MODELLING AND ANALYSIS OF SOME RANDOM PROCESS DATA FROM NEUROPHYSIOLOGY*

David R. Brillinger[1], Statistics Department, University of California, Berkeley

ABSTRACT

Models, graphs and networks are particularly useful for examining statistical dependencies amongst quantities via conditioning. In this article the nodal random variables are point processes. Basic to the study of statistical networks is some measure of the strength of (possibly directed) connections between the nodes. The coeficients of determination and of mutual information are considered in a study for inference concerning statistical graphical models. The focus of this article is simple networks. Both second-order moment and threshold model-based analyses are presented. The article includes examples from neurophysiology.

Key words: neurons, graphical model, mutual information, network, partial coherence.

RESUMEN

Modelos gráficos y redes son particularmente útiles para el examen de dependencias estadísticas entre cantidades vía su condicionamiento. En este trabajo las variables nodales aleatorias son procesos puntuales. Algunas medidas de la fortaleza de las conexiones entre nodos (posiblemente dirigidas) son básicas para el estudio de redes estadísticas. Los coeficientes de determinación y la información mutual son considerados en un estudio para inferencias concernientes a modelos gráficos estadísticos. El foco de este trabajo son las redes simples. Los dos momentos de segundo orden y sus umbrales, basados en un análisis de modelos son presentados. El trabajo incluye ejemplos de neurofisiología.

MSC: 62P10

1. INTRODUCTION

The work presented considers the use of coefficients of determination, and of coefficients of mutual information as measures of the strength of association of connections. Time-side approaches are presented for the point process case.

Two empirical examples are presented. The first refers to a time series derived from the intervals of a single point process and is by way of introducing a discussion of dependency concepts. The data involved are the firing times of a neuron of the sea hare *Aplysia californica*. The second example refers to collections of simultaneously firing neurons and the problem is that of inferring the wiring diagram amongst individual neurons and amongst particular regions of the brain. The nodes fall into groups and the structure or wiring diagram of the system is to be discovered. A box and arrow representation of the structure is given in Figure 1.

Basic books discussing statistical graphical models include Cox and Wermuth [1998], Whittaker [1990], Edwards [1995], Lauritzen [1996]. The paper has the following sections: Coefficients of Determination, Mutual Information, Networks, Results, Discussion and Extensions.

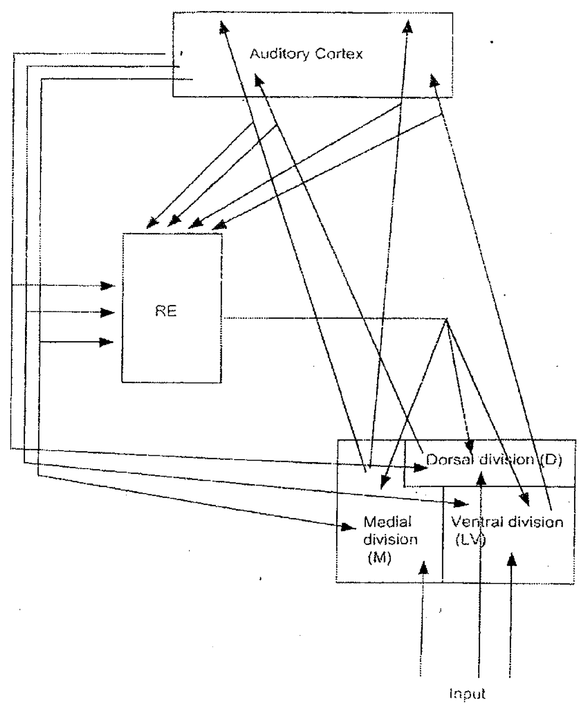

Figure 1. A schematic diagram related to the auditory regions of the brain of the cat.

*This paper was presented at VIII Conferencia Latinoamericana de Probabilidades y Estadística Matemática.
[1]The research was supported by the NSF grants DMS-9704739 and DMS-9971309.
Email:brill@stat.berkeley.edu

2. COEFFICIENTS OF DETERMINATION

2.1. Ordinary Random Variables

a) **Correlation analysis**

Let (X; Y) denote a bivariate random variable with corr{X,Y} = ρ_{XY}. The coefficient of determination is

$$\rho_{XY}^2 = |\text{corr}\{X, Y\}|^2$$

It is symmetric in X and Y and invariant under separate linear transformations of the two variates. It is a measure of independence, explained variation, and strength of linear dependence. Further it shows the uncertainty of various estimates. These points may be elaborated upon as follows:

1) *Independence*

It is the case that ρ_{XY} = 0 when X and Y are independent. It is not the case that ρ_{XY} = 0 implies independence. There are many examples of the form

$$Y = X^2 \text{ where the distribution of X is symmetric about 0}$$

Here Y is perfectly dependent on X, yet ρ_{XY} = 0. Below it will be seen that mutual information does not suffer this disadvantage.

2) *Strength of dependence*

i) Let $Y = \alpha + \beta X + \epsilon$ with $\text{cov}\{X, \epsilon\} = 0$, then

$$\rho_{XY}^2 = \text{var}\{\beta X\}/\text{var}\{Y\} = \beta^2\sigma_{XX}/(\beta^2\sigma_{XX} + \sigma_{\epsilon\epsilon})$$

and one sees that ρ_{XY}^2 increases to 1 when $|\beta|$ increases or when $\sigma_{\epsilon\epsilon}$ decreases.

ii) Suppose $Y = \beta X + \epsilon$ and $Z = \gamma Y + \eta$, with X; η; ϵ independent, then

$$\rho_{XZ}^2 \leq \rho_{XY}^2 \tag{1}$$

i.e. Y is better at linearly explaining X than Z. (An elementary proof of this result will be provided shortly.)

3) *Uncertainty of other quantities*

Suppose $Y = \alpha + \beta X + \epsilon$, with $\text{cov}\{X, \epsilon\} = 0$. Let β be the least squares estimate of β. Then

$$\text{var}\{\beta|X\} = (1 - \rho_{XY}^2)\sigma_{YY} / \sum(x_i - \bar{x})^2$$

and one sees how the estimated uncertainty of β depends on the size of ρ_{XY}^2.

A related aspect to the use of the coefficient of determination as a measure of strength of dependence is that one finds that the function of X that maximizes ρ_{XY}^2 is E{Y|X}, Rao [1965], Brillinger [1975].

b) **Partial analysis**

There is a quantity analagous to the correlation coeficient that is defined for trivariate random variables. It is of use in understanding the structure of networks such as those in Figure 2. For the variate (X, Y, Z) the partial correlation of X and Z given Y is defined as

$$\rho_{XZ/Y} = (\rho_{XZ} - \rho_{XY}\rho_{YZ}) / \sqrt{(1-\rho_{XY}^2)(1-\rho_{YZ}^2)} \tag{2}$$

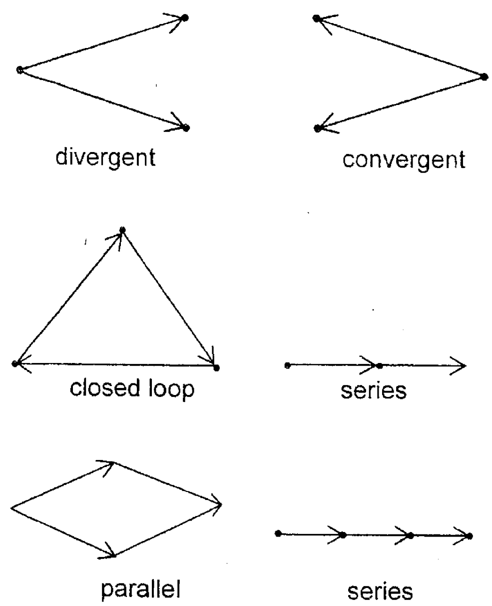

Figure 2. Some 3– and 4- node networks.

This coefficient corresponds to corr$\{\Psi,\Phi\}$ in the model

$$X = \alpha + \beta Y + \Psi. \qquad Z = \gamma + \delta Y + \Phi \tag{3}$$

with $\alpha, \beta, \gamma\ \delta$ constants and the variable Y uncorrelated with the "error" variate $\{\Psi,\Phi\}$. Partial correlation corresponds to the correlation of X and Z having removed the linear effects of the variate Y. In the model (3), β and δ are regression coefficients.

The partial correlation coefficient $\rho_{XZ|Y}$ as defined by (2) may now be used to prove the inequality (1) above. Specifically

$$\rho^2_{XZ|Y} = 0 \Rightarrow \rho_{XZ} - \rho_{XY}\rho_{YZ} = 0 \Rightarrow |\rho_{XZ}| = |\rho_{XY}||\rho_{YZ}|$$

which implies the inequality (1).

The concept of partial correlation extends directly to the case of Y vector-valued. A classic discussion of these ideas may be found in Kendall and Stuart [24]. In the case of 4 variables,

$$\rho_{XZ|YW} = \text{corr}\{X - \alpha Y - \beta W, Z - \gamma Y - \delta W\}$$

with α, β, γ, δ regression coefficients.

2.2. Processes

The next consideration will be data that are functions, in particular realizations of stationary processes. Two approaches are considered: time-side and frequency-side.

Also two types of processes, time series and point processes, are studied.

a) **Time series**

Consider a stationary time series, Y(t), t = 0, ± 1, ± 2,... . In a time-side approach, as a measure of the strength of association, one considers the coefficient of determination at lag u defined as

$$\rho_{YY}(u)^2 = |\text{corr}\{X(t+u), Y(t)\}|^2 \tag{4}$$

It may be estimated directly from an estimate of the autocorrelation function. An example will be provided later. Such a parameter has an interpretation as the proportion of variation in Y(t + u) explained by a linear function of the value Y(t).

There are direct extensions to the bivariate case and to the frequency domain. For example one has the coherence coeficient at frequency λ

$$|R_{XY}(\lambda)|^2 \approx |\text{corr}\{dZ_X(\lambda), dZ_Y(\lambda)\}|^2$$

with $dZ_X(\lambda)$ and $dZ_Y(\lambda)$ representing the components of frequency λ in the Cramér representation of the series. This parameter has an interpretation as the proportion of the variance explained at frequency λ. The coherence may be estimated by

$$|R_{XY}(\lambda)|^2 = \left|\text{corr}\left\{\sum e^{-i\lambda t} X(t), \sum e^{-i\lambda t} Y(t)\right\}\right|^2$$

In a time-side partial analysis one can consider $\rho_{XZ|Y}$ with X = Y(t), Y = Y(t + u), Z = Y(t + v) and be led to $\rho_{XZ|Y}(u,v)^2$. In a frequency-side analysis one can consider the partial coherency

$$R_{XZ|Y} = (R_{XZ} - R_{XY}R_{YZ}) / \sqrt{(1-|R_{XY}|^2)(1-|R_{YZ}|^2)}$$

suppressing the dependence on λ.

b) **Point processes**

Consider a point process $N = \{\tau_j\}$ with $\tau_{j+1} \geq \tau_j$. $j = 0, \pm 1, \pm 2, \ldots$. Define the time series of interevent intervals, $Y(j) = \tau_{j+1} - \tau_j$. $j = 0, \pm 1, \pm 2, \ldots$. In the case that N is stationary, so too will be the time series Y. One can consider the use of the quantity corf{Y(j + k), Y(j)} above for the series {Y(j)}. Alternately one might choose to work with the values log Y(j).

Another way to approach the point process case is to approximate a point process by a binary time series, Y (t) = 0, 1 having decided on an appropriate unit time interval. It may be mentioned that there are analogs of the covariance function such as, for a bivariate process {M; N},

$$E\{dN(t+u)dM(t)\} = p_{MN}(u)dtdu$$

and the coherence function

$$|R_{MN}(\lambda)|^2 = |\text{corr}\ \{dZ_M(\lambda), dZ_N(\lambda)\}|^2$$

as before. Some analogs of partial correlations were proposed for the point process case in Brillinger [1975].

3. MUTUAL INFORMATION

3.1. Discrete case

Many of the existing statistical results of information theory are for the case of discrete-valued random variables. This may be the result of the more common availablility of limit theorems for the discrete case, e.g. ergodic theorems. However in this section the focus will be on the real-valued continuous case.

3.2. Continuous case

The field of information theory provides some concepts of broad use in statistics. One of these is mutual information. It is a generalization of the coefficient of determination and it unifies a variety of problems.

For a bivariate random variable (X; Y) with density function p(x, y) the mutual information (MI) is defined as

$$I_{XY} = \int_S p(x,y)\log\frac{p(x,y)}{p_X(x)p_Y(y)}dxdy$$

where S is the region p(x, y) > 0.

As an example, for the bivariate normal the MI is given by

$$I_{XY} = \frac{1}{2}\log(1-\rho_{XY}^2)$$

and one sees the immediate connection with the coefficient of determination.

The coefficient I_{XY} has the properties of:

1) Invariance, $I_{XY} = I_{UV}$ if the transformation (X, Y) $\rightarrow$ (U, V) has the form U = f(X), V = g(Y) with f and g each 1-1 transforms.

2) Non negativity, $I_{XY} \geq 0$.

3) Measuring independence in the sense that $I_{XY} = 0$ if and only if X and Y are statistically independent.

4) Providing a measure of the strength of dependence in the senses that i) $I_{XY} = \infty$ if $Y = g(X)$, and ii) $I_{XZ} \leq I_{XY}$ if X is independent of Z given Y.

The property 3) that $I_{XY} = 0$ only if X and Y are independent stands in strong contrast to the much weaker correlation property of ρ^2_{XY}.

Brief proofs of these results run as follows:

Proof of 1) Suppose that f and g are differentiable. Set down the probability element including the Jacobian.

Proof of 2) Apply Jensen's Inequality to

$$-E\left[\log \frac{p_X(X)p_Y(Y)}{p(X,Y)}\right]$$

Proof of 3) From the only if part of Jensen's Theorem.

Proof of 4) One uses the following two expressions for the two parts of the claim.

$$I_{XY} = \int p_X(x)\,\delta(y-g(x))\log\frac{\delta(y-g(x))}{p_Y(y)}\,dxdy$$

$$I_{XZ} + I_{XY|Z} = I_{X(YZ)} = I_{XY} + I_{XZ|Y}$$

with $\delta(.)$ the Dirac delta function.

References to the present material and the proofs include: Granger and Lin (1994), Cover and Thomas [1991], and Joe [1989].

A concept related to the MI is that of the Kulback-Liebler information or entropy defined as

$$I(p,q) = \int p(x)\log\frac{p(x)}{q(x)}\,dx$$

for the possible densities p, q of the variate X. It is a measure of how close the density q is to p. Taking the variate (X, Y) and p to be p(x, y) and q to be $p_X(x)p_Y(y)$ one is led to the mutual information.

The mutual information is directly related to the concept of entropy, the entropy of a random variable Z with density $p_Z(z)$ being defined as

$$\int p_Z(z)\log p_Z(z)dz = E\{\log p_Z(z)\}$$

One has the connection

$$I_{XY} = \int_S p(x,y)\log\frac{p(x,y)}{p_X(x)p_Y(y)}\,dxdy = E\{\log p(X,Y)\} - E\{\log p_X(X)\} - E\{\log p_Y(Y)\} \quad (5)$$

The estimation of entropy

There are several methods that have been used.

Nonparametric estimate

Suppose one is considering the bivariate random variable (X, Y). Supposing further that p(x, y), is an estimate of the density p(x, y), for example a kernel estimate, then a direct estimate of the entropy is

$$\delta^2 \sum_{i,j} p(i\delta, j\delta) \log p(i\delta, j\delta) = E\{\log p(X, Y)\} \qquad \text{for } \delta \text{ small} \tag{6}$$

In the same way $E\{\log p_X(X)\}$, $E\{\log p_Y(Y)\}$ may be estimated and one can proceed to an estimate of the mutual information via expression (5). References to the type of entropy estimate just described and some statistical properties include: Joe [1989], Hall and Morton [1993], Fernandes [2000], Hong and White [2000], Granger **et al.** (2000).

Difficulties with this form of estimate can arise when $p_X(\cdot)$, $pY(\cdot)$ are small. The nonparametric form also runs into difficulty when one moves to higher dimensions.

A sieve type of estimate is presently being investigated for this situation, in particular an orthogonal function expansion employing shrunken coefficient estimates.

Parametric estimates of entropy

If the density $p(x, y|\theta)$ depends on a parameter θ that may be estimated reasonably then an immediate estimate of the entropy is provided by

$$\int p(x, y|\theta) \log(x, y|\theta)\, dxdy$$

Another form of estimate is based on the likelihood function. Suppose one has a model for the random variable (X, Y) including the parameter θ, (of dimension ν). Suppose the model has the property that X and Y are independent when $\theta = 0$. When there are n independent observations the log likelihood ratio for the hypothesis $\theta = 0$ is

$$\sum_{1}^{n} \log \frac{p(x_i, y_i|\theta)}{p_X(x_i) p_Y(y_i)}$$

with expected value

$$nI_{XY}$$

This suggests the use of the loglikelihood ratio statistic divided by n as an estimate of I_{XY}. A further aspect of the use of this statistic is that its distribution will be approximately proportional to X_ν^2 when X and Y are independent. An example is provided later in the paper.

Partial analysis

When networks are being considered the conditional mutual information is also of use. One can consider

$$I_{XY|Z} = \int\int\int p(x, y, z) \log \frac{p(x, y, z)p(z)}{p(x, z)p(y, z)}\, dxdydz$$

Its value for the trivariate normal is

$$-\frac{1}{2}\log(1 - \rho_{XY|Z}^2)$$

3.3. Processes

A disadvantage of MI as introduced above is that it is simply a scalar. As consideration turns to the process case, i.e. functions, it seems pertinent to seek to decompose its value somehow.

a) **Time series**

1. *Time-side approach*

The entropy of a process is defined by a suitable passage to the limit for example as

$$\lim_{T\to\infty} E\{\log p(x_1, x_2, \ldots, x_T)\}$$

where $p(x_1,\ldots,x_T)$ denotes the density of order T. To begin one can simply consider the mutual information of the values Y (t +u); Y(t) or of the values Y (t + u), X(t).

This leads to a consideration of the coefficients

$$I_{YY}(u), \text{ and } I_{YX}(u)$$

i.e. mutual information as a function of lag u. References to this idea include: Li [1990] and Granger and Lin [1994].

2. *Frequency side approach*

Similarly it seems worth considering the mutual information at frequency λ of two components of a bivariate stationary time series. This could be defined as the mutual information of $dZ_X(\lambda)$ and $dZ_Y(\lambda)$. Because these variates are complex valued a 4-variate random variable is involved. In the Gaussian case the MI at frequency λ is

$$-\log(1 - |R_{XY}(\lambda)|^2) \quad ,$$

with the overall information rate

$$-\int_{-\pi}^{\pi} \log(1-|R(\omega)|^2)\, d\omega$$

Granger and Hatanaka (1964).

In the general case for each frequency one might construct an estimate, $I_{XY}(\lambda)$, based on kernel estimates of the densities taking empirical FT-values near λ as the data. A difficulty that arises is that the random variables are complex-valued, i.e. the situation is 4-dimensional.

One way to estimate the MI, suggested above, is to fit a parametric model and then to use the loglikelihood ratio test statistic for a test of independence.

A novel way, also being pursued, is to use first recurrence time estimates of entropy Ornstein and Weiss (1993) and Wyner (1999).

b) **Point processes** Definitions of entropy have been set down for this case, namely

$$-E[\ln p(\{N_t : 0 \le t < T\})] = \int_0^T E[\lambda_N(t)(1-\lambda_N(t))]dt$$

where $\lambda_N(t)$ is the conditional intensity function of the process at time t. In the case of a Poisson process λ_N is simply its rate function, Snyder (1975). Daley and Vere-Jones (1988) work from the general expression

$$-\int_{\Omega} \Lambda(\omega)\log\Lambda(\omega)d\omega = E_P(\log\Lambda)$$

where Λ = dP=dμ. They also set down an intriguing expression for entropy involving Janosy densities. Brémaud [1] considers the MI on the interval [0; t] writing

$$I^t_{MN} = E\left[\log \frac{dP_{MN}}{dP_M dP_N}\right]$$

4. NETWORKS

In crude terms a network is a box (or node) and line (or edge) diagram and some of the lines may be directed. Some simple 3- and 4-node networks are shown in Figure 1. In the work here a box corresponds to a random entity, to a random variable, to a time series or to a point process. In studying such models the methods of statistical graphical models provide pertinent methodology. Typically these models are based on conditional distributions. See the books by Edwards (1995), Whittaker (1990), Lauritzen (1996), Cox and Wermuth (1998).

If A, B, C represent nodes a question may be as simple as: Is the structure A $\to$ B $\to$ C (the series case of Figure 2) appropriate or is it better described as (A, B) $\to$ C (the convergent example of Figure 2)? One the other hand the question may be as complicated as: What is the wiring diagram of the brain? See Figure 1.

Figure 1, prepared by Alessandro Villa, represents a schematic of the auditory regions of the brain of a cat. The boxes represent regions containing countless neurons. The arrows represent postulated directions of influence. A basic concern is the strength of connections between the several regions.

References include: Brillinger (1996), Brillinger and Villa (1997) and Dahlhaus **et al**. (1997).

5. RESULTS

5.1. *Aplysia californica*

The data under consideration in this first example are the intervals between the times at which an observed neuron of the sea hare (*Aplysia*) fires. Such data are considered extensively in Bryant **et al**. (1973) and Brillinger **et al**. (1970). In the present case the neuron, L10, is discharging spontaneously. The lowest panel of Figure 3 graphs the series of time intervals between successive spikes.

Many stochastic models for neurons firing spontaneously, Fienberg [1974], Holden [1976], imply that the intervals between firings are independent and identically distributed, i.e. correspond to a renewal process. The intervals' mutual information coefficient as a function of lag is identically 0 for such models.

Figure 3 provides the estimated coefficients of determination and mutual information at delay k. Also presented for these quantities are estimates of the 99 % critical null levels obtained via random permutations of the intervals. One can conclude that there is evidence against the assumption of a renewal process. Values at lag or delay k, for small k, appear related. When one looks at the trace of the data there is evidence for this as well.

The intervals appear approximately unrelated beyond delay 15 say. This is notable for the parameter $I_{YY}(k)$ has much greater implications than $\rho_{YY}(k)^2$. Were the former 0 that would imply that Y(j + k) and Y(j) are statistically independent at lag k, not just uncorrelated.

It is interesting that the top two graphs in Figure 3 are quite similar. Were the series stationary Gaussian,

$$I_{YY}(k) = -\frac{1}{2}\log(1 - |\rho_{YY}(k)^2) \approx \frac{1}{2}(|\rho_{YY}(k)^2)$$

for small $|\rho_{YY}(k)|^2$, so approximate normality is an explanation for the similarity.

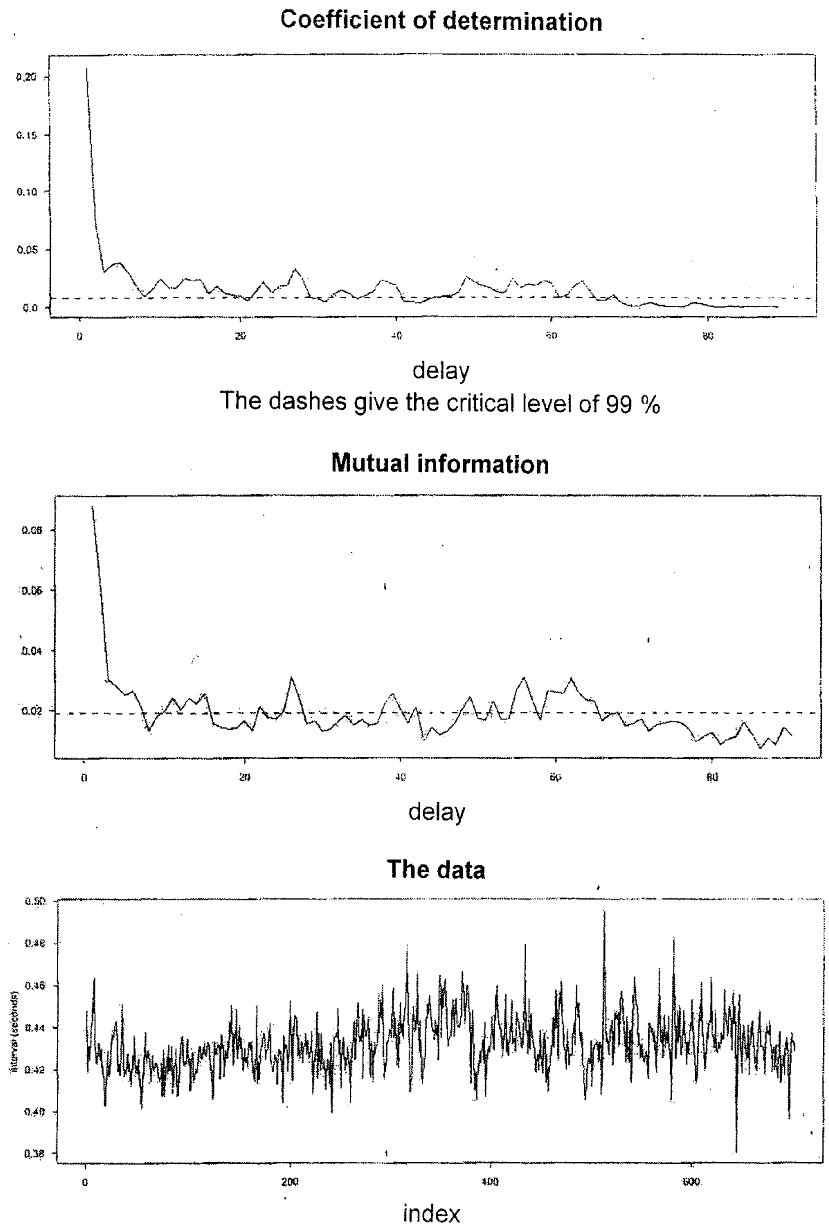

Figure 3. The bottom panel provides the interevent intervals of an *Aplysia* neuron firing spontaneously.
The top panel gives the estimated coefficient of determination as a function of lag.
The middle panel gives the corresponding estimated mutual information.
The dashed line gives the approximate upper 99 % null level.

5.2. Auditory System of the cat

Experiments were carried out in the Department of Physiology at the University of Lausanne to investigate certain aspects of the hearing system of the cat. In brief, neurons relating to the sense of hearing were identified and their firing times were recorded contemporaneously. The data were therefore of multitype point process form.

The neurons were located in different regions of the brain, in particular in the pars magnocell (M) of the medial geniculate body (MGB) and the reticular nucleus of the thalamus (R or RE). These two regions are shown in Figure 1 which provides a schematic for the conjectured way for the various regions of the brain to interact.

Questions posed by the experimenter include: Is there an association between the regions RE and M? If yes, what is its strength? What is its anatomical direction?

Another is: Is any apparent association completely the result of the stimulus?

To address these questions experiments of two different types were carried out. In the first spontaneous discharges of the neurons were recorded, i.e. no external sound stimulus was deliberately applied. In the second there was a sound stimulus lasting for 0.2 second out of each second. There were six electrodes each of which could register the firings of 2 or 3 neurons. There were a number of replicates.

Brillinger and Villa (1997) present the results of partial coherence analyses of the data. What is focused on now is how a model based analysis may be employed to estimate the mutual information between the processes of the regions RE and M.

The neurons will be viewed as a sample from the regions in which they were located. The overall data will be considered a realization of a multitype point process $[N_j(t)]$. As a model of a neuron's firing under the influence of the other neurons, a threshold model is considered. In the case of no stimulus being applied, suppose

$$dN_j(t) = \begin{cases} 1 & \text{if } \sum_{k \neq j} \int_0^{\gamma_{jt}} a_{jk}(t-u)dN_k(u) > \theta_j(t) \\ 0 & \text{if } \quad '' \qquad\qquad '' \qquad < \theta_j(t) \end{cases} \tag{7}$$

where the $\{a_{jk}(\cdot)\}$ represent the influences of the driving neurons of neuron j. The threshold processes, $\theta_j(t)$, are viewed as independent and Gaussian white noises.

Figure 4 provides a sketch of the case of the model for one neuron with the piles representing the times of influencing neurons' firings. The threshold rises to a high level after a firing. This represents the phenomenon of refractoriness. Here the $dN_j(t)$, $j = 1,\ldots, J$ are independent given the history, $H_t = \{N_j(u), u \leq t\}$. The upper limit in the integration, γ_{jt}, is the time elapsed since the neuron j last fired. The model is presented in some detail in Brillinger (1992) and Brillinger and Villa (1997).

To handle the case where the stimulus was applied S(t) is set to 1 when the stimulus is applied and to 0 otherwise. Then (7) is replaced by

$$dN_j(t) = 1 \quad \text{if} \quad \alpha_j S(t) + \sum_{k \neq j} \int_0^{\gamma_{jt}} \alpha_{jk}(t-u)dN_k(u) > \theta_j(t) \tag{8}$$

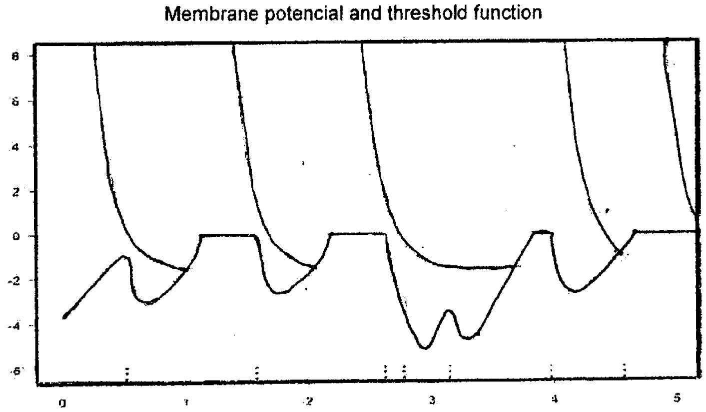

Figure 4. Representation of the firing model of a neuron under the influence of others. Shown are the membrane potential and the deterministic part of the threshold function. Piles correspond to firing times of influencing neurons.

The consideration next is estimating various coefficients of mutual information. This is done by fitting the overall models (7) and (8) and examining the change in the deviance when the α_j and certain of the α_{jk} are set to 0. The results are presented in Figure 5 with the numbers in brackets the estimates degrees of freedom. (Following likelihood ratio theory the null distributions are χ^2_ν / ν.) The estimated strengths of connections with the stimulus appear very strong. This is no surprise. One notes association in both directions between RE and M beyond that imparted by the stimulus. It is interesting though that the connection from M to RE appears stronger than that from RE to M.

6. DISCUSSION AND SECTIONS

The coefficient of mutual information is a unifying concept extending second-order quantities that have restricted applicability. Its being 0 actually implies that the quantities involved are statistically independent. Another important advantage is that the MI pays no real attention to the values of the process. They can be non-negative, integers or proportions for example.

The MI is useful when one wishes to make inferences stronger than: "The hypothesis of independence is rejected." and more of the character "The strength of connection is I."

During the work the plots of the function $I_{YY}(u)$, appeared more useful than the simple scalar I_{YY}. Both parametric model-based estimates and nonparametric estimates of mutual information have been mentioned and computed.

A number of extensions are available and some work is in progress. One can consider the cases of spatial-temporal data, local estimates, of learning, of change, of trend, and of comparative experiments.

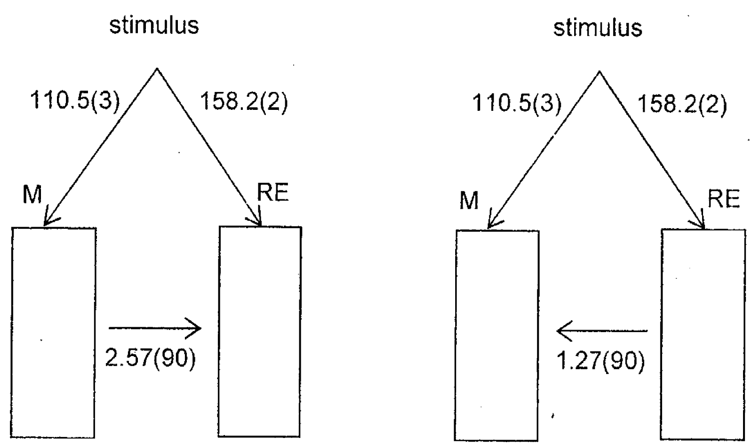

Figure 5. Estimates of the coefficients of mutual information for the links indicated. The integers in (·) are the degrees of freedom of the respective estimates.

One needs to develop the statistical properties of other estimates of MI such as the estimate based on the waiting time and the sieve estimates.

ACKNOWLEDGMENTS

The Aplysia data were collected by Professor J. P. Segundo of UCLA. The cat data were collected by Dr. Alessandro Villa of the University of Lausanne. Dr. Partha Mitra made some stimulating remarks concerning the use of mutual information.

Dr. Phil Spector helped with the preparation of one of the figures. I thank them all.

Part of the material was presented as the Opening Lecture at the XXXIème Journèes de Statistique in Grenoble in 1999.

REFERENCES

BREMAUD, P. (1981): **Point Processes and Queues, Martingale Dynamics**, Springer-Verlag, New York.

BRILLINGER, D.R. (1966): "An extremal property of the conditional expectation", **Biometrika** 53, 594-595.

_______________ (1975): "Estimation of product densities", **Proc. Computer Science and Statistics 8th Annual Symp. at the Interface** 431-438. UCLA.

BRILLINGER, D.R. (1975): **Time Series: Data Analysis and Theory**, Holt-Rinehart, New York. Republished as a SIAM Classic in Applied Mathematics (2001).

_______________ (1992): "Nerve cell spike train data analysis: a progression of technique", **J. American Stat. Assoc**. 87, 260-271.

___________ (1996): "Remarks concerning graphical models for time series and point processes", **Brazilian Review Econometrics** 16, 1-23.

BRILLINGER, D.R.; H.L. BRYANT and J.P. SEGUNDO (1976): "Identification of synaptic interactions", **Biol. Cybernetics** 22, 213-228.

BRILLINGER, D.R. and A.E.P. VILLA (1997): "Assessing connections in networks of biological neurons", 77-92 in **The Practice of Data Analysis**. (Eds. D.R. Brillinger, L.T. Fernholz and S. Morgenthaler). Princeton U. Press, Princeton.

BRYANT, H.L.; A. RUIZ MARCOS and J.P. SEGUNDO (1973): "Correlations of neuronal spike discharges prouced by monosynaptic connections and by common inputs", **J. Neurophysiol**. 36, 205-225.

COX, D.R. and N. WERMUTH (1998): **Multivariate Dependencies: Models, Analysis, and Interpretation**, Chapman & Hall, London.

COVER, T. and J. THOMAS (1991): **Elements of Information Theory**, Wiley, New York.

DAHLHAUS, R., M. EICHLER and J. SANDKHLER (1997): "Identification of synaptic connections in neural ensembles by graphical models", **J. Neuroscience Methods** 77, 93-107.

DALEY, D. and D. VERE-JONES (1988): **An Introduction to the Theory of Point Processes**, Springer-Verlag, New York.

EDWARDS, D. (1995): **Introduction to Graphical Modelling**, Springer, New York.

FERNANDES, M. (2000). Nonparametric entropy-based tests of independence between stochastic processes. Preprint, Fundacao Getulio Vargas, Rio de Janeiro.

FIENBERG, S. (1974): "Stochastic model for single neuron firing trains: a survey", **Biometrics** 30, 399-427.)

GRANGER, C.W.J. and M. HATANAKA (1964): **Spectral Analysis of Economic Time Series**, Princeton University Press, Princeton.

GRANGER, C.W.J. and J.L. LIN (1994): "Using the mutual information coefficient to identify lags in nonlinear models", **J. Time Series Anal**. 15, 371-384.

GRANGER, C. W.J.; E. MAASOUNI and J. RACINE (2000): **A dependence metric for nonlinear time series**, Preprint. Dept. of Economics, UCSD.

HALL, P. and S.C. MORTON (1993): "On the estimation of entropy", **Ann. Inst. Statist. Math**. 45, 69-88.

HOLDEN, A. (1976): **Models of the Stochastic Activity of Neurons**, Springer-Verlag, New York.

HONG, Y. and H. WHITE (2000): **Asymptotic distribution theory for nonparametric entropy measures of serial dependence**, Preprint, Economics Department, UCSD.

JOE, H. (1989). "Estimation of entropy and other functionals of a multivariate density", **Ann. Inst. Statist. Math**., 41, 683-697.

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

LAURITZEN, S. L. (1996): **Graphical Models**, Oxford University Press, Oxford.

LI, W. (1990): "Mutual information functions versus correlation functions", **J. Statistical Physics** 60, 823-837.

ORNSTEIN, D.S. and B. WEISS (1993): "Entropy and data compression schemes", **IEEE Inf. Theory** 39, 78-83.

RAO, C. R. (1965): **Linear Statistical Inference and Its Applications**, Wiley, New York.

SNYDER, D. L. (1975): **Random Point Processes**. Wiley, New York.

WHITTAKER (1990): **Graphical Models in Applied Multivariate Statistics**, Wiley, New York.

WYNER, A. J. (1999): "More on recurrence and waiting times", **Ann. App. Prob**. 9, 780-796.