M. Csörgö, D.A. Dawson, J.N.K. Rao, A.K.Md.E. Saleh (eds.)

SOME ASPECTS OF THE ANALYSIS OF EVOKED RESPONSE EXPERIMENTS

David R. Brillinger[1]

Department of Statistics
University of California
Berkeley, California
U.S.A.

Evoked response experiments provide an important class of situations in which the basic responses recorded are curves. In this paper a variety of modifications, to the usual statistical procedures, are proposed for handling such data. In particular analogs of the mean, the general linear model, robust/resistant estimates, experimental designs, analysis of variance and parametric models are investigated.

1. INTRODUCTION

A traditional, (dating back to Caton in 1875), means of studying the nervous system involves applying sensory stimuli to a subject and examining the ongoing electroencephalogram (EEG) for an evoked response (ER) . The stimulus may be auditory, visual, olfactory (an odour), somatosensory (an electric shock) or gustatory (a taste) in character. Generally the stimulus is applied for a time interval that is brief in comparison to the duration of the response. The response, if one occurs, takes place with a small delay (latency) and perhaps lasts half a second.

A general description of the evoked response technique may be found in Regan (1975). He lists as principal applications: (1) revelation of specific brain activities, (2) provision of an objective indicator of sensory function and (3) distinction of organic disorders from psychogenic ones. The technique is fast and provides an effectively risk free means of testing hearing, vision and spinal cord function that may be applied even to infants.

One specific example of the use of the procedure is related in Bergamini et al (1967). Siamese twins were joined in such a way that it was not possible to determine, by traditional means, if the peripheral nervous pathways were dependent. Before operating to separate the twins it was desired to examine for their independence. Ongoing EEG's were recorded for each twin. A series of experiments were carried out in which the twins legs were stimulated, by electrical shocks, in turn. When a leg of one twin was stimulated, EEG activity was noted only for that twin. On the basis of this information the twins were separated - successfully.

A second example of the use of the ER technique is provided by hearing exams for newborn infants, (including sleeping infants). Ongoing EEG's are recorded. These are examined for responses after loud clicks are made near the infant. Rapin and Graziani (1967) present examples of average evoked responses for an infant with hearing difficulties wearing and not-wearing a hearing aid. It was found that the aid had an objectively measurable effect.

The first and most basic obstacle to making use of the ER procedure is that of seeing the ER's in an EEG. In almost all circumstances the ER's are much smaller than the level of the continuing noise. Dawson (1951) demonstrated that one way to surmount this difficulty was to apply the stimulus periodically at well spaced time points and then to average together the EEG values that occur at the same time lag after the application of the stimulus. Specifically if $Y(t)$ is the observed series and if the stimulus is applied at the times $j\sigma$, $j = 1,\dots,M$ one computes

$$\bar{Y}(u) = \frac{1}{M}\sum_{j=1}^{M} Y(u + j\sigma) \qquad (1)$$

$0 \leq u \leq U$. This statistic is referred to as an average evoked response (AER). The interval width σ is to be taken large enough that neighboring ER's do not interfere with each other.

In fact it turns out that Laplace (1825) had earlier suggested the consideration of sums of values of a series at fixed time lags relative to the times of certain external events. Namely, in Volume 5 of his Traité de Mécanique Celeste he summed the difference between morning and evening low and high tidal heights at lags of −1, 0, 1, 2, 3, 4 days relative to the times of equinoctial syzygies. Another early example of the use of the statistic (1) is provided by the table of Buys-Ballot (1847). Buys-Ballot's concern was the detection of an effect of period σ, so that the values (1) were computed for different σ's as well as different u's. Yet another early example of a related procedure was mentioned by Professor F. N. David. In the late 1800's Galton superposed photographic negatives of faces of criminals, in a search for common features. Doing this may be viewed as analagous to superposing the separate curves $Y(u + j\sigma)$, $0 \leq u \leq U$, of (1) and performing the averaging mentally. The radar memory tube that proved so important in World War II, see Watson-Watt (1946), provides an electronic realization of Galton's procedure.

Other diverse applications of the statistic (1) include: (i) the stacking technique that exploration seismologists employ in combining the seismic traces obtained at nearby locations after letting off a series of shots, (see for example Waters (1978)), (ii) the examination of average hourly rainfall curves by Neyman (1977) in a search for an effect due to cloud seeding and (iii) the aligned activity records of animals prepared by biologists in a search for circadian rhythms (see for example Figure 4.4.1 in Pavlidis (1973)).

The computing of AER's has now become routine with fairly fleeible special purpose computers available for the analysis including the ARC (average response computer) and the CAT (computer of average transients). In a sense these computers may be viewed as providing the $\bar{X}$ button of hand-held calculators for data consisting of curves.

General references to the use and interpretation of ER's include: Donchin and Lindsley (1969), Shagass (1972), John (1977), Thatcher and John (1977), Callaway et al (1978). References going into statistical concerns in some detail are Glaser and Ruchkin (1976) and Freeman (1980).

2. SOME FORMALIZATION

In order to proceed to an investigation of the statistical properties of AER's and related quantities it is necessary to set down some notation and assumptions. It will be assumed that the experiment is not evolving in time, that the noise processes present are stationary stochastic processes and that functional transformations are time invariant.

The times of application of stimuli will be denoted by σ_j, $j = 0, \pm1, \pm2, \dots$ $M(\cdot)$ will denote the corresponding point process with the number of σ_j satisfying $0 < \sigma_j \leq t$ denoted by $M(t)$. The response observed at time t will be denoted by $Y(t)$. The time period of observation will be taken as the interval $(0,T]$. Further, for convenience, we will set $M = M(T)$. The AER

$$\bar{Y}(u) = \frac{1}{M}\sum_{j=1}^{M} Y(u + \sigma_j) = \frac{1}{M}\int_0^T Y(u + t)\, dM(t) \qquad (2)$$

may now be seen to be able to be viewed as an estimate of

$$E\{Y(u + t) \mid \text{stimulus at time } t\} \qquad (3)$$

As such it is not a system invariant, but has the distribution of the σ_j melded into it. Suppose one defines

$$E\{Y(t) \mid \text{single stimulus at } t-u\} = \mu + a(t-u) \qquad (4)$$

with $a(t)$ tending to 0 as t tends to ∞. This function $a(\cdot)$ may be viewed as a system invariant. Supposing that stimuli are applied sufficiently apart in time that their effects do not overlap, then from (4)

$$E\{Y(t) \mid M(u), u \leq t\} = \mu + \sum_{\sigma_j \leq t} a(t - \sigma_j)$$

$$= \mu + \int_{-\infty}^{t} a(t - u)\, dM(u) \qquad (5)$$

This last expression leads to a consideration of the following model for the response series when stimuli of intensity I are applied at times σ_j.

<u>Model 1</u>. Suppose $M(t)$ is a step function jumping by 1 at each σ_j. Suppose μ is a constant and that $a(\cdot)$ is a fixed function vanishing for $t < 0$. Suppose that $\varepsilon(\cdot)$ is a stationary noise process. The response series is given by

$$Y(t) = \mu + I\int_{-\infty}^{\infty} a(t-u)\, dM(u) + \varepsilon(t) \qquad (6)$$

There have been some investigations that suggest that Model 1 is reasonable in certain practical situations. Suppose that the function $a(\cdot)$ vanishes outside the interval $[0,V]$. Suppose that the stimuli are applied farther than V time units apart, then from (6),

$$\bar{Y}(u) = \mu + I\,a(u) + \bar{\varepsilon}(u) \qquad (7)$$

$0 \leq u \leq V$. Under these conditions the AER is seen to be estimating the function $a(\cdot)$ directly. The experiments of Biedenbach and Freeman (1965) may now be seen as examining the linearity in I and the superposability of the σ_j effects required in Model 1. They were concerned with responses evoked in the prepyriform cortex of the cat by olfactory tract stimulation. In some of their experiments stimuli of various intensities were applied at well separated times. The AER's appeared to be linear in the intensity I provided I was above a threshold, but not overly large. In others of their experiments the stimuli times were paired, $\sigma_{2j} - \sigma_{2j-1} = d$, for various d. The AER's obtained were compared with the result of superposing at lag d the AER's obtained from well-separated times. It was found that superposability was a reasonable assumption provided that d was not too small. (The first AER's here were computed assuming the stimulus to be the pair of pulses d units apart.)

3. INVESTIGATION OF THE AER FOR MODEL 1

Under regularity conditions large sample approximations may be derived for the mean, variance and distribution of the AER in the case that Model 1 holds. From expression (6) it is apparent that $\bar{Y}(u)$ is made up of a fixed part, $E\,\bar{Y}(u)$, and a stochastic part, $\bar{\varepsilon}(u)$. These two parts may be discussed separatly. The investigation may be carried through under the conditions of Brillinger (1973). Suppose I = 1.

These conditions include: (i) $|M(t) - M(s)| \leq A + B|t-s|$ for some finite A and B, (ii) $\hat{p} = M(T)/T$ tends to p as T tends to ∞ and (iii) for $u > 0$

$$\hat{P}(u) = \int_0^{T-u} (M(u+v) - M(v))dM(v)/T \tag{8}$$

tends to P(u) as T tends to ∞, for almost all u. These conditions require that the stimulus be applied at fixed, but stationarly distributed, time points. Now one has

$$\begin{aligned} E\,\bar{Y}(u) &= \int_0^T (\mu + \int a(u+v-w)\, dM(w))dM(v) \,/M(T) \\ &= \mu + \int a(u-v)\, d\hat{P}(v) \,/\hat{p} \\ &\rightarrow \mu + \int a(u-v)\, dP(v) \,/p \end{aligned} \tag{9}$$

as T tends to ∞.

Expression (9) is not generally the desired $\mu + a(u)$. The distribution of the times of application of the stimulus has been convolved in. Some particular cases include:
a) Purely periodic. In this case, $\sigma_j = j\sigma$ and so $p = 1/\sigma$,

$$\frac{dP(v)}{dv} = \sum_{j=-\infty}^{\infty} \delta(v - j\sigma)/\sigma$$

with $\delta(\cdot)$ the Dirac delta function, giving

$$E\,\bar{Y}(u) = \mu + \sum_j a(u - j\sigma) \quad . \tag{10}$$

This reduces to $\mu + a(u)$, for $0 \leq u \leq V$, in the case that a(u) vanishes for u outside the interval $[0,V]$ and that $\sigma > V$.
b) Poisson process rate p. Suppose that the times of stimulation are those of a Poisson process, then

$$\frac{dP(v)}{dv} = \delta(v)p + p^2$$

giving

$$E\,\bar{Y}(u) = \mu + a(u) + p\int a(v)dv \quad . \tag{11}$$

This is the desired function up to the level value. The clear advantage of this stimulation procedure is that no restrictions are placed on the support of the function $a(\cdot)$ and the rate of application of the stimulus.
c) Stationary mixing point process. Suppose that $M(\cdot)$ is taken to be a realization of a stochastic point process with rate p and autointensity function p(t), then

$$\frac{dP(v)}{dv} = \delta(v)p + p(v)$$

giving

$$E\,\bar{Y}(u) = \mu + a(u) + \int p(u-v)a(v)dv/p \quad . \tag{12}$$

A deconvolution is seen to be required before arriving at the desired $a(\cdot)$.

The variability of the estimate $\bar{Y}(u)$ is that of the stochastic part $\bar{\varepsilon}(u)$. Suppose that c(t) denotes the autocovariance function and $f(\lambda)$ the power spectrum of the series $\varepsilon(t)$. Then

$$\begin{aligned} \text{var}\,\bar{Y}(u) &= \text{var}\,\bar{\varepsilon}(u) \\ &\sim \int c(v)\, dP(v) \,/Tp^2 \end{aligned} \tag{13}$$

and the estimate is asymptotically normal, as $T \rightarrow \infty$, (using the results of Brillinger (1975).)

In the case of the purely periodic example above expression (13) becomes

$$\text{var}\,\bar{Y}(u) \doteq \sum_k f(\frac{2\pi k}{\sigma}) \,/2\pi T \quad . \tag{14}$$

The AER will have inflated variance when the spectrum of $\varepsilon(\cdot)$ has a peak at frequency $2\pi/\sigma$. Now EEG spectra do have peaks, for example corresponding to alpha rhythm. The experimenter needs to be careful not to chose a stimulus interval corresponding to such a peak.

The problem of variance estimation will be returned to later in the paper. Traditional procedures include splitting the data stretch into a number of equal disjoint segments and viewing the AER's of the segments as independent estimates and the ($\pm$) method based on the difference between the successive responses of pairs of responses (see Schimmel (1967).)

4. USES OF MODEL 1

An advantage resulting from having set down Model 1 is that a variety of hypotheses of scientific interest may be examined in a formal fashion. These include:
i) Is there an evoked response ? This is one of the questions examined in the Rapin and Graziani (1967) paper mentioned earlier. They were concerned with whether or not infants were hearing certain loud clicks. The formal hypothesis here is; is $a(t) = 0$ for all t ?
ii) Do two individuals have the same evoked response ? Lewis et al (1972) were concerned with the degree of similarity of evoked responses for monozygotic twins, dizygotic twins and non-twins. Supposing separate experiments are carried out to estimate $a_1(\cdot)$ and $a_2(\cdot)$ for two individuals. The formal hypothesis may be written; is $a_1(t) = a_2(t)$ for all t ?
iii) Are the evoked responses of an individual with respect to two different stimuli the same ? McCormack (1977) measured the visual responses evoked in an individual when different patterns were presented. This hypothesis may be formalized as in ii).
iv) Are the responses evoked at two symmetrically related locations on the skull the same ? John (1977) is concerned with this question in looking for learning disabilities. In this situation the response recorded is a vector of curves. The component curves may each be modelled as in expression (6). The possibility that the noise processes are correlated now has to be addressed.
v) If stimulus A results in a response and stimulus B results in another response is the result of simultaneously applying stimuli A and B the sum of those two responses ? Diamond (1964) is concerned with this question in the case that A is a flash of red light and B a flash of blue light. To examine this question the model (6) must be expanded to include point processes corresponding to each type of stimulus.

vi) Are the effects of the individual stimulus applications superposable or are there interactions (nonlinearities) ? Biedenbach and Freeman (1965) examined this hypothesis in one situation. They found interaction effects in the case that the σ_j were close together.

5. A FORMAL (LINEAR) APPROACH

Suppose that r separate stimuli are available. Let the counting function $M_k(t)$ provide the application times of the k-th of these stimuli. Collect the $M_k(t)$ into an r vector, $\underset{\sim}{M}(t)$. Consider the model

$$\underset{\sim}{Y}(t) = \underset{\sim}{\mu} + \int \underset{\sim}{a}(t-u)\, d\underset{\sim}{M}(u) + \underset{\sim}{\varepsilon}(t) \qquad (15)$$

with $\underset{\sim}{Y}(t)$ a vector of s response series, $\underset{\sim}{\mu}$ an s vector, $\underset{\sim}{a}(t)$ an $s \times r$ matrix (with the entry in row j and column k providing the effect of the k-th stimulus on the j-th series), and with $\underset{\sim}{\varepsilon}(t)$ a vector of s noise series.

Given data $\underset{\sim}{Y}(t)$, $\underset{\sim}{M}(t)$ $0 \leq t \leq T$, with T sufficiently large, the parameters of the model (15) may be estimated and hypotheses concerning them examined. In this connection it is easier to proceed in the frequency domain. To this end define

$$\underset{\sim}{A}(\lambda) = \int_{-\infty}^{\infty} \exp(-i\lambda t)\, \underset{\sim}{a}(t)\, dt \qquad (16)$$

$$\underset{\sim}{d}_Y^T(\lambda) = \int_0^T \exp(-i\lambda t)\, \underset{\sim}{Y}(t)\, dt \qquad (17)$$

$$\underset{\sim}{d}_M^T(\lambda) = \int_0^T \exp(-i\lambda t)\, d\underset{\sim}{M}(t) \quad . \qquad (18)$$

The transfer function $\underset{\sim}{A}(\lambda)$ may now be estimated by

$$\hat{\underset{\sim}{A}}(\lambda) = \left(\sum_j \underset{\sim}{d}_Y^T\left(\tfrac{2\pi j}{T}\right) \overline{\underset{\sim}{d}_M^T\left(\tfrac{2\pi j}{T}\right)}^{\tau} \right)\left(\sum_j \underset{\sim}{d}_M^T\left(\tfrac{2\pi j}{T}\right) \overline{\underset{\sim}{d}_M^T\left(\tfrac{2\pi j}{T}\right)}^{\tau} \right)^{-1} \qquad (19)$$

with the sums in (19) over n distinct frequencies $2\pi j/T$ near λ .

Suppose next that, among other things, the noise process $\underset{\sim}{\varepsilon}(\cdot)$ is stationary, has mean $\underset{\sim}{0}$, spectral density matrix $\underset{\sim}{f}(\lambda)$ and satisfies, for example, the mixing assumption of Brillinger (1974). Then, for $\lambda \neq 0$, π the estimate (19) may be shown to be asymptotically complex normal with mean $\underset{\sim}{A}(\lambda)$ and with vec $\hat{\underset{\sim}{A}}(\lambda)$ having covariance matrix

$$2\pi T\, \underset{\sim}{f}(\lambda) \otimes \left(\sum_j \underset{\sim}{d}_M^T\left(\tfrac{2\pi j}{T}\right) \overline{\underset{\sim}{d}_M^T\left(\tfrac{2\pi j}{T}\right)}^{\tau} \right)^{-1} \quad . \qquad (20)$$

Further estimates at distinct frequencies are asymptotically independent.

Various hypotheses of scientific interest were indicated in the previous section. In the cases i) – v) these may be written out as hypotheses concerning the entries of the matrix $\underset{\sim}{A}(\lambda)$. With the approximate distribution for $\hat{\underset{\sim}{A}}(\lambda)$ indicated above, a general linear hypothesis concerning the matrix $\underset{\sim}{A}(\lambda)$ may be examined in an analysis of variance fashion.

If desired, the function $\underset{\sim}{a}(u)$ may be estimated by an expression of the form

$$\hat{\underset{\sim}{a}}(u) = Q^{-1} \sum_{q=0}^{Q-1} \exp\left(\tfrac{2\pi u q}{Q}\right) \hat{\underset{\sim}{A}}\left(\tfrac{2\pi q}{Q}\right) \quad . \qquad (21)$$

Advantages of setting down the model (15) are now seen to include: 1) different sorts of stimuli (including steady ones) may be handled with no greater difficulty than single ones, 2) randomization analyses (where the stimulus applied is selected randomly from those available) may be developed, 3) experimental designs (eg. cross-over, systematic, rotation, repeated measurement) may be incorporated, 4) analysis of variation may be formalized, 5) best linear unbiased estimates may be constructed, approximately. (This approach was introduced in Brillinger (1978).)

6. AN ALTERNATIVE VIEWPOINT

Consider once again the single stimulus, single response series case. Suppose that the function a(u) vanishes outside the interval $[0,V]$ and that the times of application of the stimulus are more than V units apart. Suppose, further, that the autocovariance function of the noise series is essentially zero after lag V . Then the linear model may be written

$$\begin{aligned} Y_j(u) &= Y(u+\sigma_j) \\ &= \mu + a(u) + \varepsilon_j(u) \end{aligned} \qquad (22)$$

for $0 \leq u \leq V$, $j = 1,\dots,M(T)$, with the noise process $\varepsilon_j(u) = Y(u+\sigma_j)$ uncorrelated with $\varepsilon_k(v)$, $j \neq k$. The AER $\bar{Y}(u)$ is simply the mean of the separate responses. The model (22) is now seen to be the linear model of multivariate analysis and the model of growth curves with replicated observations.

Adopting this viewpoint means then that one can take over the whole apparatus and procedures from those fields. For example, a situation where a number of distinct stimuli are applied may be modelled by

$$Y_{ij}(u) = \mu_{ij} + a(u) + b_i(u) + \varepsilon_{ij}(u) \qquad (23)$$

with i indexing the various stimuli, j indexing the replicates, a(·) denoting an overall effect and $b_i(\cdot)$ providing the effect due to the i-th stimulus. The results that have been developed for MANOVA of complex experimental designs may be taken over directly.

Brillinger (1980) is a review paper on the analysis of variance of curves in the case that the noise process, ($\varepsilon_j(\cdot)$ in (22)), is stationary.

7. SUPERPOSABILITY

The next two sections examine certain effects resulting from a departure from the assumptions of Model 1 and (15). These models were motivated by a consideration of expression (3). In the superposable case it turned out to be enough to discuss the expected value (3) in the case of a single stimulus time, namely to consider expression (4).

It seems natural to move on to a consideration of characteristics like

$$E\{Y(u+t) \mid \text{stimulus at times } t,\ t-v\} \quad . \qquad (24)$$

This latter may be estimated by an expression of the form

$$(2\beta T)^{-1} \sum_{j \neq k} Y(u+\sigma_j) \left\{ |\sigma_j - \sigma_k - v| < \beta \right\} \quad . \qquad (25)$$

with β small and $\{E\}$ here defined to be 1 if the event E is true and 0 otherwise.

This statistic crops up, for example, in a consideration of;
Model 2. Suppose $M(t)$ is a step function jumping by 1 at each σ_j. Suppose that μ is a constant and that $a(\cdot)$, $b(\cdot,\cdot)$ are fixed functions.[3] Suppose that $\varepsilon(\cdot)$ is a stationary noise process. The response series is given by

$$Y(t) = \mu + \int_{-\infty}^{\infty} a(t-u)\, dM(u) + \iint_{u \neq v} b(t-u,t-v)\, dM(u)\, dM(v) + \varepsilon(t) \quad . \qquad (26)$$

This model allows an interaction between pairs of stimuli effects. If the stimulus process consists of an impulse at times 0 and d alone then (26) gives

$$Y(u) = \mu + a(u) + a(u-d) + b(u,u-d) + \varepsilon(t) \qquad (27)$$

for $u \geq d$. In the causal case $a(u)$, $b(u,v)$ will vanish for $u < 0$, $v < 0$ and (27) will hold for all u.

It is clear from expression (27) may be examined via two-pulse experiments. The computations are also direct in the case that the times of stimulus application are those of a Poisson process, (see Krausz (1975).)

8. ROBUST/RESISTANT ESTIMATES

These days research workers are very much schooled in the sensitivity of the mean to outliers. Now large transients, that are artifacts, occur commonly in EEG's – movement, eye blink, EMG (muscle electromyograph), barbituate spikes, mains pulses all occur. These all affect the AER.

To deal with this problem some researchers have computed the median response at each lag, see for example Figure 2.22 in Rosenblith (1962). Some statistical properties of this estimate have been derived, see Section 4.7 in Glaser and Ruchkin (1976). However, being nonlinear the median computation can exhibit spurious harmonics of the noise components and behave in nonelementary fashions, a case in point is illustrated in Figure 1 of Ruchkin and Walter (1975). It is clear that various trimmed means could be used in place of the median.

The median and the trimmed mean just mentioned operate separately at each time point. It seems worthwhile to develop an estimate that weights, (possibly rejects), whole ER curves differentially.

Suppose that the notation of the Alternate Viewpoint section is adopted. Let $\hat{\theta}(u)$ denote the estimate, about to be constructed, at lag u. Let $\hat{\rho}$ denote an estimate of scale and $\| Y - \theta \|$ a measure of distance, for example,

$$\| Y - \theta \|^2 = \int_0^V (Y(u) - \theta(u))^2 du \quad . \qquad (28)$$

Set

$$\hat{\theta}(u) = \sum_j W_j Y_j(u) / \sum_j W_j \qquad (29)$$

with

$$W_j = W(\| Y_j - \hat{\theta} \| / \hat{\rho}) \qquad (30)$$

$W(\cdot)$ being a non-negative weight function. (An example will be given shortly.)

Both sides of equation (29) involve the desired estimate $\hat{\theta}(\cdot)$. In practice an iterative procedure will be set up, based on an initial estimate, (perhaps the AER), with iterations carried out until the current estimate is changing little. (See below.)

As a specific example one can consider the trimmed mean

$$\hat{\theta}(u) = \Sigma' Y_j(u) / \beta M \qquad (31)$$

with Σ' demoting summation over the βM smallest $\| Y_j - \hat{\theta} \|$.

By analogy with equations (2.11), (2.12) of Huber (1977) one is led to consider parameters satisfying the relationships

$$E\{(Y(u) - \theta(u))\, W(\| Y(u) - \theta(u) \| / \rho)\} = 0 \qquad (32)$$

$$E\{U(\| Y - \theta \| / \rho)\, \| Y - \theta \|^2 / \rho^2 - V(\| Y - \theta \| / \rho)\} = 0 \qquad (33)$$

The choice

$$W(u) = (1 - u^2)^2 \quad 0 \leq u \leq 1; \qquad = 0 \quad \text{otherwise} \qquad (34)$$

gives the biweight. The choice

$$U(u) = a^2 \quad 0 \leq u \leq a; \qquad = u^2 \quad a \leq u \leq b; \qquad = b^2 \quad b \leq u \qquad (35)$$

and $V(u) = 1$ seems useful in practice.

If all of the data is available in a computer at the same time then the estimate (29) may be computed in an iterative fashion as follows;

$$\hat{\theta}_{k+1}(u) = \sum_j W_{j,k} Y_j(u) / \sum_j W_{j,k} \qquad (36)$$

$$\hat{\rho}^2_{k+1} = \sum_j U_{j,k} \| Y_j - \hat{\theta}_k \|^2 / \sum_j V_{j,k} \qquad (37)$$

$$W_{j,k} = W(\| Y_j - \hat{\theta}_k \| / \hat{\rho}_k) \qquad (38)$$

with $U_{j,k}$ and $V_{j,k}$ defined by expressions similar to (38).

9. A RECURSIVE PROCEDURE

The estimate introduced in the previous section has the disadvantage of requiring the experimenter to retain all the data. A recursive procedure requiring only the most recent estimate and the just-collected response will now be set down. The procedure is motivated by stochastic approximation and the known-scale location estimate of Martin and Masreliez (1975). Let $\hat{\theta}_j$ denote the estimate based on $Y_1, \ldots, Y_j$. Then set

$$\hat{\theta}_{j+1}(u) = \hat{\theta}_j(u) - \frac{L}{j} W(\| Y_{j+1} - \hat{\theta}_j \| / \hat{\rho}_j)\, (Y_{j+1}(u) - \hat{\theta}_j(u)) \qquad (39)$$

$$\hat{\rho}_{j+1} = \hat{\rho}_j - \frac{L}{j}\left(U(\| Y_{j+1} - \hat{\theta}_j \| / \hat{\rho}_j)\, \| Y_{j+1} - \hat{\theta}_j \|^2 / \hat{\rho}_j^2 - V(\| Y_{j+1} - \hat{\theta}_j \| / \hat{\rho}_j) \right) \qquad (40)$$

for some constants L .

The estimate of (39) will not be the same as (29), even when all the data has been collected; however the two estimates will converge to the same value, (given by (32)) as T tends to ∞ .

In the case of the 100β per cent trimmed mean the recursive equations become, for the choice L = 1/β in the first

$$\hat{\theta}_{j+1}(u) = \hat{\theta}_j(u) + \frac{L}{\beta j}(Y_{j+1}(u) - \hat{\theta}_j(u)) \quad \text{if } \|Y_{j+1} - \hat{\theta}_j\| \leq \hat{\rho}_j \tag{41}$$

$$= \hat{\theta}_j(u) \quad \text{otherwise}$$

$$\hat{\rho}_{j+1} = \hat{\rho}_j - \frac{L}{j}\left(\frac{1}{\beta} - 1\right) \quad \text{if } \|Y_{j+1} - \hat{\theta}_j\| \leq \hat{\rho}_j \tag{42}$$

$$= \hat{\rho}_j + \frac{L}{\beta j} \quad \text{otherwise}$$

The effect of employing the estimate (41) is to exclude from the averaging any (complete) evoked response that deviates substantially from the bulk of the responses. The average response computers mentioned earlier typically have circuits to detect signals above an arbitrarily established amplitude at some lag u . This procedure has the disadvantage of not rejecting responses that are not quite abnormal at any u, but that over all lags are quite abnormal. The estimates (29) and (39) are of multivariate nature.

10. RESISTANCE (FREQUENCY DOMAIN)

The above discussion refers to time domain procedures. As the example of Ruchkin and Walter (1975) shows, such procedures can have undesired effects on quasi-sinusoidal signals or noise. Some frequency-based procedures are now presented for constructing resistant estimates - estimates not substantially affected by frequency domain abnormalities.

Suppose that the various ER's are separated in time so that the data may be described by expression (22) . In the frequency domain this leads to

$$d^V_{Y_j}(\lambda) = \int_0^V e^{-i\lambda u}\, Y_j(u)\, du \tag{43}$$

$$\doteq A(\lambda) + d^V_{\varepsilon_j}(\lambda) \qquad \lambda \neq 0 \tag{44}$$

$$= a_j + ib_j = C_j \exp\{i\delta_j\} \tag{45}$$

$j = 1,\ldots,M$. In expression (45) here, dependence on λ has been suppressed in the notation. Forming the AER corresponds to averaging the a_j, b_j with respect to j .

One means of detecting abnormal values is to plot, say for $\lambda = 2\pi v/V$, $v = 0,1,\ldots, V/2$, the points (a_j,b_j), $j = 1,\ldots,M$. The point corresponding to the AER will sit in the middle of a cloud of points.

It is now apparent that one might proceed by applying scalar procedures to the individual components of (a_j,b_j) or (C_j,δ_j) or $(\log C_j,\delta_j)$. Alternatively, taking note of the Tukey (1980) procedure for a single series one might; (a) apply scalar resistance methods to δ_j , (or $\exp\{i\delta_j\}$) with result ϕ (or $d_j\exp\{i\phi_j\}$), (b) apply scalar methods to $\log C_j$ with result $\log C^*_j$, (c) take the average of the $C^*_j\exp\{i\phi_j\}$, (or of the $C^*_j d_j \exp\{i\phi_j\}$) as the estimate at frequency λ .

11. A PARTIALLY PARAMETRIC MODEL

Suppose that the model (22) is replaced by

$$Y_j(u) = \mu_j + I_j S(u + \gamma_j) + \varepsilon_j(u) \tag{46}$$

for $0 \leq u \leq V$ and $j = 1,\ldots,M$. That is, it is allowed that the response not occur always at the same time lag (latency) after the application of the stimulus and it is allowed that the intensity of the response is not always the same. One hence has the problem of estimating the γ_j and I_j as well as the desired signal S(.) .

These parameters may be estimated via a frequency domain procedure. To this end set

$$Y_{jv} = d^V_{Y_j}\left(\frac{2\pi v}{V}\right), \qquad S_v = d^V_S\left(\frac{2\pi v}{V}\right) \tag{47}$$

with a similar definition of ε_{jv} . Then expression (46) yields

$$Y_{jv} = S_v I_j \exp\{i\gamma_j 2\pi v/V\} + \varepsilon_{jv} \tag{48}$$

$v \neq 0$. Now the ε_{jv} may often be treated as independent complex normal variates with mean 0 and variance $2\pi V f(2\pi v/V)$. This last suggests setting down the (approximate) negative log-likelihood

$$\sum_j \sum_v \left(\log f\left(\frac{2\pi v}{V}\right) + |Y_{jv} - S_v I_j \exp\{i\gamma_j 2\pi v/V\}|^2 / 2\pi V f\left(\frac{2\pi v}{V}\right) \right) , \tag{49}$$

and the estimation of the unknown parameters by the minimization of (49) . In order that the model be identifiable constraints, such as,

$$\sum_j \gamma_j = 0 , \quad \sum_j I_j^2 = 1 \tag{50}$$

will have to be introduced. If an iterative procedure is employed in this minimization, then initial values for I_j, γ_j might be obtained by crosscorrelating $Y_j(u)$ with one of the resistant estimates indicating earlier.

In the case that the noise spectrum f(.) is constant the Woody (1967) adaptive filter may be seen to be one means of seeking the minimum of expression (49). The above formulation is seen to provide a maximum likelihood interpretation of Woody's procedure and an extension of it to handle autocorrelated noise.

12. A FULLY PARAMETRIC MODEL

Freeman (1975) has made substantial use of the following parametric model,

$$a(u|\theta) = \sum_k \alpha_k \exp\{-\beta_k u\} \cos(\gamma_k u + \delta_k) \qquad \text{for } u \geq u_0$$

$$= 0 \qquad \text{for } u < u_0 \tag{51}$$

with θ denoting the parameters . The model may be fit to each individual response or to the AER itself. If u_0 is known to sufficient accuracy that only $Y_j(u)$ values with $u > u_0$ are employed, then the fitting is simplified. The fitting may be carried through in either the time domain or the frequency domain. In the case that the error series $\varepsilon(\cdot)$ is stationary it seems simplest to work in the frequency domain. This is done in Bolt and Brillinger (1979) where the computations required are laid out and the asymptotic distributions of the estimates developed.

Freeman (1975) makes use of the model (51) as it represents the response for linear differential equations. In one case involving two damped sine waves, he views the larger wave as representing intracortical negative feedback and the smaller as representing other feedback loops.

13. REFERENCES

[1] Bergamini, L., Bergamasco, B., Fra, L., Gandiglio, G. and Mutani, R., Diagnostic recording of somato-sensory cortical evoked potentials in man, Electroenceph. clin. Neurophysiol. 22 (1967) 260-262.

[2] Biedenbach, M.A. and Freeman, W.J., Linear domain of potentials for the prepyriform cortex with respect to stimulus parameters, Experimental Neurology 11 (1965) 400-417.

[3] Bolt, B.A. and Brillinger, D.R., Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series, Geophys. J. r. astr. Soc. 59 (1979) 595-603.

[4] Brillinger, D.R., Estimation of the mean of a stationary time series by sampling, J. Appl. Prob. 10 (1973) 419-431.

[5] Brillinger, D.R., The Fourier analysis of stationary processes, Proc. IEEE 62 (1974) 1628-1643.

[6] Brillinger, D. R., A note on the estimation of evoked response, Biol. Cybernetics 31 (1978) 141-144.

[7] Brillinger, D. R., Analysis of variance and problems under time series models, in Krishnaiah, P.R. (ed.), Handbook of Statistics, Vol. 1 (North-Holland, Amsterdam, 1980).

[8] Buys-Ballot, C.H.D., Les Changements Periodiques de Temperature, Utrect (1847).

[9] Callaway, E., Tueting, P. and Koslow, S.H., Event-Related Brain Potentials in Man (Academic, New York, 1978).

[10] Dawson, G.D., A summation technique for detecting small signals in a large irregular background, J. Physiol. 115 (1951) 2.

[11] Diamond, S. P., Input-output relations, Ann. New York Acad. Sci. 112 (1964) 160-171.

[12] Donchin, E. and Lindsley, D. B., Average Evoked Potentials (NASA, Washington, 1969).

[13] Freeman, W.J., Mass Action in the Nervous System (Academic, New York, 1975).

[14] Freeman, W.J., Measurement of cortical evoked potentials by decomposition of their wave forms, J. Cybernetics and Inf. Sci. in press (1980).

[15] Glaser, E.M. and Ruchkin, D.S., Principles of Neurobiological Signal Analysis (Academic, New York, 1976).

[16] Huber, P.J., Robust covariances, in: Gupta, S.S. and Moore, D.S. (eds.) Statistical Decision Theory and Related Topics II (Academic, New York, 1977).

[17] John, E.R., Neurometrics (Wiley, New York, 1977).

[18] Krausz, H.I., Identification of nonlinear systems using random impulse train inputs, Biol. Cybernetics 19 (1975) 217-230.

[19] Laplace, M., Traite de Mecanique Celeste 5 (Bachelier, Paris, 1825).

[20] Lewis, E.G., Dustman, R.E. and Beck, E.C., Evoked response similarity in monozygotic, dizygotic and unrelated individuals: a comparative study, Electroenceph. clin. Neurophysiol. 23 (1972) 309-316.

[21] Martin, R.D. and Masreliez, C.J., Robust estimation via stochastic approximation, IEEE Trans. Inf. Theory 21 (1975) 263-271.

[22] McCormack, G.L., The Albino Visual Pathway Defect and Human Strabismus, Ph.D. Thesis, School of Optometry, University of California, Berkeley (1977).

[23] Neitzel, E.B., Seismic reflection records obtained by dropping a weight, Geophysics 23 (1958) 58-80.

[24] Neyman, J., A statistician's view of weather modification technology (a review), Proc. Nat. Acad. Sci. USA 74 (1977) 4714-4721.

[25] Pavlidis, T., Biological Oscillators: Their Mathematical Analysis (Academic, New York, 1973).

[26] Rapin, I. and Graziani, L.J., Auditory-evoked responses in mormal, brain-damaged and deaf infants, Neurology 17 (1967) 881-891.

[27] Regan, D., Electrical responses evoked from the human brain, Sci. Amer. 242(1979) 134-146.

[28] Rosenblith, W.A., Processing Neuroelectric Data (MIT, Cambridge, 1962).

[29] Ruchkin, D.S. and Walter, D.O., A shortcoming of the median evoked response, IEEE Trans. Biomed. Eng. 22 (1975) 245.

[30] Schimmel, H., The (±) reference: accuracy of estimated mean components in average response studies, Science 157 (1967) 92-93.

[31] Shagass, C., Evoked Brain Potentials in Psychiatry (Plenum, New York,1972).

[32] Thatcher, R.W. and John, E.R., Foundations of Cognitive Processes (Wiley, New York, 1977).

[33] Tukey, J.W., Can we predict where "time series" should go next?, in: Brillinger, D.R. and Tiao, G.C. (eds.), Directions in Time Series (Institute of Mathematical Statistics, Hayward, 1980).

[34] Waters, K.H., Reflection Seismology (Wiley, New York, 1978).

[35] Watson-Watt, R., The evolution of radar location, J. Inst. Elec. Eng. 93 (1946) 6-20.

[36] Woody, C.D., Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals, Med. and Biol. Eng. 5 (1967) 539-553.

-0-0-0-

[1]Prepared with the support of the National Science Foundation Grant MCS-772986.

STATISTICS AND RELATED TOPICS
M. Csörgö, D.A. Dawson, J.N.K. Rao, A.K.Md.E. Saleh (eds.)

MEAN RESIDUAL LIFE

W. J. Hall and Jon A. Wellner

Statistics Department
University of Rochester
Rochester, New York
U.S.A.

The mean residual life function e at age x is defined to be the expected remaining life given survival to age x; it is a function of interest in actuarial studies, survivorship analysis, and reliability. Here we characterize those functions which can arise as mean residual life functions, present and study an "inversion formula" which expresses the survival function in terms of e, and collect a variety of facts about e and other residual moments: inequalities for e, new characterizations of the exponential distribution, inequalities for coefficients of variation, and limiting behavior of e at 'great age'. We also discuss applications to parametric modelling.

1. INTRODUCTION

Let X be a non-negative random variable with right continuous distribution function (df) F, and survival function $\overline{F} = 1 - F$, on $\mathbb{R}^+$ and suppose that $F(0) = 0$ and $\mu \equiv E(X) = \int_0^\infty x dF(x) = \int_0^\infty \overline{F}(x)dx < \infty$; write $T \equiv T_F \equiv \inf\{x: F(x) = 1\} \leq \infty$. The mean residual life (MRL) function or remaining life expectancy function at age x is defined as

(1.1) $$e(x) = e_F(x) \equiv E(X-x|X>x) = \int_x^\infty \overline{F}dI/\overline{F}(x), \qquad \text{for } x\geq 0,$$

and $e(x) \equiv 0$ whenever $\overline{F}(x) = 0$. We use I to denote the identity function and Lebesgue measure on $\mathbb{R}^+$.

The discretized version of the MRL function e has had considerable use in life table analysis (see e.g. Chiang, 1968, pages 189 and 213-214; Gross and Clark, 1975, page 25ff), and estimation of $e = e_F$ on the basis of samples from F has

AMS 1970 Subject classifications: Primary 62P05, 62N05, 62E10; Secondary 62G99
Key words and phrases: life expectancy, characterizations, residual variance, coefficient of variation, new better than used in expectation, renewal theory, regular variation, inequalities.
Research of the second author supported by National Science Foundation Grant MCS 77-02255.