Statistical Analysis of Synaptic Transmission: Model Discrimination and Confidence Limits

Christian Stricker,* Stephen Redman,* and Daryl Daley‡
*Division of Neuroscience, John Curtin School of Medical Research, and ‡Stochastic Analysis Group, School of Mathematical Sciences, Australian National University, Canberra ACT 0200 Australia

ABSTRACT Procedures for discriminating between competing statistical models of synaptic transmission, and for providing confidence limits on the parameters of these models, have been developed. These procedures were tested against simulated data and were used to analyze the fluctuations in synaptic currents evoked in hippocampal neurones. All models were fitted to data using the Expectation-Maximization algorithm and a maximum likelihood criterion. Competing models were evaluated using the log-likelihood ratio (Wilks statistic). When the competing models were not nested, Monte Carlo sampling of the model used as the null hypothesis (H_0) provided density functions against which H_0 and the alternate model (H_1) were tested. The statistic for the log-likelihood ratio was determined from the fit of H_0 and H_1 to these probability densities. This statistic was used to determine the significance level at which H_0 could be rejected for the original data. When the competing models were nested, log-likelihood ratios and the χ^2 statistic were used to determine the confidence level for rejection. Once the model that provided the best statistical fit to the data was identified, many estimates for the model parameters were calculated by resampling the original data. Bootstrap techniques were then used to obtain the confidence limits of these parameters.

INTRODUCTION

The use of statistical techniques for analyzing synaptic transmission dates from quantal analysis at the neuromuscular junction (del Castillo and Katz, 1954). Since then, similar techniques have been applied to many different types of synapses, including synapses in the central nervous system. These techniques have evolved to cope with the special conditions which prevail at central synapses (reviewed in Korn and Faber, 1987; Redman, 1990).

A largely unresolved problem in quantal analysis has been one of discriminating between different statistical models of synaptic transmission. Questions relating to release probabilities at different release sites (uniform or nonuniform), quantal amplitudes from different active sites (with or without variability) would best be answered by direct measurements. In practice, the answers have to be obtained by comparing the ability of different models of transmission to match the statistics of observed responses. When the competing models are "nested", the ratio of the likelihood of the fits to the two models has asymptotically a χ^2 distribution. Two models are nested when one is a sub-hypothesis of the other and when one model can be transformed to the other by a smooth parametric transition. This statistic (known as the Wilks statistic) provides a more powerful test for rejecting an alternative model than does the χ^2 goodness-of-fit. When the competing models are not nested, such as occurs when the number of components is different in the two models, the likelihood ratio need not be asymptotically like a χ^2 distribution (Horn, 1987; Titterington et al., 1985). Monte Carlo techniques can be used to estimate the distribution of the likelihood ratio. This distribution can then be used to calculate the level of significance (α) for rejection of an alternative model.

In this paper, the techniques described above for nested and non-nested models have been developed; tested against a simulated model of synaptic transmission and applied to data obtained from synaptic transmission at excitatory synapses on pyramidal cells in the hippocampus. The unifying concept throughout is the use of the likelihood measure of goodness-of-fit. All models of transmission were fitted to the data using the maximum likelihood criterion. The Expectation-Maximization (EM) algorithm has been used throughout, because this algorithm guarantees convergence. The application of this algorithm to all of the models of transmission that have been considered is the subject of the subsequent paper (Stricker and Redman, 1994).

Another procedure that has been missing from statistical analyses of synaptic transmission data has been a simple method of calculating confidence limits for the estimated parameters. Not only are these confidence limits important indicators of reliability, but they are of crucial importance in assessing the significance or otherwise of apparent changes in parameters occurring after synaptic conditioning procedures. Potential errors in parameter estimation using deconvolution techniques have been evaluated in numerous publications using Monte Carlo simulations (Wong and Redman, 1980; Jack et al., 1981; Ling and Tolhurst, 1983; Clements et al., 1987; Kullmann, 1989; Sayer et al., 1989; Redman, 1990; Solodkin et al., 1991; Clamann et al., 1991; Voronin et al., 1992). One recent approach to this problem has been to use Bayesian inference techniques (Turner and West,

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.
© 1994 by the Biophysical Society
0006-3495/94/08/532/16 $2.00

1993). This scheme delivers a conditional probability for the number of components in the mixture distribution, and uncertainty estimates for the amplitude and probability of each component in the mixture. In another scheme, Smith et al. (1991) used the Fisher information matrix to obtain the variance of each of the model parameters. The standard errors for n and p in a binomial release process were provided in McLachlan (1978).

This paper shows how resampling methods can be used to obtain confidence limits. This approach, developed by Efron (1979), involves resampling (with replacement) of the original data to generate new samples. The unknown parameters are calculated from the original data, and from each set of resampled data. By this means (commonly referred to as "bootstrapping"), many estimates of each parameter are calculated (usually 50 to 200), allowing confidence limits to be determined. We have combined this technique for calculating confidence limits with the procedures outlined above for model discrimination to provide confidence limits on the parameters for acceptable models of synaptic transmission. We have also examined how confidence limits depend upon sample size and signal-to-noise conditions.

MATERIALS AND METHODS

The primary sample

In experiments, this sample is taken from an unknown amplitude probability density. However, to evaluate the reliability of the procedures developed for model discrimination, and for calculating confidence limits of parameter values, it is necessary to have a sample that closely represents a known probabilistic model of synaptic transmission. If $P(x)$ is the density function of the responses generated by this model, then the sample $(x_1, x_2, \cdots, x_j, \cdots, x_N)$ that defines $P(x)$ is calculated from

$$\frac{j}{N} = \int_{x_{min}}^{x_j} P(y)\,dy \qquad \text{for} \quad j = 1, 2, \cdots, N, \tag{1}$$

where N is the sample size and x_{min} is the lower bound of $P(x)$. The x_js are the values of x that correspond to increments of $1/N$ in the cumulative distribution of $P(x)$.

Monte Carlo sampling

The cumulative distribution defined in (1) is calculated for all values of x_i. Random numbers between zero and one were generated, and the value of x_i corresponding to each of these random values was interpolated. These values formed the sample of size N.

Balanced resampling

The sample obtained from either sampling procedure described above can be resampled randomly, with replacement. Each resampling generates a sample equal in size to the original sample. These sample sets will contain repeats of some values in the original sample. To ensure that the sample sets obtained in this way are not biased, a balanced resampling scheme is used (Davison et al., 1986). In this scheme, the probability density formed by combining all of the sample sets obtained from resampling is identical with the density of the original sample. The resampling procedure was checked to ensure that the original probability density was recovered. We used 250 resamplings to calculate the SD of each model parameter. According to Efron and Tibshirani (1993, Chapter 6), 50 to 200 resamplings are sufficient to estimate moments of a statistic, such as a SD.

The noise probability density

Baseline noise in intracellular recordings is usually skewed, because of the presence of spontaneous synaptic potentials or currents. This skewness can be accommodated by using a mixture of two normal distributions (Kullmann, 1989; Sayer et al., 1989). Simulated noise was generated using a mixture of two normal distributions. Experimental noise was fitted in the same way.

Amplitude probability density

The data sample must be converted to a probability density. To do this, each value in a sample was convolved with a normal density having a SD that depended on the SD of the baseline noise and the anticipated quantal amplitude (Silverman, 1986). The required amplitude density $D(y)$ is the sum of all of the normal densities with means equal to the sample values y_i

$$D(y) = \frac{1}{N}\sum_{i=1}^{N} G(y_i, \sigma), \tag{2}$$

where $G(y_i, \sigma)$ is a normal density and $(y_1, y_2, \cdots, y_N)$ is the sample. The choice of σ is of crucial importance in density estimation. This issue is discussed in Silverman (1986). When the mixture consists of only one normal density,

$$\sigma = 1.06\,N^{-1/5}\sigma_n,$$

where N is the sample size and σ_n is the SD of the noise (Silverman, 1986, Eq. 3.28). When the mixture has more than one component, separated by 2.5 σ_n, then σ as calculated from the above equation is reduced by a factor of approximately 0.8. The scaling factor for other separations is given in Silverman (1986, Fig. 3.3).

The noise density was represented by the sum of two normal densities, usually separated by less than σ_n, and the appropriate value of σ (for N = 600) is approximately 0.3 σ_n. The densities for EPSCs evoked in CA1 pyramidal neurones usually had at least four components in their mixture representation, with peak separations of between 2 and 5 σ_n, making the appropriate value of σ approximately 0.2 σ_n. Because the noise density and the EPSC density must be formed using the same kernel, we used an intermediate value of σ (0.25 σ_n) for both densities when N = 600.

The density function $D(y)$ (Eq. 2) must be calculated at successive intervals Δy that depend on σ. The 3-db frequency (f_c) for the normal distribution $G(y_i, \sigma)$ is (Colquhoun and Sigworth, 1983)

$$f_c = \frac{(\ln 2)^{1/2}}{2\pi\sigma} = \frac{0.133}{\sigma}.$$

The appropriate sample rate for a gaussian kernel, when no interpolation is used, is about 20 f_c, or 2.66/σ (Colquhoun and Sigworth, 1983). Because σ depends on σ_n, the sample interval Δy will be $\approx 0.1\sigma_n$ when σ = 0.25 σ_n.

Model comparison

The Wilks test provides a basis for comparison of the adequacy of one model to fit the observed probability density compared with the fit obtained using an alternative model, provided the two models are nested.

In what follows, we assume that a data set X can be described as a family of random variables yielding the likelihood function

$$L(X; \theta_1, \cdots, \theta_{k+\nu}) \tag{3}$$

for some parameters $\theta_1, \cdots, \theta_{k+\nu}$ lying in some specified regions. Call this model B for short. Model A for the same data set X is *nested* in model B

if it has a likelihood

$$L(X; \theta_1, \cdots, \theta_k, \theta'_{k+1}, \cdots, \theta'_{k+\nu}),$$

where the primed parameters are specified, lie in the range of the parameters in the likelihood expression at Eq. 3, and $L(\cdot)$, as a function of its parameters, is such that the maximum likelihood estimators are asymptotically normal and efficient. Then the expression

$$W \equiv -2 \ln(\hat{L}_A/\hat{L}_B),$$

where $\hat{}$ denotes the likelihood evaluated at its maximum, so that with $\hat{\theta}_j$ denoting maximum likelihood estimators,

$$\hat{L}_A = L(X; \hat{\theta}_1, \cdots, \hat{\theta}_k, \cdots, \theta'_{k+1}, \cdots, \theta'_{k+\nu}), \qquad \hat{L}_B = L(X; \hat{\theta}_1, \cdots, \hat{\theta}_{k+\nu})$$

is asymptotically distributed as a χ^2 random variable with ν degrees of freedom (Wilks, 1938).

Wilks' result here depends on model A being nested within model B, and under this condition it yields a statistical test for comparing the fit of the two models ostensibly describing the data. When we are given two competing models described by their likelihood functions but without their being nested, we could still compute the likelihood estimates and thereby calculate the statistic W for our data set, but we would not usually know anything about its distribution. By choosing (say) model A as constituting the null hypothesis, we can estimate the distribution of W by Monte Carlo sampling (as described below).

According to this definition of nested models, classifying data into either K or $K + 1$ categories (corresponding to models A and B, respectively) would not necessarily yield model A nested in model B. But if the union of the two categories of model B yields model A, then we could properly regard them as being a nested pair of models, and a statistic asymptotically following the χ^2 distribution might hold because of the theory of the Pearson goodness-of-fit χ^2 statistic (of $\sum$(observed − expected)^2/expected). But in the case of K itself being a discrete-valued parameter, describing the number of components of the distribution, we do not have nested models.

When the models are nested, the Wilks statistic is calculated for the fits to the data, and this statistic together with the difference in the number of degrees of freedom between the two models can be used to find a level of significance for rejection of one of the models.

When the two models are not nested, one model is chosen as the null hypothesis (H_0), and the other as the alternative hypothesis (H_1). We use the EM algorithm to maximize the log-likelihoods for the best fit to the data, producing L_0 and L_1, say. The model yielding the best fit to the data under H_0 becomes the parent probability density from which Monte Carlo sampling occurs. Each sample is fitted assuming H_0 and H_1, and the Wilks statistic ($W = -2(L_0 - L_1)$) for the best fit to each model is calculated. This procedure is repeated 250 times (McLachlan and Basford, 1988, p. 25), and the Wilks statistics are then rank-ordered to create a cumulative distribution. The Wilks statistic for the fit of H_0 and H_1 to the original data is then located on this cumulative distribution, which gives the level of significance at which H_0 is rejected. If H_0 cannot be rejected at $\alpha \leq 0.05$, we then accept H_0. In this case, model selection is based on the principle of parsimony, i.e., the less complex of the two models is chosen.

Bootstrap method for calculating SDs of model parameters

This procedure is illustrated schematically in Fig. 1. The EM algorithm for the model chosen was applied to the density formed by each balanced resampling, as well as to the original sample, and a parameter set was obtained. This process was repeated 250 times. The mean of each parameter corresponded to the result from the original sample. The SD of each parameter was calculated from the set of optimized values obtained for that parameter.

RESULTS

Analysis of a simulated release process

A model of release was simulated in which four release sites each released transmitter independently with the same probability of 0.3. The quantal current was 3.0 pA, and a zero offset (caused by a field potential) of 0.2 pA was present. Furthermore, the axon that gave rise to the four release sites was intermittently excited, with a probability of 0.6 of being activated. A noise process was generated from the sum of two normal distributions, with mean and SD 0.0/0.9 and 1.0/1.0 pA, respectively, and with associated probabilities of 0.7 and 0.3.

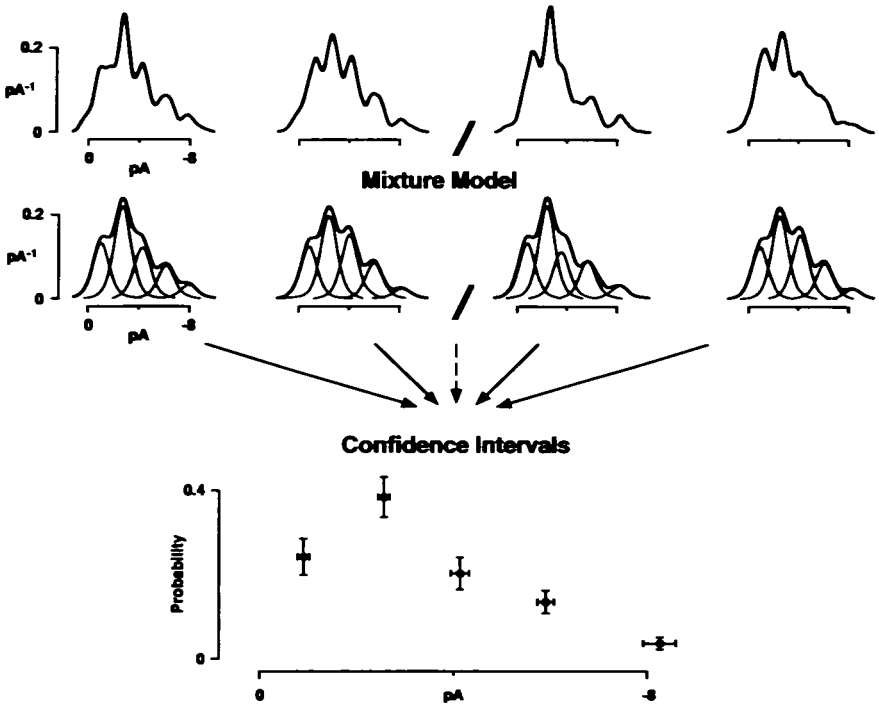

FIGURE 1 Schematic illustration of how confidence limits were calculated. The top row is a sequence of probability densities obtained by resampling the data set. Each density was then analyzed to find the amplitude and probability of each component in the mixture distribution that was used to model synaptic transmission. The SDs were then calculated for each amplitude and probability.

The probability densities of the release process and the noise process were generated. These two densities were convolved to create a probability density that simulated the amplitudes of evoked responses with additive baseline noise. A primary sample (as described in Materials and Methods) was obtained for this density function. Twenty sample sets (N = 600) were then taken from this primary sample using the balanced sampling procedure (see Materials and Methods (Balanced resampling)). A random draw from 1 to 20 resulted in the fifth density function being used as the sampled distribution for subsequent analysis. A probability density was calculated for this sample using a Gaussian kernel, with $\sigma = 0.25\ \sigma_x$. The noise process was also sampled by the same procedure to provide a probability density for the baseline noise.

Noise analysis

The EM algorithm was used to obtain a mixture of two normal distributions to represent the baseline noise. The result is given below, with P_j being the relative contribution of each normal distribution to the mixture and with mean and SD (μ_j, σ_j)

$$P_0 = 0.83, \quad \mu_0 = 0.5 \text{ pA}, \quad \sigma_0 = 0.92 \text{ pA}$$
$$P_1 = 0.17, \quad \mu_1 = 1.47 \text{ pA}, \quad \sigma_1 = 0.84 \text{ pA}.$$

The density function for the noise and its mixture distribution are shown in Fig. 2 *A*.

Unconstrained mixture model

The probability density of the synaptic current with noise added and simulated by the procedures outlined in Analysis of a simulated release process is shown in Fig. 2 *B*. The EM algorithm, with no constraints on the probability or the mean amplitude of any component in the mixture, was applied to this density function for different numbers of components (K) in the mixture. The results for K = 4, 5, and 6 are given in Table 1. The log-likelihood became much more negative for values of K smaller than 4. The model with K = 4 (H_0) was tested against the model with K = 5 (H_1). Using the

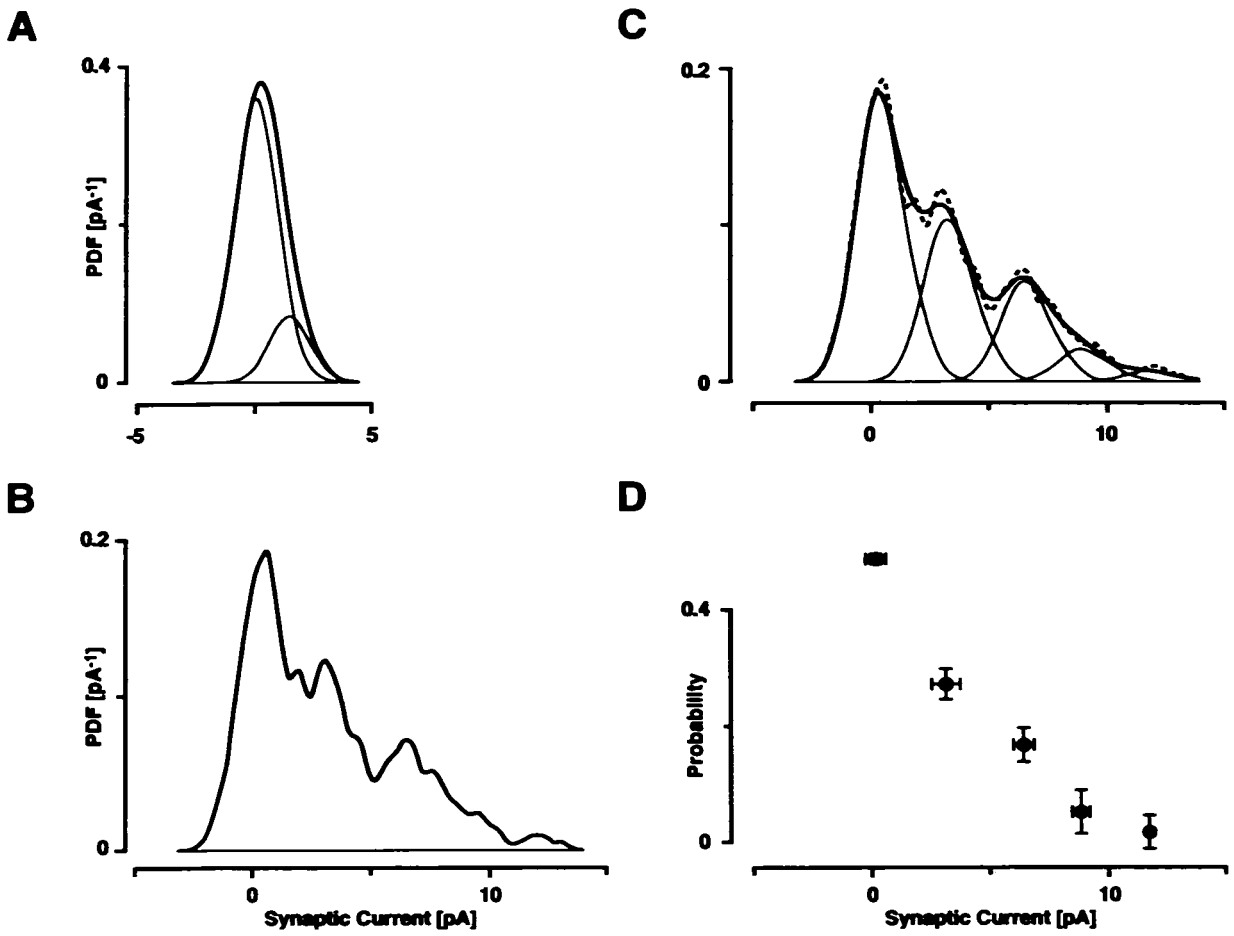

FIGURE 2 (*A*) The heavy line is the probability density of the noise sample, with sample size 600. This density was resolved to be a mixture of two normal distributions, shown as thin lines. (*B*) The probability density of the sample obtained from the simulated synaptic current, convolved with the noise. (*C*) The dashed line is the density shown in *B*. The heavy line is the sum of the five component distributions (shown as light lines) and is the best fit to the data obtained using the EM algorithm, assuming five components in the mixture. No constraints were placed on the amplitudes or on the probabilities of the components in the mix. (*D*) The filled circles indicate the amplitude and the probability of each component in the mixture. The bars correspond to ±1.0 SD for the amplitude and probability.

TABLE 1 The results for the best fitting mixture comprised of 4, 5, and 6 components

	$K = 4$		$K = 5$		$K = 6$	
Component	P	μ	P	μ	P	μ
0	0.49	0.23	0.49	0.22	0.49	0.22
1	0.28	3.23	0.27	3.14	0.27	3.14
2	0.19	6.71	0.17	6.38	0.17	6.38
3	0.04	10.33	0.05	8.79	0.03	8.75
4			0.02	11.69	0.03	8.82
5					0.01	11.69
L	−1449.31		−1445.24		−1445.24	

In each case μ is the mean of each component in the mixture (in pA) and P is the probability associated with that component. L corresponds to the log-likelihood for the best fit, obtained using the EM algorithm.

values shown in Table 1 for $K = 4$, a parent density function was generated, and 250 sets of data ($N = 600$) were drawn from this density using Monte Carlo sampling as described in Materials and Methods. The EM algorithm was applied to each of the density functions obtained in this way, with $K = 4$ and $K = 5$. The Wilks statistic was calculated for each pair of mixture models. These values were rank-ordered and plotted as a cumulative distribution, as shown in Fig. 3 *A*. (The rank ordering of 1 to 250 has been rescaled from 0 to 1.0.) The Wilks statistic for the original data was 8.14, which occurs for $\alpha < 0.01$. That is, we reject $K = 4$ with a probability of less than 0.01 of being incorrect. The statistic in Fig. 3 *A* indicates that for about one-half of the Monte Carlo sample sets, the quality of the fits with $K = 4$ and $K = 5$ was almost identical. For the other half, the $K = 5$ model was superior.

A similar procedure was used to discriminate between the models with $K = 5$ (H_0) and $K = 6$ (H_1). The $K = 5$ model (Table 1) was the parent distribution for Monte Carlo sampling, and the distribution formed by the Wilks statistic when generated by this process is shown in Fig. 3 *B*. The Wilks statistic obtained when these two models were applied to the original data was 0, and on this basis $K = 5$ cannot be rejected; also, well over half of the samples gave equally good fits for $K = 5$ and $K = 6$. Discrimination between the models with $K = 5$ and $K = 6$ must be based on other criteria. Because the mixture with $K = 6$ has two additional degrees of freedom over the mixture with $K = 5$, we accept the $K = 5$ model and reject the $K = 6$ model on the grounds of parsimony. The mixture distribution with $K = 5$ is shown in Fig. 2 *A*.

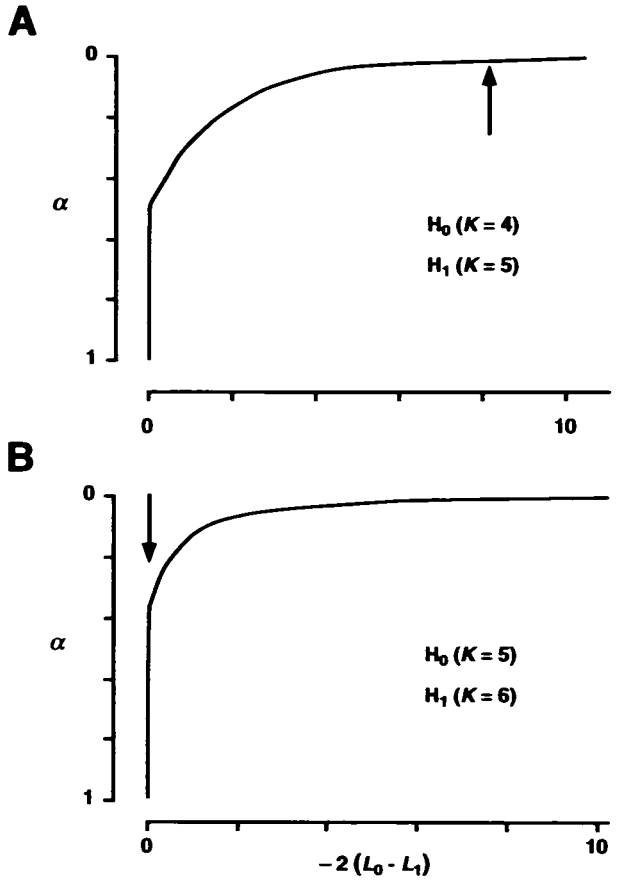

FIGURE 3 (*A*) The Wilks statistic was calculated for each of 250 data sets drawn from the best fit to a mixture with $K = 4$, when each data set was fitted to a mixture with $K = 4$ (H_0) and $K = 5$ (H_1). The Wilks statistic was rank-ordered, and the ordinate was scaled from 1.0 to 0 (corresponding to $\alpha = 1 - p$). The arrow indicates the Wilks statistic for the original data. The mixture with $K = 4$ could be rejected with $\alpha < 0.01$. (*B*) The Wilks statistic for testing $K = 5$ (H_0) against $K = 6$ (H_1). It was generated in a similar manner to the statistic in *A*.

Confidence limits of discrete amplitudes and their probabilities

The data sample was then resampled in a balanced manner as described in Materials and Methods. Each new sample contained the same number of observations as the original data (600) and was formed into a probability density. The EM algorithm, with $K = 5$, was applied to each probability density, and a set of parameters corresponding to those in Table 1, $K = 5$, was calculated. This process was repeated 250 times, and provided 250 values for each of the model parameters. These were used to calculate the SD for each parameter. The result is shown in Fig. 2 *C*, where the error bars correspond to ±1.0 SD.

The relative error in probability increases as the probability of a component decreases, because the sample size from which estimates of that probability are obtained is smaller than for components with a higher probability. The confidence limits on the amplitude of the components are more

complicated to interpret. They will be narrow if the peak in the parent density function corresponding to that amplitude is well defined and has a width that is comparable to the SD of the noise. They will also be narrow even if they correspond to a poorly defined peak in the parent density, provided well defined peaks exist for the two components on either side of it. This has the effect of locking the possible amplitudes of the intermediate component into a narrow range. When peaks become indistinct because of poor signal-to-noise, the error in amplitude will increase. These issues are considered more fully later in Results.

Quantal separation of amplitudes

The EM algorithm for the constraint that equal separation occurs between adjacent discrete amplitudes was applied to the data (Stricker and Redman, 1994, Eqs. A12–A15). In this case, a zero offset is needed to allow for the possibility that a field potential or a stimulus artefact adds to the recorded current. The result is shown in Fig. 4 *A*, together with the original density. The offset was 0.24 ± 0.08 pA, the quantal amplitude 2.97 ± 0.16 pA, and the probabilities attached to the five amplitudes (including failures) are indicated in Fig. 4 *B*. The confidence limits for amplitudes and probabilities were obtained from 250 subsamples and are also shown in Fig. 4 *B*. These were obtained by the same procedure as that described for Fig. 2 *C*. The confidence limits on the currents with quantal separation are narrower than for the unconstrained model, because the quantal constraint allows less freedom in the values than these currents can take. The confidence limits on the larger currents are greater, because the error in offset and quantal size is cumulative for components having increasing numbers of quanta.

This result has a log-likelihood of −1447.5, compared with −1445.24 for the unconstrained result. The difference in the degrees of freedom is 3 (9 − 6). Because the two models are nested, the Wilks statistic (4.52) is distributed as $\chi^2_{(3)}$, and this statistic indicates that the quantal model cannot be rejected for $\alpha < 0.05$. On the basis of parsimony, the quantal model must be accepted.

Binomial models

Two binomial models were considered. One assumed that all the failures were failures of transmitter release. The other allowed for failures to be caused by a combination of failure to stimulate the axon and failure to release transmitter. The EM algorithm for the first model (Stricker and Redman, 1994, Eqs. A13–A18) was used to obtain the result shown in Fig. 5 *A*. The Wilks statistic for the comparison of this model with the quantal model in Fig. 4 is 43.4, with 3 (6 − 3) degrees of freedom. The two models are nested. This model can clearly be rejected at a level of significance $\alpha < 0.005$.

The parameters for the second model (with stimulation failures) were obtained using Eqs. A39 and A40 in the companion paper. The best fit to this model is shown in Fig. 5 *C*, with release probability = 0.32 ± 0.03, offset = 0.24 ± 0.08 pA, quantal amplitude = 3.03 ± 0.13 pA, and the probability of stimulation failure = 0.36 ± 0.05. The Wilks statistic for comparison with the quantal model (Fig. 4) is 1.6 with 2 (6 − 4) degrees of freedom. This model cannot be rejected and has to be accepted on grounds of parsimony. As stated at the beginning of Results, this model is the one from which the data used in the analysis were sampled. The parameter values recovered from the data using this model are within ±1.0 SD of the known parameter values. The small differences between these results and the actual parameters used in the simulation can be explained by the randomness of sampling, rather than by any inaccuracies introduced by the algorithm. The confidence limits on the individual peak currents and probabilities are shown in Fig. 5 *D*.

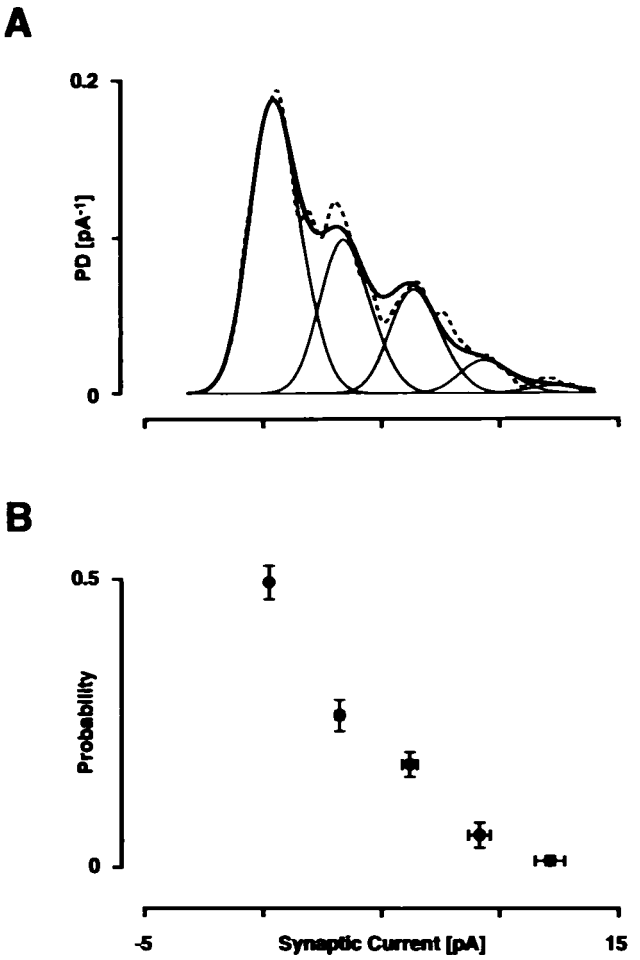

FIGURE 4 (*A*) The heavy line is the best fit to the density derived from simulated data (- - - -) using a mixture with $K = 5$. The components of the mixture are also shown. The separation of the amplitudes of these components was constrained to be quantal, and the amplitude of the first component can be offset from zero. (*B*) The filled circles indicate the probability and amplitude of each of the five components. The bars indicate ±1.0 SD in amplitude and probability.

Compound binomial models

The two versions of a compound binomial process considered were models with and without failures to stimulate the

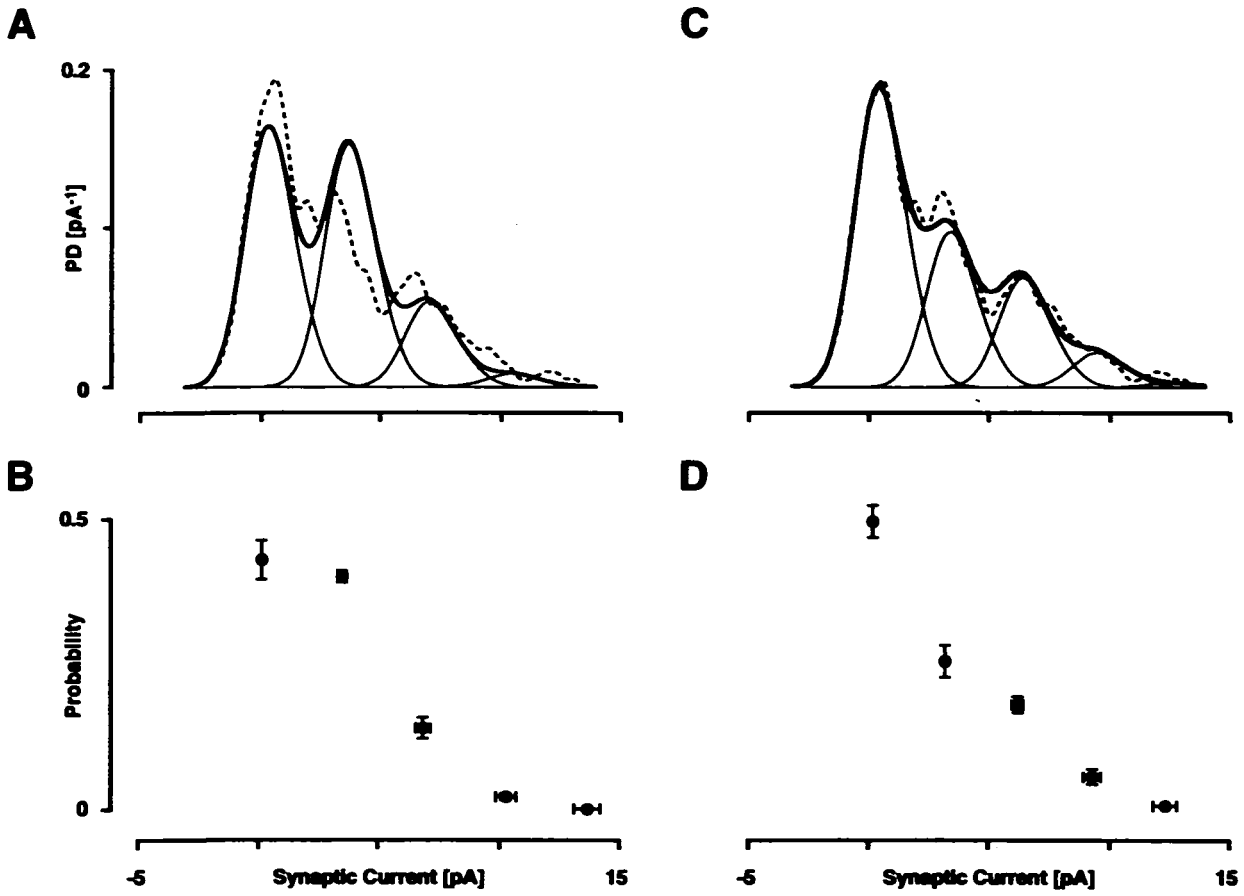

FIGURE 5 (*A*) The solid line is the best fitting density to the simulated data (- - - -) when the amplitudes of the underlying components were constrained to have quantal separation, and the probabilities of the components were constrained to a binomial release process. An offset from zero for the first component was also allowed. (*B*) The amplitudes, probabilities, and ±1.0 SD in these measures are shown for the five components used to fit the data in *A*. (*C*) The solid line is the best fit to the data when the same constraints as in *A* are imposed, but with the freedom to allow some of the failures to be stimulation failures. (*D*) The probability and amplitude of each component of the mixture in *C* is indicated, together with the confidence limits (±1.0 SD) for each probability and amplitude.

axon (Stricker and Redman, 1994, Eqs. A25, A41, and A42). Results that were almost identical to those obtained for the corresponding uniform binomial model were obtained using these models. When stimulation failures were allowed, the assumption of different release probabilities at the four release sites resulted in all of the release probabilities being the same (0.32). The quantal current, offset, and failure to excite the axon were very similar to those obtained for the binomial model.

Skewed unimodal distribution of synaptic current combined with failures

These distributions were introduced to test the hypothesis that such peaks as are present in the density for the current are sampling artefacts, and that the true density describing the synaptic responses is a skewed unimodal distribution. This could result from a quantal transmission process in which the quantal variance was sufficiently large compared with the quantal amplitude so as to obscure the discrete nature of the transmission process. The distributions considered here are the gamma and Weibull distributions. The EM algorithms for fitting these distributions were derived in Stricker and Redman (1994, Eqs. A28–A37).

The best fit to a gamma distribution is illustrated in Fig. 6 *A*, with scale factor $\lambda = 0.76 \pm 0.08$, shape parameter $\beta = 4.04 \pm 0.50$, the probability of failure $= 0.48 \pm 0.03$, and the offset $= 0.21 \pm 0.08$ pA. The gamma distribution in Fig. 6 *A* was sampled (Monte Carlo), and the best fits to a gamma distribution (H_0) and to an unconstrained model (H_0 with $K = 5$) to each sample were obtained. The Wilks statistic was calculated for the fits to these two models. This process was repeated 250 times, and the Wilks statistic was rank-ordered and converted to a cumulative distribution as shown in Fig.6 *B*. The Wilks statistic for the fits of these two models to the original density was 19.5, allowing the gamma distribution to be rejected with $\alpha < 0.01$. A similar result was obtained for the Weibull distribution. The best fit to the data (Fig. 6 *C*) was found with $\beta = 0.0094 \pm 0.0022$ and $\gamma = 2.53 \pm 0.14$, and a probability of failure of 0.51 ± 0.03. The Wilks statistic for the fit of the Weibull distribution (H_0) and the unconstrained mixture model with $K = 5$ (H_0) to the data was 27.9. When Monte Carlo sampling from the Weibull density in Fig. 6 *C* was used to obtain the distribution of the

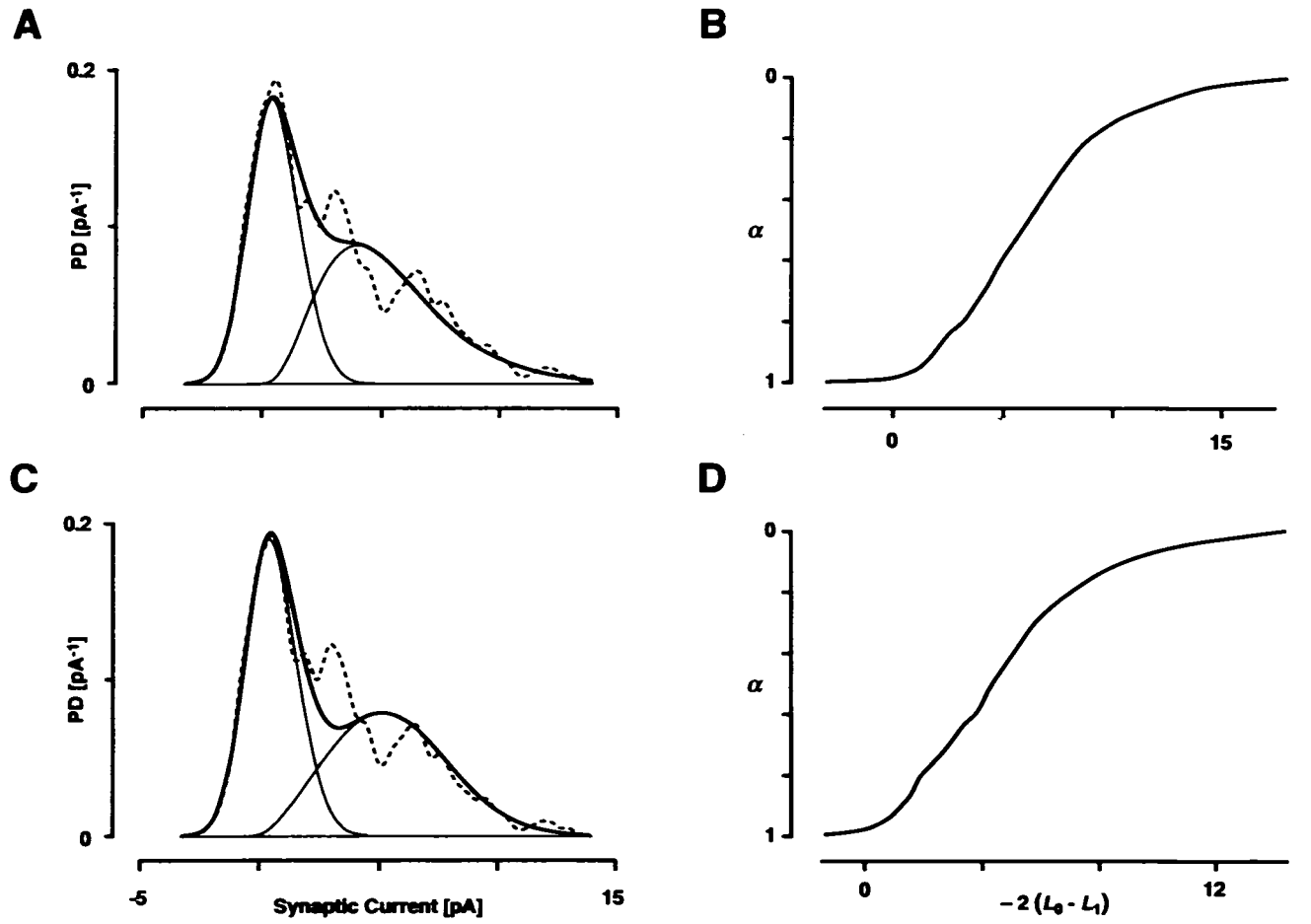

FIGURE 6 (*A*) The heavy line is the best fit to the density obtained from simulated data (- - - -) for a mixture of two distributions. One corresponds to failures in release, and this distribution has the same density as the simulated noise density. The second is a gamma distribution. (*B*) The Wilks statistic was calculated from data sets obtained by Monte Carlo sampling from the fit shown in *A*. H_0 was the sum of a noise distribution and a gamma distribution, whereas H_1 was a mixture model with $K = 5$ and unconstrained separation. The Wilks statistic for the simulated data was 19.5, allowing the mixture with the gamma distribution to be rejected with $\alpha < 0.01$. (*C*) The best fit of the simulated data to the sum of a Weibull distribution and the density representing the noise. (*D*) The Wilks statistic for the mixture shown in *C* (H_0) and a mixture of five components, with unconstrained separation (H_1).

Wilks statistic for these two models (Fig. 6 *D*), the Weibull distribution could also be rejected with $\alpha < 0.004$.

Analysis of experimental data

Here we apply the techniques described in the previous section to a transmitter release process with unknown statistical properties. The EPSCs were evoked in a CA1 pyramidal neurone, by stimulating axons in stratum radiatum at 1 Hz, using a hippocampal slice preparation and whole cell recording techniques. The EPSCs to be analyzed were taken from an experiment in which the average EPSC amplitude (calculated from every 60 sequential responses), access resistance, and baseline noise remained constant. Peak amplitudes and baseline noise were calculated using narrow averaging time windows, as described in Sayer et al. (1989). 600 evoked responses were recorded, of which 11 were subsequently rejected because large spontaneous synaptic currents contaminated the baseline region or the evoked response.

The probability densities of the evoked EPSC and the baseline noise are illustrated in Fig. 7 *A*. These densities were formed from 589 records, using a Gaussian kernel with $\sigma = 0.225$ pA. Resolution of the noise (with $\sigma_n = 0.9$ pA) into two normal distributions gave

$$P_0 = 0.86, \quad \mu_0 = -0.2 \text{ pA}, \quad \sigma_0 = 0.76 \text{ pA}$$
$$P_1 = 0.14, \quad \mu_1 = 1.35 \text{ pA}, \quad \sigma_1 = 0.46 \text{ pA}.$$

Unconstrained mixture model

The EM algorithm was applied to the EPSC density, with no constraints on the probabilities or the amplitudes of the discrete components. No allowance was made for quantal variance. The results are shown in Table 2, for $K = 4$, 5, 6, and 7. The log-likelihood values for these solutions are similar, and it was necessary to generate a Wilks statistic to discriminate between these mixtures. This was done by letting H_0 correspond to $K = 4$ and H_1 to $K = 5$. Monte Carlo samples were taken from the density defined by the parameters in Table 2 for $K = 4$ and then using the EM algorithm to fit the density obtained from each sample to a mixture with $K = 4$ and to another mixture with $K = 5$. The Wilks

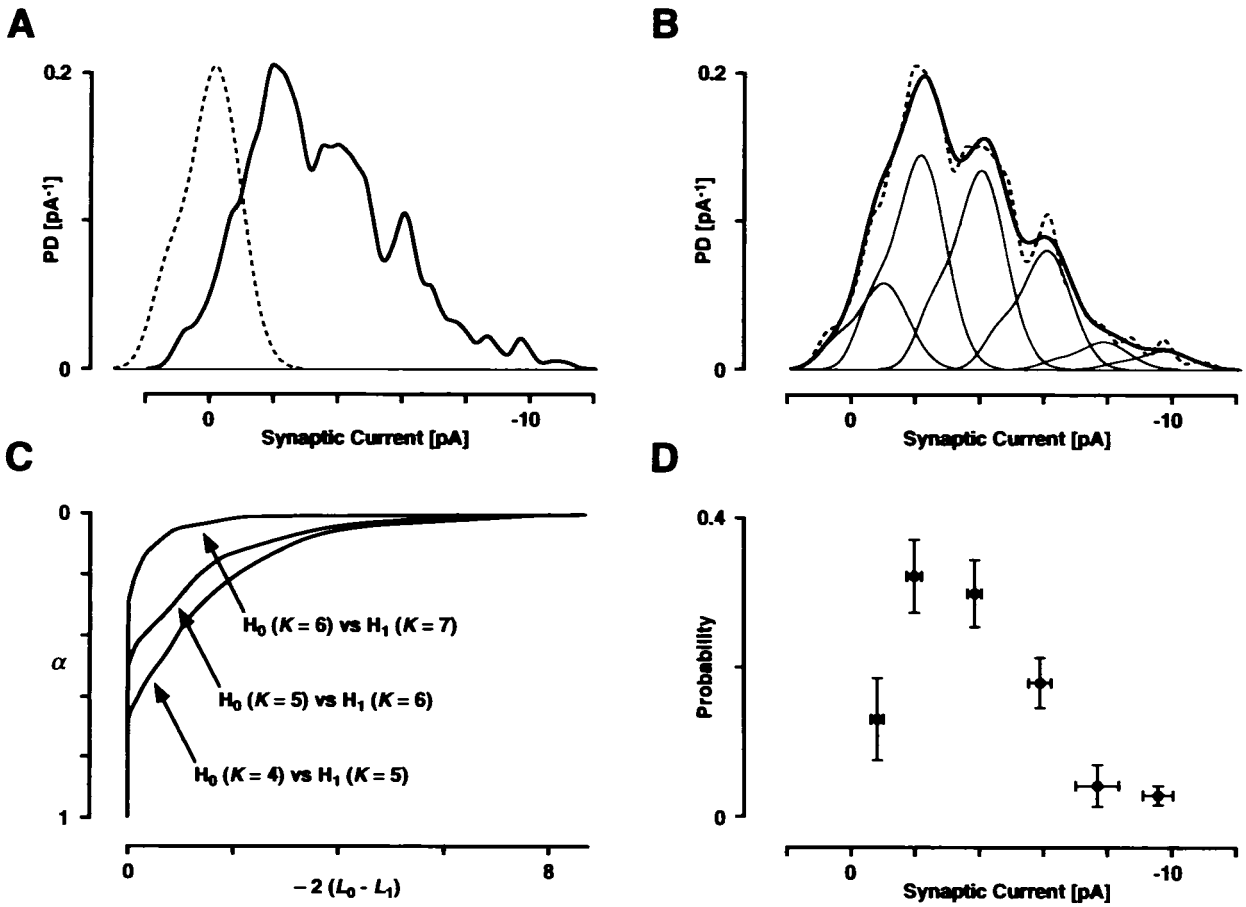

FIGURE 7 (*A*) The heavy line is the probability density for the EPSC, and the dashed line is the measured noise density. (*B*) The heavy line is the best fit for a mixture of six components, without constraints on the amplitude and probability of each component in the mixture. The dashed line is the data, and the underlying mixture components are also shown. (*C*) Wilks statistic calculated for three different mixture models for the probability density of the EPSC shown in *A*. When H_0 corresponded to $K = 4$, and H_1 corresponded to $K = 5$, the Wilks statistic for the data in *A* was 10.5, which allowed H_0 to be rejected with $\alpha < 0.004$. When H_0 corresponded to $K = 5$ and H_1 corresponded to $K = 6$, the Wilks statistic for the data was 4.02, allowing H_0 to be rejected with $\alpha < 0.04$. The Wilks statistic was zero when H_0 corresponded to $K = 6$ and H_1 corresponded to $K = 7$. The model with $K = 6$ could not be rejected. (*D*) The probability and amplitude of each component in the mixture in *B* is indicated, together with confidence limits (±1.0 SD) for each probability and amplitude.

TABLE 2 The results for the best fitting mixture models to the experimental data, using mixtures with 4, 5, 6, and 7 components

	$K = 4$		$K = 5$		$K = 6$		$K = 7$	
Component	P	μ	P	μ	P	μ	P	μ
0	0.38	−1.42	0.13	−0.83	0.13	−0.83	0.13	−0.83
1	0.37	−3.69	0.32	−1.98	0.32	−1.98	0.32	−1.98
2	0.20	−6.09	0.31	−3.90	0.30	−3.86	0.30	−3.86
3	0.05	−9.17	0.19	−6.13	0.18	−5.91	0.18	−5.90
4			0.05	−9.18	0.04	−7.69	0.04	−7.61
5					0.03	−9.59	0.02	−9.21
6							0.01	−9.59
L	−1307.22		−1301.96		−1299.95		−1299.91	

P and μ [pA] are the probability and mean respectively, for each component. L corresponds to the log-likelihood for the fit of each mixture to the data.

statistic was calculated for each sample, and the process was repeated 250 times. These statistics were rank-ordered and formed a cumulative distribution, as shown in Fig. 7 *C*. The Wilks statistic for the original data was 10.52, which allows the $K = 4$ model to be rejected with $\alpha < 0.004$.

The same procedure was used with $K = 5$ (H_0) and $K = 6$ (H_1). The Wilks statistic for the data was 4.02, allowing $K = 5$ to be rejected with $a < 0.04$ (Fig. 7 *C*). When H_0 was the mixture with $K = 6$ and H_1 was the mixture with $K = 7$, the Wilks statistic (Fig. 7 *C*) indicated that for 75% of the sample sets from the $K = 6$ mixture model, there was

no difference in the likelihood obtained by fitting mixture models with $K = 6$ and $K = 7$. For the other 25% of sample sets, a mixture with $K = 7$ gave a better fit. Because the Wilks statistic for the data with the $K = 6$ and $K = 7$ mixture models was 0.08, H_0 cannot be rejected. On this basis, the mixture with $K = 6$ was accepted because it is the more parsimonious model. The fit to the original density is shown in Fig. 7 *B*. The error bars associated with the amplitudes and probabilities of the mixture components are shown in Fig. 7 *D*.

Quantal model

The EM algorithm was used to find the best fit to the EPSC density when a quantal separation was imposed on the discrete currents. No quantal variance or models of release probabilities were assumed. The result is shown in Fig. 8 *A*, and the model parameters are given in Table 3. Comparing the fit to this model with that for the unconstrained model ($K = 6$), the Wilks statistic is 5.94 with four degrees of freedom. The quantal model cannot be rejected for $\alpha < 0.05$ and, therefore, should be accepted.

When quantal variance was introduced to the quantal model, the fit is shown in Fig. 8 *B*. The quantal variance was 0.04 pA^2. The other parameters appear in Table 3. There was very little change in the quantal size or in the probabilities of the discrete amplitudes caused by introducing quantal variance. The coefficient of variation for this result was 0.10. The Wilks statistic was very similar to that for the quantal model, and with three degrees of freedom in this case. Again, this model cannot be rejected for $\alpha < 0.05$. However, because this model has an extra degree of freedom compared with the quantal model with no quantal variance, we can reject this model on the grounds of parsimony.

Binomial and compound binomial models

A different result was reached when both the binomial and compound binomial constraints were applied to the probabilities of the discrete components. The best fits for both models are shown in Fig. 8, *C* and *D*, and the parameters for these models are given in Table 3. Both models could be rejected with $\alpha < 0.005$. When the possibility of failure to stimulate the axon was introduced to the model, the probability of this outcome was zero, and the release probabilities were unaltered.

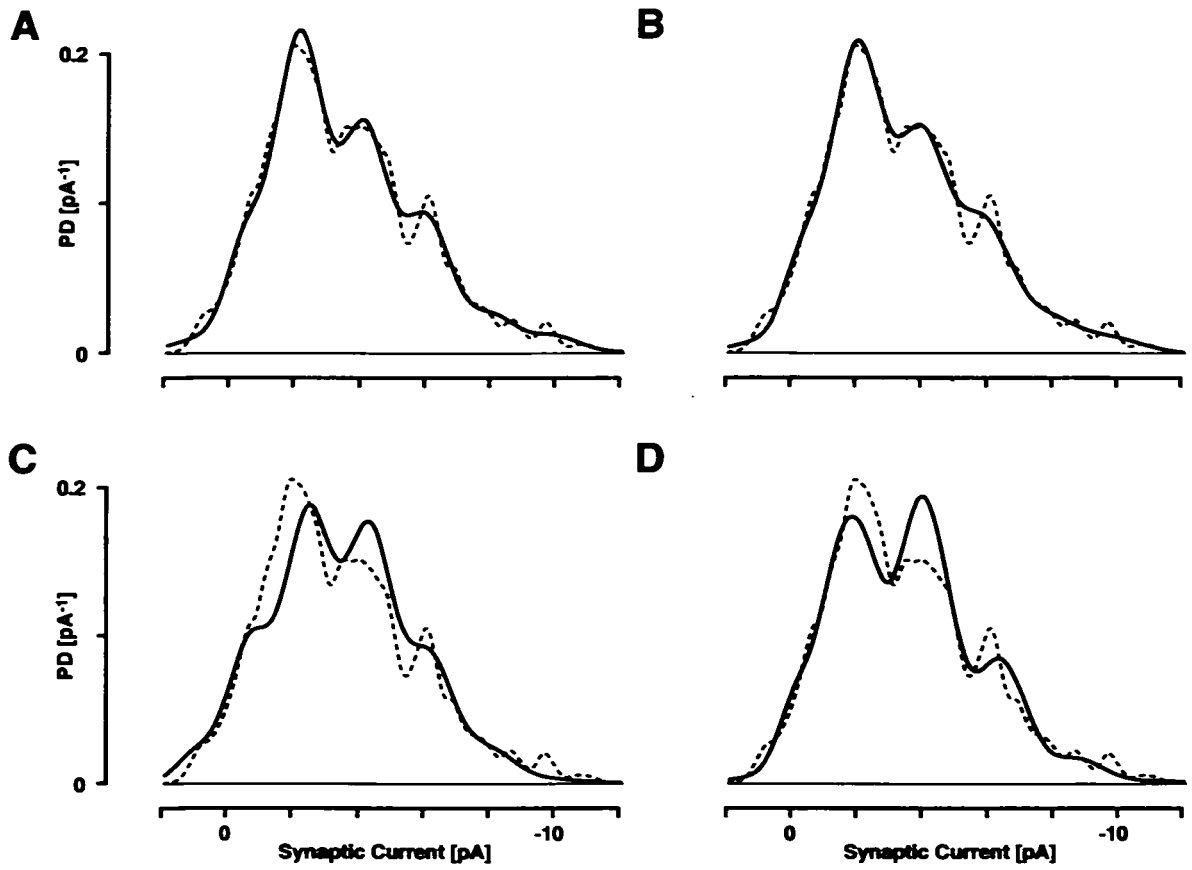

FIGURE 8 (*A*) The dashed line indicates the probability density of the evoked EPSC. The continuous line is the best fit of a mixture model with $K = 6$ and with a quantal separation imposed on the components of the mixture. (*B*) As in *A*, but with the additional freedom to include quantal variance in the model. (*C*) As for *A*, except that the release probabilities at each of the five release sites have been made equal. This constrains the probabilities of each of the six components to obey a binomial distribution. (*D*) As for *A*, except that the probabilities at each release site must be real and positive, and the release process must be statistically independent at each release site. This constrains the probabilities of each component to obey a compound binomial distribution.

TABLE 3 The results for the best fit of various models that all involve quantal separation of components

	Unconstrained quantal model	Unconstrained quantal model with quantal variance	Binomial model with quantal separation	Compound binomial model with quantal separation
Component				
0	0.04	0.03		
1	0.41	0.40		$p_1 = 0.98$
2	0.29	0.30	$p = 0.35$	$p_2 = 0.23$
3	0.20	0.20		$p_3 = 0.23$
4	0.04	0.05		$p_4 = 0.23$
5	0.02	0.02		$p_5 = 0.23$
Quantal size	−1.98	−1.99	−1.91	−2.38
Offset	0.16	0.24	−0.25	0.93
Quantal variance		0.04		
CV		0.10		
L	−1277.19	−1276.88	−1317.51	−1310.05

The first two columns give the probabilities for each component, the quantal size [pA], the zero offset [pA], and the coefficient of variation (CV) in the case of the model with quantal variance [pA^2]. Release probabilities are given for the binomial and compound binomial models. L is the log-likelihood for the fit of each model to the data.

Skewed unimodal distributions combined with failures

The peaks in the EPSC density might have arisen by chance, as a result of finite sampling. To test for this possibility, various skewed unimodal distributions were combined with a normal distribution (to account for failures in the responses) and fitted to the data (Stricker and Redman, 1994, Eqs. A28–A37). Fig. 9*A* shows the best fit to a gamma distribution. The probability of failure was 0.15 ± 0.04, the failure peak was offset from zero by 0.67 ± 0.19 pA, and scale factor $\lambda = 0.86$ ± 0.08 and the shape parameter $\beta = 3.45 \pm 0.40$. The Wilks

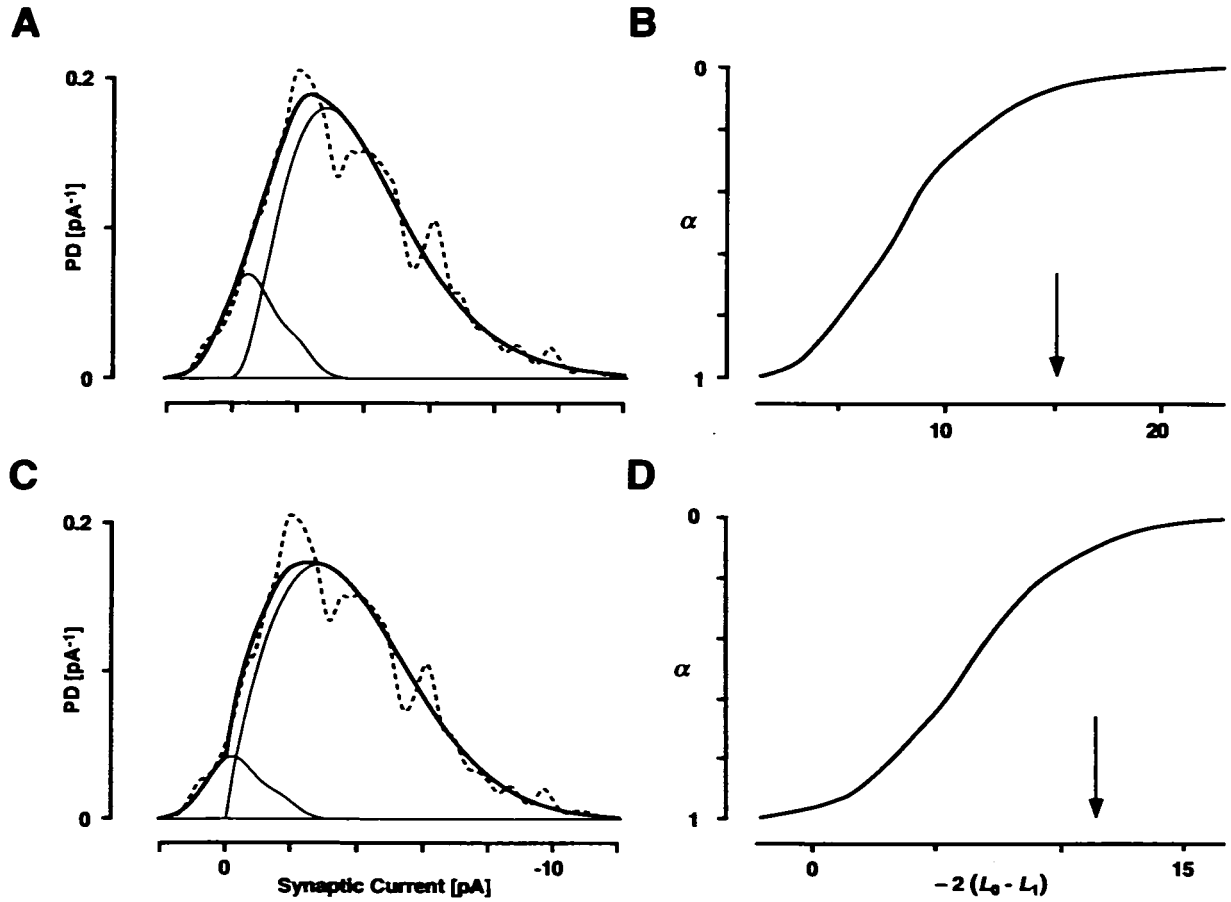

FIGURE 9 (*A*) The dashed line is the probability density of the evoked EPSC. The heavy continuous line is the best fit of a mixture of two distributions: a gamma distribution and a normal distribution. The contribution of each of these two distributions is shown. (*B*) The Wilks statistic obtained by taking 250 sets of samples from the fitted distribution in *A*, fitting a mixture with $K = 6$ (H_1) and a mixture of a normal distribution plus a gamma distribution (H_0), and calculating the Wilks statistic for each sample. The arrow indicates the Wilks statistic for the data (*dashed line* in *A*). (*C*, *D*) These two results were obtained in a similar manner to those in *A* and *B*, except that the mixture distribution fitted in *C* was the sum of a normal and a Weibull distribution.

statistic for comparing the gamma distribution with the unconstrained mixture distribution ($K = 6$) was 15.1. This value was tested using the cumulative distribution of the Wilks statistic obtained from fitting both models (gamma and mixture with $K = 6$) to densities formed by Monte Carlo samples from the gamma distribution in Fig. 9 *A*. The gamma distribution could not be rejected at $\alpha < 0.05$ (Fig. 9 *B*, *arrow*, $\alpha = 0.06$).

A different result was obtained for the Weibull distribution. The best fit for this distribution is shown in Fig. 9 *C*, with a failure probability of 0.09 ± 0.03, an offset from zero of 0.38 ± 0.22 pA, $\beta = 0.067 \pm 0.030$, and $\gamma = 1.85 \pm 0.094$. The Wilks statistic for this fit, compared with the fit to the unconstrained mixture with $K = 6$, was 12.78. The distribution of the Wilks statistic for these two models obtained by Monte Carlo sampling from the Weibull distribution in Fig. 9 *C* showed that the Weibull distribution could be rejected for $\alpha < 0.05$

These procedures were applied to two other unimodal distributions for the nonfailure responses. The first was a normal distribution, and this could be rejected compared with the $K = 6$ mixture distribution, with $\alpha \ll 0.05$. The second was a distribution based on a cubic transformation of a normal variable (Jonas et al., 1994). This distribution could also be rejected for $\alpha < 0.05$.

Confidence limits depend on sample size and signal-to-noise conditions

The confidence limits on model parameters obtained by resampling provide a measure of reliability of those parameters for the conditions pertaining to that data sample. These conditions are the sample size, the signal-to-noise ratio (Q/σ_n, where Q is the quantal amplitude), the number of components in the mixture distribution representing the sample density, and the probabilities of these components.

The effect of some of these variables was studied by simulating a release process that resulted in five discrete amplitudes, equally separated by quantal amplitude Q, and having symmetrical probabilities of 0.12, 0.23, 0.3, 0.23, and 0.12. The effect of varying Q/σ_n was examined by setting the sample size to 500. Fig. 10 *A* illustrates how the confidence limits for the quantal amplitude and the probability of the second amplitude (0.23) varied with Q/σ_n. The normalized error in the probability of the second amplitude is the SD of the probability divided by 0.23. The normalized error in the amplitude is the SD of the amplitude of the second component divided by Q. When $Q/\sigma_n > 3$, there is little improvement in the reliability of the estimates for both amplitude and probability. When $Q/\sigma_n < 2$, the estimates rapidly became unreliable. There was some improvement in reliability when the sample size was increased to 1000, as indicated in Fig. 10 *B*. The results in Fig. 10 *B* apply when $Q/\sigma_n = 3$. There was a rapid deterioration in reliability when the sample size was less than 250.

These results apply for a mixture with five discrete amplitudes. They will be conservative if there are fewer than five components in the mixture, whereas they will be optimistic if there are more than five mixture components. The reliability of probability estimates will also differ for components having different probabilities. Most EPSCs recorded in hippocampal pyramidal neurones have five or more components in their mixture representation (Stricker et al., 1994, in press).

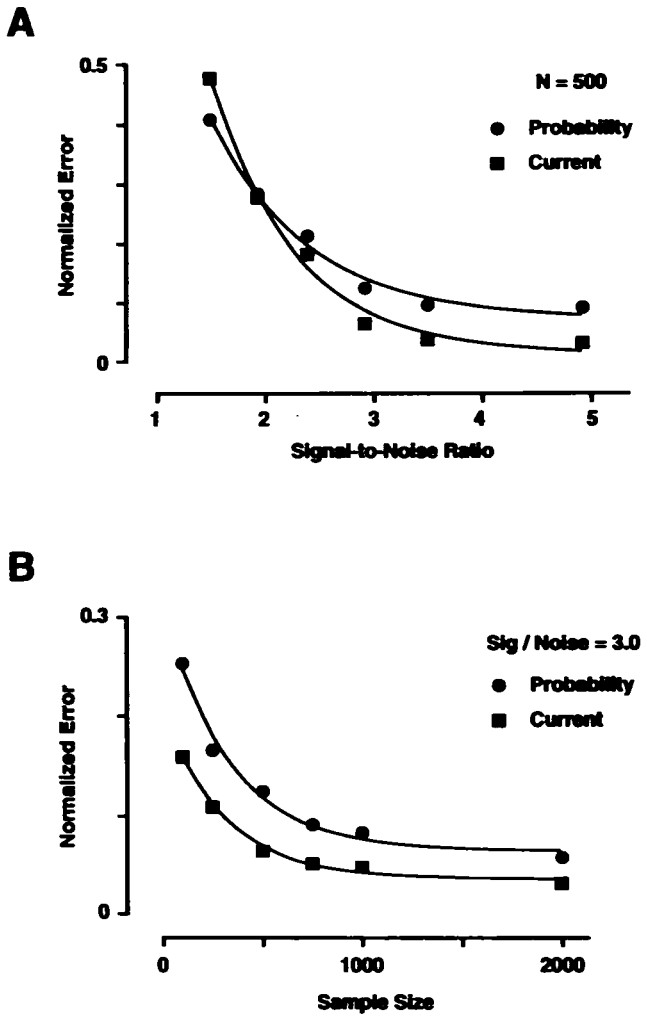

FIGURE 10 How the SDs of the amplitudes and probabilities of the components in a mixture model depend on sample size and signal-to-noise conditions. Details are given in the text.

These examples were not meant to provide an exhaustive treatment of how confidence limits are altered by different quantal sizes, SDs of the baseline noise, sample size, and mixture content. By calculating these confidence limits for each set of data, the reliability of the results can always be assessed. Thus, simulations of the kind illustrated in Fig. 10 are no longer necessary.

DISCUSSION

The methods described in this paper provide a rigorous statistical approach to quantal analysis. There has been a clear need for a powerful statistical test that discriminates between different models of transmission and also a need for confidence intervals of model parameters. Our methods allow

competing models to be tested using one model as a null hypothesis. Once the best model is identified, a method has been described that allows confidence limits to be calculated for the parameters of that model.

Assumptions

There are a number of implicit assumptions in the procedures we have implemented. The first is that the bootstrap principle can be applied to data obtained when measuring the statistical fluctuations of evoked responses in neurones. This can only be established by testing using simulations. We have done this and provided one example to show that the bootstrap principle applies. The second assumption arises when the null hypothesis cannot be rejected, in which case the model having the smaller number of parameters is chosen on the grounds of parsimony. The third assumption is as much a potential source of bias as an assumption. Parameters for quantal models can only be evaluated reliably from data for which the probability density is clearly multimodal. By selecting data of this kind, for which a unimodal distribution can be rejected, we are biasing our analysis towards models of transmission that have a small quantal variance. We are unable to rule out the possibility that the nonmultimodal data resulted from a quantal transmission process with large quantal variance and/or small quantal amplitude.

The number of components in a mixture distribution

The most difficult problem with finite mixture models is to determine the number of components in the mixture. As the mixture dimension is increased, the EM algorithm will provide an asymptotically better fit to the data, and there is a danger of overfitting the data. One approach is to build a penalty factor into the likelihood of the fit to the data that increases as the number of components increases (Horn, 1987; Smith et al., 1991) such as was done by Akaike (1974). The distribution of the statistic calculated in this way (called the Akaike information criterion, or AIC) must be estimated using bootstrap techniques (Horn, 1987), because as Titterington et al. (1985) pointed out, the χ^2 distribution is inappropriate for the AIC measure. There are similarities between this approach and the one used in this paper. The AIC criterion involves an explicit penalty in terms of the number of fitted parameters. The rationale for it has now been derived from at least two distinct intuitively based arguments, and its usefulness has been well tested, especially in the area of autoregression.

Our approach provides a sound statistical basis for rejection of an alternative model without using an *ad hoc* criterion. When a model cannot be rejected at an appropriate level of significance, the decision to accept or reject is made on the grounds of parsimony, thereby imposing an implicit penalty for complexity.

Another method that was designed to avoid overfitting (Kullmann, 1992) involves a trade-off between maximizing both the likelihood and the entropy. When entropy is maximized as done by Kullmann, all components in the mixture have equal probabilities. The trade-off between likelihood and entropy is made by adjusting the weighting of these two criteria in a modified EM algorithm until an acceptable significance level is achieved (using χ^2 or Kolmogorov-Smirnov goodness-of-fit test) that the data were drawn from the mixture model. Because the probabilities of the components generated by a release process are usually less dispersed than for a binomial process (Walmsley et al., 1988), the maximum entropy limit as used by Kullmann is unphysiological and, for this reason, we have not compared the power of this procedure with the likelihood ratio tests.

Unimodal distributions of synaptic responses

An important application of model comparison arises from the debate over whether multimodal densities of EPSC peak amplitudes represent genuine quantal levels of transmission or whether they are artefacts arising from finite sampling. Distributions of EPSCs that are "peaky," or multimodal, have been published (Larkman et al., 1991; Liao et al., 1992; Kullmann, 1992; Jonas et al., 1994), and one appears in Fig. 7. Clements (1991) suggested that these multimodal densities could be artefacts of finite sampling from a unimodal distribution. When this hypothesis was tested against published data using the χ^2 goodness-of-fit test, it could not be rejected. Unimodal distributions of synaptic currents would arise if a quantal release process was associated with a large quantal variance (Bekkers et al., 1990; Raastad et al., 1992). They would also occur if there was negligible quantal variance, but σ_a was large compared with the quantal amplitude. If the quantal amplitudes from different active sites were very different, a unimodal distribution could also occur even with negligible quantal variance. This paper provides a powerful statistical test for examining the null hypothesis that the underlying transmission process gave rise to a unimodal distribution. In practice, this test should be applied to the data before attempting to refine the model beyond an unconstrained mixture model. There would be no point in continuing the analysis if a unimodal distribution could not be rejected. The EPSC analyzed in Results was chosen deliberately to provide an example that failed the multimodal test by a narrow margin. We continued the analysis to illustrate the procedures for evaluating models of quantal transmission. The analysis of a large number of EPSCs will appear elsewhere.

The test for multi-modality is based on the competing claims of an unconstrained mixture model and a unimodal distribution of amplitudes combined with response failures. It does not matter that there might be many more degrees of freedom in the unconstrained mixture model than in the unimodal model. This difference is taken into account when generating the cumulative distribution of the Wilks statistic for comparison of the two models. It would be wrong to examine the ability of a unimodal distribution to represent the data against a highly restrictive model with few degrees of

freedom, such as a binomial release process with quantal amplitudes, because both models might provide equally poor fits to genuinely multimodal data. By making such a comparison, the unimodal distribution might not be rejectable.

In another approach, Larkman et al. (1992) used an autocorrelation measure calculated from the multimodal distribution (based on earlier work by Magleby and Miller, 1981) to test the hypothesis that the multimodal distribution was an artefact caused by finite sampling. We have not assessed the relative power of this autocorrelation method compared with the likelihood ratio test.

Four unimodal distributions have been considered. Many others (e.g., beta or log-normal) could have been included. Why were the four distributions chosen to test for unimodality? The gamma distribution is positively skewed, which makes it an appropriate density for fitting many published densities of evoked responses. Positive skew in evoked responses arises if the release probabilities are small. The cubic transform of a normal variable is also positively skewed. This density function was used by Jonas et al. (1994) to examine unimodality of evoked EPSCs in CA3 pyramidal cells. The normal distribution was used by Clements (1991) when arguing a case that multi-modality could be a sampling artefact. This case was based on data published by Larkman et al. (1991), where the release probabilities were raised by using 4 mM Ca^{2+}, resulting in a roughly symmetric density for the evoked responses. Under these circumstances, a symmetrical distribution, such as a normal distribution, is suitable. The Weibull distribution can be either positively or negatively skewed, and is thus able to cope with all ranges of release probabilities. For this reason, we favor the Weibull distribution as the test distribution for unimodality.

Uniform or nonuniform release statistics

Another important example of model discrimination concerns the probabilities of release at the active sites formed between a neurone and a presynaptic axon. They could all be different, or some or all could be identical. There are many examples of analyses where uniform release probabilities have been assumed at the outset, and the release probability, the quantal amplitude, and the number of release sites have been calculated by fitting the observed density to a binomial model, or more simply, by using the first and second moments of the observations to extract parameters (Korn et al., 1982; Grantyn et al., 1984; Bekkers and Stevens, 1990; Malinow and Tsien, 1990; Foster and McNaughton, 1991; Larkman et al., 1992; Kuhnt et al., 1992; Voronin et al., 1992a–c). In none of these analyses was any attempt made to establish whether a model with uniform release probabilities was superior to one assuming nonuniform release probabilities. It might be that at some of the synapses investigated, uniform release probabilities existed, but if this assumption were incorrect, the parameter values extracted from analysis could contain serious errors.

When models with uniform and nonuniform release statistics have been compared (Jack et al., 1981; Walmsley et al., 1988; Smith et al., 1991) using the χ^2 goodness-of-fit, the uniform release model has been rejected and the nonuniform release model has been accepted. Use of the Wilks criterion led to both models being rejected for the EPSC analyzed in this paper. Note that a χ^2 goodness-of-fit test (strictly speaking, a Pearson χ^2 goodness-of-fit test) is a global test for the adequacy of a specific model describing the data. On the other hand, in the context of two specified models, one nested in the other, with the likelihood functions available for both, the Wilks criterion tests whether the more detailed model does in fact provide a significantly better description of the data.

Problems can occur when searching for the (nonuniform) release probabilities, because local maxima can be encountered (Smith et al., 1991). In our experience, the equations derived in the accompanying paper (Stricker and Redman, 1984) do find the global maximum reliably if the multimodality of the probability density of the responses can be readily seen. If the density approaches a unimodal form, local maxima become a problem. To overcome this difficulty, we used a three-step optimization. In the first step, the quantal amplitude and zero offset were determined, with no constraints on probability. The second step was to use a simplex optimization on the release probabilities with fixed quantal amplitude, as described by Walmsley et al. (1988). The values for quantal amplitude and release probabilities were then used as the starting values for a final optimization on all parameters using the EM algorithm.

The choice of the null hypothesis within a non-nested scheme

Within a non-nested scheme, we have consistently chosen the more parsimonious model description as H_0 and, in doing so, we have introduced a systematic "prejudice" towards it. We have chosen $\alpha = 0.05$, a widely accepted value in biological sciences, and based on this value, the rigor of the tests can be illustrated in Fig. 9, *B* and *D*. Although the Monte Carlo samples were drawn from the gamma or Weibull distribution, in over 90% of the samples the alternative hypothesis ($K = 6$) provided the better fit. This is not surprising given the many more parameters used for H_1 and the probabilistic nature of the sampling procedure that can result in bunching of observations. It also illustrates the nature of the implicit penalty for complexity (see discussion on the number of components in the mixture).

Based on the finding that about half of our data yield multimodal densities at a significance level of 0.05 (Stricker et al., 1994, in press), we are confident that the data sets that allowed the rejection of a unimodal distribution are genuinely multimodal.

The decision pathway

The process used to evaluate competing models has been "top down", in that we began with the model having the largest number of degrees of freedom. If unimodality could

be rejected, we proceeded to eliminate free variables until a model was reached where any further simplification resulted in rejection of the model. Apart from the test for unimodality, this testing was done within the confines of nested models. In particular, we kept K constant. The decision as to which model is H_0 and which is H_1 is not important in this procedure. It might be that the difficulties we have experienced in fitting the binomial and compound binomial models stem from this restriction. The appropriate way to proceed for each of the binomial and compound binomial models might be to find the value of K that gives the best fit for each release model, and then to evaluate the competing (non-nested) models with different values of K. This procedure involves much more computation than is required when working within a nested scheme, and we are currently evaluating these ideas.

An alternative decision pathway is one of evaluating competing models using a "bottom up" approach, whereby model complexity is added progressively until a model is reached where there is no advantage in having an additional parameter. This would need a non-nested approach, because K would need to be varied. Although this is an appropriate pathway for determining K within the confines of the same model, the danger with this approach for comparisons of different models is that a decision to accept a model might be made before all parameters have been tested, and an important parameter might be missed.

Confidence limits

It is surprising that after almost 40 years of quantal analysis, methods for providing confidence limits on quantal parameters (apart from n and p in the binomial model; McLachlan, 1978) have only recently received attention. These confidence limits are important because not only do they indicate the reliability of the model parameters, but they allow meaningful comparisons of apparent changes in these parameters when synaptic strength is altered. The methods developed in this paper are based on resampling and bootstrapping procedures. Recently, Turner and West (1993) used Bayesian inference techniques to calculate confidence limits. At present, their method cannot be attached to mixture models with constraints of quantal separation of amplitudes and with binomial constraints on the release process. Smith et al. (1991) pointed out that the Fisher information matrix can be used to obtain the variance of each of the estimated model parameters. They have calculated the confidence limits of the parameters in several mixture models.

It is important to interpret these confidence limits correctly. They are a measure of the uncertainty in the estimates of the model parameters caused by a finite sample, the signal-to-noise ratio (Q/σ_n), the number of components in the underlying mixture distribution, the probabilities of these components, and the errors in the maximum likelihood solution caused by the algorithm finding local maxima. In previous analyses, the effects of these variables on the resolution of model parameters have been inferred for representative experimental conditions using Monte Carlo simulations (Jack et al., 1981; Wong and Redman, 1980; Sayer et al., 1989; Solodkin et al., 1991; Clamann et al., 1991; Voronin et al., 1992). Claims that quantal amplitudes have been overestimated because of poor signal-to-noise levels have been made (Solodkin et al., 1991; Clamann et al., 1991; Voronin et al., 1992; Dityatev and Clamann, 1993). The provision of confidence limits removes this uncertainty about the resolvability of the model parameters, and it is an important step in introducing rigor to the analysis. However, the confidence limits convey no information on the reliability of the original sample. When only one data sample can be obtained from the experiment, there is no way to obtain a measure of how well this sample represents the underlying process.

The likelihood measure

The procedures developed in this paper use the likelihood measure as a unifying framework. The maximum likelihood was found for all models using the EM algorithm, which guarantees convergence. The maximum likelihood was then used as the statistic on which all model comparisons were based. The use of the likelihood measure avoids the use of *ad hoc* procedures such as autocorrelation measures and maximum entropy to cope with special problems. The methods have been tested against simulated data and have been shown to be reliable. The use of the EM algorithm to find the maximum likelihood for all the models relies on recursive equations derived in the following paper (Stricker and Redman, 1994).

We are grateful to Dr. John Clements for lively discussions and comments on this paper. We appreciate the help we received from Dr. Andrew Wood. Some of the referees comments were very helpful.

This work was supported by fellowships of the Swiss National Science Foundation and the Schweizerische Stiftung für Medizinisch-Biologische Stipendien to C. Stricker.

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19:716–723.

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.

Bekkers, J. M., and C. F. Stevens. 1990. Presynaptic mechanism for long-term potentiation in the hippocampus. *Nature.* 346:724–729.

Clamann, H. P., M.-S. Rioult-Pedotti, and H.-R. Lüscher. 1991. The influence of noise on quantal EPSP size obtained by deconvolution in spinal motoneurons in the cat. *J. Neurophysiol.* 65:67–75.

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

Clements, J. D., I. D. Forsythe, and S. J. Redman. 1987. Presynaptic inhibition of synaptic potentials evoked in cat spinal motoneurones by impulses in single group Ia axons. *J. Physiol.* 383:153–169.

Colquhoun, D., and F. J. Sigworth. 1983. Fitting and statistical analysis of single-channel records. *In* Single-Channel Recording. 1st Ed. B. Sakman and E. Neher, editors. Plenum Press, New York. 191–263.

Davison, A. C., D. V. Hinkley, and E. Schechtman. 1986. Efficient bootstrap simulation. *Biometrika.* 73:555–566.

Del Castillo, J., and B. Katz. 1954. Quantal components of the end-plate potential. *J. Physiol.* 124:560–573.

Dityatev, A. E., and H. P. Clamann. 1993. Limits of quantal analysis reliability: quantal and unimodal constraints and setting of confidence intervals for quantal size. *J. Neurosci. Methods.* 50:67–82.

Efron, B. 1979. Bootstrap methods: another look at the jacknife. *Ann. Statist.* 7:1–26.

Efron, B. 1990. More efficient bootstrap computations. *J. Am. Stat. Ass.* 85:79–89.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* 1:54–77.

Efron, B., and R. Tibshirani. 1993. An Introduction to the Bootstrap. Chapman and Hall, New York. 436 pp.

Foster, T. C., and B. L. McNaughton. 1991. Long-term enhancement of CA1 synaptic transmission is due to quantal size, not quantal content. *Hippocampus.* 1:79–91.

Grantyn, R., A. I. Shapovalov, and B. I. Shiriaev. 1984. Relation between structural and release parameters at the frog sensory-motor synapse. *J. Physiol.* 349:459–474.

Hestrin, S. 1992. Activation and desensitization of glutamate-activated channels mediating fast excitatory synaptic currents in the visual cortex. *Neuron.* 9:991–999.

Horn, R. 1987. Statistical methods for model discrimination. Applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophys. J.* 51:255–263.

Jack, J. J. B., S. J. Redman, and K. Wong. 1981. The components of synaptic potentials evoked in cat spinal motoneurones by impulses in single group Ia afferents. *J. Physiol.* 321:65–96.

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

Korn, H., and D. S. Faber. 1987. Regulation and significance of probabilistic release mechanisms at central synapses. *In* Synaptic Function. G. M. Edelman, W. E. Gall, and W. M. Cowan, editors. John Wiley & Sons, New York. 57–108.

Korn, H., A. Mallet, A. Triller, and D. S. Faber. 1982. Transmission at a central inhibitory synapse. II. Quantal description of release, with a physical correlate for binomial *n*. *J. Neurophysiol.* 48:679–707.

Kuhnt, U., G. Hess, and L. L. Voronin. 1992. Statistical analysis of long-term potentiation of large excitatory postsynaptic potentials recorded in guinea pig hippocampal slices: binomial model. *Exp. Brain Res.* 89: 265–274.

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.

Kullmann, D. M., and R. A. Nicoll. 1992. Long-term potentiation is associated with increase in quantal content and quantal amplitude. *Nature.* 357:240–244.

Larkman, A., T. Hannay, K. Stratford, and J. J. B. Jack. 1992. Presynaptic release probability influences the locus of long-term potentiation. *Nature.* 360:70–73.

Larkman, A., K. Stratford, and J. Jack. 1991. Quantal analysis of excitatory synaptic action and depression in hippocampal slices. *Nature.* 350: 344–347.

Léger, C., D. N. Politis, and J. P. Romano. 1992. Bootstrap technology and applications. *Technometrics.* 34:378–398.

Liao, D., A. Jones, and R. Malinow. 1992. Direct measurement of quantal changes underlying long-term potentiation in CA1 hippocampus. *Neuron.* 9:1089–1097.

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.

Magleby, K. L., and D. C. Miller. 1981. Is the quantum of transmitter release composed of subunits? A critical analysis in the mouse and frog. *J. Physiol.* 311:267–287.

Malinow, R., and R. W. Tsien. 1990. Presynaptic enhancement shown by whole-cell recordings of long-term potentiation in hippocampal slices. *Nature.* 346:177–180.

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.

Raastad, M., J. F. Storm, and P. Andersen. 1992. Putative single quantum and single fibre excitatory postsynaptic currents show similar amplitude range and variability in rat hippocampal slices. *Eur. J. Neurosci.* 4: 113–117.

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

Sayer, R. J., S. J. Redman, and P. Andersen. 1989. Amplitude fluctuations in small EPSPs recorded from CA1 pyramidal cells in the guinea pig hippocampal slice. *J. Neurosci.* 9:840–850.

Silverman, B. W. 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London. 175 pp.

Smith, B. R., J. M. Wojtowicz, and H. L. Atwood. 1991. Maximum likelihood estimation of non-uniform transmitter release probabilities at the crayfish neuromuscular junction. *J. Theor. Biol.* 150:457–472.

Solodkin, M., I. Jiménez, W. F. Collins III, L. M. Mendell, and P. Rudomin. 1991. Interaction of baseline synaptic noise and Ia EPSPs: evidence for appreciable negative correlation under physiological conditions. *J. Neurophysiol.* 65:927–945.

Stricker, C., and S. J. Redman. 1994. Statistical models of synaptic transmission evaluated using the expectation-maximization algorithm. *Biophys. J.* In press.

Titterington, D. M., A. F. M. Smith, and U. E. Makov. 1985. Statistical Analysis of Finite Mixture Distributions. John Wiley & Sons, Chichester, England. 243 pp.

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

Voronin, L. L., U. Kuhnt, and A. G. Gusev. 1992a. Analysis of fluctuations of "minimal" excitatory postsynaptic potentials during long-term potentiation in guinea pig hippocampal slices. *Exp. Brain Res.* 89:288–299.

Voronin, L. L., U. Kuhnt, A. G. Gusev, and G. Hess. 1992b. Quantal analysis of long-term potentiation of "minimal" excitatory postsynaptic potentials in guinea pig hippocampal slices: binomial approach. *Exp. Brain Res.* 89:275–287.

Voronin, L. L., U. Kuhnt, G. Hess, A. G. Gusev, and V. Roschin. 1992c. Quantal parameters of "minimal" excitatory postsynaptic potentials in guinea pig hippocampal slices: binomial approach. *Exp. Brain Res.* 89: 248–264.

Voronin, L. L., U. Kuhnt, N. V. Ivanov, and A. G. Gusev. 1992d. Reliability of quantal parameter estimates and their changes during long-term potentiation in guinea pig hippocampal slices. *Neurosci. Lett.* 146:111–114.

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

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.

Wilks, S. S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* 9:60–62.

Wong, K., and S. Redman. 1980. The recovery of a random variable from a noisy record with application to the study of fluctuations in synaptic potentials. *J. Neurosci. Methods.* 2:389–409.