ASSESSING MECHANISMS OF NEURAL SYNAPTIC ACTIVITY

Guoliang Cao
ISDS, Duke University
&
Mike West
ISDS, Duke University

DP# 91-A18
Assessing Mechanisms of Neural Synaptic Activity

By

MIKE WEST & GUOLIANG CAO

Institute of Statistics & Decision Sciences
Duke University
Durham, NC 27706. USA

---


This work is part of a recently initiated project with neuroscientist Dr Dennis Turner of the Department of Neurosurgery at Duke University Medical Center. Guoliang Cao, graduate student in ISDS, has been partially supported by the Department of Neurosurgery. The research of Mike West was partially financed by the National Science Foundation under grants DMS-8903842 and DMS-9024793. The authors may be reached via email at mw@isds.duke.edu and cgl@isds.duke.edu
ABSTRACT
Masses of data relating to the synaptic activity of stimulated nerve tissues is routinely generated in neurophysiological experiments designed to investigate the stochastic behaviour of neurons, individually and collectively. Common experiments result in observed time series measuring evoked neural responses following various levels of electrical stimulus of the nerve tissue, and analyses typically focus on simple numerical summaries, such as estimates of maximum levels of response under given stimuli. Issues of wide interest in the neurosciences are being addressed using representations of such data using mixtures of uncertain numbers of normal distributions. Focuses of analysis are on mixture deconvolution, and on the numbers and weights of components. One approach currently under study uses Dirichlet process mixtures of normals. Some of the scientific issues, together with technical aspects of data analysis, modelling, and the use of prior information, are described and exemplified in this paper.

Keywords: Deconvolution of mixtures; Mixtures of normal distributions; Neural response mechanisms
1. Neurological Background

The mechanisms driving neuronal activity in nervous systems are inherently stochastic. Explaining and predicting the nature of such mechanisms is central to neurological issues of wide import, especially in contributing to our understanding of nervous disorders and diseases. Physiological studies of nervous systems designed to provide insight into the nature of such mechanisms typically focus on measuring synaptic transmission signals at neural, neuromuscular or glandular junctions in response to electrical stimuli, either naturally occurring or experimentally induced. Commonly performed experiments result in observed time series measuring evoked neural responses following various levels of chemically induced electrical stimulus of nerve tissue, and analyses typically focus on simple numerical summaries, such as estimates of maximum levels of resulting response under given stimuli. With respect to a specific neural region, variations in the levels of electrical response may be used to assess physiological theories of neural mechanism, and to predict neural function independently of the support for such theories.

The probabilistic nature of the mechanisms of chemical transmission across synaptic junctions in the mammalian (and other) central nervous system has been the subject of acute study in recent years. Many individual neurons may be involved in determining transmission at any given junction, and their collective response to varying levels of input stimulus determines the overall response of the nerve tissue. A single nerve axon contacts a neural, muscular or glandular cell at a synapse. Electrical potential, either spontaneously occurring in the neural terminal or induced through chemical changes, promotes release of chemical transmitter from synaptic vesicles. This transmitter causes a conductance change in the postsynaptic cellular membrane and hence a change in the electrical potential of the membrane. These evoked synaptic responses in electrical potential may be measured via intracellular probes. With respect to the mechanisms governing overall neural responses, there are two basic and key features: (i) the determined levels of transmitter release, measured through the evoked response, to a given electrical stimulus of individual neurons or synaptic vesicles, and (ii) the numbers of synaptic vesicles responding to any given stimulus.

On the first of these, it is widely posited that transmitter release occurs in basic, quantal units. If so, overall responses will be in terms of multiples of a base or quantal unit, and experimental evidence may be used, in principle, to assess this quantal hypothesis. Of subsequent interest are the possible patterns of variation in quantal size with respect to alterations in synaptic transmission characteristics – naturally occurring alterations in transmission characteristics are, for example, associated with variation in recent levels of stimulus and with disorders and diseases of the nervous system. On the second issue of numbers of synaptic release sites, various simple stochastic models have been entertained, notable among which is the obvious binomial hypothesis. Here it is
hypothesised that any given neuromuscular site has a fixed number of transmitter release sites that may respond to stimulus at some receptor(s), that the individual vesicles release transmitter independently, and that there is a common chance of release across vesicles. Of subsequent interest here, too, is variation in the release probability with respect to changes in the input stimulus and transmission characteristics. The simple quantal/binomial explanation has recently been the subject of extensive scrutiny amongst neurophysiologists, and seems to be increasingly viewed as a grossly oversimplified description of nerve functioning. Direct generalisations to varying release probabilities and varying levels of transmitter clearly provide a much more realistic, flexible and unconstrained class of statistical models.

An archetype experiment designed to investigate these issues involves recording the changes in potential of postsynaptic membranes across a range of levels of electrical stimuli induced by small variations in ionic makeup of the neural environment, including no stimulus. Typically, a section of tissue identified and isolated for experimentation is subject to repeat stimulation with concurrent intracellular recording of postsynaptic response potentials. Stimuli based on square wave-form currents evoke responses in terms of postsynaptic electrical potentials that typically exhibit an initial and rapid rise, then slow levelling-off followed by exponential decay to base or background potentials. There are often significant fluctuations in background potentials that obscure the assessment of response characteristics – responses to spontaneous transmitter release at, typically, low (quantal?) levels. Recordings with no direct stimulus are made to provide prior information about base levels of background transmission that may be incorporated into inferences about the effects of non-zero stimulus. These effects are measured in several ways, one key measure being the maximum level of response attained for a given input stimulus. Figure 1 schematically displays digitised response recordings pre-stimulus (noise alone due to spontaneous neural activity) and post-stimulus (evoked signal plus noise); the spike just after 50 milliseconds is characteristic of the immediate reactive response at the time of stimulus. The time span of about 50 milliseconds for the pre- and post-stimulus periods is typical, the stimulus response decaying essentially completely towards the end of the post-stimulus recording. Note the typical zero-mean levels of background variation pre-stimulus. The peak levels of stimulus response vary widely across tissue type and according to levels of controlled stimulation. Estimated peak levels in a variety of such data series available to the authors range from (commonly) low values in the 1–2 millivolts ranges, to much higher values around 9–12 millivolts, for example.

A single experiment will be set up to record postsynaptic responses at a given neural, neuromuscular or neuroglandular junction, the recordings being replicated possibly several hundred (or more) times across ranges of input electrical stimulus. Thus, at a single junction, recorded data consists of possibly many replications of the sort of time
series exhibited in Figure 1. Here we are concerned with analysis of the resulting data without regard to the input levels of stimulus. This is partly a reflection of the early stage of development of this work, and is clearly a restriction since interesting neurophysiological issues arise in connection with changing patterns of response across varying levels of stimulus, and these will be properly addressed using statistical models that incorporate the collateral information about stimulus levels. A second, and over-riding reason for the restricted focus at this stage is that particular forms of analysis of response data, elaborated below, have become popular and accepted in the neurophysiology literature in recent years, and our initial aim is to assess the scientific issues addressed in these approaches using the same data reduction but alternative (Bayesian) concepts and models. In beginning to develop novel (Bayesian) views of the scientific issues in a field where specific models and methods of (non-Bayesian) statistics are routinely applied, and more or less accepted as norms, it is often wise to initiate development by building on and adapting existing approaches, and to identify potential for improving data analysis and inference without the appearance of radically changing established protocol. Once such potential is identified and capitalised upon (often in small but significant ways), the stage is set for more radical conceptual and modelling advances, with the benefit of (partial) acceptance of the new approach. So much for strategy, now back to statistics.

The large amount of data generated in a single experiment is subject to drastic reduction and summarisation prior to currently popular and accepted statistical analyses designed to investigate response mechanisms. A particular approach that has gained wide acceptance is to simply consider maximum levels of stimulus response, compared with the background (nominally) zero-mean levels, and to estimate the maximum levels obtained from the time series records. This is typically done by identifying a window in time that covers most of the peak in response, and then simply averaging the response values across that window. This results in a single post-response summary measure — Window 2 in Figure 1 identifies a typical region for this replication; the maximum response level is estimated as the mean of the 50 or 60 observations in this window. For each such measurement, a pre-stimulus counterpart is evaluated by using a window of the same length located at a fixed lag from the post-stimulus region, providing a single estimate of pre-stimulus (nominally zero) level. Across replications, this leads to two (assumedly independent and random) samples of (estimates of) pre- and post-stimulus neural levels. The latter are assumed as measures of the extent of response of the collective of neurons in the experimental region, and variation in these measures is expected to be informative about the scientific hypotheses of neural response. Figure 2 displays one (of many) example datasets, displaying histograms of the pre-stimulus (noise) levels and post-stimulus (signal plus noise) levels; here the sample size is 159 in each case, rather smaller than the norm. Data such as these have been widely studied using a variety of (non-
Bayesian) statistical methods, the level of statistical sophistication being typically fairly high. In developing models and methods to address the neuroscientific issues, previous investigators have been led to models involving mixtures of normal distributions for the signal plus noise data, and subsequent analyses have focused on methods of mixture deconvolution. Some rationale for this follows in the next section, where we also develop Bayesian approaches to mixture modelling and analysis that represent our early progress in this area.

Key references to the underlying neuroscience, and including currently popular and accepted methods of statistical analysis, include McClachlan (1978), Turner and Schlieckert (1990), Walmsley, Edwards and Tracey (1987), and Wong and Redman (1980). Much of the detail of early development of the quantal/binomial hypothesis, covering statistical and neurological issues in great detail, appears in McClachlan (1978), with more recent developments and investigations reported in Walmsley, Edwards and Tracey (1987). Turner and Schlieckert (1990) describe in detail the current state of the art in experimental protocol and data recording in this area.

2. Mixture Modelling of Response Data

The following rationale underlies accepted approaches to the modelling of data such as displayed in Figure 2 (Turner and Schlieckert 1990, Walmsley et al 1987, and Wong and Redman 1980). The effect of stimulus is to induce transmitter release from some (uncertain) number of synaptic vesicles, the levels of individual responses also being uncertain and differing across vesicles. Higher levels of input stimulus are expected to lead to greater numbers (up to a maximum determined by the physical extent of the neural region under study) of neurons responding, and possibly higher levels of response from each up to maximum possible levels. The quantal hypothesis posits transmitter release at integer multiples of a quantal unit, so that summing across cells leads to a discrete and finite range of possible values for the stimulus response level. Were it possible to observe actual maximum levels of response for any given stimulus, then we could directly assess the quantal hypothesis. The presence of physiological and experimentally induced noise obscures this assessment; attempts to model the noise effects have proceeded by assuming the noise level recordings to be independent, zero mean and approximately normally distributed. The noise variance $\nu$ represents basic, background variation that must be considered in assessment of post-stimulus responses. Under the quantal hypothesis, or one of many variations on such an hypothesis, the signal measurements then approximately follow a mixture of normals, the components of which are located at or near differing levels that derive from the different number and individual responses of activated neurons.

The analysis then reduces, at least as a starting point, to the assessment of the number of components of the mixture, together with estimation of their locations and
weights. Traditional methods of mixture deconvolution, using non-linear least squares and maximum likelihood estimation via the EM algorithm, are reasonably well established in the neurosciences (e.g. Ling and Tolhurst 1983), though the technical difficulties associated with them, and the difficulties of assessing uncertainties about the parameters of a mixture estimated these ways, are considerable and well-known (Titterington, Smith and Makov 1985). By comparison, Bayesian approaches to mixture estimation (West 1991, and references therein) seem more directly geared to some of the inferential problems arising here, and (as always) provides formal means for provision of realistic assessments of the uncertainties about parameter values. We have begun to develop methods based on Dirichlet mixtures of normals (Escobar and West 1991; Ferguson 1983; West 1990, 1991), and illustrate some of the results of that application here.

A typical signal dataset of size \( n \), say, one to some few hundred observations, is assumed to come from a mixture of an unknown number of normal distributions, with uncertain means and variance \( \nu \). This variance is assumed to be that of the signal-free background noise. Denote the signal data by \( y_i \) and write \( \mu_i \) for their means; thus 

\[
(y_i | \mu_i, \nu) \sim N(\mu_i, \nu), \quad i = 1, \ldots, n.
\]

The mixture structure simply imposes the constraint that, for some positive integer \( k \), there exist \( k \) distinct numbers \( \theta = \{\theta_1, \ldots, \theta_k\} \) such that, for each \( i = 1, \ldots, n \), \( \mu_i = \theta_j \) for some \( j = 1, \ldots, k \). We refer to the elements of \( \theta \) as the distinct means of the mixture. A suitable way of generating such a mixture uses a Dirichlet process for the prior distribution of the original data means; full technical details can be found in Ferguson (1983), Escobar and West (1991), or West (1990). Technical development here closely follows the details in West (1990). We suppose that the \( \mu_i \) are independently drawn from an uncertain prior distribution \( G(.) \) which is modelled as a Dirichlet process, \( G \sim D(\alpha G_0) \), defined by parameters \( \alpha \), a positive scalar, and \( G_0(.) \), a specified distribution function over the real line; \( G_0(.) \) is an initial prior for each of the \( \mu_i \); marginally so satisfies \( E[G(\mu_i)] = G_0(\mu_i) \), and \( \alpha \) is a precision parameter measuring the concentration of the prior for \( G(.) \) about \( G_0(.) \). A key feature of this model is that the \( \mu_i \) concentrate on a set \( \theta = \{\theta_1, \ldots, \theta_k\} \) of distinct means, for some number \( k \leq n \). Also, conditional on \( k \), the \( \theta_j \) are initially independently distributed according to \( G_0(.) \). Further, the \( \mu_j \) are allocated amongst the elements of \( \theta \) according to a uniform multinomial, resulting in some \( n_i \) taking the value \( \theta_j \). See the above references for further discussion. As a consequence, we have a normal mixture model; given \( k \) and \( \theta \) and the \( n_i \) (in addition to \( G_0 \) and \( \nu \)), a further observation \( Y \) has predictive distribution

\[
\pi_0 N(\theta_0, \nu) + a_n \sum_{j=1}^{n} N(\mu_j, \nu) \equiv \pi_0 N(\theta_0, \nu) + \sum_{i=1}^{k} \pi_i N(\theta_i, \nu) \tag{1}
\]

where \( a_n = 1/(\alpha + n) \), \( \pi_0 = \alpha a_n \), \( \pi_i = n_i a_n \) for \( i = 1, \ldots, k \), and \( \theta_0 \) is a further draw from \( G_0 \). Thus we have a mixture with an uncertain number of components, a class of
priors (parametrised by $\alpha$) for the number of components, and a prior that treats the locations of the components (however many there may be) as exchangeable. In this latter sense, the framework is not informative about possible structured relationships amongst the locations and so provides a suitably neutral standpoint from which to assess possible theories, such as variations on the basic quantal hypothesis, about any such structure. The prior for $k$ is implicit in the Dirichlet process construction, and is determined by $\alpha$ and $n$ alone. In most practical contexts, and certainly in the neurological problems here, $k$ is expected to small relative to $n$ and this implies small values of $\alpha$, so (1) is similar to a more traditional mixture model in which $k$ is independent of $n$ (eg Lavine and West 1991). Were $k, \theta$ and $\pi = \{\pi_1, \ldots, \pi_k\}$ known, equation (1) would provide a Bayesian density estimate; otherwise, data analysis will produce the posterior distribution for $k, \theta$ and $\pi$ conditional on the observed data $D = \{y_1, \ldots, y_k\}$, suitable for inference about the number, weights and locations of components, and predictive inferences/density estimation follow by averaging (1) with respect this posterior (Escobar and West 1991).

Our analyses adopt the normal prior $G_0(\mu) \equiv N(m, \tau v)$, and assumes that the three additional parameters $m, \tau$ and $v$ are uncertain. Thus data analysis involves computing the posterior distributions for $k, \theta, \pi, m, \tau, v$ given data $D$, extending the above discussion to include these three extra quantities. Here $\tau$ is a smoothing parameter; large value lead to greater dispersion amongst the $\theta_i$ which induces more erratic behaviour in the density (1), whilst lower values induce smoother predictions. The scaling of the prior variance $\tau$ by $v$ allows prior assessment of suitable ranges of values of $m$ and $\tau$ to be performed without regard to the value of the noise variance $v$; conditional only on $m$, $\tau$ and $v$, the prior predictive distribution for any measurement is just $y \sim N(m, (1 + \tau)v)$. Thus $m$ is the prior mean, and $1 + \tau$ determines the scale inflation expected post-stimulus. A complete prior specification involves prescription of a normal prior for $m$ and independent, inverse gamma priors for $v$ and $\tau$. In practice, there is little gain from working hard at specifying the former and analyses assume a uniform, reference prior for $m$. The pre-stimulus noise level recordings provide independent data to base the prior for $v$ upon. On the assumptions that the noise data are a normal random sample of size $n$, the reference analysis leads to posterior (to the noise data) $v^{-1} \sim G(s/2, S/2)$ where $s = n - 1$ and $S$ is the corrected sum of squares of the noise recordings. Currently, this is taken as determining the prior for $v$ for analysis of post-stimulus response levels. The prior for $\tau$ has a similar form, $\tau^{-1} \sim G(w/2, W/2)$, for some specified $w$ and $W$. Given the informative prior for $v$ (and $n$ is typically in the hundreds), it is convenient to assume the initial degrees of freedom $w$ to be small, hence specifying a relatively uninformative prior on $w$, and taking $W/w$ as an initial point estimate of $\tau$.

Further technical discussion appears in Escobar and West (1991) and West (1991), together with details of the Gibbs sampling based techniques of computation of posterior
distributions for the parameters $k$, $\mu$, $m$, $\tau$ and $v$. The Monte Carlo analysis provides approximate samples from the joint posterior distribution $p(k, \mu, m, \tau, v|D)$. The sampled values of any component may be used to directly estimate posterior quantities, and marginal posteriors for most of the parameters can be approximated as averages of conditional posteriors (Escobar and West 1991) as is now standard practice with iterative sampling methods generally (Gelfand and Smith 1990). Inference about $\theta$ and $\pi$ requires further thought and theory. The dimension of each is $k$, which itself is uncertain, so the issue of identifiability arises. The way ahead is to focus instead on the ordered values of the component locations; whatever $k$ may be, $\theta(1) < \theta(2) < \ldots < \theta(k)$ where $\theta(j)$ denotes the $j^{th}$ largest mean of the $k$ in $\theta$. With respect to the quantal/binomial hypothesis of neural responses, it is these ordered values that are of direct interest in any case; roughly equal (= quantal) spacing of the ordered levels is consistent with the hypothesis, so display of posteriors for the ordered values, and summary inferences for the spacings $\theta(j+1) - \theta(j)$, directly addresses this scientific issue. In the Monte Carlo analysis, each sampled vector from $p(\mu|D)$ contains some number $k$ of distinct $\theta_j$ values with explicit identification of each of the $\mu_i$ with one $\theta_j$, plus the corresponding numbers $n_j$ and hence the mixture weights $\pi$. The $k$ distinct means can be ordered to give a vector of draws from the posterior of the $k$ smallest locations; the reordering applies also, of course, to the vector of weights $\pi$. Thus we can repeatedly sample from the (approximate) posterior for the entire parameter vector $\{k, \theta, \pi, m, \tau, v\}$. Resulting inferences for the $\theta(j)$ may be based on simple Monte Carlo means or, more efficiently, on averages of conditional distributions. Here new theoretical issues arise, as follows. Conditional on any (sampled) values of $\{k, \pi, m, \tau, v\}$ and the identification of each $\mu_i$ with a particular normal mean $\theta_j$, the $\theta_j$ are conditionally independently normally distributed with moments that are easily computed as functions of $\{k, \pi, m, \tau, v\}$ and the data $D$. The corresponding conditional distributions of the ordered means $\theta(j)$ are then those of the order statistics in independent but non-identically distributed normal samples. The Monte Carlo estimates of posteriors $p(\theta(j)|D)$ are averages of the densities of such conditional distributions. Theoretical and computational issues arising in the evaluations required here will be discussed elsewhere (Cao and West, in preparation). It suffices here to note that such evaluations may be performed quite efficiently to deduce the required inferences, as illustrated in the next section.

3. Data Analysis Example

An analysis of the small dataset in Figure 2 is summarised here. This example has sample sizes $n = 159$ for each of the noise and signal samples. The former has a sample variance of 0.047 on 158 degrees of freedom. The prior (to signal data analysis) for $v$ based on these values was determined by $s = 150$ and $S = 150 \times 0.0467 = 7.0$. Since
the prior predictive distribution of any observation is \( N(m, (1 + \tau)v) \), then, with \( v \) close to 0.047, we may assess plausible values of \( \tau \) (and \( m \)) by requiring the experimenter to identify plausible ranges of outcome responses; levels of proposed stimuli will guide the scientist's assessments here. In later stages of this project, this will become a routine part of analysis. At this early stage, we are still developing and refining the model and prior structure and have not yet proceeded to this more interactive assessment, and simplify as follows. The average \( m \) is assigned a uniform reference prior, not anticipating the levels of evoked responses at all. For \( \tau \), note that values between 3-8 lead to between 2 and 3-fold increases in predictive standard deviations; a prior with \( W/w = 5 \) and low degrees of freedom, say \( w = 1 \), is very diffuse and provides appreciable support across this range – the dashed curve in Figure 3(e) is the resulting prior inverse gamma density, assumed here. The remaining parameter to be specified is the prior precision \( \alpha \) of the Dirichlet process; this is critical in determining the initial view of likely numbers of mixture components, \( k \). Figure 3(d) displays (as dashed lines) the prior distribution \( p(k) \) based on \( \alpha = 0.1 \) (so \( p(k) \equiv p(k|\alpha = 0.1, n = 159) \), formally). These probabilities decrease with \( k \), giving reasonable chances to two or three components; under this prior, \( P(k \geq 4) \approx 0.017 \). Typically the view is that \( k \) will be in the low (2-6) integer range, though this is case dependent and another area where more detailed assessments with the neuroscientists may assist in refining the specification of \( \alpha \).

Figure 3(a) displays a histogram of the signal data plus the approximate predictive density \( p(y|D) \) from an analysis using the Gibbs sampling techniques briefly described and referenced earlier. The Monte Carlo sample size used here was 20,000; thus the predictive density is the average of 20,000 versions of equation (1), each corresponding to a different value of the vector parameter \( \{k, \pi, m, \tau, v\} \) that is an approximate draw from the posterior for the parameters. (The resampling iterations were run for an initial 500 draws to 'burn-in' from starting values; see Escobar and West 1991). Some idea as to the posterior uncertainty associated with the density estimate is gleaned from Figure 3(b) which displays just one hundred of the sampled predictives. The average density appears again in Figure 3(d), together with graphical illustration of the deconvolution into mixture components. The blocks displayed here are symmetrically centered at the approximate posterior means of the \( \theta(j) \), with widths set at twice the corresponding posterior standard deviations. For example, the first mean has approximate moments \( E(\theta_{(1)}|D) = 9.36 \) and \( V(\theta_{(1)}|D) = 0.08^2 \); thus the first block is located at \( 9.36 \pm 0.08 \). Increasing uncertainty about the larger locations is apparent in this figure. Also, the heights of the blocks are the approximate posterior means of the component weights; thus the first block has height \( E(\pi_{3}|D) = 0.11 \), and so forth.

Figure 3(d) plots prior (dashed lines) and posterior (full lines) probabilities over values of \( k \), illustrating a major shift to values around 4 or 5 that were deemed as
very unlikely initially; the posterior mode is evidently $k = 4$ with $P(k = 4|D) \approx 0.58$. As discussed in Escobar and West (1991) mixture models will naturally tend to rather over-estimate the number of components in attempts to produce fidelity to even minor features of non-normality in the data, so that the posterior mass on larger values of $k$ might be informally discounted somewhat in summary inferences; here, clearly, four or five components are indicated even with this in mind. Priors (dashed lines) and posteriors (full lines) for $\tau$ and $\nu$ appear in Figures 3(e) and 3(f) respectively. The former display is consistent with the fact that, quite typically, likelihood functions for the smoothing parameter $\tau$ tend to be flat, the data being relatively uninformative about $\tau$ (West 1990). The latter suggests the noise based prior for $\nu$ favours rather smaller values than are appropriate for the signal data mixture model. However, the assumption of constant variance $\nu$ across mixture components may be suspect and underlie this feature; larger variances in the ‘tail’ components of the mixture, particularly the upper tail for this dataset, are often appropriate (Escobar and West 1991). Any such extra-variation in the observed signal data will feed through to shift $p(\nu|D)$ to larger values, whereas a more appropriate model would separate out the variance components, allowing a different value for each component.

Figure 4 displays graphs relevant to inference about the component means $\theta$, providing plots of the approximate posterior densities for ordered values $p(\theta_{(j)}|D)$. Figure 4(a) gives these densities for $j = 1, \ldots, 5$. The first three locations are clearly distinguished. The posteriors for $\theta_{(4)}$ and $\theta_{(5)}$ are apparently substantially similar; although $P(k = 5|D) \approx 0.30$ is appreciable, the posteriors here indicate that a fifth component location would be very close to the fourth. Bearing in mind the earlier comment about over-estimation of $k$, this strongly suggests a model with $k = 4$ to be reasonable. Incidentally, the posterior for $\theta_{(6)}$, having probability $P(k = 6|D) \approx 0.08$, similarly overlaps that for $\theta_{(5)}$. As a result, we condition on $k = 4$ to produce Figure 4(b); in the Gibbs sampling iterations, this conditioning is trivially effected by simply rejecting posterior samples unless $k = 4$ (the posterior $P(k = 4|D) \approx 0.58$ indicates that about 11,600 of the 20,000 posterior samples were saved). Figure 4(b) indicates the consequent sharpening-up of the inferences on the assumed four levels, and similar conditional inferences could as easily be summarised for the other parameters $k$, $\tau$, etc., as could the effects on prediction, though this is not explored further here.

Major practical advantages of the approach illustrated here stem from the ease with which posterior inferences involving apparently rather complex calculations can be reduced to routine using the posterior sampling approach. Existing approaches using maximum likelihood or least squares techniques can be given approximate Bayesian interpretations, perhaps with some modification, but have absolutely no hope of producing complete posterior inferences about the parameter $k$, $\tau$ and $\theta$ that are fundamental to the
neuroscientific issues. In addition, exploring changes in inferences under various forms of imposed conditioning, such as fixing \( k \) above, is of major benefit for model simplification in the light of observed data and also for sensitivity analyses. Displays of posteriors for \( k, \theta \) and \( \pi \), possibly subject to additional conditioning, provide the neurologists with the basic tools to explore changes in neural response mechanisms induced by changes in neural stimulation or physical changes in neural tissue. Much more is needed, however, and current and proposed investigations to further the scientific goals lie in the following areas. Concerning the 'testing' of the neurological quantal/binomial hypotheses, and other possible theories, the posteriors for \((\pi|D)\) and \((\theta|D)\) are central. For example, we may explore the posterior of the ordered values in \( \theta \) to assess support for approximately constant spacings. The margins in Figure 4 give some indication of this, though the joint posterior for the actual spacings \( |\theta_{(j+1)} - \theta_{(j)}| \) is really needed. Evaluating and summarising this raises theoretical issues, though simple Monte Carlo estimates of posterior moments and probabilities for the spacings are easily deduced from the existing analysis. More formally testing the quantal hypothesis is more difficult, as is the related issue of assessing consistency of \( p(\pi|D) \) with (one of many) possible binomial distributions. The conceptual and theoretical basis for such assessments need to be formulated.

Other developments under way, and that are expected to have immediate and marked practical effects, include extensions to include (practically reasonable) variations in the noise variance \( v \) across mixture components (Escobar and West 1991) already noted above. Assessment of priors for the resulting variances, say, \( v_f \), will be naturally based on the post-noise data prior \( p(v) \), though probably subject to some form of discounting of that prior information. Probably most important currently is the extension to include the final model parameter \( \alpha \), the precision of the Dirichlet process, as uncertain. It has recently become clear that the Gibbs sampling computations can be very easily extended to incorporate prior distributions over a discrete (or, via interpolation) continuous range for \( \alpha \), and hence to produce posteriors for the key parameters unconditional on \( \alpha \). Rather uninformative priors will probably suffice, and the anticipated resulting robustness of analysis will be important in practice; not least in practical importance is the removal of this final parameter from the inputs required of the neurologist. This will contribute substantially to the development and automation of a routine class of models suitable for repeat applications across the very many datasets that are being generated in this area. Another area that bears study is the assumed and basic normality of the underlying noise distribution which (Figure 2), which may be suspect in some cases; use of a mixture to model the noise too is one obvious and attractive generalisation under consideration.
References


Titles and legends for figures

Figure 1. Pre- and post-stimulus neural potentials
Figure 2. Neural noise and signal histograms
Figure 3. Summaries of posterior and predictive distributions
Figure 4. Posteriors for component means
Figure 1. Pre- and post-stimulus neural potentials

Figure 2. Neural noise and signal histograms
Figure 3. Summaries of posterior and predictive distributions

(a) Predictive pdf $p(y|D)$

(b) Sampled pdfs $p(y|D)$

(c) Mixture deconvolution

(d) $p(k)$ and $p(k|D)$

(e) Pdfs $p(\tau)$ and $p(\tau|D)$

(f) Pdfs $p(v)$ and $p(v|D)$
Figure 4. Posteriors for component means

(a) Unconditional posterior densities

(b) Posterior densities given $k=4$