated spectrum.

It is, of course, not possible to make a choice between the two models on the basis of these results, although the semi-Markov model seems to be more internally consistent with the data than the other model. However, the analysis does, I think, underscore Professor Cox's point that an analysis of the data should be based on all the properties and characteristics of the process.

With regard to the clustering process, I think there is a special case for estimating the variance–time curve, since its slope is independent of the distribution of the random variable Y. Estimates of the parameters λ_c, α and c can be obtained from this, the estimated mean of the intervals, and the slope of the tail of the logarithm of the estimated survivor function. One is then in a much better position to use Professor Bartlett's spectral analysis. In particular, in analysing computer failure data, it is difficult to make reasonable assumptions about the distribution of the random variable Y. The spectral analysis presented in this paper may be very useful in resolving this difficulty.

Professor G. A. BARNARD: Professor Cox has mentioned an exact test of significance for the bunching of the sort we have to deal with here, and while he is, of course, quite right to say that such a thing is relatively unimportant in this case, I think it may be worth while to point out that in a situation such as this, where the hypothesis tested is simple, or may be made so by suitable conditioning (in this case conditioning on the total number of cars observed in the time interval in question), provided one has access to a reasonable amount of time on a reasonably powerful computer, an exact test of significance is something one never need be without. To illustrate the procedure in this case, let us suppose that we are prepared to work at the 1 in 20 level of significance. Then we get the computer to construct for us nineteen sets of points uniformly distributed in the unit interval, and it will be convenient to plot them, if possible, on a graph paper in cumulative form, in the same way as the original data relating to cars are plotted —the cumulative number of cars being plotted against the cumulative length of the interval. We have then in principle to determine an appropriate measure of "bunchiness" (for example, the mean length of the five or six longest intervals between vehicles could serve). Whatever the choice of test criterion, it may appear that the value of this criterion is largest for the observed set, the nineteen artificially generated sets giving a smaller value for the criterion. It will then be *exactly* true that the probability of such occurrence, on the null hypothesis of no bunching in the experimental data, will be 1 in 20. Further details will be published elsewhere, but it may be noted that one advantage of this method is that once the 19 sets of data are available a whole set of test criteria may be used and it may be verified that any one of the set may give the same answer—or perhaps it may not be so verified, in which case the issue as to whether there is or is not significant bunching may be seen to depend on just what criterion of "bunchiness" is used.

Professor BARTLETT thanked all the speakers in the discussion for their contributions, and in reply made some brief comments which he amplified in writing as follows:

It is evident from the discussion that the paper has raised a number of problems that merit attention, including various "loose ends" that need tidying up. Many of these problems will no doubt be considered further in due course—by others, I hope, as well as by myself. Here I shall deal mainly with points that suggest an immediate reply.

In answer to Professor Whittle's query, equation (15) does not imply quite what he says, as it does not say that $J_\Lambda(\omega)$ is asymptotically normal, though we know this is so for $\Lambda(t)$ a linear process. My wording after equation (16) was rather careless in the proof, as my reference at this stage to normality mainly had in mind the independent component additional to $J_\Lambda(\omega)$ implied by the additive character of (15), and the normality of $J(\omega)$ only follows *if* $J_\Lambda(\omega)$ is normal. Equation (15) is of course to be taken with equation (17), etc. Professor Whittle's extension of relation (6) appears to me very useful. It may assist in unifying the conditions for which the asymptotic sampling theory implied by (15), (17), etc. holds, if employed in place of the separate approach I used for clustering processes. Dr Priestley's alternative approach may also be helpful with this problem; but the snag I mentioned in my verbal reply—that the sampling properties of continuous (I do not like Dr Priestley's term *real*) stationary processes are known only in certain cases, for example, for linear processes—does not seem to be met by his further remarks. What is meant by a linear process in this context is one where the "impulses" generating the process are independent and not merely uncorrelated. In this connection I agree with Professor Whittle that sampling theory, even for stochastic processes, is tied more closely than we would like to independence in one form or another.

Professor Cox is right to stress that the spectral analysis of point processes merely looks at one aspect, and there may be other aspects to consider in any situation. Dr Miller points out that the density functions defined for point processes specify all their stochastic properties; but a correlation analysis of intervals involves three points and cannot therefore be covered by a spectral analysis of points, involving, as this does, only pairs of points at a time, as indeed Professor Cox indicates. Nevertheless, just as for continuous processes second-order moments are used to describe at least partially the departure from simple independence, so for point processes. For the latter, the second-order "product density" is closely related to the conditional probability density, which is often of direct interest, for example, in physical problems. In cases where *a priori* knowledge of the theoretical model is available, this should of course be used in deciding what analysis to use. It would also determine any particular relation between interval properties and product densities: thus, for renewal processes the interval distribution may be inferred from the second-order probability density and conversely. One feature about spectral analysis, that rather attracted me, is its immediate extension in principle to more than one dimension; whereas intervals between successive points in one dimension represent an ordering property with no such extension.

I think Dr Miller's analyses very interesting, especially his direct estimates of the conditional probability density. My own concentration on the frequency domain was linked with the simpler sampling theory, and I wonder whether Dr Miller has investigated the sampling errors for his estimates. The spectral analysis he shows in his first diagram is the one I anticipated in my acknowledgements. The earlier history of this analysis is rather amusing, as when the periodogram values were first shown to me by Mr Walley, I was momentarily appalled at the huge peak, having forgotten the different character of the Birmingham traffic data in comparison with the Swedish. In fact, the results are an excellent illustration of the analysis of data containing a discrete spectral component, though in view of this I would prefer to work with the unsmoothed values in the neighbourhood of the peak, the position of which provides, of course, a very good estimate of the signal cycle time. One incidental remark on traffic data I hope to amplify when I discuss two-dimensional processes: that is, a full picture of the traffic flow even along one road constitutes a process with various possible descriptions, some more complete than others. For example, a two-dimensional point process exists in position–velocity phase-space, or in time–velocity phase-space, but the specification in terms of position and time is not a point process, but what I will term a *line process*.

We can, of course, consider particular features of such a process, such as the bivariate point process in time at two points in space. This is one of the examples mentioned by Dr Jenkins in his valuable suggestions on multivariate point processes, and on mixed continuous and point processes.

I am indebted to Mr T. Lewis for drawing attention to the problem of end effects in equation (33). A similar problem was mentioned in the discussion following equation (24), but I admit I had overlooked it in (33), and have taken the liberty of adding a brief phrase before the equation to cover myself against his criticism. With regard to the use of spectral analysis, say, for estimating the parameters of a process, the possible value of other methods has been raised also by some of the other speakers, including Mr Lewis's namesake, Mr P. A. W. Lewis. I have noted above that any *a priori* knowledge of the model should be exploited, and this could particularly apply to the estimation of parameters in the model. With regard to the model I used in my paper for the Swedish traffic data, I would not deny that it could be improved on, but it seemed sufficient for my purpose. This was rather confirmed by its providing an adequate fit for the spectrum of a second set of Swedish traffic data kindly shown to me by Dr Miller, with the *same* values for the parameters. Mr P. A. W. Lewis's example of failure patterns in computers is an intriguing one that I hope he will discuss some time at greater length.

I see no difficulty in principle in extending this type of analysis to processes of the type defined by Mr Walker. The (uncorrected) second-order product density is merely multiplied by the (uncorrected) second-order product moment for $X(t)$, though I have not looked to see whether there are any complications in the sampling theory.

Professor Barnard's proposed test of significance of the null hypothesis, while as he says of less relevance here, appears a promising one; but it would seem important to decide on the "bunching" criterion to be adopted *before* looking at the graphs.