MEASURING THE ASSOCIATION OF POINT PROCESSES: A CASE HISTORY

DAVID R. BRILLINGER

1. Introduction. Modern applied statistics typically involves elements of computation, probability theory, statistical theory and collaboration with specialists in the subject matter of some substantive field. In this article I shall describe part of a continuing experience of collaboration with two neurophysiologists from U.C.L.A., H. L. Bryant Jr. and J. P. Segundo. In formal terms, the problem considered is one of measuring the degree of association of points of two different sorts distributed along a straight line in an irregular manner. In real terms, the problem is one of investigating the behavior of a simple nerve cell network in a sea slug (*Aplysia californica*). The paper discusses a summary measure of association that has proved useful in assessing whether two nerve cells are behaving in a related manner or are behaving independently. The experiments by means of which the data were collected are described in Bryant, Ruiz Marcos and Segundo **[4]**, as are the results of preliminary statistical analyses. The paper **[4]** is representative of the extent to which quantification is now occuring in the life sciences.

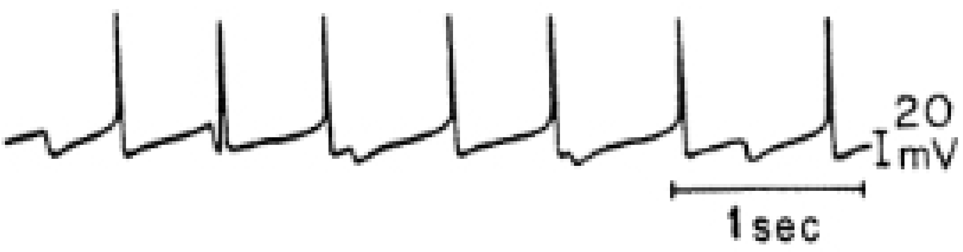

FIG. 1. A typical record of the changing voltage level of a nerve cell.

2. Some neurophysiology. The nerve cell (or neuron) is the basic unit of the animal involved in the transmission of information. Described schematically, it consists of a central cell body (or soma), branches (called dendrites) carrying impulses to the body and a long outgrowth (the axon) conducting impulses from the body. One way information is transmitted through the dendrites and axon is through changes in electrical activity. Figure 1 is an example of the changing voltage recorded when a microelectrode is inserted into a nerve cell. The record is seen to be made up of pulses of large amplitude compared to their duration. Because of its appearance, such a record is often called a spike train.

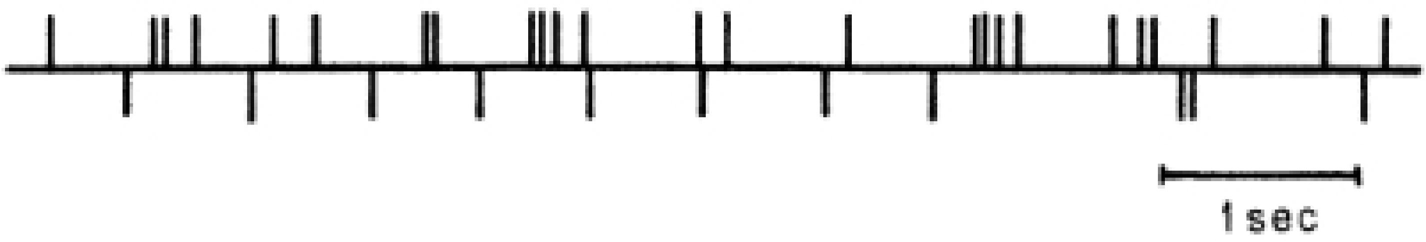

FIG. 2. A record of the times of spikes of two simultaneously firing nerve cells.

The junction whereby one neuron may influence another is called the synapse. When a pulse reaches the terminal point of an axon it provokes the release of a transmitter substance which alters the permeability of the dendrite of the next cell to certain ions. The resulting flow of ions generates a small electric current which moves down the dendrite to the soma. If the junction is excitory, the spike activity of the second cell is increased, if inhibitory it is decreased. Figure 2 is an example of the times of spikes for two nearby cells, the times for one cell corresponding to spikes above the line and for the other corresponding to spikes below. In practise, given two neurons, it may not be known whether either is influencing the other and it may be of interest to determine if there is some influence or association. Doing this by eye from records such as those of Figure 2 can be very difficult. Researchers have therefore been led to compute summary values from the records (see Griffith and Horn [9] for example) and this is the concern of the present paper.

The data discussed is recorded simultaneously on cells L3 and L10 of the sea hare. This particular animal and these particular cells were used because the cells may be identified in different specimens and consequently experiments may be repeated. The experimental methods are described in detail in [4]. Further information concerning neurons and synapses may be found in Eccles [7].

3. Some probability theory. Commonly in his work on applied problems, a statistician brings the apparatus of probability theory into use. This involves his asking the experimentalists and himself whether or not it is reasonable to talk about random outcomes and probabilities of events connected with outcomes. The statistician seeks to bring probability theory into a problem because it provides a precise means of defining parameters and models and it allows him to interpret and assess various manipulations of experimental data. Not all problems of data analysis require the introduction of probability theory, but many seem to benefit from its appearance—among the latter are problems concerning nerve cell spike trains.

The branch of probability theory concerned with entities like irregular spike trains is that of stochastic point processes. A stochastic point process is a random, non-negative, integer-valued measure. If I is an interval of the real line and ω is a random element, then the values of this measure may be denoted by $N(I, \omega)$, with $N(I, \omega)$ denoting the number of points in the interval I for the realization corresponding to ω. Here the atoms of the measure $N(I, \omega)$ correspond to the times of spikes of a particular spike train. Repeating the experiment would most likely yield a different set of spike times and consequently a different measure $N(I, \omega')$. In this sense N is a random measure. (We remark that in many problems one can suppress the dependence of N on ω, however, it is an essential element of the approach.) Point processes were considered recently in the MONTHLY by Chung [5] and are discussed in Cox and Lewis [6] and in a volume [11] edited by Lewis, for example.

For nerve cell trains, it is appropriate to assume that the point process is without multiple points; that is, the spike times are isolated, separated by positive distances. Because the spikes proceed from no inherent origin, it also seems appropriate to assume that the point process is stationary in time in the sense that the probability distribution of the random vector

$$\{N(I_1, \omega), \cdots, N(I_k, \omega)\}$$

is the same as that of the shifted vector

$$\{N(I_1 + t, \omega), \cdots, N(I_k + t, \omega)\}$$

for all t and $k = 1, 2, \cdots$, where $I + t$ denotes the interval $(a + t, b + t)$ if $I = (a, b)$.

Important parameters of a stationary point process N include the **mean intensity,** p_N, and the **second-order product density,** $p_{NN}(u)$, given by

$$p_N = \lim_{h \downarrow 0} \text{Prob}\{\text{point in the interval } (t, t + h)\}/h \tag{1}$$

and

$$p_{NN}(u) = \lim_{h, h' \downarrow 0} \text{Prob}\{\text{point in } (t + u - h, t + u + h) \text{ and point in } (t - h', t + h')\}/(4hh') \tag{2}$$

$-\infty < u < \infty$, respectively when these limits exist.

In fact we shall be concerned with two different types of points, say M points and N points, with $M(I, \omega)$ referring to the number of M points in the interval I and $N(I, \omega)$ the number of N points in the interval I. We denote the mean intensity of M points by p_M and the second-order product density of M points by $p_{MM}(u)$. We also define a **cross-product density,** $p_{MN}(u)$, by

$$p_{MN}(u) = \lim_{h, h' \downarrow 0} \text{Prob}\{M \text{ point in } (t + u - h, t + u + h) \text{ and } N \text{ point in } (t - h', t + h')\}/(4hh'). \tag{3}$$

The parameters in (1), (2), (3) do not depend on t because the process is stationary.

The first thing that one tends to notice when examining a spike train is whether there are a lot of spikes or only a few. The mean intensity of the process gives information in this connection. Expression (1) implies that the probability of there being an N point in a small interval of length h is approximately $p_N h$. The next thing that one tends to notice is the relative positioning of pairs of spikes of a single train or from one train to another. Expressions (2) and (3) give information in this connection. From expression (3), for example, we have

$$\text{(4)} \qquad \operatorname{Prob}\{M \text{ point in } (t+u-h, t+u+h) \text{ and } N \text{ point in } (t-h', t+h')\} \sim p_{MN}(u) 4hh'$$

for h, h' non-negative and small. Using the definition of conditional probability and (1), this implies that

$$\text{(5)} \qquad \operatorname{Prob}\{M \text{ point in } (t+u-h, t+u+h) \text{ given an } N \text{ point at } t\} \sim 2h\, p_{MN}(u)/p_N.$$

In the case that the M points are distributed independently of the N points, the probability referred to in expression (5) is just $\operatorname{Prob}\{M \text{ point in } (t+u-h,\ t+u+h)\}$ and so

$$\text{(6)} \qquad p_{MN}(u)/p_N = p_M \quad \text{or} \quad p_{MN}(u) = p_M p_N$$

for all u. This last suggests that the function $p_{MN}(u)$, and related functions such as

$$\text{(7)} \qquad \frac{p_{MN}(u)}{p_M p_N} \quad \text{or} \quad \sqrt{\frac{p_{MN}(u)}{p_M p_N}}$$

might prove useful measures of the degree of association of points of the M process with points of the N process. They are identically 1.00 in the case of independence.

We remark that, since we have assumed the points of the processes to be isolated, we can replace the probabilities of expressions (1)–(3) by expected values, for example we could write for (3)

$$p_{MN}(u) = \lim_{h,h' \downarrow 0} E\{N(t+u-h, t+u+h)\, N(t-h', t+h')\}/(4hh').$$

4. Some statistical theory. The preceding section described a mathematical idealization that could be of use in examining the degree of relationship of two given spike trains. The idealization suggested the definition of parameters p_M, p_N, $p_{MN}(u)$ based on the probabilities of certain events. In order to make concrete use of these parameters we need to have some idea of their values for the spike trains at hand.

Statistical theory has long been concerned with the problem of estimating the probability of an event given experimental results. In elementary situations one estimates the probability of an event A by n_A/n, where n_A denotes the number of times the event A occurred out of n times when it might have occurred. Let us use this approach to construct estimates of $p_M, p_N, p_{MN}(u)$.

Suppose that spike trains M and N are observed throughout the time interval $(0, T)$. Let $s_1 < s_2 < \cdots < s_{M(T)}$ be the observed times of M spikes and $t_1 < t_2 < \cdots < t_{N(T)}$ be the observed times of N spikes where we have observed $M(T)$ M spikes and $N(T)$ N spikes in all. Let h be small and imagine the interval $(0, T)$ divided into T/h intervals of length h. The number of times the event "M spike in small interval of length h" occurred is $M(T)$. It might have occurred T/h times. This suggests estimating $p_M h$ by $M(T)/(T/h)$ and so estimating p_M by

$$\text{(8)} \qquad \hat{p}_M = M(T)/T.$$

Likewise we could estimate p_N by $\hat{p}_N = N(T)/T$.

Next we consider the estimation of $p_{MN}(u)$. For small h, let