Biophysical Journal Volume 67 August 1994 656–670

Statistical Models of Synaptic Transmission Evaluated Using the Expectation-Maximization Algorithm

Christian Stricker and Stephen Redman
Division of Neuroscience, John Curtin School of Medical Research, Australian National University, Canberra ACT 0200 Australia

ABSTRACT Amplitude fluctuations of evoked synaptic responses can be used to extract information on the probabilities of release at the active sites, and on the amplitudes of the synaptic responses generated by transmission at each active site. The parameters that describe this process must be obtained from an incomplete data set represented by the probability density of the evoked synaptic response. In this paper, the equations required to calculate these parameters using the Expectation-Maximization algorithm and the maximum likelihood criterion have been derived for a variety of statistical models of synaptic transmission. These models are ones where the probabilities associated with the different discrete amplitudes in the evoked responses are a) unconstrained, b) binomial, and c) compound binomial. The discrete amplitudes may be separated by equal (quantal) or unequal amounts, with or without quantal variance. Alternative models have been considered where the variance associated with the discrete amplitudes is sufficiently large such that no quantal amplitudes can be detected. These models involve the sum of a normal distribution (to represent failures) and a unimodal distribution (to represent the evoked responses). The implementation of the algorithm is described in each case, and its accuracy and convergence have been demonstrated.

GLOSSARY

β	shape parameter for gamma distribution
ϵ	offset (from zero) of distribution
g_{1j}	probability density of first noise component convolved with the jth amplitude component
g_{2j}	probability density of second noise component convolved with the jth amplitude component
γ	shape parameter for Weibull distribution
δ	scale factor for Weibull distribution
f_i	frequency of observation x_i
$K + 1$	number of components
L	log-likelihood
λ	scale factor for gamma distribution
μ_j	mean of jth component
μ_n	mean of the noise distribution
μ_{n1}	mean of first noise component
μ_{n2}	mean of second noise component
$M(x, \theta)$	mixture distribution
n	number of release sites in the binomial and compound binomial model
N	sample size
p	binomial release probability
p_r	release probability at rth site
PD	probability density
P_j	probability associated with the jth component
π_0	proportion of failures caused by failure to stimulate
π_n	probability of first noise component
q_{ij}	density of x_i at jth component
Q	quantal amplitude
σ_j	SD of jth component
σ_n	SD of the noise distribution
σ_{n1}	SD of first noise component
σ_{n2}	SD of second noise component
σ_Q	SD of the quantal amplitude
θ	parameter vector
x_i	ith observation

Received for publication 29 December 1993 and in final form 4 May 1994.

Address reprint requests to Stephen Redman, Division of Neuroscience, John Curtin School of Medical Research, Australian National University, GPO Box 334 Canberra ACT 0200 Australia. Tel.: 61-6-249-2602; Fax: 61-6-249-2687; E-mail: beth@eccles.anu.edu.au.

0006-3495/94/08/656/15 $2.00

INTRODUCTION

Trial-to-trial variations in the amplitude of evoked synaptic currents (or voltages) can be used to extract information about the underlying mechanisms of synaptic transmission. The parameters of interest are the probabilities of transmitter release at the different release sites involved in transmission, the number of these release sites, the amplitudes of the synaptic current generated at each active site, and the variability associated with these amplitudes. The usual approach to estimating these parameters is to assume that the evoked currents are distributed as a mixture of normal distributions. Each normal distribution is centered on a discrete amplitude, and these amplitudes (in increasing order) correspond to response failures, the response caused by the release at any one release site, the response caused by simultaneous release at any combination of two release sites, and so on. The relative contribution that each normal distribution makes to the mixture is the total probability of release for the appropriate combination of release sites. Obviously, for this scheme to result in a mixture distribution with clear peaks that are approximately equally spaced, the synaptic responses arising from transmission at different release sites must be roughly equal when recorded at the soma, and the variability associated with these responses must not be large compared with their average amplitudes. The important assumptions are that the noise and the EPSC are statistically independent random variables, that they are stationary processes, and that the noise and the EPSC add linearly.

The purpose of this paper is to provide procedures for determining the parameters of a mixture model of synaptic responses. One method of calculating these parameters is to adjust them until an optimal fit to the measured probability density is achieved, using a maximum-likelihood criterion. An algorithm was developed by Hasselblad (1966) for this purpose. A more formal development of this method, called the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) has been extended to apply to finite mixtures of a wide class of density functions (reviewed in Redner, 1984; McLachlan and Basford, 1988). The EM algorithm has also been applied to quantal analysis by Ling and Tolhurst (1983) and Kullmann (1989).

However, when all parameters of the mixture are allowed to vary without constraints, too many free parameters exist for this fitting procedure to provide reliable estimates (Kullmann, 1992). In most preparations where statistical analysis of fluctuations in evoked responses is being attempted, constraints can be applied to the mixture model to reduce the number of free parameters. One constraint is to require the mean amplitudes to be separated by equal increments (the quantal amplitude), and for the variances associated with each quantal increment in amplitude to also increment by the quantal variance in addition to the noise variance (which is the same for each amplitude). This is the conventional model of quantal release. Further constraints can be imposed on the probabilities associated with the different mean amplitudes including a binomial distribution (uniform release probabilities) or a compound binomial distribution (unequal release probabilities). In this paper, we use the EM approach to derive the recursive equations for quantal amplitude, quantal variance, the (uniform) probability of release for the binomial model and the nonuniform release probabilities for the compound binomial model. These equations differ from those provided by Kullmann (1989), and they converge rapidly, reliably, and with high accuracy.

An alternative model of the transmission process is to assume that there is so much variability associated with transmission at different combinations of release sites that the evoked synaptic current is best described by a skewed unimodal distribution (Bekkers et al., 1990; Clements, 1991; Jonas et al., 1994). We have referred to these models as nonquantal models, because their amplitude distribution will not be multimodal even in the absence of recording noise. An additional distribution must be added to this unimodal density to account for failures in transmission. We have used the noise distribution for this purpose, with a variable offset for the mean. The EM algorithm can also be used to obtain the best fit to the measured probability density under these circumstances. We have derived recursive equations to obtain the parameters for a mixture of a normal distribution (representing failures) with 1) a normal distribution (Clements, 1991), 2) a gamma distribution (McLachlan, 1978), 3) a Weibull distribution, and 4) a cubic transformation of a normal random variable (Bekkers et al., 1990).

Finally, we have used the EM algorithm, together with the binomial or compound binomial constraint on probabilities, to allow for two different types of failures in evoked response: 1) a failure caused by irregular stimulation of afferent axon(s), and 2) a genuine failure to release transmitter despite the presence of a presynaptic action potential. In this situation, the algorithm can be used to calculate the proportion of failures caused by failure to stimulate.

In this paper, we show how the parameters for the different models can be found, and demonstrate that the equations return the correct parameters for each model. The previous paper (Stricker et al., 1994) presents methods for calculating confidence limits on the model parameters obtained from the experimental sample. It also shows how the maximum log-likelihood obtained when fitting a model using the EM algorithm can be used in a unified approach to model comparison.

THEORY AND METHODS

The parameters of a mixture of $(K + 1)$ normal distributions are the means and the variances $\{\mu_0, \sigma_0^2; \mu_1, \sigma_1^2; \cdots; \mu_K, \sigma_K^2\}$ associated with them and the probabilities associated with each amplitude $\{P_0, P_1, \cdots, P_K\}$. Let the vector θ contain all these parameters. The mixture distribution is then defined as

$$M(x, \theta) = \sum_{j=0}^{K} P_j \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-(x-\mu_j)^2/2\sigma_j^2}, \tag{1}$$

where

$$\sum_{j=0}^{K} P_j = 1.$$

An evoked response is measured N times, with values $\{x_1, x_2, \cdots, x_N\}$. It is assumed that this sample is drawn from the probability density $M(x, \theta)$. The problem is to find the parameters of the mixture $M(x, \theta)$ that gives the best fit to the observations by using the likelihood criterion. To do this we define the log-likelihood function L for this sample as

$$L = \sum_{i=1}^{N} f_i \ln(M(x_i, \theta)), \tag{2}$$

where f_i is the frequency of the observation x_i. The next step is to find the values of the parameters that maximize the log-likelihood function L. This is done by differentiating L with respect to each of the unknown parameters, setting the derivatives equal to zero, and solving these equations in a recursive manner. In the case of Eq. 1, this means solving $(\partial L/\partial P_j) = 0$; $(\partial L/\partial \mu_j) = 0$; $(\partial L/\partial \sigma_j) = 0$ for P_j, μ_j, and σ_j, respectively. The equations require initial guesses for the optimal parameters, and these values are used in the equations to modify the initial guesses. This process is then repeated many times. The essence of the problem is to ensure convergence of the procedure to the parameter values that maximize the likelihood. This is guaranteed by the EM algorithm. All of the derivations of the recursive equations are given in the Appendix. Here we discuss the implementation of the algorithm for each model.

Unconstrained model

In this model, there are no constraints on the P_js, μ_js, and σ_js, except that $\sum_{j=0}^{K} P_j = 1$. An assumption is made about K, and initial estimates of P_j, μ_j, and σ_j are chosen. The μ_js must all be different. Equation A3 is used to calculate a new set of P_js. These new values of the P_js are used in Eqs. A4 and A5 to calculate new sets of μ_js and σ_js (Kullmann, 1989). Eventually, after numerous iterations of this procedure, the P_js are only altered by insignificant amounts, and the procedure is terminated.

Unless there exists very little overlap of the normal distributions in the mixture, this model will provide too many degrees of freedom to the fitting procedure, leading to unreliable estimates. In some instances, it might be appropriate to assume negligible quantal variance, in which case the variance associated with each amplitude is identical and equal to that of the recording noise. In this case, only Eqs. A3 and A4 are used.

If the correct value for K is not obvious, as will often be the case with a small sample size or a poor signal-to-noise ratio (quantal amplitude divided by noise SD), the procedure should be repeated for progressively larger values of K. Eventually, some of the P_js will be zero, or some of the μ_js will become very similar (with smaller separation than the noise variance). The questions of how to determine the most reliable value of K, and the confidence limits of the parameter estimates for a given sample size and signal-to-noise ratio, have been discussed in the preceding paper (Stricker et al., 1994).

One important application of this model is to find the parameters of two normal distributions whose sum is used to represent the recording noise. Often the density function for the recording noise is skewed (Kullmann, 1989). It can be represented as the sum of two normal distributions (Eqs. A6 and A7). To find the parameters $\{\pi_n; \mu_{n1}, \sigma_{n1}; \mu_{n2}, \sigma_{n2}\}$ set $K = 1$, use Eq. A3 to update π_n, and use the updated value of π_n in Eqs. A4 and A5 to update μ_{n1}, μ_{n2}, and σ_{n1}, σ_{n2}. An example of this procedure is given in Results.

Quantal models

Unconstrained probabilities

In this model, it is assumed that the separation between μ_j and μ_{j+1} is Q, the quantal amplitude. It is further assumed that the variance associated with μ_j, i.e., $\sigma_j^2 = \sigma_n^2 + j\sigma_Q^2$, where σ_n^2 is the noise variance and σ_Q^2 is the quantal variance (see Eq. A12). Because the μ_js cannot be adjusted independently, any offset in the average amplitude of the failures will be transferred equally to all the μ_js. An offset can arise from a stimulus artefact or a field potential. An offset parameter (ϵ) can be incorporated into the parameter set to be estimated. When this is done, the free parameters in this model are the P_js ($j = 0, 1, \cdots, K - 1$), Q, σ_Q^2, and ϵ. The equations for calculating these parameters are Eqs. A3, A9, A10, and A11, respectively, when the noise is represented by a single normal distribution, and Eqs. A3, A13, A14, and A19 when the noise is represented by the sum of two normal distributions. Once the appropriate parameters for the noise have been determined, initial guesses are made for the P_js, Q, σ_Q^2, and ϵ. In the case of a single gaussian distribution, the initial guesses are applied to Eq. A3 to obtain a revised set of P_js. These new P_js, and the initial guesses for the other parameters are applied to Eq. A9 to obtain a revised value for Q, to Eq. A10 to obtain σ_Q^2, and to Eq. A11 to obtain ϵ, and then the cycle is repeated until the changes in the P_js are negligible. One variant of this model is to assume no quantal variance. In this case $\sigma_j^2 = \sigma_n^2$ and the equation for σ_Q^2 is omitted from the recursive procedure.

Binomial probabilities

In this model, the P_js are constrained to satisfy the equation

$$P_j = {}^K C_j (1 - p)^{K-j} p^j, \tag{3}$$

where p is the release probability at each of the K release sites and ${}^K C_j$ is the binomial coefficient, sometimes written as $\binom{K}{j}$. The equation needed to calculate p is Eq. A18. Otherwise, the equations for calculating Q, σ_Q^2, and ϵ are the same as for the quantal model, except that the P_js must be calculated using Eq. 3 above. There are four free parameters associated with this model, once K has been fixed. The procedure is to assume initial values of p, Q, σ_Q^2, and ϵ, calculate the P_js from Eq. 3, and then proceed as for the quantal model.

Compound binomial probabilities

This model allows the release probabilities to be different for each release site, i.e., $p_1, p_2, \cdots, p_r, \cdots, p_K$. The relationship between these release probabilities and the P_js is given by Eq. A23. The equation for p_r is Eq. A22. The equations for Q, σ_Q^2, and ϵ are the same as for the quantal model. There are $K + 3$ free parameters with this model. After an initial guess at the values for $p_1, \cdots, p_r, \cdots, p_K$, Q, σ_Q^2, and ϵ, Eq. A23 is used to calculate the P_js, and Eq. A21 is needed to calculate the function $P_{r,j}$ defined by Eq. A20. When these values have been calculated, they are then used in Eq. A22 to calculate the p_rs. The P_js can then be calculated using Eq. A23. When this is completed, the quantal parameters Q, σ_Q^2, and ϵ are calculated as for the case of unconstrained probabilities and the appropriate equations from the quantal model. The constraints of a quantal variance and a quantal separation could be relaxed, in which case the appropriate equation for the μ_js is Eq. A4.

A special case has to be considered if there are no detectable failures. As shown in the Appendix, the derivation only holds if none of the p_rs is 1. However, as shown by Walmsley et al. (1988), this might not always be the case. When no failures occur, the offset bears the information about the number of additional release sites where the release probability is 1. The offset determines the magnitude of the smallest amplitude response. Scaling the offset by the quantal size gives the number of release sites with a release probability of 1 plus a remaining offset, which is again caused by stimulus artefact or field potential. The number of contributing release sites K^* is therefore $K^* = K + \mathrm{int}(\epsilon/Q)$, where $\mathrm{int}(\epsilon/Q)$ is the lower integer value of (ϵ/Q).

Non-quantal models

An alternative hypothesis to the previous model is based on the premise that the noise-free evoked responses are drawn from a distribution that is effectively continuous. In this scheme, the intrinsic variability in the responses originating at the same active site and the differences in amplitudes (at the soma) of the responses originating from different active sites combine to create a continuous distribution of amplitudes. Any apparent peaks in the observed amplitude distribution must be caused by finite sampling. The amplitude distribution of the measured currents is the sum of two probability densities. One represents the "failures," with a mean of zero, and has the same distribution as the recorded noise. The other represents the responses to released transmitter. Various density functions can be used for this continuous distribution. We have derived the equations when the continuous distribution is 1) a normal distribution, 2) a gamma distribution, 3) a Weibull distribution, and 4) a cubic transformation of a normal random variable. Here, Eq. 1 is modified to

$$M_i = (1 - P)q_{i1} + Pq_{i2}, \tag{4}$$

where q_{i1} is the probability density for the failure responses, $(1 - P)$ is the probability of failure to release transmitter, and q_{i2} is the probability density for the responses to released transmitter. q_{i1} is given by Eq. A27 when the noise is represented by the sum of two normal distributions. Otherwise, it can be a single normal density function. The subscript i denotes dependence on the ith sample (x_i).

Normal distribution

Here

$$q_{i2} = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-(x_i-\mu_2)^2/2\sigma_2^2} \tag{5}$$

and q_{i1} is given by Eq. A27. The unknown parameters are P, μ_1, μ_2, and σ_2. (The parameters describing the double gaussian noise (π_n; μ_{n1}, σ_{n1}; μ_{n2}, σ_{n2}) will have been obtained previously.) μ_1 (in Eq. A27) is an offset parameter to allow for the possibility that a stimulus artefact or a field contaminates the failure records. μ_1 is different from the previously introduced offset parameter because it only affects the failures peak, whereas ϵ offsets the entire distribution. Any offset required in the continuous distribution q_{i2} is embedded in μ_2. P is obtained from Eq. A3, with $K = 1$. μ_2 and σ_2 are obtained from Eqs. A4 and A5, and μ_1 is obtained from Eq. A8. The disadvantage of using a normal distribution to represent the nonfailures is that these responses are usually skewed towards larger amplitudes.

Gamma distribution

Here

$$q_{i2} = \frac{\lambda^\beta}{\Gamma(\beta)} x_i^{\beta-1} e^{-\lambda x_i}. \tag{6}$$

The parameters λ and β are commonly referred to as scale factor and shape parameter, respectively. This distribution is positively skewed. The unknown parameters are P, μ_1, λ, and β, and these are obtained using Eqs. A3, A8, A30, and A31, respectively. This function was previously used to describe the distribution of the amplitudes of spontaneous synaptic potentials in autonomic ganglion cells (McLachlan, 1975, 1978).

Weibull distribution

Here

$$q_{i2} = \gamma\delta x_i^{\delta-1} e^{-\gamma x_i^\delta} \tag{7}$$

This distribution (Freund, 1992, p. 233) depends on two parameters (γ and δ). It is very flexible in fitting distributions of any skewness. The unknown parameters are P, μ_1, γ, and δ, and these are found using Eqs. A3, A8, A33, and A34, respectively.

Cubic transform of a normal random variable

$$q_{i2} = \frac{1}{3\sqrt{2\pi}\,\sigma_2} x_i^{-2/3} e^{-((x_i^{-1/3}-\mu_2)^2/2\sigma_2^2} \tag{8}$$

This transform was used by Bekkers et al. (1990) for the amplitude distribution of spontaneous miniature synaptic currents and by Jonas et al. (1994) for evoked responses. The use of this distribution is predicted from a linear relationship between transmitter concentration at the postsynaptic receptors and the synaptic current. The reason behind the use of this transformed normal variable is that synaptic current was assumed to be proportional to the contents of a vesicle determined by volume. Because vesicle diameters are normally distributed (Bekkers et al., 1990), the current is distributed as the third power of a normal variable, giving rise to Eq. 8. The unknown parameters are P, μ_1, μ_2, and σ_2, and these may be obtained from Eqs. A3, A8, A36, and A37, respectively.

Failure to stimulate the afferent axon(s) and failure to release transmitter

Extracellular stimulus currents can be adjusted to be just above threshold for a single afferent axon or a small group of afferent axons. If the threshold current increases, or if the spread of the current from the stimulating electrode into the axon is altered by swelling or damage to the tissue surrounding the electrode, some of the failures to respond to a stimulus might be failures to stimulate the axon. It is possible to introduce another parameter to the analysis that will separate the two types of failures, but only when the release process is constrained to an explicit probabilistic model. Here we consider the partitioning of failures for both the binomial and compound binomial models, and Eq. 1 becomes

$$M_i = \pi_0 q_{i0} + (1 - \pi_0) \sum_{j=0}^{K} P_j q_{ij}, \tag{9}$$

where π_0 is the proportion of failures caused by failure to stimulate the axon. q_{ij} is defined by Eq. A12 or A16.

Binomial distribution

The P_js are defined by Eq. A17. The unknown parameters are π_0, probability of release p, quantal size Q, quantal variance σ_Q^2, and offset ϵ. To find π_0, we consider the probability density of all responses (including both types of failure) as a mixture of two distributions, as indicated in Eq. 9. In this case, we make use of the unconstrained model, with $K = 1$, and Eq. A3 is the appropriate equation for π_0, but written as

$$\pi_0^* = \sum_{i=1}^{N} \frac{f_i}{M_i} \pi_0 q_{i0}. \tag{10}$$

The equation for the release probability is altered from Eq. A18 by allowing some of the failures to be stimulation failures. The correct equation for p is Eq. A39. The offset is also altered by this modification, and ϵ is now given by Eq. A40 or A41. The equations for quantal size and quantal variance are independent of the way in which the failure peak is partitioned. These parameters are the same as for the binomial model *binomial probabilities* above.

Compound binomial distribution

The P_js in Eq. 9 are defined by Eq. A19; otherwise, the problem is identical to the one for a uniform binomial distribution. The equation for π_0 is as above (Eq. 10), as are the equations for the offset, quantal size, and quantal variance. The only equation to change is the one for release probabilities, and this is now given by Eq. A42.

Testing the convergence of the algorithms and the accuracy of parameter recovery

The algorithms described above were implemented as double precision IGOR Pro (WaveMetrics, Lake Oswego, OR) functions on a Macintosh computer. They were tested on probability density functions appropriate for each model. These density functions were calculated from the appropriate equations using parameter values that are representative for synapses on central neurones. The density functions were evaluated over a sufficient range of the random variable such that the integral of the function over this range was between 0.999 and 1.0. Thus, we tested the algorithms against a completely specified distribution, because this permits the accuracy with which the parameters are evaluated to be determined. To calculate the likelihood, the number of observations from which the density function was formed is required. In all simulations, this number was assumed to be 500.

We terminated the optimizations when the sum of all of the absolute changes in the P_js after an iteration was less than 10^{-6}. Other termination conditions dependent on minimal changes of other parameters (μ_js and σ_js) did not improve the accuracy and, therefore, were not implemented. For the models that required the numerical solution of a recursive equation, we required an accuracy of 10^{-7} in the estimation of the root.

RESULTS

The aim of this section is to demonstrate the convergence of the algorithm for each model discussed in the previous section, and to determine the accuracy with which the parameters can be estimated for a completely defined density function. Except for the continuous models, we have chosen a signal-to-noise ratio of 2.5. The number of components (K) was set to be 5. The same skewed noise distribution was used for all the simulations and was described by the sum of two normal density functions. We have chosen initial estimates of parameters to deviate from the known values by a minimum of 20% and a maximum of 200%. To test for the sensitivity of the starting values, 100 random estimates within the boundaries given were used for each probability density. All recursive schemes were insensitive to starting values, and the error introduced by the randomized estimates was small compared with the accuracy of the algorithm. Therefore, we defined the relative accuracy as the ratio of the absolute difference of the result and the correct value divided by the correct value. As predicted by theory, convergence of the log-likelihoods was monotonic for all models tested.

Deconvolution of skewed noise

A fully unconstrained deconvolution was applied to recover the parameters of a simulated noise distribution, which consisted of the sum of two normal probability densities. Equations A3, A4, and A5 were applied to the data that provided the distribution shown in Fig. 1 *A* ($\pi_n = 0.8$; $\mu_{n1} = -0.1$, $\sigma_{n1} = 0.8$; $\mu_{n2} = 0.4$, $\sigma_{n2} = 0.9$). To recover the parameters when signal-to-noise is small, as it is in this example, we set the convergence criterion to 10^{-7} (instead of 10^{-6} as described in Theory and Methods). Convergence in this case was obtained after more than 160,000 iterations with a relative accuracy of <0.17 for the P_js, <0.14 for the μ_js, and $<4 \times 10^{-3}$ for the σ_js. This example illustrates the known behavior of the EM algorithm when the signal-to-noise ratio is small. However, with even more rigorous convergence criteria as used here, it is possible to obtain more accurate values.

Unconstrained deconvolution

In Fig. 1 *B*, a deconvolution is shown where the P_js and μ_js are unconstrained and negligible quantal variance is as-

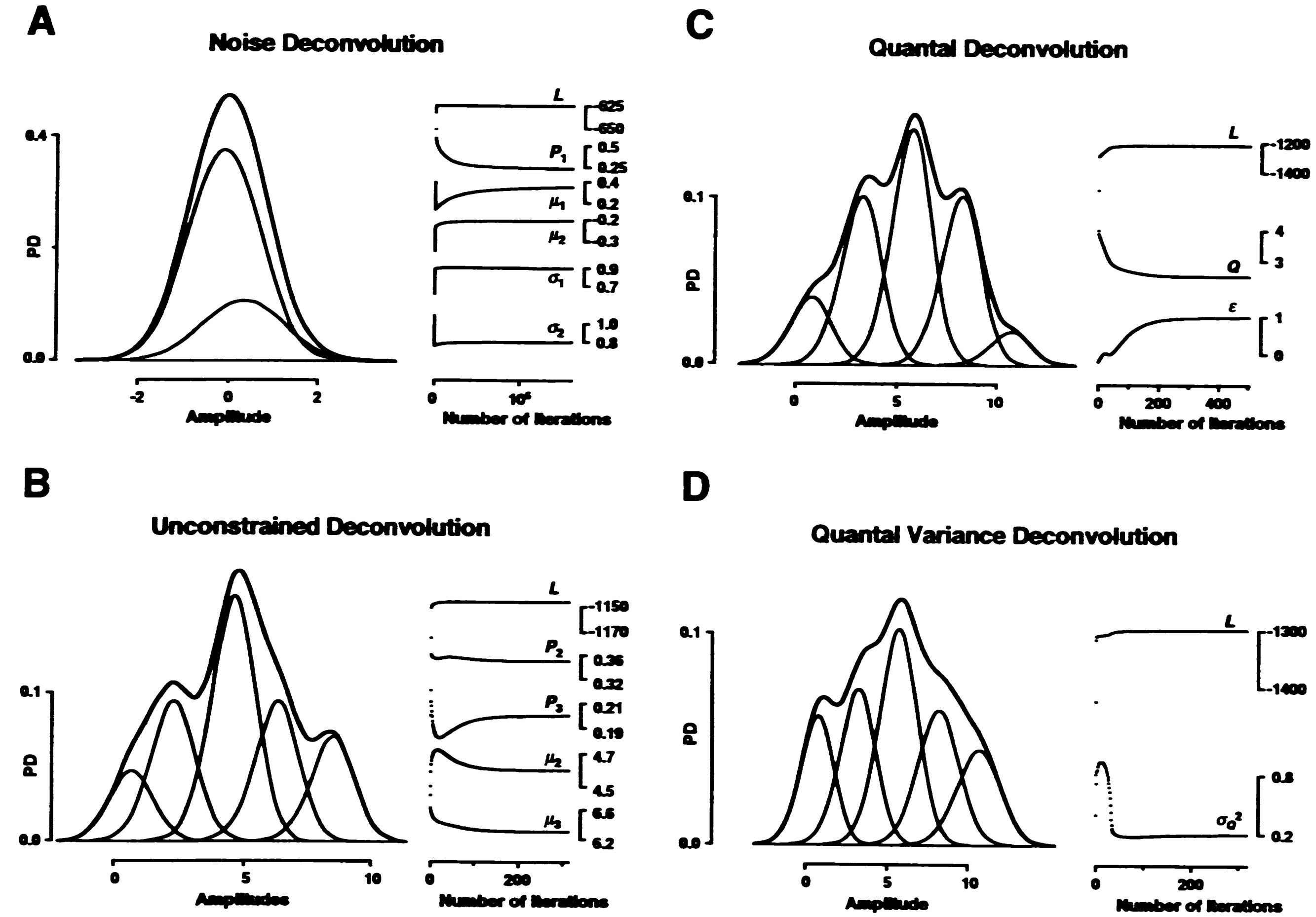

FIGURE 1 Unconstrained probability deconvolutions. The insets to the right of the figures show the convergence characteristics for particular parameters. L corresponds to log-likelihood, and PD corresponds to probability density. Data and reconvolutions superimpose in all figures. (*A*) Unconstrained deconvolution for a mixture of two gaussian components. (*B*) Deconvolution based on a noise with two gaussian components (*A*) allowing for unconstrained means. Insets show four out of eight parameters. (*C*) Deconvolution as in *B*, but where the means are constrained to show equal spacing (Q) and the multimodal distribution is shifted to the right (ϵ). (*D*) Deconvolution as in C with the inclusion of quantal variance (σ_Q^2).

sumed. The reason for presenting this example is that we can compare subsequent models with respect to speed of convergence and accuracy. In this example, the P_js were 0.1, 0.2, 0.35, 0.2, and 0.15; the μ_js 0.7, 2.3, 4.6, 6.3, and 8.5, and the noise was as described in the previous section. Starting conditions for the probabilities were chosen that differed by >20% from the known parameter values. We used Eq. A3 to calculate the P_js and Eq. A8 to obtain the μ_js. Convergence was reached after 314 iterations with a relative accuracy of $<1 \times 10^{-4}$ for the P_js and $<4 \times 10^{-4}$ for the μ_js.

Quantal deconvolution

In this simulation (Fig. 1 *C*), the probabilities were unconstrained, but the quantal amplitudes were constrained to show a constant separation with zero quantal variance. An offset of the density function (ϵ) was included. The P_js were 0.1, 0.2, 0.35, 0.2, and 0.15; Q and ϵ were 2.5 and 1, respectively. The noise was as described in Skewed Noise. We applied Eqs. A13 for Q and A14 for ϵ. Convergence was reached after 504 iterations from starting conditions well away from the known values. Relative accuracy was $<3 \times 10^{-4}$ for the probabilities, $<4 \times 10^{-4}$ for ϵ and $<6 \times 10^{-6}$ for Q.

Quantal deconvolution with quantal variance

Fig. 1 *D* shows a deconvolution that had the same parameters as the quantal deconvolution in the previous section but for which a quantal variance (σ_Q^2) was included and equalled 0.2. We used Eq. A13 for Q, Eq. A14 for ϵ, and Eq. A15 for σ_Q^2. Convergence was reached after 318 iterations with a relative accuracy for Q and ϵ as for the example in the previous section. Relative accuracy for quantal variance was 8×10^{-4}, which was well within the accuracy for the other quantal parameters. Note that within the first 40 iterations, σ_Q^2 increased and then decreased. We also tested the algorithm for the case when no quantal variance was included but when the starting conditions assumed a non-zero value. The algorithm for σ_Q^2 converged to zero. The rate, however, was usually slow, and up to 1000 iterations were required. The algorithm also works for negative variances, as long as $\sigma_K^2 > 0$.

Binomial deconvolution

In this simulation, all of the other parameters were the same as in the previous section except that the P_js were constrained to a binomial process with $p = 0.4$ (Fig. 2 *A*). Equation A13 was used to obtain Q, A14 for ϵ, A15 for σ_Q^2, and A18 for p. Although quantal separation and quantal variance were incorporated, it took only 294 iterations to converge to the final values, which is less than used in the previous section. Faster convergence could be expected, because the binomial assumption imposes a tight constraint on the P_js. Relative accuracy for p was $< 1 \times 10^{-4}$, and for the remaining parameters the accuracy was comparable to that obtained in the previous section.

Binomial deconvolution including stimulus failures

In this example, we illustrate Eq. A38 to obtain p when some of the failures were not caused by release failures (Fig. 2 *B*). We used the same parameters as in the previous section together with a parameter $\pi_0 = 0.2$ to account for the non-release failures, which was calculated according to Eq. A3. The final result was obtained after 264 iterations with a relative accuracy of $<3 \times 10^{-4}$ for π_0 and p. The reason convergence was faster in the present simulation than in the previous one relates to the distinct failure peak, which quickly forces the convergence of ϵ and Q into the correct range. We tested the algorithm for the case where there were no extra failures in the data but some were assumed in the initial estimates. The final result was obtained after about 1000 iterations with the correct answer of zero for the probability of additional failures.

Compound binomial deconvolution

In Fig. 2 *C*, we simulated a density function that had an underlying compound binomial release process. The p_rs were 0.8, 0.6, 0.4, and 0.2. For this particular case, the results were obtained using Eqs. A13 for Q, A14 for ϵ, A15 for σ_Q^2, and A22 for the p_rs. The final result was obtained after 1860 iterations, which was significantly more than was required for the binomial cases. Relative accuracy was $<2 \times 10^{-3}$ for

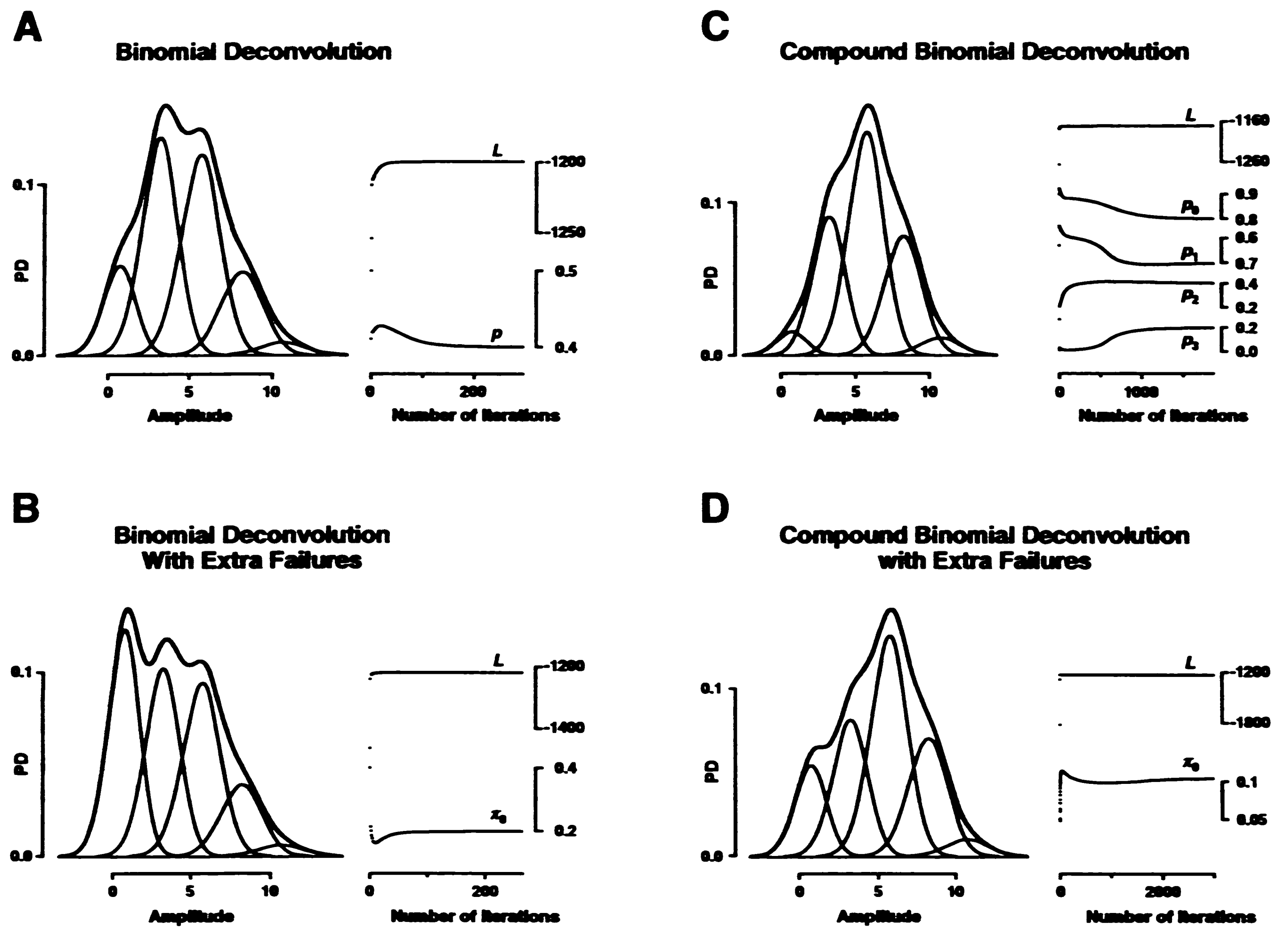

FIGURE 2 Deconvolutions with constrained probabilities. Insets to the right show convergence characteristics. All Figures are based on quantal spacing (Q) with offset (ϵ) and quantal variance (σ_Q^2). (*A*) Binomial probability constraint (p). (*B*) Binomial constraint including extra failures (π_0). (*C*) Compound binomial constraint (p_1, p_2, p_3, p_4). (*D*) Compound binomial constraint including extra failures (π_0).

the quantal parameters and $<8 \times 10^{-3}$ for the p_rs. Significantly faster convergence was observed when the signal-to-noise ratio was >2.5.

Compound binomial deconvolution including stimulus failures

In the example illustrated in Fig. 2 *D*, we used Eqs. A13 for Q, A40 for ϵ, A15 for σ_Q^2, A42 for the p_rs, and A3 for π_0. The parameter values were the same as for the previous section, and $\pi_0 = 0.1$. Monotonic convergence was reached after 2985 iterations, providing a relative accuracy of <0.06 for the p_rs and <0.04 for π_0. The other parameters were resolved with the same precision as in the previous section. Convergence to the final value took considerably longer than for the preceding examples. The accuracy for the p_rs was less than in the previous section because the precision with which π_0 can be resolved determines the accuracy with which the compound binomial parameters can be found. We tested this recursive scheme for an example where no extra failures were included but where the starting estimates assumed additional failures. Under these conditions, convergence was slower, and the recursive scheme finally reached a value for π_0 close to zero.

Non-quantal models

Because the amplitude distribution of the failure records corresponds to the noise distribution, we have chosen the same skewed noise distribution as that used in the previous examples. The amplitude distribution for the responses was assumed to be a continuous unimodal density function, and the four distributions considered in Theory and Methods were used.

Normal distribution

Fig. 3 *A* shows the two components in this mixture. The parameters for the response distribution were $P = 0.6$, $\mu_2 = 5$, $\sigma_2^2 = 2.5$, and $\mu_1 = 2$. In principle, the problem is not different from the noise deconvolution illustrated in Fig. 1 *A*. P, μ_2, σ_2^2, and μ_1 were calculated using Eqs. A3, A4, A5, and A8, respectively. The final results were obtained after

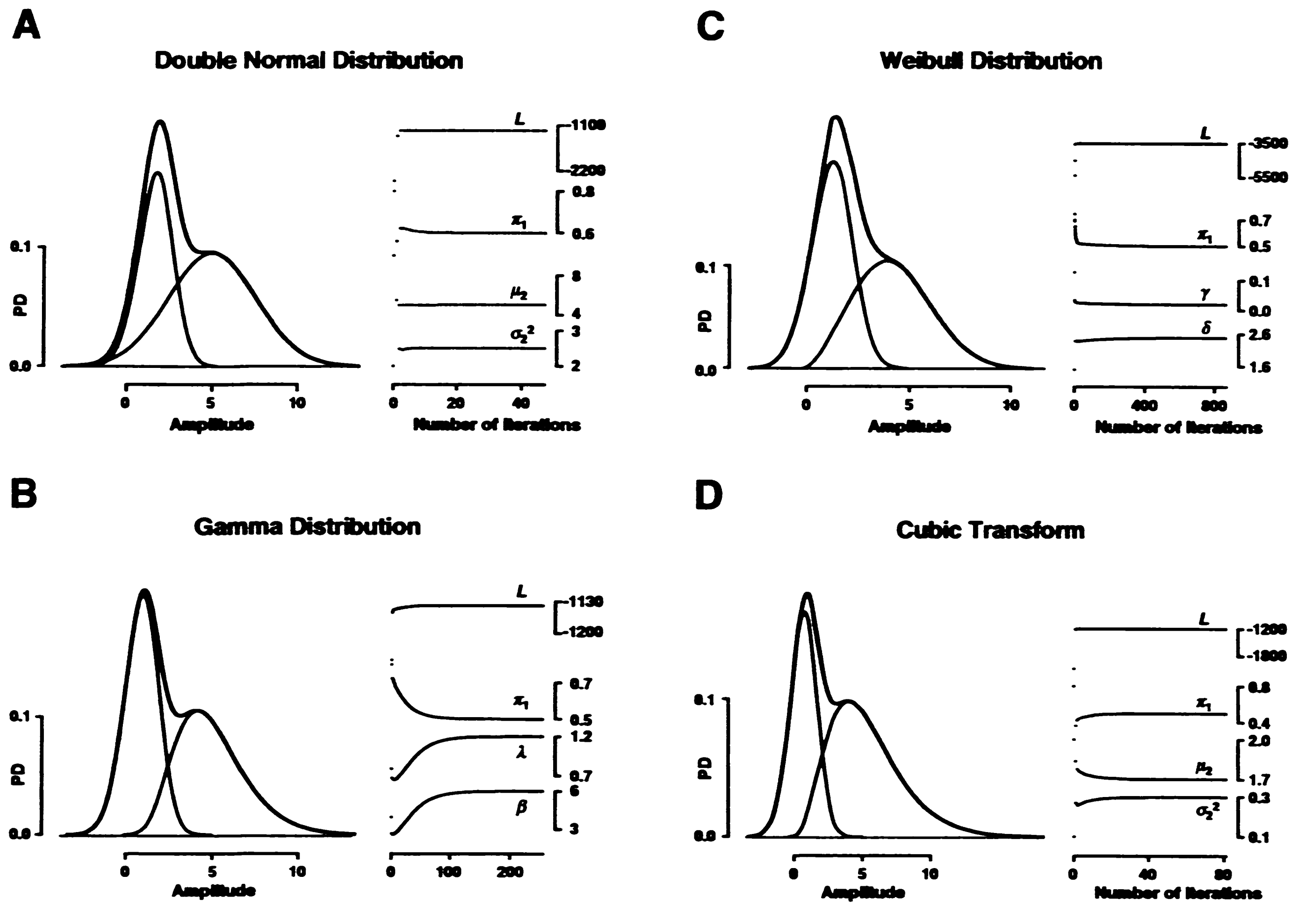

FIGURE 3 Nonquantal deconvolutions. Insets to the right show convergence characteristics. Failures assumed to obey noise represented by the sum of two normal distributions (Fig. 1 *A*) with variable offset μ_1. (*A*) Normal distribution to represent evoked unimodal distribution (π_1, μ_2, σ_Q^2). (*B*) Gamma distribution representing evoked responses (π_1, λ, β). (*C*) Evoked responses described by a Weibull distribution (π_1, γ, δ). (*D*) Cubic transform of a normal random variable representing the evoked unimodal distribution (π_1, μ_2, σ_2^2).

47 iterations with a relative accuracy of $<1.5 \times 10^{-3}$ for P, $<1 \times 10^{-3}$ for μ_2, and $<2.5 \times 10^{-3}$ for σ_2^2.

Gamma distribution

For this model, the constraint that the response distribution is only defined for positive values of the density function has to be implemented. The parameter values used were $P = 0.5$, $\mu_1 = 1.2$, $\beta = 6$, and $\lambda = 1.2$. We used Eqs. A3 to obtain P, A8 to optimize μ_1, A30 for λ, and A31 to obtain β. Convergence was reached after 253 iterations. Relative accuracy was $<2 \times 10^{-3}$ for P, $<9 \times 10^{-3}$ for β, and $<2 \times 10^{-3}$ for λ. Because the implementation of the iterative scheme is based on an approximation of $\Gamma(\beta)$ as well as a numerical solution for β, the accuracy is still acceptable but not as high as it was for other models.

Weibull distribution

The distribution is defined for positive values of the continuous distribution only. As a rule of thumb, the parameter γ describes the skewness of the density function and typically takes values <0.1 for a positively skewed distribution. The parameter δ describes the width of the distribution and takes values >1. (Note that for $\delta = 1$ the Weibull distribution becomes an exponential distribution.) The parameter values used were $P = 0.5$, $\gamma = 0.02$, $\delta = 2.5$, and $\mu_1 = 1.5$. The values of γ and δ required Eqs. A33 and A34, respectively. For the example illustrated in Fig. 3 C, the results were obtained after 867 iterations. The parameters were recovered with high precision (relative accuracy $<4 \times 10^{-4}$ for γ and $<3 \times 10^{-4}$ for δ). In practice, although γ and δ must maintain a certain proportionality to describe adequately a positively skewed response distribution, the optimization based on these recursive equations is well behaved.

Cubic transform of a normal random variable

As for the two previous density functions, the cubic transform of a normal random variable is only defined for positive responses. Typical parameter values are $1 < \mu < 3$ and $\sigma < 1$. The values used in Fig. 3 *D* were $P = 0.5$, $\mu_2 = 1.7$, $\sigma_2^2 = 0.3$, and $\mu_1 = 1$. We used Eqs. A36 and A37 to obtain final values of μ_2 and σ_2^2, respectively. The results were obtained after 81 iterations, with a relative accuracy of 7×10^{-4} for μ_2 and 6×10^{-3} for σ_2^2. In practice, the optimization based on Eqs. A36 and A37 is well behaved.

DISCUSSION

The results show that the parameters for each model can be recovered with high accuracy. Convergence graphs also indicated that the EM algorithm converged monotonically. This was true for all of the models examined. The main advantage gained by using the EM algorithm is that the log-likelihood function is guaranteed to increase monotonically with each iteration. Successive iterations will always converge to a local maximum. Convergence may be to one of many local maxima, but in practice these local maxima will form a compact connected set over which the parameter values will not differ significantly (Boyles, 1983; Wu, 1983; Redner, 1984). It is important to establish how sensitive the solution is to the starting estimates of the parameters for a given model, sample size, and signal-to-noise conditions. It was shown in Results that for the conditions simulated, the solutions showed negligible sensitivity to the initial estimates of the parameter values. Effects of sample size and signal-to-noise ratio were discussed in the companion paper. Other algorithms, such as the Newton-Raphson method, are faster and can be applied with little modification to the different models, but they are not as robust as the EM algorithm (see discussion in Dempster et al., 1977; Ling and Tolhurst, 1983; Redner, 1984; McLachlan and Basford, 1988).

A completely different approach to parameter identification using Bayesian estimation techniques has recently been published by Turner and West (1993). This procedure has advantages and disadvantages over the EM algorithm. Its advantages are that the solution is not dependent on the initial guesses of the parameter values. In addition, it provides confidence limits as well as the means of the discrete amplitudes and their probabilities. Its disadvantages are that a knowledge of the initial priors is essential. Currently, it can only be applied to an unconstrained model and cannot handle skewed noise distributions.

The equations for the quantal parameters and for the binomial and compound binomial release probabilities differ from those given by Kullmann (1989). The equations in this paper have been derived by strictly following the procedures outlined by Hasselblad (1966), and they ensure convergence to the correct solution. In contrast, the equations provided by Kullmann for the quantal parameters do not use the EM approach and, consequently, convergence cannot be guaranteed. This is particularly true if the noise distribution is represented by the sum of two normal probability densities. When convergence does occur, the accuracy achieved can be unacceptable. The equations for the binomial and the compound binomial model given here allow the optimization in a single step, whereas the scheme proposed by Kullmann relies on a two-step procedure: in the first step, the unconstrained probabilities are estimated, and in a second step, they are optimized for the binomial or compound binomial constraint. Our procedures are more efficient in reaching the final results.

We have also provided equations for recovering the parameters of bimodal distributions. These distributions have two applications. One is to provide an alternative hypothesis that apparent regularly occurring peaks in the amplitude distribution of evoked responses are caused by finite sampling, and that a large variability in the discrete amplitudes that arise from variations in transmission at individual active sites, or through variations in the amplitudes of responses originating at different sites does not allow the resolution of discrete amplitudes (Bekkers et al., 1990; Clements, 1991). The other application is one where failure to stimulate the afferent axon occurs, and an unknown proportion of the failures in the response can be attributed to this problem. The equations allow this proportion to be determined.

The application of these algorithms to finite samples has been introduced in the preceding paper (Stricker et al., 1994). The paper uses methods based on balanced resampling and shows how to obtain confidence limits for parameter estimates of the models of synaptic transmission for which the equations have been derived in this paper.

We have been assisted in this work by many helpful discussions with Dr. Daryl Daley. We are also grateful to Drs. John Bekkers and Dennis Turner for their comments on the manuscript.
This work was supported by fellowships of the Swiss National Science Foundation and the Schweizerische Stiftung für Medizinisch-Biologische Stipendien to C. Stricker.

APPENDIX

Unconstrained model

In this model, the amplitudes $\{\mu_0, \mu_1, \cdots, \mu_K\}$ and the variances associated with them $\{\sigma_0^2, \sigma_1^2, \cdots, \sigma_K^2\}$ are unconstrained. The only constraint on the probabilities is that they must be positive and sum to one, i.e.,

$$\sum_{j=0}^{K} P_j = 1 \quad \text{and} \quad P_j \geq 0 \qquad \text{for all } j.$$

The equations for updating P_j, μ_j, and σ_j^2 in this model have been derived by Hasselblad (1966) and were presented by Kullmann (1989, Eqs. R1, R2, and R3). We present them here for completeness and to provide consistency with the notation used in all the following models. The observations to be fitted to a particular model are $(x_1, x_2, \cdots, x_N)$. Throughout this Appendix, the updated parameters are indicated by a * superscript.

Noise represented by a single gaussian probability density

Let

$$q_{ij} = \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-(x_i-\mu_j)^2/2\sigma_j^2} \tag{A1}$$

and

$$M_i = M(x_i, \theta) = \sum_{j=0}^{K} P_j q_{ij}. \tag{A2}$$

Probability;

$$P_j^* = \frac{1}{N} \sum_{i=1}^{N} \frac{f_i}{M_i} P_j q_{ij}, \qquad j = 0, 1, \cdots, K-1 \tag{A3}$$

and

$$P_K^* = 1 - \sum_{j=0}^{K-1} P_j,$$

where P_j^* is the updated probability associated with μ_j and P_j is its value from the previous iteration, f_i is the frequency of the observation x_i.
Amplitude:

$$\mu_j^* = \frac{1}{NP_j} \sum_{i=1}^{N} \frac{f_i}{M_i} q_{ij} x_i, \qquad j = 0, 1, \cdots, K \tag{A4}$$

Variance:

$$\sigma_j^{2*} = \frac{1}{NP_j} \sum_{i=1}^{N} \frac{f_i}{M_i} q_{ij}(x_i - \mu_j)^2, \qquad j = 0, 1, \cdots, K. \tag{A5}$$

μ_j^* and σ_j^{2*} are the updated mean and variance. The previous value of μ_j does not appear explicitly in Eq. A4, but it is implicit in q_{ij}. Similarly, the previous value of σ_j^2 is implicit in q_{ij}. (Eq. A1).

In some situations, it might be appropriate to restrict σ_j^2 to be the variance of the noise, in which case Eq. A5 will not be required. If this is not done, too many degrees of freedom will be introduced. However, when seeking the parameters for a mixture of two normal distributions to represent the recording noise, the variances can be different, and Eq. A5 is used together with Eqs. A3 and A4, with $K = 1$.

Noise represented by the sum of two gaussian probability densities

Here

$$q_{ij} = \pi_n g_{1j}(x_i, \mu_{n1}, \sigma_{n1}) + (1 - \pi_n) g_{2j}(x_i, \mu_{n2}, \sigma_{n2}), \tag{A6}$$

where

$$g_{1j} = \frac{1}{\sqrt{2\pi}\,\sigma_{n1}} e^{-(x_i-\mu_j-\mu_{n1})^2/2\sigma_{n1}^2} \tag{A7}$$

and similarly for g_{2j}. π_n is the probability attached to the first component of the noise. Note that this representation constrains the variance associated with μ_j to be the variance of the sum of the two gaussian distributions (Kullmann, 1989, Eq. 3). The same procedure as outlined by Hasselblad (1966) can be used to obtain the equations for this form of q_{ij}.
Probability: See Eq. A3.
Amplitude:

$$\mu_j^* = \frac{\sum_{i=1}^{N} \frac{f_i}{M_i} P_j \left(\frac{g_{1j}\pi_n}{\sigma_{n1}^2}(x_i - \mu_{n1}) + \frac{g_{2j}(1-\pi_n)}{\sigma_{n2}^2}(x_i - \mu_{n2}) \right)}{\sum_{i=1}^{N} \frac{f_i}{M_i} P_j \left(\frac{g_{1j}\pi_n}{\sigma_{n1}^2} + \frac{g_{2j}(1-\pi_n)}{\sigma_{n2}^2} \right)} \tag{A8}$$

Variance: No formula is given because this would provide too many free parameters. Instead, the variance associated with each amplitude is the noise variance.

Quantal models

In these models, the amplitudes are separated by equal increments, with all amplitudes shifted equally by an offset. This offset could be introduced by stimulus artefact and/or field potentials. The variance could be constrained to be the noise variance, or it could be the noise variance added to a quantal variance that increments with the quantal amplitude.

Unconstrained probabilities

Noise represented by a single gaussian probability density:

$$q_{ij} = \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-(x_i-jQ-\epsilon)^2/2\sigma_j^2}$$

where

$$\sigma_j^2 = \sigma_n^2 + j\sigma_Q^2,$$

Q = quantal size, ϵ = offset, σ_n^2 = noise variance, and σ_Q^2 = quantal variance.

Let

$$M_i = \sum_{j=0}^{K} P_j q_{ij}$$

and the log-likelihood

$$L = \sum_{i=1}^{N} f_i \ln(M_i).$$

Probability: P_j is determined by Eq. A3.
Quantal size:

$$\frac{\partial L}{\partial Q} = \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j \frac{\partial q_{ij}}{\partial Q} = \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j q_{ij} \frac{j(x_i - jQ - \epsilon)}{\sigma_j^2}.$$

When

$$\frac{\partial L}{\partial Q} = 0, \qquad Q^* = \frac{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} P_j q_{ij} \frac{j(x_i - \epsilon)}{\sigma_j^2}}{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} P_j q_{ij} \frac{j^2}{\sigma_j^2}}. \tag{A9}$$

Quantal variance:

$$\frac{\partial L}{\partial \sigma_Q} = \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} P_j q_{ij} \frac{j\sigma_Q}{\sigma_j^4} ((x_i - jQ - \epsilon)^2 - \sigma_Q^2) \tag{A10}$$

When $(\partial L/\partial Q) = 0$, the RHS cannot be solved explicitly for σ_Q, because σ_Q is embedded in σ_j. The roots of Eq. A10 must be obtained numerically. We used an algorithm described by Brent (1973, Chapter 4) for this purpose.
Offset:

$$\frac{\partial L}{\partial \epsilon} = \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j q_{ij} \frac{x_i - jQ - \epsilon}{\sigma_j^2}$$

When

$$\frac{\partial L}{\partial \epsilon} = 0, \qquad \epsilon^* = \frac{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j q_{ij} \frac{x_i - jQ}{\sigma_j^2}}{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j q_{ij} \frac{1}{\sigma_j^2}}. \tag{A11}$$

Note that if quantal variance is assumed to be zero, or very small compared with the noise variance, then $\sigma_j^2 = \sigma_n^2$, and the σ_j^2 terms cancel in Eqs. A9 and A11.

Noise represented by the sum of two gaussian probability densities: Here

$$q_{ij} = \pi_n g_{1j}(x_i, \mu_{n1}, \sigma_{1j}) + (1 - \pi_n) g_{2j}(x_i, \mu_{n2}, \sigma_{2j}),$$

where

$$g_{1j} = \frac{1}{\sqrt{2\pi}\sigma_{1j}} e^{-(x_i - jQ - \epsilon - \mu_{n1})^2/2\sigma_{1j}^2}, \qquad \sigma_{1j}^2 = \sigma_{n1}^2 + j\sigma_Q^2, \tag{A12}$$

and similarly for g_{2j}. The procedures outlined above can be followed to obtain the following results:
Probability: P_j is determined from Eq. A3.
Quantal size:

$$Q^* = \frac{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} P_j \left(g_{1j}\pi_n \frac{j(x_i - \epsilon - \mu_{n1})}{\sigma_{1j}^2} + g_{2j}(1 - \pi_n) \frac{j(x_i - \epsilon - \mu_{n2})}{\sigma_{2j}^2} \right)}{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} P_j \left(g_{1j}\pi_n \frac{j^2}{\sigma_{1j}^2} + g_{2j}(1 - \pi_n) \frac{j^2}{\sigma_{2j}^2} \right)} \tag{A13}$$

Offset:

$$\epsilon^* = \frac{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j \left(g_{1j}\pi_n \frac{x_i - jQ - \mu_{n1}}{\sigma_{1j}^2} + g_{2j}(1 - \pi_n) \frac{x_i - jQ - \mu_{n2}}{\sigma_{2j}^2} \right)}{\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j \left(g_{1j}\pi_n \frac{1}{\sigma_{1j}^2} + g_{2j}(1 - \pi_n) \frac{1}{\sigma_{2j}^2} \right)} \tag{A14}$$

Quantal variance: The following equation must be solved numerically for σ_Q^2:

$$\sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} P_j j \left\{ \frac{g_{1j}\pi_n}{\sigma_{1j}^4} \{(x_i - jQ - \epsilon - \mu_{n1})^2 - \sigma_{1j}^2\} + \frac{g_{2j}(1 - \pi_n)}{\sigma_{2j}^4} \{(x_i - jQ - \epsilon - \mu_{n2})^2 - \sigma_{2j}^2\} \right\} = 0. \tag{A15}$$

Note that if σ_Q is assumed to be zero, $\sigma_{1j}^2 = \sigma_{n1}^2$; $\sigma_{2j}^2 = \sigma_{n2}^2$; but this does not lead to significant simplification of Eqs. A13 and A14.

Binomial probabilities

In this model, all of the release probabilities are identical.

Noise represented by a single gaussian probability density:

$$q_{ij} = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-(x_i - jQ - \epsilon)^2/2\sigma_j^2}, \tag{A16}$$

where

$$\sigma_j^2 = \sigma_n^2 + j\,\sigma_Q^2,$$

Q = quantal size, ϵ = offset, σ_n^2 = noise variance, and σ_Q^2 = quantal variance.
Let

$$M_i = \sum_{j=0}^{K} P_j q_{ij},$$

where

$$P_j = {}^K C_j (1 - p)^{K-j} p^j, \tag{A17}$$

p is the release probability at all release sites, and P_j is the binomial coefficient. The log-likelihood is $L = \sum_{i=1}^{N} f_i \ln(M_i)$.
Probability: To maximize L with respect to p, we note that

$$\frac{\partial L}{\partial p} = \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j q_{ij} \left(\frac{j - Kp}{p(1 - p)} \right).$$

Putting $\partial L/\partial p = 0$ and because $0 \neq p \neq 1$, we have

$$pK \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j q_{ij} = \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} j P_j q_{ij}.$$

This equation provides a recursive procedure for updating p, which appears explicitly in the LHS and implicitly (in P_j) in the RHS (see Hasselblad, 1966, Eqs. 18–20)). Note that $M_i = \sum_{j=0}^{K} P_j q_{ij}$ and $\sum_{i=1}^{N} f_i = N$. The equation above becomes

$$p^* = \frac{1}{NK} \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} j P_j q_{ij}. \tag{A18}$$

Quantal size: Because q_{ij} in this model is the same as in the quantal model, the appropriate equation for Q is given by Eq. A9, with P_j now determined by Eq. A17.
Offset: See Eq. A11, with P_j defined by Eq. A17.
Quantal variance: See Eq. A15.

Noise represented by the sum of two gaussian probability densities:
Equation A12 defines q_{ij}. The recursive equation for p proceeds exactly as above for this new definition of q_{ij}, and Eq. A18 is the appropriate equa-

tion. The recursive equations for the quantal size, quantal variance, and offset are given by Eqs. A13, A15, and A14, respectively, remembering that P_j is defined by Eq. A17.

Compound binomial probabilities

In this model, the release probabilities at the K release sites are $p_1, p_2, \cdots, p_K$. The probabilities $P_0, P_1, \cdots, P_K$ associated with the amplitudes $\mu_0, \mu_1, \cdots, \mu_K$ are defined by the coefficients of the polynomial

$$G_K(z) = \prod_{r=1}^{K} (q_r + p_r z) = P_0 + P_1 z + P_2 z^2 + \cdots + P_r z^r + \cdots + P_K z^K, \quad \text{(A19)}$$

where $q_r = 1 - p_r$, $r = 1, 2, \cdots, K$ and z is a dummy variable.
Let

$$G_K(z) = G_{r,K}(z)(q_r + p_r z)$$

and

$$G_{r,K}(z) = P_{r,0} + P_{r,1}z + P_{r,2}z^2 + \cdots + P_{r,j}z^j + \cdots + P_{r,K-1}z^{K-1} \quad \text{(A20)}$$

for $r = 1, 2, \cdots, K - 1$. When comparing Eqs. A19 and A20, it can be seen that

$$P_j = P_{r,j}(1 - p_r) + P_{r,j-1}p_r \qquad \text{for } r = 1, 2, \cdots, K - 1,$$

$$P_0 = (1 - p_r)P_{r,0}, \quad \text{(A21)}$$

and

$$P_K = p_r P_{r,K-1}.$$

Noise represented by a single gaussian probability density function:

$$q_{ij} = \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-(x_i - jQ - \epsilon)^2/2\sigma_j^2} \qquad \text{and} \qquad M_i = \sum_{j=0}^{K} P_j q_{ij}.$$

Probability: To find the release probability p_r, we need to maximize the log-likelihood function with respect to p_r

$$\frac{\partial L}{\partial p_r} = \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} q_{ij} \frac{\partial P_j}{\partial p_r}.$$

The expression for $\partial P_j/\partial p_r$ can be obtained from Eq. A21.

$$\sum_{i=1}^{N} \frac{f_i}{M_i} \left(-q_{i0} \frac{P_0}{1 - p_r} + \sum_{j=1}^{K-1} q_{ij} \frac{P_j - P_{r,j}}{p_r} + q_{iK} \frac{P_K}{p_r} \right).$$

Setting this equation to zero, and rearranging terms gives

$$p_r \sum_{i=1}^{N} \frac{f_i}{M_i} \left(P_0 q_{i0} + \sum_{j=1}^{K-1} P_j q_{ij} + P_K q_{iK} \right) = \sum_{i=1}^{N} \frac{f_i}{M_i} \left(P_K q_{iK} + \sum_{j=1}^{K-1} (P_j - P_{r,j}(1 - p_r))\, q_{ij} \right).$$

The term inside the brackets on the LHS is M_i (see Eq. A2) and the sum on i of f_i is N. This equation provides a recursive procedure for p_r (as for the binomial distribution above), and we obtain

$$p_r^* = \frac{1}{N} \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} P_{r,j-1} p_r q_{ij}, \quad \text{(A22)}$$

where p_r^* is the update on p_r. We have used the identities in Eq. A21 to simplify the RHS. The P_js can be calculated from the p_rs and the recurrence relationship

$$P_j(K) = P_{j-1}(K-1)p_K + P_j(K-1)(1 - p_K) \qquad j = 1, 2, \cdots, K - 1$$

$$P_0(K) = P_0(K-1)(1 - p_K) \quad \text{(A23)}$$

$$P_K(K) = P_{K-1}(K-1)p_K$$

where $P_j(K)$ is the probability associated with the jth amplitude when there are K release sites. This recurrence relationship can be derived from Eq. A19. Using the notation from Eq. A23, Eq. A19 gives

$$G_K(z) = \prod_{j=1}^{K} (q_j + p_j z) = P_0(K) + P_1(K)\,z + P_2(K)\,z^2 + \cdots + P_j(K)\,z^j + \cdots + P_K(K)\,z^K \quad \text{(A24)}$$

and

$$G_{K-1}(z) = \prod_{j=1}^{K-1} (q_j + p_j z) = P_0(K-1) + P_1(K-1)z + P_2(K-1)z^2 + \cdots + P_j(K-1)z^j + \cdots + P_{K-1}(K-1)z^{K-1} \quad \text{(A25)}$$

Because $G_K(z) = G_{K-1}(z)(q_K + p_K z)$, multiply Eq. A25 by $(q_K + p_K z)$ and equate the result with Eq. A24. Equation A23 is obtained by equating the coefficients of z^j. The procedures for calculating Q, ϵ, and σ_Q are unchanged from the quantal model (Eqs. A9–A11) except that the P_js must now satisfy Eq. A19. The compound binomial constraint could be applied without the constraints of quantal spacing, and quantal increments in variance, in which case the μ_js and σ_js can be calculated from Eqs. A4 and A5.

Noise represented by two gaussian probability density functions:
If the noise is represented by two normal distributions, the calculation for P_js is unaltered. Previously derived equations apply for Q, ϵ, and σ_Q.

Non-quantal models

We have derived the equations using the EM algorithm to find the best fit to the recorded data for a normal distribution, a gamma distribution, a Weibull distribution, and a cubic transformation of a normal random variable.

We let

$$M_i = (1 - P)\, q_{i1} + P q_{i2}, \quad \text{(A26)}$$

where

$$q_{i1} = \frac{1}{\sqrt{2\pi}} \left(\frac{\pi_n}{\sigma_{n1}} e^{-(x_i - \mu_1 - \mu_{n1})^2/2\sigma_{n1}^2} + \frac{1 - \pi_n}{\sigma_{n2}} e^{-(x_i - \mu_1 - \mu_{n2})^2/\sigma_{n2}^2} \right), \quad \text{(A27)}$$

and q_{i2} is the continuous probability density representing the distribution of synaptic currents (excluding failures), and $(1 - P)$ is the probability of failures. Equation A27 is the sum of two normal distributions representing the recording noise. The mean μ_1 has been introduced to allow an offset for the failures peak, possibly caused by stimulus artefact or an extracellular field.
Log-likelihood

$$L = \sum_{i=1}^{N} f_i \ln(M_i).$$

Normal distribution

$$q_{i2} = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-(x_i - \mu_2)^2/2\sigma_2^2} \quad \text{(A28)}$$

The equations for P, μ_1, μ_2, and σ_2^2 are identical to those used in the unconstrained deconvolution but with the failures peak represented by the sum of two normal distributions and the (sole) response peak by a single normal distribution. P is obtained from Eq. A3, with $K = 1$. μ_1 is obtained from Eq. A8. μ_2 and σ_2^2 are obtained from Eqs. A4 and A5, respectively.

Gamma distribution

Put

$$q_{i2} = \frac{\lambda^{\beta}}{\Gamma(\beta)} x_i^{\beta-1} e^{-\lambda x_i} \tag{A29}$$

in Eq. A26. We require the iterative equations for P, μ_1, λ, and β. The equations for P and μ_1 are Eqs. A3 and A8, respectively. To obtain the equation for λ, we require $\partial q_{i2}/\partial\lambda$. This derivative is

$$\frac{\partial q_{i2}}{\partial\lambda} = q_{i2}\left(\frac{\beta}{\lambda} - x\right).$$

Setting $\partial L/\partial\lambda = 0$ and solving for λ we get

$$\lambda^* = \beta \frac{\sum_{i=1}^{N} \frac{f_i}{M_i} q_{i2}}{\sum_{i=1}^{N} \frac{f_i}{M_i} q_{i2} x_i}. \tag{A30}$$

To obtain the equation for β, we need to solve (for β)

$$\frac{\partial L}{\partial\beta} = \sum_{i=1}^{N} \frac{f_i}{M_i}\frac{\partial M_i}{\partial\beta} = 0,$$

and this requires the derivative $\partial q_{i2}/\partial\beta$. To obtain this derivative, we use Stirling's approximation for $\Gamma(\beta)$ (Abramowitz and Stegun, 1965, p. 251)

$$\frac{1}{\Gamma(\beta)} = \frac{1}{\sqrt{2\pi}} e^{\beta} \lambda^{(1/2-\beta)} \times \left[1 + \frac{1}{12\beta} + \frac{1}{288\beta^2} - \frac{139}{51840\beta^3} - \frac{571}{2488320\beta^4} + \cdots\right]^{-1}$$

and the following derivatives:

$$\frac{\partial}{\partial\beta}(\beta^{-(\beta-1/2)}) = \left(\frac{1}{2\beta} - 1 - \ln(\beta)\right)\beta^{-(\beta-1/2)}$$

$$\frac{\partial}{\partial\beta}(\lambda^{\beta}x^{\beta-1}) = \lambda^{\beta}x^{\beta-1}(\ln(x) + \ln(\lambda)).$$

The large number of terms in the approximation for $\Gamma^{-1}(\beta)$ is necessary if $\beta < 1$, and high accuracy is required when solving Eq. A31 for β. Using the identities above, and differentiating by parts, we obtain

$$\frac{\partial q_{i2}}{\partial\beta} = \frac{\lambda^{\beta}}{\Gamma(\beta)} x_i^{\beta-1}e^{-\lambda x_i} \times \left(\ln(x_i) + \ln\left(\frac{\lambda}{\beta}\right) + \frac{1}{2\beta} + \frac{1}{12\beta^2} - \frac{1}{120\beta^4} - \frac{59}{62208\beta^5} - \cdots\right).$$

The equation for β then becomes

$$\sum_{i=1}^{N} \frac{f_i}{M_i} x_i^{\beta-1} e^{-\lambda x_i} \times \left(\ln(x_i) + \ln\left(\frac{\lambda}{\beta}\right) + \frac{1}{2\beta} + \frac{1}{12\beta^2} - \frac{1}{120\beta^4} - \frac{59}{62208\beta^5} - \cdots\right) = 0. \tag{A31}$$

This equation must be solved numerically using an algorithm such as the one described by Brent (1973, Chapter 4).

Weibull distribution

$$q_{i2} = \gamma\delta x_i^{\delta-1}e^{-\gamma x_i^{\delta}} \tag{A32}$$

As under Gamma Distribution above, P and μ_1 are obtained from Eqs. A3 and A8, respectively. To obtain γ and δ, we require the solution for $\partial L/\partial\gamma = 0$ and $\partial L/\partial\delta = 0$ where $L = \sum_{i=1}^{N} f_i \ln(M_i)$.

$$\frac{\partial L}{\partial\gamma} = \sum_{i=1}^{N} \frac{f_i}{M_i} P \frac{\partial q_{i2}}{\partial\gamma}$$

Using the identity$(\partial/\partial\gamma)(e-\gamma x i\delta) = -x_i^{\delta}e^{-\gamma x_i^{\delta}}$ and differentiation by parts

$$\frac{\partial L}{\partial\gamma} = \sum_{i=1}^{N} \frac{f_i}{M_i} P q_{i2}\left(\frac{1}{\gamma} - x_i^{\delta}\right).$$

And $\partial L/\partial\gamma = 0$ when

$$\gamma^* = \frac{NP}{\sum_{i=1}^{N} \frac{f_i}{M_i} q_{i2} x_i^{\delta}}, \tag{A33}$$

where $\sum_{i=1}^{N} (f_i/M_i)q_{i2} = NP$ has been used to simplify Eq. A33. To obtain $\partial L/\partial\gamma$, we use the identities

$$\frac{\partial}{\partial\delta}(e^{-\gamma x_i^{\delta}}) = -\gamma \ln(x_i)x_i^{\delta}e^{-\gamma x_i^{\delta}}$$

$$\frac{\partial}{\partial\delta}(x_i^{\delta}) = \ln(x_i)x_i^{\delta}$$

and

$$\frac{\partial L}{\partial\delta} = \sum_{i=1}^{N} \frac{f_i}{M_i} P q_{i2}\left(\frac{1}{\delta} + \ln(x_i)(1 - \gamma x_i^{\delta})\right) = 0.$$

This equation becomes

$$NP + \delta \sum_{i=1}^{N} \frac{f_i}{M_i} q_{i2}(\ln(x_i)(1 - \gamma x_i^{\delta})) = 0, \tag{A34}$$

and it must be solved numerically to find the root for δ.

Cubic transform of a normal random variable

$$q_{i2} = \frac{1}{3\sqrt{2\pi}\sigma_2} x_i^{-2/3} e^{-(x_i^{1/3}-\mu_2)^2/2\sigma_2^2} \tag{A35}$$

$$\frac{\partial L}{\partial\mu_2} = \sum_{i=1}^{N} \frac{f_i}{M_i} P \frac{\partial q_{i2}}{\partial\mu_2}$$

When $\partial L/\partial\mu_2 = 0$,

$$\mu_2^* = \frac{1}{N}\sum_{i=1}^{N} \frac{f_i}{M_i} q_{i2} x_i^{1/3}. \tag{A36}$$

Similarly, $\partial L/\partial\sigma_2^2 = 0$ when

$$\sigma_2^{2*} = \frac{1}{N}\sum_{i=1}^{N} \frac{f_i}{M_i} q_{i2}(x_i^{1/3} - \mu_2)^2. \tag{A37}$$

P and μ_1 are obtained from Eqs. A3 and A8, respectively.

Failure to stimulate the afferent axon(s) and failure to release transmitter

In these models, there are more failures than would be expected from an uniform or compound binomial release process. Fitting the observed density function to the nonfailure peaks specifies the probability of failures expected for the parameters obtained. The excess failures are then attributed to failure to excite the axon(s).

Binomial distribution

Let

$$M_i = \pi_0 q_{i0} + (1 - \pi_0) \sum_{j=0}^{K} P_j q_{ij} \tag{A38}$$

where π_0 is the proportion of failures caused by failure to stimulate the axon. q_{ij} is defined by Eq. A16 and P_j by Eq. A17. Again,

$$L = \sum_{i=1}^{N} f_i \ln(M_i).$$

We require iterative equations for the proportion of failures π_0, probability of release p, quantal size Q, quantal variance σ_Q^2, and offset ϵ.

Proportion of failures: Equation A3 is the appropriate recursive expression to use with $K = 1$.

Release probabilities:

$$\frac{\partial L}{\partial p} = \sum_{i=1}^{N} \frac{f_i}{M_i} (1 - \pi_0) \sum_{j=0}^{K} P_j q_{ij} \frac{\partial P_j}{\partial p}$$

$$= \sum_{i=1}^{N} \frac{f_i}{M_i} (1 - \pi_0) \sum_{j=0}^{K} P_j q_{ij} \left(\frac{j - Kp}{p(1 - p)} \right)$$

Set $\partial L/\partial p = 0$, and because $0 \neq p \neq 1$, the equation becomes

$$pK(1 - \pi_0) \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=0}^{K} P_j q_{ij} = (1 - \pi_0) \sum_{i=1}^{N} \frac{f_i}{M_i} \sum_{j=1}^{K} j P_j q_{ij}.$$

Substitute Eq. A38 in the LHS of this equation and obtain

$$pK \sum_{i=1}^{N} \frac{f_i}{M_i} M_i = \sum_{i=1}^{N} \frac{f_i}{Q_i} \left(pK\pi_0 q_{i0} + (1 - \pi_0) \sum_{j=1}^{K} j P_j q_{ij} \right).$$

Because the LHS sums to pKN, this equation provides a recursive procedure for p (see Hasselblad, 1966, Eqs. 18–20); i.e.,

$$p^* = \frac{1 - \pi_0}{KN} \sum_{i=1}^{N} \frac{f_i}{M_i} \left(\frac{pK\pi_0}{1 - \pi_0} q_{i0} + \sum_{j=1}^{K} j P_j q_{ij} \right), \tag{A39}$$

where p^* is the update on p, and p appears explicitly in the RHS, as well as implicitly in P_j.

Quantal size: The expression for the quantal size is not altered by the introduction of stimulus failures (Eq. A9).

Quantal variance: Equation A15 remains appropriate for determining quantal variance.

Offset: We require the solution to

$$\frac{\partial L}{\partial \rho} = \sum_{i=1}^{N} \frac{f_i}{M_i} \frac{\partial M_i}{\partial \rho} = 0.$$

Noise represented by a single gaussian probability density function:
q_{ij} is given by (A1) and $\partial q_{ij}/\partial \epsilon = q_{ij}((x_i - jQ - \epsilon)/\sigma_j^2)$.

$$\frac{\partial M_i}{\partial \epsilon} = \pi_0 \frac{\partial q_{i0}}{\partial \epsilon} + (1 - \pi_0) \sum_{j=0}^{K} P_j \frac{\partial q_{ij}}{\partial \epsilon}$$

Then

$$\frac{\partial L}{\partial \epsilon} = \sum_{i=1}^{N} \frac{f_i}{M_i} \left(\pi_0 q_{i0} \frac{x_i - \epsilon}{\sigma_n^2} + (1 - \pi_0) \sum_{j=0}^{K} P_j q_{ij} \frac{x_i - jQ - \epsilon}{\sigma_j^2} \right).$$

Setting this equation to zero, rearranging terms, and solving for ϵ, we obtain

$$\epsilon^* = \frac{\sum_{i=1}^{N} \frac{f_i}{M_i} \left(\pi_0 q_{i0} \frac{x_i}{\sigma_n^2} + (1 - \pi_0) \sum_{j=0}^{K} P_j q_{ij} \frac{x_i - jQ}{\sigma_j^2} \right)}{\sum_{i=1}^{N} \frac{f_i}{M_i} \left(\pi_0 q_{i0} \frac{1}{\sigma_n^2} + (1 - \pi_0) \sum_{j=0}^{K} P_j q_{ij} \frac{1}{\sigma_j^2} \right)}. \tag{A40}$$

Noise represented by two normal probability density functions:
Here, q_{ij} is given by Eq. A12 and

$$\frac{\partial q_{ij}}{\partial \epsilon} = \pi_n g_{1j}(x_i, \mu_{n1}, \sigma_{n1}) \frac{x_i - \mu_{n1} - jQ - \epsilon}{\sigma_{1j}^2}$$
$$+ (1 - \pi_n) g_{2j}(x_i, \mu_{n2}, \sigma_{n2}) \frac{x_i - \mu_{n2} - jQ - \epsilon}{\sigma_{2j}^2}.$$

Again,

$$\frac{\partial M_i}{\partial \epsilon} = \pi_0 \frac{\partial q_{i0}}{\partial \epsilon} + (1 - \pi_0) \sum_{j=0}^{K} P_j \frac{\partial q_{ij}}{\partial \epsilon}.$$

Let

$$f_{10} = \frac{\pi_n g_{10}(x_i, \mu_{n1}, \sigma_{n1})}{\sigma_{n1}^2}$$

$$f_{1j} = \frac{\pi_n g_{1j}(x_i, \mu_{n1}, \sigma_{1j})}{\sigma_{1j}^2}$$

$$f_{20} = \frac{(1 - \pi_n) g_{20}(x_i, \mu_{n2}, \sigma_{n2})}{\sigma_{n2}^2}$$

$$f_{2j} = \frac{(1 - \pi_n) g_{2j}(x_i, \mu_{n2}, \sigma_{2j})}{\sigma_{2j}^2},$$

where the terms on the RHS are defined by Eq. A12. The solution for ϵ obtained from $\partial L/\partial \epsilon = 0$ is given by

$$\epsilon^* = \frac{\mathcal{A}}{\sum_{i=1}^{N} \frac{f_i}{M_i} \left(\pi_0 (f_{10} + f_{20}) + (1 - \pi_0) \sum_{j=0}^{K} (f_{1j} + f_{2j}) \right)} \tag{A41}$$

where

$$\mathcal{A} = \sum_{i=1}^{N} \frac{f_i}{M_i} \Bigg(\pi_0 (f_{10}(x_i - \mu_{n1}) + f_{20}(x_i - \mu_{n2}))$$
$$+ (1 - \pi_0) \sum_{j=0}^{K} P_j (f_{1j}(x_i - \mu_{n1} - jQ) + f_{2j}(x_i - \mu_{n2} - jQ)) \Bigg).$$

Compound binomial distribution

M_i is defined by Eq. A38, but in this case the P_js are defined by Eq. A19. q_{ij} can either be a normal distribution or the sum of two normal distributions. Quantal separation and quantal increments in variance can be included as required, following the derivation in Failure to Stimulate Afferent Axons. The introduction of stimulation failures only influences the equations for the release probabilities and the offset. As for the binomial model, Eq. A3 is the appropriate recursive expression to obtain π_0.

Release probabilities: Use Eq. A38 in the formula for log-likelihood, differentiate with respect to p_r, the release probability at the rth release site, and set the derivative to zero. The procedure closely follows that used in Failure to Stimulate Afferent Axons. We obtain

$$p_r(1-\pi_0)\sum_{i=1}^{N}\frac{f_i}{M_i}\left(P_0q_{i0}+\sum_{j=1}^{K-1}P_jq_{ij}+P_Kq_{iK}\right)$$
$$=(1-\pi_0)\sum_{i=1}^{N}\frac{f_i}{M_i}\left(P_Kq_{iK}+\sum_{j=1}^{K-1}(P_j-P_{r,j}(1-p_r))q_{ij}\right).$$

Note that

$$(1-\pi_0)\left(P_0q_{i0}+\sum_{j=1}^{K-1}P_jq_{ij}+P_Kq_{iK}\right)=M_i-\pi_0q_{i0}.$$

This identity allows the development of the recursive equation for p_r, i.e.,

$$p_r^*=\frac{p_r(1-\pi_0)}{N}\sum_{i=1}^{N}\frac{f_i}{M_i}\left(\frac{\pi_0}{1-\pi_0}q_{i0}+\sum_{j=1}^{K}P_{r,j-1}\,q_{ij}\right), \qquad \text{(A42)}$$

where p_r^* is the update on the previous p_r.

Offset: The equation for offset is the same as for the binomial model (Eq. A41), except that the P_js for the compound binomial must be used. If a quantal model is not wanted, let $jQ = \mu_j$. The variance can be incremented as in the quantal model, or left as the noise variance, in which case $\sigma_j^2 = \sigma_n^2$ for a single normal distribution or $\sigma_{1j}^2 = \sigma_{n1}^2$ and $\sigma_{2j}^2 = \sigma_{n2}^2$ in the case of the sum of two normal distributions.

Availability of source code

Source code can be made available upon written request to C. Stricker at the address above. The code can be provided either as IGOR Pro functions or as ANSI C routines. Requests should be accompanied by an E-mail address.

REFERENCES

Abramowitz, M., and I. A. Stegun. 1965. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Dover Publications, New York. 1046 pp.

Bekkers, J. M., G. B. Richerson, and C. F. Stevens. 1990. Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proc. Natl. Acad. Sci. USA.* 87:5359–5362.

Boyles, R. A. 1983. On the convergence of the EM algorithm. *J. R. Stat. Soc. Ser. B.* 45:47–50.

Brent, R. P. 1973. Algorithms for Minimization Without Derivatives. Prentice Hall, Englewood Cliffs, NJ. 195 pp.

Clements, J. 1991. Quantal synaptic transmission? *Nature.* 353: 396.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the *EM* algorithm. *J. R. Stat. Soc. Ser. B.* 39: 1–22.

Freund, J. E. 1992. Mathematical Statistics. Prentice Hall, Englewood Cliffs, NJ. 658 pp.

Hasselblad, V. 1966. Estimation of parameters for a mixture of normal distributions. *Technometrics.* 8:431–446.

Jonas, P., G. Major, and B. Sakmann. 1994. Quantal components of unitary EPSCs at the mossy fibre synapse on CA3 pyramidal cells of rat hippocampus. *J. Physiol.* 472:615–663.

Kullmann, D. M. 1989. Applications of the expectation-maximization algorithm to quantal analysis of postsynaptic potentials. *J. Neurosci. Methods.* 30:231–245.

Kullmann, D. M. 1992. Quantal analysis using maximum entropy noise deconvolution. *J. Neurosci. Methods.* 44:47–57.

Ling, L., and D. J. Tolhurst. 1983. Recovering the parameters of finite mixtures of normal distributions from a noisy record: an empirical comparison of different estimating procedures. *J. Neurosci. Methods.* 8: 309–333.

McLachlan, E. M. 1975. An analysis of the release of acetylcholine from preganglionic nerve terminals. *J. Physiol.* 245:447–466.

McLachlan, E. M. 1978. The statistics of transmitter release at chemical synapses. *Int. Rev. Physiol. Neurophysiol. III.* 17:49–117.

McLachlan, G. J., and K. E. Basford. 1988. Mixture Models; Inference and Applications to Clustering. Marcel Dekker, New York. 253 pp.

Redner, R. A. 1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* 26:195–239.

Stricker, C., S. J. Redman, and D. Daley. 1994. Statistical analysis of synaptic transmission: model discrimination and confidence limits. *Biophys. J.* 67:532–547.

Turner, D. A., and M. West. 1993. Bayesian analysis of mixtures applied to post-synaptic potential fluctuations. *J. Neurosci. Methods.* 47:1–21.

Walmsley, B., F. R. Edwards, and D. J. Tracey. 1988. Nonuniform release probabilities underlie quantal synaptic transmission at a mammalian excitatory central synapse. *J. Neurophysiol.* 60:889–908.

Wu, C. F. J. 1983. On the convergence properties of the EM algorithm. *Ann. Statist.* 11:95–103.