Estimation of the Second-order Intensities of a Bivariate Stationary Point Process

By DAVID R. BRILLINGER

The University of California, Berkeley

[Received October 1974. Final revision May 1975]

SUMMARY

We consider histogram and smoothed histogram type estimates of the auto and cross intensity functions of a bivariate stationary point process. The asymptotic distributions are found to be multiples of Poissons in the histogram case and linear combinations of Poissons in the smoothed case. These asymptotic distributions suggest the plotting of the square roots of the estimates in order to stabilize the variance and to make the distributions more nearly normal. Two examples of such plots are presented in the paper.

Keywords: AUTOINTENSITY FUNCTION; CROSSINTENSITY FUNCTION; POINT PROCESS; SQUARE ROOT TRANSFORMATION; STATIONARY PROCESS

1. INTRODUCTION

LET $\{N_1(t), N_2(t)\}$, $-\infty<t<\infty$, be a bivariate stationary point process with $N_1(t)$ being the number of events of Type 1 that occurred in the time interval $(0,t]$ and $N_2(t)$ the number of events of Type 2 in the same interval. Suppose that the process is orderly in the sense that there is zero probability that events occur simultaneously. The intensity of events of type a is defined by

$$p_a = \lim_{h\downarrow 0} \Pr\{\text{type } a \text{ event in } (t,t+h]\}/h \tag{1.1}$$

for $a=1,2$. The existence of the limit (1.1) was shown by Khintchine (1960). Korolyuk showed that, with orderliness,

$$E\{dN_a(t)\} = p_a\,dt \tag{1.2}$$

(see Khintchine, 1960). The second-order product density function of events of type a with events of type b is defined by

$$p_{ab}(u) = \lim_{h,h'\downarrow 0} \Pr\{\text{type } a \text{ event in } (t+u,t+u+h] \text{ and type } b \text{ event in } (t,t+h']\}/(hh') \tag{1.3}$$

for $a,b=1,2$ and $u\neq 0$. The second-order intensity function of events of type a, given events of type b, is defined by

$$\begin{aligned} m_{ab}(u) &= \lim_{h\downarrow 0} \Pr\{\text{type } a \text{ event in } (t+u,t+u+h]\,|\,\text{type } b \text{ at } t\}/h \\ &= p_{ab}(u)/p_b \end{aligned} \tag{1.4}$$

for $a,b=1,2$ and $u\neq 0$. In this paper we are concerned with large sample properties of estimates of $p_{ab}(u)$, $m_{ab}(u)$ that have the form proposed in Griffith and Horn (1963), Cox (1965) and Cox and Lewis (1972). We shall propose a modified form of these estimates and, in the light of the large sample properties, recommend the application of a square root transformation. Numerous practical examples of estimates of the original form may be found in Bryant *et al.* (1973) for bivariate processes consisting of the input and output spike trains of nerve cells. Two examples of the modified estimates are presented in this paper. Numerous additional examples are given in a paper by Brillinger, Bryant and Segundo which is in preparation.

Suppose that the process $\{N_1(t), N_2(t)\}$ is given for $0<t\leqslant T$, that is, the times at which events occurred in the interval $(0,T]$ are known. Let the times of events of type a be $s_1, s_2, \ldots$, and the times of events of type b be $t_1, t_2, \ldots$. Let $\beta>0$ denote a scale parameter. Next, let $\#\{A\}$ denote the number of elements in a set A. Then the estimates of $p_{ab}(u)$ and $m_{ab}(u)$, considered in Cox and Lewis (1972), are based on the counting variate

$$J_{ab}^T(u) = \#\{(j,k) \text{ such that } u-\beta<s_j-t_k<u+\beta \text{ and } s_j\neq t_k\}. \tag{1.5}$$

$J_{ab}^T(u)$ counts the number of a events falling in a cell of bin width 2β and midpoint u time units along from a b event. It is a histogram type statistic. Cox and Lewis (1972) show that

$$\begin{aligned} EJ_{ab}^T(u) &\doteq (T-u)\int_{u-\beta}^{u+\beta} p_{ab}(v)\,dv \\ &\doteq 2\beta T p_{ab}(u) \end{aligned} \tag{1.6}$$

for large T, small β and moderate u, suggesting the estimates

$$\left.\begin{aligned} \hat{p}_{ab}(u) &= J_{ab}^T(u)/(2\beta T), \\ \hat{m}_{ab}(u) &= J_{ab}^T(u)/\{2\beta N_b(T)\}. \end{aligned}\right\} \tag{1.7}$$

We shall determine the asymptotic distributions of these estimates under certain regularity conditions. In addition we shall propose the use of the following modified estimates

$$\left.\begin{aligned} \hat{p}'_{ab}(u) &= \hat{p}_{ab}(u)+|u|\,N_a(T)\,N_b(T)/T^3, \\ \hat{m}'_{ab}(u) &= \hat{m}_{ab}(u)+|u|\,N_a(T)/T^2 \end{aligned}\right\} \tag{1.8}$$

for $|u|\leqslant T$. Under the regularity conditions mentioned, these appear to have better overall mean-squared error properties. Their definition will be motivated in Section 3. In the case that $|u|$ is not large compared to T, there is little difference between the estimates of (1.7) and (1.8). Their asymptotic distributions are the same.

2. The Asymptotic Distributions

Many random processes that occur in practice seem to satisfy some form of mixing condition, that is, functionals of the process that are well separated in time are only weakly dependent. We will make use of the following condition of that character.

Definition. A stationary bivariate process $\{N_1(t), N_2(t)\}$, $-\infty<t<\infty$, is called strong mixing when

$$\alpha(\tau) = \sup\{|P(AB)-P(A)P(B)|: A\in\mathcal{M}_{-\infty}^t, B\in\mathcal{M}_{t+\tau}^\infty\}\to 0 \tag{2.1}$$

as $\tau\to\infty$. Here $P(\cdot)$ denotes the probability measure of the process and $\mathcal{M}_u^v$ denotes the σ-algebra of events generated by events of the form

$$\{N_{a_1}(v_1)-N_{a_1}(u_1)\leqslant h_1, \ldots, N_{a_K}(v_K)-N_{a_K}(u_K)\leqslant h_K\},$$

where $a_k = 1,2$; $u<u_k<v_k\leqslant v$; h_k is a non-negative integer for $k=1,2,\ldots,K$ and $K=1,2,\ldots$.

This condition appears in Volkonskii and Rozanov (1959) for example. We shall also require that the second- to fourth-order moments of the process have the following forms:

$$\left.\begin{aligned} E\{dN_a(t+u)\,dN_b(t)\} &= p_{ab}(u)\,dt\,du, \\ E\{dN_a(t+u)\,dN_b(t+v)\,dN_c(t)\} &= p_{abc}(u,v)\,dt\,du\,dv, \\ E\{dN_a(t+u)\,dN_b(t+v)\,dN_c(t+w)\,dN_d(t)\} &= p_{abcd}(u,v,w)\,dt\,du\,dv\,dw \end{aligned}\right\} \tag{2.2}$$

for $a,b,c,d=1,2$ and $u,v,w,0$ distinct. Finally, let $P(\mu)$ denote a Poisson variate with mean μ.

Theorem 1. Let $\{N_1(t), N_2(t)\}$, $-\infty < t < \infty$, be a stationary bivariate point process that is strong mixing, $\alpha(\tau+u) = O\{\alpha(\tau)\}$ as $\tau \to \infty$, and such that $p_{ab}(u)$, $p_{abc}(u,v)$, $p_{abcd}(u,v,w)$ of (2.2) are finite and continuous for $a,b,c,d = 1,2$. Then for $u_k^T \to u_k$, $|u_k^T - u_{k'}^T| \geqslant 2\beta$, $1 \leqslant k < k' \leqslant K$ and $\beta = L/T$, L constant, the variates $J_{a_1 b_1}^T(u_1^T), \ldots, J_{a_K b_K}^T(u_K^T)$ are asymptotically independent $P\{2Lp_{a_k b_k}(u_k)\}$, $k = 1, \ldots, K$ for $a_k, b_k = 1,2$ as $T \to \infty$.

This result is proved in Section 4 of the paper assuming a direct variant of Theorem 1.3 of Volkonskii and Rozanov (1959). We may take $K = 1$ and $u_1^T = u$, here, and so see that $J_{ab}^T(u) \to P(2Lp_{ab}(u))$ for $a, b = 1,2$. We have allowed the arguments u_k^T to depend on T in order to be able to handle the case of a number of bins in the neighbourhood of a given lag u. The restriction on $|u_k^T - u_{k'}^T|$ means that the counting variates refer to distinct bins. In connection with the estimates of $p_{ab}(u)$, $m_{ab}(u)$ we have:

Corollary 1. Under the conditions of Theorem 1, $\hat{p}_{ab}(u)$, $\hat{p}'_{ab}(u)$, given by (1.7), (1.8), are asymptotically distributed as $(2L)^{-1} P\{2Lp_{ab}(u)\}$.

Corollary 2. Under the conditions of Theorem 1, $\hat{m}_{ab}(u), \hat{m}'_{ab}(u)$, given by (1.7), (1.8), are asymptotically distributed as $(2L)^{-1} p_b^{-1} P\{2Lp_{ab}(u)\}$.

Had we so desired, we could have considered collections of estimates, at lags u_k^T, in the manner of the theorem, here. The asymptotic distributions of the estimates of (1.7) are not affected by the modification to (1.8) because of the convergence of the correction terms to zero, in probability. In both cases, the variance of the asymptotic distribution is seen to be proportional to the parameter being estimated. This occurrence suggests the application of a square root transformation to the estimates. We will return to this comment in the next section.

The estimates discussed here are histogram type estimates, involving a rectangular smoothing function. Cox (1965) remarks that one might want to consider other smoothing functions. For example, we might base estimates on

$$\sum_{i=-I}^{I} w_i J_{ab}^T(u - 2i\beta), \tag{2.3}$$

where $\sum w_i = 1$. From Theorem 1, the asymptotic distribution of this variate is seen to be that of $\sum w_i P_i$, where the P_i are independent $P\{2Lp_{ab}(u)\}$ variates. The mean of this asymptotic distribution is $2Lp_{ab}(u)$. The variance is $(\sum w_i^2)\, 2Lp_{ab}(u)$, a result that again suggests a square root transformation.

3. Some Further Considerations and Practical Examples

The second-order product density, $p_{ab}(u)$, and the intensity function, $m_{ab}(u)$, both provide measures of the degree of statistical dependence of increments of the process $N_a(\cdot)$ that are u time units ahead of corresponding increments of the process $N_b(\cdot)$. In the case that these increments are independent, $p_{ab}(u) = p_a p_b$ and $m_{ab}(u) = p_a$. In the case that the process is stong mixing

$$E\{dN_a(t+u)\, dN_b(t)\} - E\{dN_a(t+u)\}\, E\{dN_b(t)\} = O\{\alpha(u)\} \to 0$$

as $|u| \to \infty$, see Volkonskii and Rozanov (1959) and so

$$\lim_{|u|\to\infty} p_{ab}(u) = p_a p_b \quad \text{and} \quad \lim_{|u|\to\infty} m_{ab}(u) = p_a. \tag{3.1}$$

This suggests that graphs of estimates of the functions $p_{ab}(\cdot)$ or $m_{ab}(\cdot)$ should also contain estimates of the constant levels $p_a p_b$ or p_a, as the case may be.

The relations of (3.1) suggest the source of the estimates $\hat{p}'_{ab}(u)$, $\hat{m}'_{ab}(u)$ of (1.8). For many processes, the covariance, $\operatorname{cov}\{dN_a(t+u), dN_b(t)\}$ will be near 0 for large $|u|$. In the case of an ordinary bivariate stationary process $\{X_1(t), X_2(t)\}$ the covariance function

$$c_{ab}(u) = \operatorname{cov}\{X_a(t+u), X_b(t)\}$$

is likewise near 0 for large $|u|$ for mixing processes. This has led workers to feel that for many purposes, $c_{ab}(u)$ is best estimated by

$$\hat{c}_{ab}(u) = T^{-1} \sum_{0 \leqslant t, t+u \leqslant T-1} \{X_a(t+u) - \bar{X}_a\}\{X_b(t) - \bar{X}_b\} \tag{3.2}$$

given values for $t = 0, 1, \ldots, T-1$ with $\bar{X}_a$, $\bar{X}_b$ the sample means $a, b = 1, 2$. The estimate (3.2) has the property of being near 0 for $|u|$ near T. The estimate (3.2) suggests estimating the second-order product moment, $E\{X_a(t+u)\,X_b(t)\}$ by

$$\hat{c}_{ab}(u) + \bar{X}_a \bar{X}_b = T^{-1} \sum_{0 \leqslant t, t+u \leqslant T-1} X_a(t+u)\,X_b(t) + |u|\,\bar{X}_a\,\bar{X}_b/T. \tag{3.3}$$

The estimates (3.2), (3.3) appear to have better overall mean-squared error properties than the corresponding "unbiased" estimates with the divisor T replaced by $(T-|u|)$, see Parzen (1961, p. 139). The estimate $\hat{p}'_{ab}(u)$ is the point process analogue of (3.3). It and the corresponding $\hat{m}'_{ab}(u)$ may be expected to have better overall mean-squared error for mixing processes.

The conclusions of Corollaries 1 and 2 suggest applying a square root transformation to the estimates. This is a common procedure for counting variates. In the cases of $\hat{m}_{ab}(u)$ and $\hat{m}'_{ab}(u)$, the large sample variances of $\sqrt{\{\hat{m}_{ab}(u)\}}$ and $\sqrt{\{\hat{m}'_{ab}(u)\}}$ are approximately $(8Lp_b)^{-1}$ which may be estimated by $\{8\beta N_b(T)\}^{-1}$. Confidence limits may be set by using either a Poisson or a normal approximation. A particularly simple approximation to 95 per cent limits is to add $\pm\{2\beta N_b(T)\}^{-\frac{1}{2}}$ to the estimate. In a case where the weighted estimate (2.3) was employed, a further factor $(\sum w_i^2)^{\frac{1}{2}}$ would be included.

In practice we have found it exceedingly useful to graph the following four curves on the same plot,

$$\sqrt{\{\hat{m}'_{ab}(u)\}}, \quad \sqrt{\hat{p}_a}, \quad \sqrt{\hat{p}_a} + \{2\beta N_b(T)\}^{-\frac{1}{2}}, \quad \sqrt{\hat{p}_a} - \{2\beta N_b(T)\}^{-\frac{1}{2}},$$

where $\hat{p}_a = N_a(T)/T$, $a = 1, 2$. Figs 1 and 2 give two examples of this. Fig. 1 is based on a point process corresponding to the times of 1,355 consecutive beats of a human heart. The

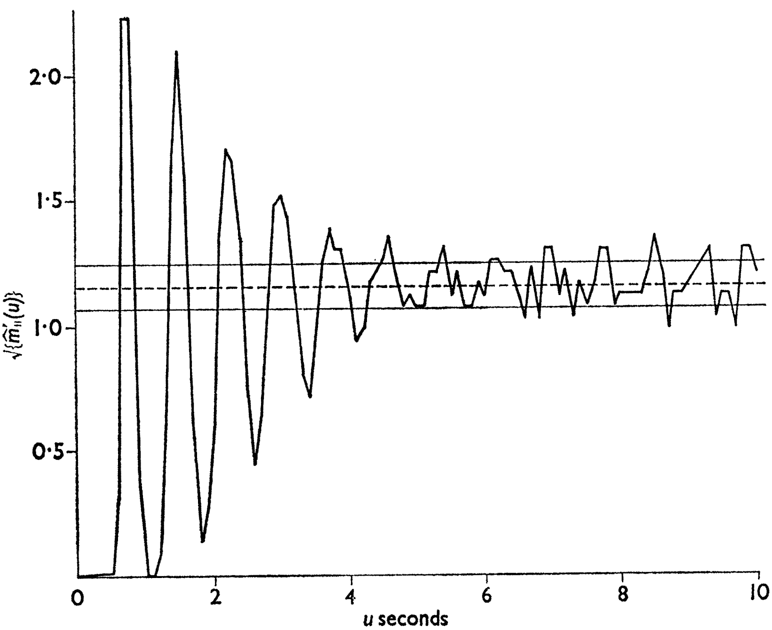

FIG. 1. Square root of the autointensity histogram of a sequence of times of heartbeats.

estimated intensity is $\hat{p}_1 = 1{\cdot}33$ beats per sec. The central horizontal line in the Figure corresponds to $\sqrt{\hat{p}_1}$. The Figure suggests that increments of the corresponding process are approximately independent when more than 4 or 5 sec apart. The estimate is essentially 0 for the first 0·4 sec because no heart beats occurred closer together than that interval. This behaviour is characteristic of point processes generated by mechanisms with dead times. The equi-spaced spikes appear in this estimate because of the periodic character of heartbeats. Fig. 2 is based upon the times of the 187 world-wide earthquakes of magnitude 7·9 or greater which occurred in the years 1900–71. The estimated intensity is $\hat{p}_1 = 0{\cdot}22$ major earthquakes per month. Such a process is often thought to be near Poisson. The function of Fig. 2 suggests that there is

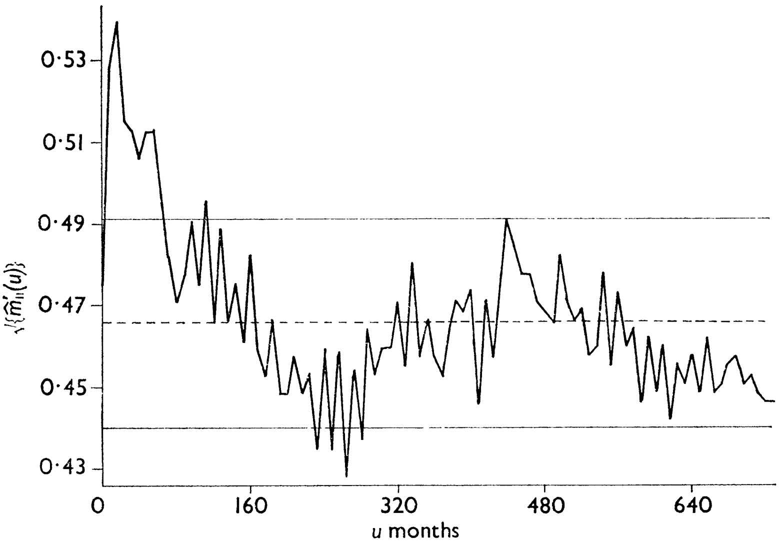

FIG. 2. Square root of the autointensity histogram of a sequence of times of major earthquakes.

some degree of clustering present. The central horizontal line is at the level $\sqrt{\hat{p}_1}$. The times of these earthquakes may be found in Richter (1958) and "Seismological Notes" appearing in the *Bulletin of the Seismological Society of America*.

4. PROOFS

The proof of Theorem 1 will be based on Theorem 2.

Theorem 2. Let $\mathbf{M}^T(t) = \{M_1^T(t), \ldots, M_K^T(t)\}$, $-\infty < t < \infty$, $T = 1, 2, \ldots$ be a sequence of stationary K-variate point processes. Suppose the process $\mathbf{M}^T$ has mixing coefficient $\alpha^T(\tau)$, where $\alpha^T(\tau) \to 0$ uniformly as $\tau \to \infty$, $\alpha^T(\tau) \approx \alpha(\tau)$. Suppose that the process M_k^T has intensity $\mu_k \varepsilon_T$ where $\varepsilon_T \to 0$ as $T \to \infty$ and that

$$\left.\begin{aligned} E[M_k^T(t/\varepsilon_T)\{M_k^T(t/\varepsilon_T) - 1\}] &= o(t), \\ E\{M_k^T(t/\varepsilon_T)\, M_{k'}^T(t/\varepsilon_T)\} &= o(t) \end{aligned}\right\} \tag{4.1}$$

as $t \to 0$, $T \to \infty$ for $1 \leqslant k < k' \leqslant K$. Then $M_1^T(t_1/\varepsilon_T), \ldots, M_K^T(t_K/\varepsilon_T)$ are asymptotically independent $P(\mu_1 t_1), \ldots, P(\mu_K t_K)$ as $T \to \infty$.

This theorem is a simple variant of Theorem 1.3 of Volkonskii and Rozanov (1959) and is not proved here. A related result is discussed in Section 5 of Leadbetter (1969). The condition that the mixing coefficients of all the processes are of the same order of magnitude means, that with regards degree of mixing, the processes retain the same time scale. The intensities are assumed to tend to 0, meaning that events are becoming rare as $T \to \infty$. The first condition of (4.1) prevents the processes from having too many events in small intervals.

The second condition is what leads to the limit process having independent components. The time scaling, t_k/ε_T leads to the limit process having non-zero intensities.

It is convenient to prove Theorem 1 by applying Theorem 2 to a particular sequence of K variate processes associated with the counting of events in intervals of width 2β derived from the given process $\{N_1(t), N_2(t)\}$. Specifically define the process $M_k^T(\cdot)$ to have j events at time t if N_{b_k} has an event at time t and if N_{a_k} has j events, at times other than t, between $t+u_k^T-\beta$ and $t+u_k^T+\beta$. In differential notation this corresponds to writing

$$dM_k^T(t) = \int_A dN_{a_k}(s+t)\, dN_{b_k}(t), \tag{4.2}$$

where A is the set $\{s\colon u_k^T-\beta < s < u_k^T+\beta,\ s\neq 0\}$. The counting variates of (1.5) and Theorem 1 are now given by allowing $dM_k^T(t)$ to range over the interval $(\beta, T-\beta)$ whose end points are the centres of the first and last bins. Hence

$$J_{a_k b_k}^T(u_k^T) = M_k^T(T-\beta) - M_k^T(\beta) \sim M_k^T(T).$$

Proof of Theorem 1. From (4.2) we note that

$$E\, dM_k^T(t) \sim 2\beta p_{a_k b_k}(u_k)\, dt$$

as β is small. This gives the mean values as stated in Theorem 1 for

$$EJ_{a_k b_k}^T(u_k^T) \sim E\, M_k^T(T) \sim 2L\, p_{a_k b_k}(u_k).$$

It also suggests defining ε_T of Theorem 2 to be $1/T$. The process M_k^T has mixing coefficient

$$\alpha^T(\tau) = \sup\{|P(AB) - P(A)P(B)|\},$$

where A ranges over events involving the variates $dN_{a_k}(s)$, $dN_{b_k}(s)$ with $s < \min(t, t+u_k^T+\beta)$ and B ranges over events involving these variates with $s > \max(t+\tau, t+\tau+u_k^T-\beta)$. It follows that $\alpha^T(\tau) = \alpha(\tau - |u_k^T| - \beta) \to 0$ uniformly as $\tau\to\infty$, and that $\alpha^T(\tau) \approx \alpha(\tau)$ as it has been assumed that $\alpha(\tau+u) = O\{\alpha(\tau)\}$ for $\tau\to\infty$.

Next, following expressions (3.15), (3.16) of Brillinger (1972) the expected values of (4.1) may both be written

$$\int_0^{t/\varepsilon_T}\int_0^{t/\varepsilon_T} P_{kk'}^T(v_1-v_2)\, dv_1\, dv_2, \tag{4.3}$$

whether $k=k'$ or not, where $P_{kk'}^T$ is a second-order product density of the M_T process. Now from the representation (4.2) and Theorem 3.1 of Brillinger (1972), expression (4.3) hence has the value

$$\begin{aligned}\int_{A_1}\int_{A_2}\int_{B_1}\int_{B_2} [&\delta\{a_k-b_{k'}\}\,\delta(v_1+t_1-v_2-t_2)\, p_{a_k b_k a_{k'}}(v_1+t_1-v_2-t_2, v_1-v_2-t_2)\\ &+\delta\{a_k-a_{k'}\}\,\delta(v_1+t_1-v_2-t_2)\, p_{a_k b_k b_{k'}}(v_1+t_1-v_2, v_1-v_2)\\ &+\delta\{b_k-a_{k'}\}\,\delta(v_1-v_2-t_2)\, p_{a_k b_k b_{k'}}(v_1+t_1-v_2, v_1-v_2)\\ &+\delta\{a_k-b_{k'}\}\,\delta\{b_k-a_{k'}\}\,\delta(v_1+t_1-v_2)\,\delta(v_1-v_2-t_2)\, p_{a_k b_k}(t_1)\\ &+p_{a_k a_{k'} b_k b_{k'}}(v_1+t_1-v_2, t_2, v_1-v_2)]\, dt_1\, dt_2\, dv_1\, dv_2,\end{aligned}$$

where A_j denotes the set $\{u_k^T-\beta < t_j < u_k^T+\beta,\ t_j\neq 0\}$, $j=1,2$, B_j denotes the set $\{0<v_j<t/\varepsilon_T\}$, $j=1,2$, $\delta\{\cdot\}$, denotes the Kronecker delta function, $\delta(\cdot)$ denotes the Dirac delta function and where the domain of integration excludes the values $v_1-v_2=0$. Making use of the boundedness of the densities of (2.2) we find that (4.3) is $O(t)+O(t^2)+O(t/T)$. This tends to 0 as $t\to 0$, $T\to\infty$. The conditions of Theorem 2 are therefore satisfied and so Theorem 1 follows.

ACKNOWLEDGEMENTS

I would like to thank Professor J. P. Segundo and Dr H. L. Bryant, Jr. for raising questions which led to the preparation of this paper. I would like to thank Dr J. G. Fryer for providing the data analysed in Fig. 1. I also would like to thank the referee for comments leading to improved clarity of the paper. This work was partially supported by the National Science Foundation Grant GP–31411.

REFERENCES

BRILLINGER, D. R. (1972). The spectral analysis of stationary interval functions. *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, **1**, 483–513.

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

COX, D. R. (1965). On the estimation of the intensity function of a stationary point process. *J. R. Statist. Soc.* B, **27**, 332–337.

COX, D. R. and LEWIS, P. A. W. (1972). Multivariate point processes. *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, **2**, 401–448.

GRIFFITH, J. S. and HORN, G. (1963). Functional coupling between cells in the visual cortex of the unrestrained cat. *Nature, Lond.* **199**, 876; 893–895.

KHINTCHINE, A. YA. (1960). *Mathematical Methods in the Theory of Queueing*. London: Griffin.

LEADBETTER, M. R. (1969). On certain results for stationary point processes and their application. *Bull. Int. Statist. Inst.*, **43**, Book 1, 309–320.

PARZEN, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics*, **3**, 167–190.

RICHTER, C. F. (1958). *Elementary Seismology*. San Francisco: Freeman.

VOLKONSKII, V. A. and ROZANOV, YU. A. (1959). Some limit theorems for random functions, I. *Theory Prob. Appl.*, **4**, 178–197.