Deconvolution of Mixtures in Analysis of Neural Synaptic Transmission

Mike West; Dennis A. Turner

The Statistician, Vol. 43, No. 1, Special Issue: Conference on Practical Bayesian Statistics, 1992 (3). (1994), pp. 31-43.

Stable URL:
http://links.jstor.org/sici?sici=0039-0526%281994%2943%3A1%3C31%3ADOMIAO%3E2.0.CO%3B2-W

The Statistician is currently published by Royal Statistical Society.

The Statistician (1994)
43, *No.* 1, *pp.* 31–43

Deconvolution of mixtures in analysis of neural synaptic transmission

By MIKE WEST† and DENNIS A. TURNER

Duke University, Durham, USA

[Received July 1992. Revised July 1993]

SUMMARY

Neurophysiologists investigating mechanisms underlying neural responses to stimuli have, in recent years, developed substantial interest in modelling certain types of neural response data by using simple mixture distributions. Techniques of mixture deconvolution using likelihood-based techniques have become popular. This paper reports on novel Bayesian approaches using (uncertain) mixtures of (uncertain numbers of) noise distributions to model data measuring maximum levels of evoked neural responses following various levels of electrical stimulus of nerve tissue. We discuss some of the key scientific issues, including physiological hypotheses of 'quantal' levels of neuronal transmissions, together with technical aspects of data analysis, modelling and the use of prior information in addressing these issues within an appropriate Bayesian framework. Illustration of neural response deconvolution analysis using this approach is presented.

Keywords: Bayesian mixture models; Dirichlet process mixtures; Markov chain Monte Carlo method; Synaptic potential analysis

1. Introduction

Just before the third 'Practical Bayesian statistics' meeting, the journal *Nature* published two articles concerned with *long-term potentiation* of electrical transmission at synapses in mammalian brains, a phenomenon that is believed to be critical to memory and learning (Kullman and Nicoll, 1992; Malgaroli and Tsien, 1992). Central to the potentiation process are changes in the characteristics of chemical transmitter release at synaptic junctions. The *Nature* articles are concerned with, among other things, detailing such changes—one corner of the huge field of neurophysiological research into mechanisms of neural response in mammalian, and other, nervous systems (Walmsley *et al.*, 1987, 1988). Statistical mixture analysis and deconvolution have become a central technique across this field, as the review of Redman (1990) indicates. Turner and West (1993) give numerous additional and recent references, and introduce Bayesian approaches to address the difficulties of inference in this field. This modelling is further developed, applied and illustrated here.

A single nerve cell, or neuron, can be viewed as a black box that converts input electrical signals to outputs. The box is the cell body, or soma. Inputs are received through a root-like system of dendrites that contact other neurons at nerve junctions, or synapses. Chemical transmitter releases from individual release sites on these 'input' neurons induce electrical conductance changes throughout the membrane of the receiving neuron which may then 'trigger' or 'excite' an action potential change at the soma. This 'evoked potential response' passes through as output along the axon of the cell which ends in branch-like structures contacting further synapses. So the 'message' is communicated throughout the nervous

† *Address for correspondence*: Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, USA.
E-mail: mw@isds.duke.edu

 0039–0526/94/43031

system. A basic and critical hypothesis under investigation at many different neural junctions is that, at an individual transmitter release site, transmission (when it occurs) is in a basic quantal unit; if so, this results in equally spaced or quantal levels of potential response, the equally spaced levels representing the sum of quanta across the combinations of numbers of transmitting sites. A related 'binomial' hypothesis posits a small and roughly constant probability of transmitter release at each site. Investigations of neural function concern questions about quantal size and release probability, variation in these characteristics across release sites and across synapses, evidence for variation in quantal size and release probability due to physiological changes as exemplified by the potentiation process in Kullman and Nicoll (1992) and Malgaroli and Tsien (1992), pre- and post-synaptic physiological influences on release characteristics, and so forth.

In some studies of central nervous systems, it has recently become possible to isolate effectively the collective of release sites contacting the dendrites of individual central neurons, and these large cells are also amenable to recording of electrical potential changes throughout the soma (Turner and Schlieckert, 1990). With some 1–15 or more possible release sites, the variation in levels of evoked potential responses due to chemically induced experimental stimulus provides information that is relevant to the study of quantal size and release probabilities, and to the investigation of deviations from the hypotheses of constancy of these characteristics. Measurements of synaptic potentials are, naturally, corrupted by noise—physiological noise due to spontaneous transmitter release and unidentified synaptic contacts, and experimentally induced noise in recording of potentials and post-recording data processing. Thus the discrete levels of potential response, and their respective chances of occurrence, are convolved with noise, resulting in a class of mixture distributions for the measured data. Inference about levels and chances of response thus involves modelling this mixture structure and deconvolution to identify the noise-free process.

2. Data and models

Data ensembles are created by discrete sampling of somatic electrical potential signals measured via an intracellular probe. The synaptic junction is stimulated to produce a response time curve exhibiting a fast rise then exponential decay to pre-stimulus levels, across a time interval of around 50–100 ms. The cell potential is rapidly sampled in this time interval and the resulting recordings reduced to an estimate of the maximum level of evoked response by averaging potentials in an interval around the peak of the response curve, and then subtracting a similar average of pre-stimulus levels. The process is repeated to produce a sample of usually 300–1000 observations on response levels (Turner and Schlieckert, 1990). Observations on levels of synaptic noise are obtained concurrently by precisely the same process but without stimulus, usually interleaving one noise measurement between each signal–response measurement. The differencing process implies zero mean for the noise distribution. Stimulation intensity is controlled to low, near threshold, levels, attempting to bias responses to low levels consistent with few mixture components, say 3–10, represented in the signal data. Noise and signal histograms of one data set, coded c2epsp10 and obtained from intracellular recordings from a CA1 pyramidal neuron (Turner, 1987; Turner and West, 1993), appear in Fig. 1. Analysis assumes the noise data to be an approximate normal random sample and the signal data to derive from a mixture of normals with base-line variation that of the noise. Preliminary screening of the series to assure approximate concordance with the implicit assumptions of independence and stationarity of the noise and signal recording (separately) is performed. The noise sample here has two or three values in each tail that, relative to normality, are overextreme. These may be removed as outliers before analysis though repeat analyses confirm that they have negligible influence in the deconvolution of the signal data. There are other aspects of this noise sample indicative of non-normality that are not central to the analysis here but which are mooted in discussion of current research directions in Section 5. The signa'

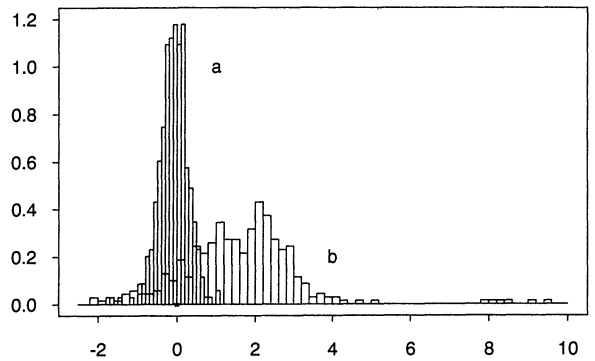

Fig. 1. Histograms of c2epsp10 noise and signal recordings: histogram a, noise data; histogram b, signal data

histogram shows evidence of several components assuming a normal mixture description. Simply noting the range and shape of the signal histogram indicates that, with a base-line component variance equal to that of the noise, the signal data will not be adequately described with fewer than four or five components.

The basic model supposes the noise data to be independently and identically distributed (IID) $N(0, v)$ and the signal data to arise from some discrete mixture of normals, generally $\Sigma_{j=1}^{k} p_j N(\theta_j, v_j)$. Ideally, $v_j = v$ and the means θ_j represent noise-free response levels. Under the quantal–binomial hypothesis of transmitter release, the ordered values of these levels are equally spaced (the first being 0) and the corresponding weights p_j binomial. Plausible and physiologically important deviations from this ideal include

(a) different quantal amplitude at different sites—the means are then unequally spaced,
(b) differing release chances across sites—the chances are then not binomial,
(c) a combination of (a) and (b), and
(d) small stochastic variations in site-specific quantal levels—inducing differences in component variances.

These in combination, and other possible features, indicate the need for flexibility and scientific neutrality in modelling the data to permit posterior assessment of the plausibility of constant spacings, binomial chances, equal variances, etc. Inferences about mixture structure in an unconstrained model may have an unforeseen but viable neurological interpretation or be suggestive of features of neural functioning that may be further investigated experimentally. This contrasts critically with more traditional approaches in this area that impose quantal–binomial constraints from the outset.

Our approach to date uses Dirichlet process mixtures (Escobar and West, 1992) generalizing earlier work using constant variance components (West and Cao, 1993). The central structural features of such models are as follows.

(a) The signal data $\{y_i\}_{i=1}^{n}$ arise from a mixture of some $k \geqslant 1$ normal distributions $N(\theta_j, v_j)$, $j = 1, \ldots, k$, with distinct means and variances. Inference about the numbers of mixture components is addressed via k, the number of components actually represented in the observed data.
(b) The number of components, or groups, k has a prior determined by a single scalar parameter $\alpha > 0$ (the precision of the Dirichlet process); large values of α are consistent with larger values of k, and α will typically be sufficiently small that realized values of k are much less than n. The prior $P(k|\alpha)$ is such that $x = k - 1$ has a Poisson-like distribution, and $E(k|\alpha) \approx \alpha \log(1 + n/\alpha)$ (West, 1992).
(c) Given k, the signal data $\{y_i\}_{i=1}^{n}$ are allocated to components or groups according to a uniform multinomial distribution. As in West (1990) write $c_i = j$ if y_i comes from

component j (or belongs to group j) and set $c = (c_1, \ldots, c_n)$ and denote the allocation, or *configuration*, by $\mathscr{C}_{n,k}$.

(d) Given k and $\mathscr{C}_{n,k}$, the distinct pairs (θ_j, v_j) are initially a random sample from a continuous bivariate prior distribution $G_0(\ ,\ )$.

One interpretation is that nature generates k then determines the configuration—just a one-way normal layout of the data with distinct means and variances across groups. Knowledge of $\mathscr{C}_{n,k}$ implicitly determines the group sizes $n_j = \#\{c_i = j\}_{i=1}^n$ for each $j = 1, \ldots, k$. Further we have the following.

(e) Given $\mathscr{C}_{n,k}$ and the moments $\{\theta_j, v_j\}_{j=1}^k$, the moments (θ_0, v_0) of the normal distribution of a *further* signal have posterior distribution

$$G_n(\theta_0, v_0) = p_0\, G_0(\theta_0, v_0) + \sum_{j=1}^{k} p_j\, \delta(\theta_0 - \theta_j, v_0 - v_j)$$

where $\delta(z, u)$ is the delta function, $\delta(z, u) = 1$ if $z = u = 0$ and $\delta(z, u) = 0$ otherwise, and the mixture has weights $p_0 = \alpha/(\alpha + n)$ and, for $j = 1, \ldots, k$, $p_j = n_j/(\alpha + n)$. This posterior, a mixture of the continuous prior $G_0(\ ,\ )$ and point masses on the k distinct pairs, is also the posterior mean of the uncertain mixing distribution $G(\ ,\ )$.

(f) As a direct consequence of (e), a further signal is generated by the mixture distribution $\Sigma_{j=0}^k p_j\, N(\theta_j, v_j)$.

The precision α of the Dirichlet is usually interpreted in terms of 'strength' of belief in $G_0(\ ,\ ) = E\{G(\ ,\ )\}$ as a prior guess at the form of the mixing distribution—large α is consistent with strong belief. This has been viewed by some as being at odds with the fact that large α induces larger numbers k of distinct parameters and so apparently increased disparity between the posterior mean $G_n(\ ,\ )$ and $G_0(\ ,\ )$. The point is clarified in item (e) above; large α implies a strong belief in $G_0(\ ,\ )$ as the distribution generating *future* (θ_0, v_0) pairs, whatever the values generated in the past.

A discussion of the priors for k induced by α appears in Turner and West (1993) in some generality. For c2epsp10 noise with $n = 349$, Table 1 exhibits prior probabilities over k (to two decimal places—note that $P(k|\alpha) > 0$ for all $k > 0$ and $\alpha > 0$) for $\alpha = 0.01, 0.1$ and 1.0. Larger α-values induce diffuse priors over larger ranges of k. The smaller values, although concentrating on small k, are still positive for all k and provide more conservative initial positions on the number of components, as will be discussed later.

The model requires a specification for $G_0(\ ,\ )$. Mixtures of conjugate normal–inverse gamma priors lead to conditional conjugacy, though others (e.g. finite uniform priors) are possible. Currently, a single conjugate form is implemented, under which $\theta_j|v_j \sim N(m, wv_j)$ and $s/v_j \sim \chi_r^2$, for some hyperparameters m, w, r and s. As discussed in Escobar and West (1992), conjugate priors for all four quantities may be incorporated in analysis. Typically, and here, the reference prior $p(m, w) \propto w^{-1}$ is used for the prior mean m and the 'smoothing parameter' w (West, 1990). For the v_j, a prior guess s is based on the variance of the observed

TABLE 1
$P(k|\alpha)$ for c2epsp10 noise ($n = 349$)

α	*$P(k\|\alpha)$-values for the following values of k:*												
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*$\geqslant 13$*
0.01	0.94	0.06											
0.1	0.53	0.34	0.11	0.02									
1.0		0.02	0.06	0.11	0.16	0.18	0.17	0.13	0.08	0.05	0.02	0.01	0.01

noise sample, and the degrees-of-freedom parameter is chosen to reflect uncertainty—large r is consistent with the view that the component variances are likely to be close to the noise value v, uncontaminated by extra noise sources of variation (including possible increases due to small degrees of variation in quantal sizes at any or each transmission site). Routine analyses are based on r between 5 and 10, with sensitivity to values in the range explored. With n in the hundreds and k in the low integers, the numbers per component n_j are likely to, and tend to, outweigh such low r-values; the resulting likelihoods for the v_j then dominate the prior. This expectation is often not valid in tails of the signal data, however, where a component will generate rather few observations, so that the proper prior on v_j is critical.

Finally on model and prior structure, we may introduce priors over the remaining, and critical, parameter α. West (1992) discusses the use of mixtures of gamma priors. Here we restrict attention to a single gamma prior, $\alpha \sim G(b, b/a)$ with $E(\alpha) = a$ and shape parameter b, as implemented and illustrated in Escobar and West (1992).

3. Simulation analysis summaries

Given the above specification, posterior analysis may be performed by using iterative posterior simulations (Escobar and West, 1992) to produce sequences of samples from the joint posterior of all model parameters: the primary parameters k, $\{\theta_i, v_i, p_i\}_{i=1}^k$ (and note that the dimension of the parameter space is uncertain); the configuration $\mathscr{C}_{n,k}$ through the quantities c_i; the hyperparameters m, w and α. The Gibbs sampling algorithms used are very efficient and have been empirically explored and applied for a long time now. (Fortran, C and S software are available from the authors for such analyses.) Convergence results have been established in Escobar and West (1992). Each Gibbs iteration generates a single set of values that, if the Markov chain simulation had converged, would be drawn from the full joint posterior; assuming approximate convergence, these draws provide the basis for approximate posterior inference. The important 'complete conditional' posteriors are noted below (and detailed completely in Escobar and West (1992)), and the sampling process results in replicates of the following single-sample summary:

(a) a sampled value from the posterior for k;
(b) given this k, a sampled configuration $\mathscr{C}_{n,k}$ determined by allocation indices c_i (the clustering process implicitly determined by the posterior distribution over configurations (West, 1990; Escobar and West, 1992) naturally groups neighbouring observations together);
(c) implicitly, a sample from the posterior for component weights $\{p_i\}_{i=1}^k$ determined by the sampled values $n_i/(\alpha + n)$;
(d) given this k, a set of distinct sampled moments $\{\theta_i, v_i\}_{i=1}^k$;
(e) a sampled value of each hyperparameter m, w and α (West (1992) shows how data augmentation ideas provide easy sampling of α with mixtures of gamma priors).

In addition to the summary samples, various conditional posteriors are of standard analytical form and may be used for posterior inference. Summarizing:

(i) given k, $\mathscr{C}_{n,k}$ and hyperparameters, the moment pairs have independent normal–inverse gamma posteriors, $\theta_i | v_i \sim N(m_i, w_i v_i)$ and $s_i/v_i \sim \chi^2_{r_i}$ with defining parameters that are simple functions of the conditioning quantities;
(ii) given k, $\mathscr{C}_{n,k}$ and α, the weights $\{p_i\}_{i=0}^k$ have a conditional Dirichlet posterior with parameters $\{\alpha, n_1, \ldots, n_k\}$;
(iii) given k, $\mathscr{C}_{n,k}$ and $\{\theta_i, v_i\}_{i=1}^k$, the hyperparameters (m, w) have a normal–inverse gamma posterior with trivially computable defining parameters;
(iv) given k, the conditional posterior for α is a mixture of gamma distributions.

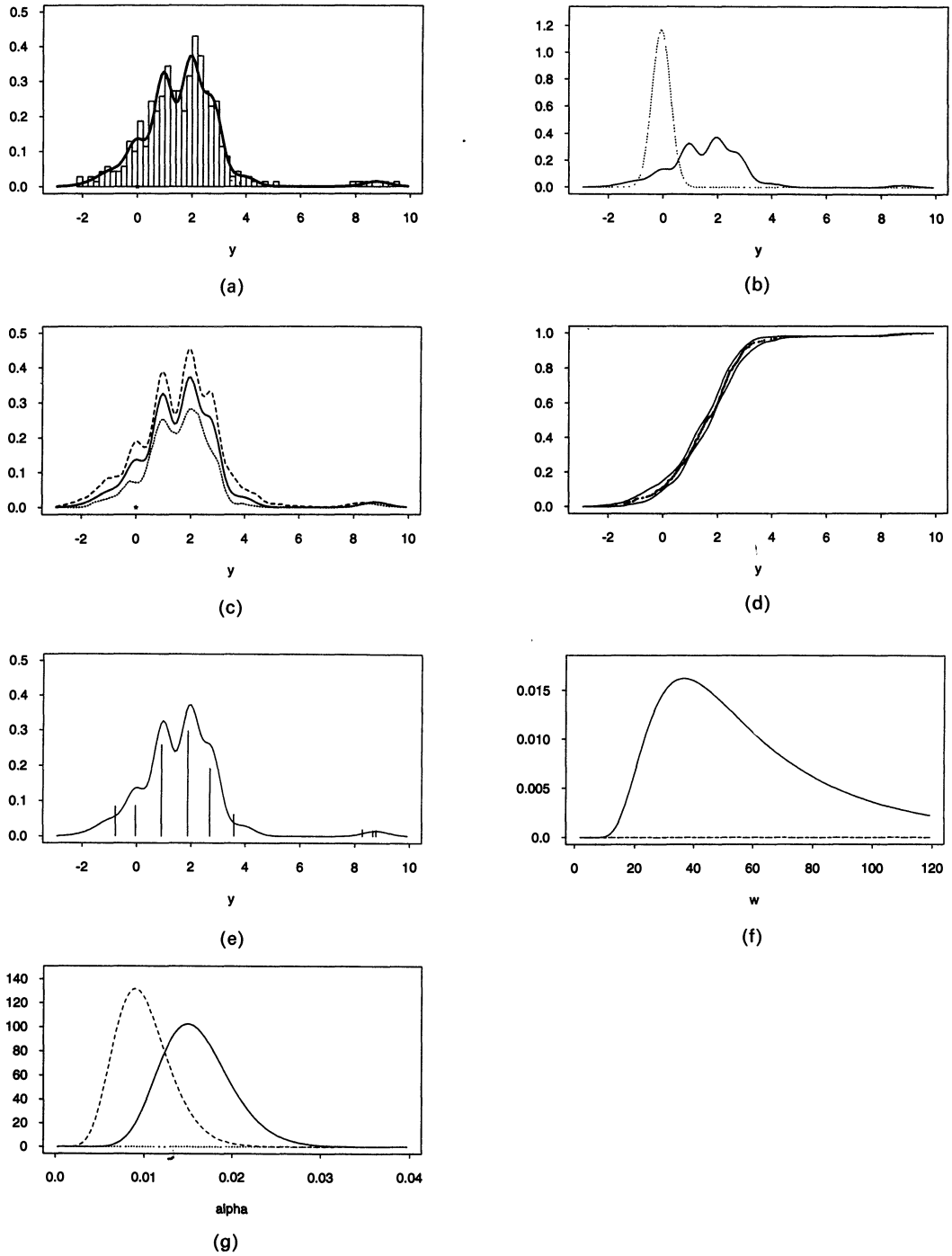

Fig. 2. Summary of c2epsp10 noise analysis: (a) signal data and predictive densities; (b) predictive densities for the signal (———) and noise (·····); (c) predictive probability density function (PDF) for the signal, with range bands; (d) predictive cumulative density function for the signal, with range bands; (e) predictive PDF for the signal with estimated mixture weights and locations; (f) posterior for the smoothing parameter w; (g) prior (---------) and posterior (———) for the precision parameter α

If possible, inference is best based on Monte Carlo averages of such conditional posteriors (rather than on direct Monte Carlo samples). Consider, for example, inference about the smoothing parameter w. As above, each Gibbs iterate generates a 'sampled' χ^2-posterior—the shape and scale typically differ in each sample depending on the changing conditioning variates. Across Monte Carlo samples this implies sampled χ^2-posteriors that may be averaged to produce an approximation to the marginal posterior for w. One analysis of c2epsp10 noise leads to the margin displayed in Fig. 2(f), for example. Similar calculations provide margins for α (Fig. 2(g)), k and m.

Inference about component means, variances and weights has a complication: their number is uncertain. Our approach to this problem is via the ordered means, as follows. At any simulation stage, say indexed by r, denote by π_r the set of current simulated values of k, the configuration, the variances v_j and all hyperparameters. The configuration implicitly determines the number $k = k_r$ at this iteration; further, given π_r, the quantities $\theta_1, \ldots, \theta_k$ are conditionally independent normal, though not IID, from (i) above. Consider their ordered values $\theta_{(1)} < \ldots < \theta_{(k)}$. The cumulative density function (CDF) and probability density function (PDF) of the conditional distributions $(\theta_{(i)}|D, \pi_r)$ may be evaluated via recursions derived in Cao and West (1992); see also West and Cao (1993). Then an approximate evaluation of the marginal posteriors follows by the usual averaging of conditional posteriors over simulated conditioning quantities, i.e.

$$p(\theta_{(i)}|D) \approx \sum_{r=1}^{R} p(\theta_{(i)}|D, \pi_r)/R$$

where R is the Monte Carlo sample size. This surmounts the problem of uncertain number of components; fewer terms will appear in this sum for higher values of i, implicitly averaging with respect to the posterior distribution for k. Further the analysis leads easily into 'what if?' exploration of inferences about the component moments across ranges of *a posteriori* plausible values of k. An illustration in the analysis of c2epsp10 noise follows.

4. Analysis of c2epsp10 data

Specific priors underlying the analyses reported are summarized as follows.

(a) Reference analysis of the c2epsp10 noise sample yields the noise variance posterior $s/v \sim \chi^2_{348}$ where s is the noise sample variance. The estimate s is used as the base-line prior guess for the component variances, assumed to have degrees of freedom $r = 10$.
(b) Hyperparameters (m, w) have the usual reference prior w^{-1}.
(c) Dirichlet precision parameter $\alpha \sim G(10, 10/a)$ for some specified prior mean a.

An analysis summarized in detail uses $a = 0.01$. From Table 1, values of α this small induce priors that are concentrated on one or two components, giving small but non-zero mass to higher values. The simulation analysis is based on a Gibbs sample of size 5000 consecutive sequential draws, following burn-in of 2000 iterations: 1000 initial iterations with α fixed at 0.01 are run to burn-in on all parameters but α, followed by 1000 including α with the specified prior. This approach has been empirically determined to improve convergence when learning α. Some summary inferences for k appear in the first row of Table 2, quoting direct Monte Carlo sample frequencies representing the posterior for k. In spite of the extremely high prior mass on $k = 1$, the posterior is evidently overwhelmingly in favour of $k = 7$ or $k = 8$, most likely $k = 7$. Consistency with the raw data is apparent from the fitted density estimate interpolating the histogram in Fig. 2(a). This is simply the Monte Carlo average of the conditional posterior predictive densities $\Sigma_{j=0}^{k} p_j N(\theta_j, v_j)$ of item (f) in Section 2, denoted $p(y|D)$ where D denotes the observed data. Fig. 2(b) redisplays this density with that of the noise data analysis (just a Student t-density) for comparison. From the viewpoint of density estimation, simply smoothing the data, the model incorporates data-based estimation of the

TABLE 2
$P(k|D)$ for c2epsp10 noise ($n = 349$)

$E(a)$	$P(k\|D)$-values for the following values of k:												
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*≥13*
0.01							0.90	0.10					
0.1						0.41	0.39	0.16	0.04				
1.0							0.01	0.03	0.08	0.11	0.14	0.16	0.47

key smoothing parameters k and w, and also addresses issues of uncertainty in the density estimation—Fig. 2(c) again displays the predictive PDF but now with range bands indicative, in a broad sense, of posterior uncertainties; Fig. 2(d) displays similar range bands for the cumulative predictive distribution, superimposed on the empirical CDF of the data. Such plots provide useful visual checks on goodness of fit.

Fig. 2(e) exhibits the predictive density estimate with approximate posterior means of the component weights as vertical bars located at the approximate means of the ordered $\theta_{(j)}$. This gives one graphical summary of mixture deconvolution (that might be refined by adding indications of posterior uncertainty about the weights and locations). At this crude level, locations of the second to sixth components appear roughly consistent with quantal spacing, with $\theta_{(2)}$ the zero-response level. Beyond this, however, are an apparent component, with some appreciable mass, at negative levels near -1, and some activity at much higher levels 8–9. The former is a clear single component; investigating the experimental records associated with such low signal levels is now in order to investigate this feature. This was done, retrospectively evaluating the actual time traces of the original experiment; it became clear in retrospect that a large subset of the negative signals recorded were clustered in time around patterns of *inhibitory*, rather than excitary, neural activity. At these time stages the neural responses were quite atypical, for reasons as yet unknown, and so contribute to a component with negative mean that is irrelevant to the assessment of excitary response mechanisms. One suggestion might be to remove these identified inhibitory responses from the data set before a reanalysis. However, the mixture framework is robust and sufficiently flexible to accommodate the phenomenon by adding a mixture component that directly isolates the discrepant data, and simply recognizing this in summary inferences is appropriate. Similarly, the cluster at very high levels involving just very few—six out of the 349—observations is essentially discardable as erroneous measurements though necessarily subject to the same retrospective investigation of experimental records. We are then left with precisely five 'real' components with locations spaced roughly one (quantal?) unit of 1 meV apart from zero to five units.

Additional posterior analysis allows a more formal assessment of values of component locations, their spacing, variances and component weights. We note here some summary inferences for the $\theta_{(j)}$ directly, commenting further on these issues in the next section. As we have briefly detailed earlier, posteriors $p(\theta_{(j)}|D)$ may be approximately evaluated. Fig. 3 gives density plots for the unconditional (on k) margins $\theta_{(j)}|D$ for $j = 1, \ldots, 9$ in this analysis. Note the high precision with which the first seven components are estimated. Note also the multimodality that arises with intermediate components. This is common (West and Cao, 1993) and natural; if there are precisely seven components, for example, then $\theta_{(6)}$ is very probably around 3.7–4.4, whereas having eight components would shift $\theta_{(6)}$ downwards to the region of 2.5–3.0; hence the bimodality of the density for $\theta_{(6)}$ in Fig. 3. Note also that this phenomenon induces posterior means $E(\theta_{(i)}|D)$ that tend to be biased downwards, as can be seen by comparing the identified means in Fig. 2(e) with the corresponding densities in Fig. 3. The densities more strongly indicate support for the quantal hypothesis, i.e. that of equal spacing of the means, than is evident from the cruder summary in Fig. 2(e). Finally, note the overlap of posteriors for locations $\theta_{(i)}$ for $i = 7$, 8 and 9; this indicates that a

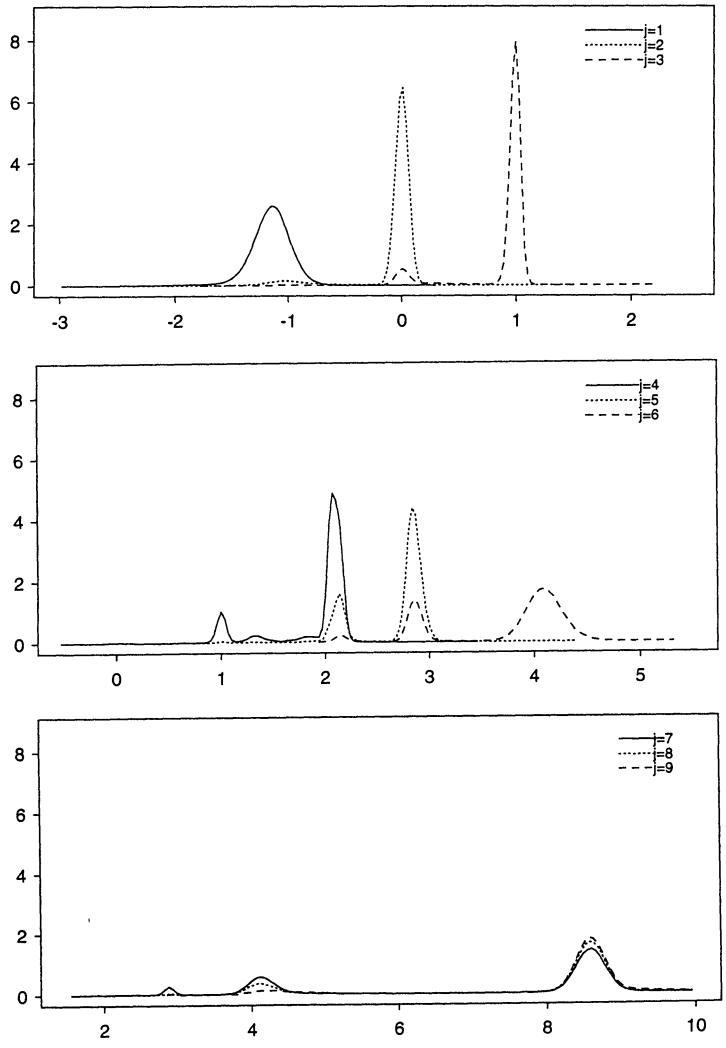

Fig. 3. Posteriors for ordered component means $\theta_{(j)}, j = 1, \ldots, 9$

seven-component model is appropriate, which is consistent with the high posterior probability on $k = 7$.

From Fig. 3, ignoring the inhibitory component around 0 and the outlier component around 8–9, the posteriors isolate the five roughly equally spaced component locations. Under the hypothesis that synaptic sites release transmitter independently and with equal chances, the apparently constant spacing of response levels is consistent with a binomial–quantal mechanism and with exactly four release sites at this neural junction. The component mean at zero represents no release, that at one unit represents that just one of the four sites releases, and so on up to all four sites releasing at the average level near four units. With a constant release probability π at any site, the level of $x = 0, \ldots, 4$ is achieved with binomial chance $\binom{4}{x}\pi^x(1-\pi)^{4-x}$, which should match the heights of the component weights displayed in Fig. 2(e) (after rescaling the weights to sum to 1 over the five locations of interest by ignoring the inhibitory and outlying component). The rescaled weights (as estimated by simulation averages displayed in Fig. 2(e)) are (very) approximately 0.09, 0.29, 0.34, 0.21 and 0.07.

Matching the 0.09 here to the binomial chance at 0, just $(1 - \pi)^4$, leads to $\pi \approx 0.45$. The corresponding binomial chances across $x = 0, \ldots, 4$ are then 0.09, 0.30, 0.37, 0.20 and 0.04. At this exploratory level, there is therefore remarkably close agreement with the (completely unconstrained) component weights in the mixture analysis and those implied by a binomial distribution with release chance near 0.45. Hence, we deduce apparently quite strong indications that the underlying physiology is characterized by

(a) exactly four synaptic transmission sites,
(b) independent release across sites with a constant release probability of around 0.45 and
(c) quantal transmitter release measured in terms of an induced potential of around 1 meV.

Additional analysis may be performed conditionally on any specified value of k rather easily. For example, in the simulation of posteriors described earlier suppose that we create a file containing all sampled parameters for which the sampled value of k is $k = 7$. Then approximate posterior and predictive inferences conditional on assuming a mixture of exactly seven components are available by using the ideas above but based on just this slice through the simulated posterior. In this sense, the Dirichlet mixture framework encompasses more traditional models that assume a fixed and specified value of k. In the current data analysis, conditioning on $k = 7$ refines inference about deconvolution minimally in this example. The analogue of Fig. 2(e) conditional on $k = 7$ shows essentially the same density function (and no discernible difference in predictive CDFs). In the constrained analysis, the posterior for $\theta_{(7)}$ simply reconciles the three densities in the lower frame of Fig. 3 for the unconstrained analysis. The deconvolution analysis is essentially unaffected otherwise.

5. General discussion and comments on current developments

Some further features of analysis, and comparison with other analyses, are now mentioned, as are issues of practical import in deconvolution analysis and some model extensions.

5.1. *Overfitting and parameter dimension*

Traditional statistical measures of goodness-of-fit and associated significance tests are commonly used to assess model analyses and often purportedly to validate the resulting inferences. The comfort derived from seemingly acceptable fits may be illusory. Mixture modelling with a facility for data-based estimation of the number of mixture components is a case in point: increasing the number of components allows the model to obtain closer fidelity to the observed data and to run the risk of overfitting. A finite sample histogram of data from a low component mixture will exhibit sampling variation that mixtures with larger numbers of components will more adequately imitate; hence the potential for overfitting. The problem is parameter dimension—generally, the likelihood function will favour larger values of k than will, from subjective assessment of graphical fit (e.g. Figs 2(a) and 2(d)), provide adequate interpolation of observed data configurations. In the Dirichlet mixture context, the prior for k is driven by that for α, and exploration of likelihood functions for α across ranges of data analyses has empirically confirmed this theoretically implied phenomenon. West (1990) illustrated this in developing, among other things, an algorithm for approximate maximum likelihood estimation of the number of components. Escobar and West (1992) also discussed estimation of the number of possible modes in a population distribution in the context of providing lower bound inferences on the number of components.

For the c2epsp10 analysis, a prior with α near 1 implies, from Table 1, moderate support across the range of 3–10 components, consistent with subjective assessment of the noise and signal histograms. Repeating the analysis of Section 4 with a single change—taking $E(\alpha)$ equal to 1.0 rather than to 0.01—leads to posterior probabilities for k in the third row of Table 2. Note the support for much larger k-values due to the overfitting fidelity factor. In this

analysis, several apparent components are interleaved across the range 0–6 of the data, each with smaller weight. As a result, the posteriors for the component locations are very much more diffuse, whereas the resulting predictive density is essentially similar to that of preceding analyses (and hence traditional goodness-of-fit tests will not distinguish the analyses). There is no purely data-based resolution of this issue: without additional control (through prior distributions or direct intervention), overfitting is essentially implied whatever the information in the data set. Traditional approaches have been to reject individual components on the basis of some kind of hypothesis test. This can be incorporated in the current analyses—Fig. 3 indicates the irrelevance of more than seven components, for example. Also, using priors for k—induced by priors for α—that strongly favour rather small values, such as that with $E(\alpha) = 0.01$ in the c2epsp10 analysis above, attempts to control for overfitting. The notion is that controlling α this way will lead to parsimony in parameterization of the model, restricting the posteriors to favour smaller numbers of components, and hence smaller numbers of parameters. In this sense, the resulting inferences will provide lower bound implications for the numbers of components. Thus the conclusion for the $k = 7$ posterior above determines a conservative, minimum acceptable description. It may also be effective to constrain partially separation between components to exceed some minimum threshold value—such as has been traditionally used with some maximum likelihood estimation methods (Redman, 1990) to overcome the noise bias limitations. Simple threshold constraints are very specific cases of more flexible stochastic constraints in which the prior distribution for component locations are modified to embody the view that spacings are, in some sense, unlikely to be very small. In the context of quantal hypotheses about spacings, this makes sense and also addresses the difficulties of resolving close components. Work in this direction will be reported elsewhere.

5.2. *Mixture noise distributions*

Normality of synaptic noise is very constraining and highly questionable (Kullman, 1989). In particular, spontaneous synaptic transmission from single or small numbers of sites will skew noise to higher values, so that observed noise ensembles often exhibit considerable skewness and even, though less often, multimodality. Resulting biases in overestimating component numbers may be severe. To model this, we can simply adapt the mixture modelling technology for the noise as well as signal data analysis, with appropriate links between the two. Explicitly, current developments of the approach involve the following modelling extension.

Write e_i for a noise observation. The set e_i $(i = 1, \ldots, n^*)$ is assumed to be a random sample from some normal mixture

$$F_e(e) = \int N(\phi, z)\, \mathrm{d}G_e(\phi, z)$$

where $G_e(\ ,\ )$ has a Dirichlet process prior. The signal response values y_i $(i = 1, \ldots, n)$ are modelled as $y_i = \mu_i + e_{n+i}$ where e_i $(i = n^* + 1, \ldots, n^* + n)$ is a further set of n independent draws from $F_e(e)$ and the locations μ_i are drawn from an uncertain prior $G(\ )$ which has an independent Dirichlet process prior. With this structure, the resampling, simulation-based analysis of Escobar and West (1992) may be generalized.

5.3. *Assessment of hypotheses*

Our modelling approach explicitly adopts a neutral position on the issue of possible scientific hypotheses about the mechanisms generating the mixture structure. This permits, in principle, a posterior assessment of posited hypotheses about restrictions on the parameters. The simplest example of constant quantal increments of the component spacing is a case in point. If such a hypothesis turns out to be at odds with the posterior distributions describing

the observed data from the neutral standpoint, then the hypothesis is discredited, at least in so far as it fails to describe observed patterns of variation in the specific data set under analysis adequately.

To address these issues the posterior distributions arising in a given analysis must be summarized and post-processed to investigate consistency with posited parameter constraints appropriately. Some of this may be performed subjectively via direct display of posterior summaries. Fig. 3, for example, suggests consistency with constant quantal spacing for components 2–6. Other hypotheses, including specific forms of irregular spacing of primary components and the inclusion of necessary sums of components with appropriate probabilities, may be explored similarly. It is of further interest to attempt to quantify consistency of posterior inferences with a hypothesis more formally. To do this will require the development of theoretical issues to provide tools to measure the support, under any given posterior distribution, for possibly very complicated sets of parameter constraints. In the constant quantal response scenario, for instance, the constraints involve equal spacing of component means but the spacing is unknown; how to measure formally the consistency of such a hypothesis with an arbitrary, unconstrained posterior distribution for the spacings determined is an open question. The peculiarities of the simulation-based Bayesian analysis provide a large amount of summary information about the posterior distribution, in terms of simulated draws of spacing values as well as theoretical estimates, that may eventually be drawn on in exploring approaches to the problems of formal assessment of hypotheses. Additional implications include the assessment of hypotheses about extra-noise variation, such as is explicit in quantal response hypotheses with individual levels of random variation in the quantal amplitudes.

5.4. *Comparisons of mixtures*

Evaluations of differences between distributions of synaptic response over time and with induced experimental alterations (such as long-term potentiation or depression) also require more formal methods for the comparison of two or more separate mixture distributions. Physiological examples of experimental manipulations include alterations in signal content with potentiation and changes in synaptic signals with pharmacological agents. Several hypotheses are currently being evaluated regarding mechanisms of potentiation, including an increased probability of release at presynaptic terminals, increased release of neurotransmitter, changes in post-synaptic sensitivity to the released transmitter and the formation of new synapses. An additional example concerns the use of agents which change release parameters or post-synaptic sensitivity to the neurotransmitter substance. Detecting these changes and concluding that a change has occurred are difficult inference questions.

5.5. *Non-stationarity of signal records*

Preliminary work is under way to extend mixture models to a time series context. In essence, the data originate as a time series—actually a bivariate interleaved time series of noise and signal recordings. The physiological background is that the stimulation of synaptic signals results in changes over time, depending on the frequency and intensity of the stimulation. Thus a stationary segment of data is a small subsample of a larger process which is probably varying in both unexpected and experimentally induced ways. A very general and highly global statistical model of synaptic potentials would include a dynamic description of small subsegments of the signal over time, particularly as a function of recent experience. To model this we are studying a synthesis of mixture models with Bayesian dynamic modelling and time series concepts (West and Harrison, 1989), with a global goal of incorporating possible time variation in mixture distributions and allowing for non-stationarity, in a general sense, in observed data ensembles by permitting defining parameters to be unstable over time.

5.6. *Simulation studies*

Simulation studies that are currently under way enable the bench-marking and tuning of model analyses, especially with respect to the issue of performance characteristics in detecting and estimating components. Within this, key issues that can be addressed via simulation include the assessment of noise bias limitations and the sensitivity to non-normal noise distributions.

5.7. *Physiological validation*

In these and other development areas, interaction between the theoretical development and physiological data, sometimes involving feed-back to experimental studies to provide physiological validation of statistical inferences, is a critical feature of the research.

Acknowledgements

Mike West was partially supported by the National Science Foundation under grants DMS-8903842 and DMS-9024793. Dennis Turner was supported by a merit review award from the Durham Veteran's Administration Medical Center, a Burroughs–Wellcome travel award and National Institute for Neurological Disorders and Stroke grant RO1 NS29482-01. Guoliang Cao assisted with programming and computation for some of the figures. We are grateful for the comments of a referee on an earlier version of the paper.

References

Cao, G. and West, M. (1992) Computing distributions of order statistics. *Discussion Paper 92-A02.* Institute of Statistics and Decision Sciences, Duke University, Durham.

Escobar, D. M. and West, M. (1992) Bayesian density estimation and inference using mixtures. *J. Am. Statist. Ass.*, to be published.

Kullman, D. M. (1989) Application of the expectation-maximisation algorithm to quantal analysis of postsynaptic potentials. *J. Neursci. Meth.*, **30**, 231–245.

Kullman, D. M. and Nicoll, R. A. (1992) Long-term potentiation is associated with increases in quantal content and quantal amplitude. *Nature*, **357**, 240–244.

Malgaroli, A. and Tsien, R. W. (1992) Glutamate-induced long-term potentiation of the frequency of miniature synaptic currents in cultured hippocampal neurons. *Nature*, **357**, 134–139.

Redman, S. (1990) Quantal analysis of synaptic potentials in neurones of the central nervous system. *Physiol. Rev.*, **70**, 165–198.

Turner, D. A. (1987) Identification of components of evoked composite EPSPs in CA1 hippocampal pyramidal neurons. *Neursci. Abstr.*, **13**, 155.

Turner, D. A. and Schlieckert, M. (1990) Data acquisition and analysis system for intracellular neuronal signals. *J. Neursci. Meth.*, **35**, 241–251.

Turner, D. A. and West, M. (1993) Statistical analysis of mixtures applied to postsynaptic potential fluctuations. *J. Neursci. Meth.*, **47**, 1–23.

Walmsley, B., Edwards, F. R. and Tracey, D. J. (1987) The probabilistic nature of synaptic transmission at a mammalian excitatory central synapse. *J. Neursci.*, **7**, 1037–1046.

——— (1988) Nonuniform release probabilities underlie quantal synaptic transmission at a mammalian excitatory central synapse. *J. Neur. Physiol.*, **60**, 889–908.

West, M. (1990) Bayesian kernel density estimation. *Discussion Paper 90-A02.* Institute of Statistics and Decision Sciences, Duke University, Durham.

——— (1992) Hyperparameter estimation in Dirichlet process mixture models. *Discussion Paper 92-A03.* Institute of Statistics and Decision Sciences, Duke University, Durham.

West, M. and Cao, G. (1993) Assessing mechanisms of neural synaptic activity. In *Bayesian Statistics in Science and Technology: Case Studies* (eds C. Gatsonis, J. Hodges, R. Kass and N. Singpurwalla).

West, M. and Harrison, P. J. (1989) *Bayesian Forecasting and Dynamic Models.* New York: Springer.