Downloaded from rsta.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Distributions of the Apparent Open Times and Shut Times in a Single Channel Record when Brief Events Cannot Be Detected

A. G. Hawkes, A. Jalali and D. Colquhoun

Phil. Trans. R. Soc. Lond. A 1990 **332**, 511-538
doi: 10.1098/rsta.1990.0129

References

Article cited in:
http://rsta.royalsocietypublishing.org/content/332/1627/511#related-urls

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. A* go to: **http://rsta.royalsocietypublishing.org/subscriptions**

This journal is © 1990 The Royal Society

Downloaded from rsta.royalsocietypublishing.org

The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected

By A. G. Hawkes[1], A. Jalali[1] and D. Colquhoun[2]

[1] *European Business Management School, University College of Swansea, Singleton Park, Swansea SA2 8PP, U.K.*

[2] *Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, U.K.*

The openings and shuttings of individual ion channel molecules can be described by a Markov process with discrete states in continuous time. The predicted distributions of the durations of open times, shut times, bursts of openings, etc., are all described, in principle, by mixtures of exponential densities. In practice it is usually found that some of the open times, and/or shut times, are too short to be detected reliably. If a fixed dead-time τ is assumed then it is possible to define, as an approximation to what is actually observed, an 'extended opening' or e-opening which starts with an opening of duration at least τ followed by any number of openings and shuttings, all the shut times being shorter than τ; the e-opening ends when a shut time longer than τ occurs. A similar definition is used for e-shut times. Several authors have derived approximations to the distribution of durations of e-openings and e-shuttings. In this paper the exact distributions are derived. They are defined piecewise over the intervals τ to 2τ, 2τ to 3τ, ..., etc., the distribution in each interval being a sum of products of polynomials in t with exponential terms. The number of terms is finite, but increases as intervals get further from $t = \tau$. An asymptotic form for large t (for which the exact solution becomes difficult to compute) is given for the two state case. The exact solution is compared with several approximations, some of which are shown to be good enough for use in most practical applications.

1. The nature of the problem

Currents through ion channels

Electric currents that flow through cell membranes are of great importance for a wide range of biological functions, for example conduction of nerve impulses, transmission of nerve impulses across synaptic junctions, the control of secretion and many others. These currents are carried by a flow of ions (such as Na^+, K^+ and Ca^{2+}), but the cell membrane itself is impermeable to such charged particles. The ions flow through specialized protein pores embedded in the cell membrane; these are known as *ion channels*. The opening and shutting of ion channels is controlled, for example, by the membrane potential or by the concentration of a neurotransmitter substance such as acetylcholine. Until recently it was possible, as in most of chemistry and physics, to measure only the responses of a large ensemble of molecules (channels). However, in

Phil. Trans. R. Soc. Lond. A (1990) **332**, 511–538
Printed in Great Britain

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions: Physical Sciences and Engineering.* www.jstor.org

Downloaded from rsta.royalsocietypublishing.org

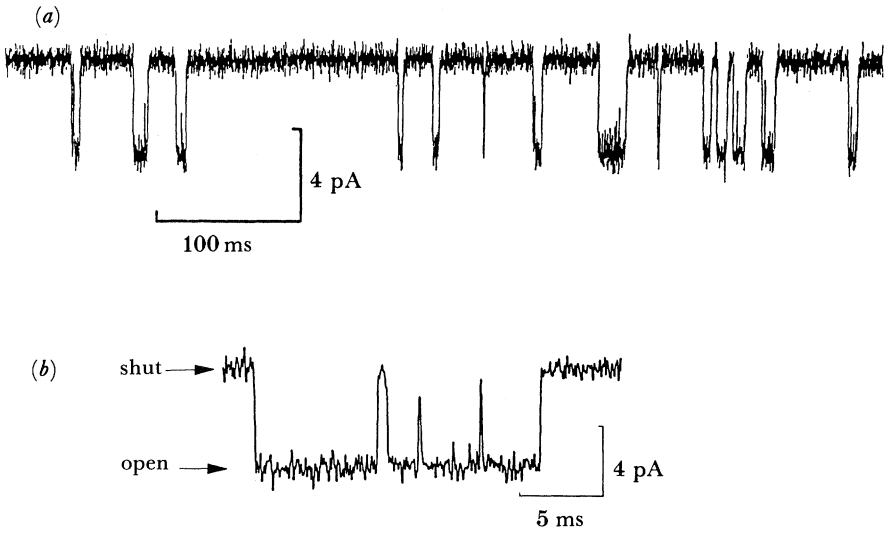

Figure 1. (*a*) A short section from a recording of nicotinic acetylcholine receptor-channels in frog muscle endplate, activated by 100 nM suberyldicholine. Reproduced from Colquhoun & Sakmann (1985) with permission. (*b*) An individual channel activation on an expanded timescale.

1976 Neher & Sakmann invented a method of measuring the current through one individual channel molecule. This *patch clamp* method (Hamill *et al.* 1981) has revolutionized the study of ion channel function.

Behaviour of single molecules

Individual molecules do not behave in the deterministic manner expected (as an excellent approximation) of large ensembles of molecules. In conventional deterministic chemical kinetics the law of mass action usually describes observed phenomena well. It is implicit in the application of this law that the system being studied can exist in a small number of discrete interconvertible states, the rate of conversion from one state to another being proportional to the products of the concentrations (or activities) of the reactants involved. On the scale of individual molecules this is equivalent to saying that molecular species behave like a Markov process with discrete states in continuous time. Figure 1*a* shows a short section from a long experimental record; openings of the channel, during which a current of about 4 pA flows, are shown as downward deflections. Notice that the openings all have much the same amplitude (in this case), but randomly variable durations.

In macroscopic kinetics it is usual to apply a further constraint known as the principal of microscopic reversibility or detailed balance. On the single-molecule scale this is equivalent to the stochastic process being time-reversible (see, for example, Colquhoun & Hawkes 1982, pp. 24–25).

The problem of resolution

Single channel recording has a much better time resolution than previous methods, so it has revealed rapid events in channel function that were previously unsuspected (Colquhoun & Sakmann 1985). Movements within protein molecules take place on a very wide range of timescales, from femtoseconds to seconds (McCammon & Harvey 1987). It is not, therefore, surprising that experimental records always seem to show

Downloaded from rsta.royalsocietypublishing.org

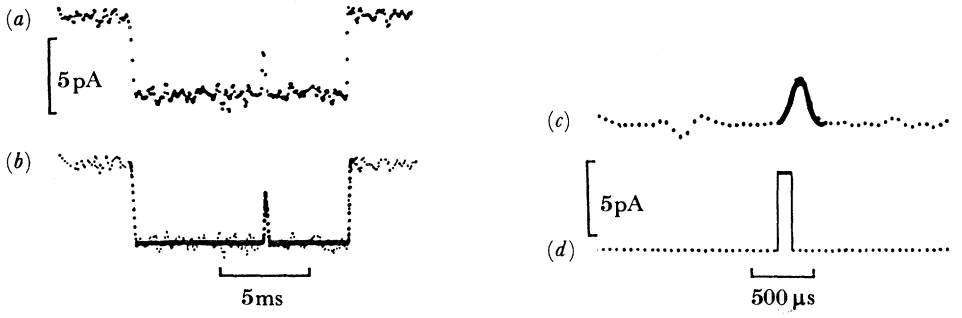

Figure 2. (*a*) A digitized channel activation; (*b*) the same record with calculated response superimposed on it (heavy dots); (*c*) the brief shutting from (*b*) shown expanded; (*d*) the input, a complete shutting of length 104 µs, that gives rise to the fit shown in (*c*). Reproduced from Colquhoun & Sakmann (1981) with permission.

phenomena that are too rapid to be resolved easily, whatever efforts are made to increase the resolution. Figure 1*b* shows one channel activation on an expanded timescale; it consists of at least four openings, separated by three brief shut periods that are clearly resolved, though there may well be briefer undetected shut periods too.

The filtering effect of the recording apparatus is such that the rise-time (10–90 %) of the observed signal, in response to a square input, is at best 30–35 µs. Thus an opening of the ion channel shorter than 20–25 µs will not be detectable given the noise which is present in the recording. The resolution is very often worse than this, up to 500 µs or more, depending on the signal-to-noise ratio in the experimental record, and on the method used for its analysis (see Colquhoun & Sigworth 1983). Figure 2*a* shows a (digitized) channel opening that contains a single obvious brief shutting. In figure 2*b* the heavier points show a fit of the event with a function generated by convolution of measured step-response function of the apparatus; this fit is shown on an expanded timescale in figure 2*c*, together with the input that would give rise to it (figure 2*d*). The effect of filtering is obvious.

Events (openings or shuttings) of the channel that have a mean duration much shorter than the resolution will not be detected at all, and nothing can be done about this. Events with mean durations roughly in the range 5 µs to 1 ms will be detected sometimes and not others. For example if channel open times have a simple exponential distribution with a mean of 50 µs and if the resolution is also 50 µs then 63 % of events will be too short to be detected, only the 37 % that have durations greater than the mean being observed. This can seriously distort the results. For example if channel openings, of mean duration 10 ms, were separated by 50 µs shut periods of which 63 % are undetected, the mean open time would appear to be not 10 ms, but about 27 ms, a serious error. The problems get worse when more complex distributions are considered; for example, attempts to look at the distribution of the length of the first shut period in a burst of openings would be doomed in a case like this.

In cases like that above, when the mean open time is long and few openings are undetected, one can correct for missed shut times retrospectively, by extrapolating the observed distribution of shut times back to $t = 0$ to estimate the number of missing shut periods. This allows a correction to the observed mean open time (but not to the distribution of open times), as was done, for example, by Colquhoun &

Downloaded from rsta.royalsocietypublishing.org

Sakmann (1985). The usefulness of this approach is limited, however. It will not work unambiguously if the distribution of open times has several components; are the missed shut times lost from the 'short' or from the 'long' openings? It obviously will not work if substantial numbers of both open and shut periods are too short to be resolved. Furthermore a retrospective correction of this sort prevents the fitting of a specified model to the data because the model makes predictions, of open and shut time distributions for example, that do not cater for the experimental limitations. In some cases simulation may provide a limited solution (see, for example, Marshall *et al.* 1990), but what is needed is a method of calculating the distributions of what is actually observed, rather than those of what would be observed if there were no resolution problem, as found by methods such as those of Colquhoun & Hawkes (1982).

Several attempts have been made to approximate the distributions that allow for missed events, and these will be referred to below. Sine *et al.* (1990) used such an approximation very effectively to fit experimental results directly with a specified model. In this paper we derive distributions which are exact (given the usual theoretical definition of what is 'observable'), and the exact results are compared with several of the approximations that have been proposed.

Defining the resolution

In what follows it is supposed that all events shorter than some fixed resolution or dead-time (denote τ) are not detected, while all events that are longer than τ are detected and measured accurately. This is itself an approximation. For example, if event durations are measured from points where the signal crosses a 50% threshold, then τ is the event duration needed to just reach the threshold. However, the durations must be corrected for the pulse shape near the threshold and noise in the record makes this imprecise because some events, though actually shorter than τ may cross the threshold because of superimposed noise, and conversely (Colquhoun & Sigworth 1983). When event durations are measured by time-course fitting (see figure 2) the resolution is not well defined, so it must be imposed retrospectively on the measurements by concatenating any shut time below τ with the open times on each side of it to make one long 'apparent opening'. This assumes that all events longer than τ are located and fitted correctly. There is a problem in principle here, because the theory supposes that the resolution is imposed on a perfect record, whereas in fact it is imposed on a record that has, in effect, already had a finite resolution imposed on it by the recording apparatus, and this will not produce exactly the same result.

The problem of defining the resolution gets much harder when the record contains openings to more than one conductance level, especially when there are direct transitions from one open level to another without any (detectable) sojourn in a shut state between them. A brief 'closure' may in fact not represent a complete shutting of the channel but a brief sojourn in one of the lower conductance levels. The problem of the optimum setting of the resolution in such cases is discussed by Howe *et al.* (1990).

2. Distribution of open and shut times

The principles and notation are those used by Colquhoun & Hawkes (1982). The rate constants for transitions between states i and j ($i \neq j$) are the elements, q_{ij}, of the transition rate matrix Q, having the dimensions of reciprocal time, and the diagonal

Downloaded from rsta.royalsocietypublishing.org

elements, q_{ii}, are defined so that the rows sum to zero, so $-1/q_{ii}$ is the mean lifetime of a sojourn in state i. Note that Q is singular and has $k-1$ non-zero eigenvalues, the eigenvalues of $-Q$ being the observed rate constants for the $k-1$ exponential components that describe macroscopic relaxations or fluctuations (noise) (Colquhoun & Hawkes 1977). The transition matrix $T(t)$ has elements defined as

$$T_{ij}(t) = \text{Prob}\,[\text{state } j \text{ at time } t \,|\, \text{state } i \text{ at time zero}] \tag{2.1}$$

and is given by

$$T(t) = \mathrm{e}^{Qt}. \tag{2.2}$$

It is often useful to represent this in terms of the spectral resolution of the matrix Q (see, for example, Colquhoun & Hawkes 1982), so, if the eigenvalues are distinct,

$$T(t) = \sum_{i=1}^{k} A_i \mathrm{e}^{-\lambda_i t}, \tag{2.3}$$

where $\lambda_i, i = 1$ to k, are the eigenvalues of $-Q$ and A_i are the spectral matrices of Q. The Laplace transform of $T(t)$ is

$$T^*(s) = (sI - Q)^{-1}. \tag{2.4}$$

If the states are divided into subset $\mathscr{A}$ which contains the open states, $k_{\mathscr{A}}$ in number, and subset $\mathscr{F}$ which contains the shut states, $k_{\mathscr{F}}$ in number so $k_{\mathscr{A}} + k_{\mathscr{F}} = k$, then (2.4) may be partitioned thus

$$T^*(s) = \begin{bmatrix} sI - Q_{\mathscr{A}\mathscr{A}} & -Q_{\mathscr{A}\mathscr{F}} \\ -Q_{\mathscr{F}\mathscr{A}} & sI - Q_{\mathscr{F}\mathscr{F}} \end{bmatrix}^{-1}. \tag{2.5}$$

This may be written in a more convenient form by using well-known results on the inverse of a partitioned matrix, giving

$$T^*_{\mathscr{A}\mathscr{A}}(s) = [sI - Q_{\mathscr{A}\mathscr{A}} - Q_{\mathscr{A}\mathscr{F}}(sI - Q_{\mathscr{F}\mathscr{F}})^{-1} Q_{\mathscr{F}\mathscr{A}}]^{-1}, \tag{2.6}$$

$$T^*_{\mathscr{A}\mathscr{F}}(s) = T^*_{\mathscr{A}\mathscr{A}}(s)\, Q_{\mathscr{A}\mathscr{F}}(sI - Q_{\mathscr{F}\mathscr{F}})^{-1}. \tag{2.7}$$

A semi-Markov process is embedded in the process at the instants at which the system enters the set $\mathscr{A}$ or enters set $\mathscr{F}$. The intervals between these points have probability densities given by the matrix

$$G(t) = \begin{bmatrix} 0 & \exp(Q_{\mathscr{A}\mathscr{A}} t)\, Q_{\mathscr{A}\mathscr{F}} \\ \exp(Q_{\mathscr{F}\mathscr{F}} t)\, Q_{\mathscr{F}\mathscr{A}} & 0 \end{bmatrix}. \tag{2.8}$$

Thus each event is, alternately, the start of an open period or the start of a closed period. The element $g_{ij}(t)$ of $G(t)$ is the probability density of the time to the next entry into a new subset and the probability that the state entered is j, conditional on starting in state i. The Laplace transform of this matrix will be denoted by

$$G^*(s) = \begin{bmatrix} 0 & G^*_{\mathscr{A}\mathscr{F}}(s) \\ G^*_{\mathscr{F}\mathscr{A}}(s) & 0 \end{bmatrix}, \tag{2.9}$$

where

$$G^*_{\mathscr{A}\mathscr{F}}(s) = (sI - Q_{\mathscr{A}\mathscr{A}})^{-1} Q_{\mathscr{A}\mathscr{F}}, \quad G^*_{\mathscr{F}\mathscr{A}}(s) = (sI - Q_{\mathscr{F}\mathscr{F}})^{-1} Q_{\mathscr{F}\mathscr{A}}. \tag{2.10}$$

Downloaded from rsta.royalsocietypublishing.org

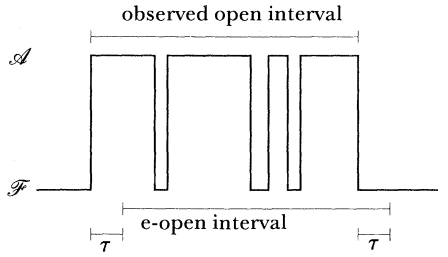

Figure 3. Illustration of the definition of an e-opening for a channel with two levels of conductance and dead-time τ. The e-open interval is equivalent to an observed open interval shifted by an amount τ.

From these transition densities the open and closed time distributions are readily found. For example, the equilibrium distribution of open times has probability density

$$f(t) = \phi_o \exp(Q_{\mathscr{A}\mathscr{A}} t) Q_{\mathscr{A}\mathscr{F}} u_{\mathscr{F}} = \phi_o \exp(Q_{\mathscr{A}\mathscr{A}} t)(-Q_{\mathscr{A}\mathscr{A}}) u_{\mathscr{A}}, \tag{2.11a}$$

where

$$\phi_o = p_{\mathscr{F}}(\infty) Q_{\mathscr{F}\mathscr{A}} / p_{\mathscr{F}}(\infty) Q_{\mathscr{F}\mathscr{A}} u_{\mathscr{A}}. \tag{2.11b}$$

Here $p_{\mathscr{F}}(\infty)$ is the $\mathscr{F}$ partition of the vector of equilibrium probabilities and the vector, ϕ_o, contains equilibrium probabilities of an opening starting in each of the open states; $u_{\mathscr{A}}$ and $u_{\mathscr{F}}$ are vectors of units. Similar results hold for shut times. Using the spectral expansion of the matrices $\exp(Q_{\mathscr{A}\mathscr{A}} t)$ and $\exp(Q_{\mathscr{F}\mathscr{F}} t)$, these distributions may be represented as mixtures of exponentials (Colquhoun & Hawkes 1982). The numbers of components in the mixtures are, respectively, the numbers of open and shut states $k_{\mathscr{A}}$ and $k_{\mathscr{F}}$. Fitting mixtures of exponentials to observed histograms has therefore been used to obtain lower bounds for the numbers of open and shut states.

As described earlier, these distributions are much distorted by an inability to detect small intervals. We shall assume a constant critical gap or dead-time, τ, such that open or shut times less than τ are missed. It is possible to take different dead-times for open and closed times, but it is not necessary in practice and adds a little confusion to the theory. We suppose, after Colquhoun & Sigworth (1983), that an observable open time begins with a sojourn in the $\mathscr{A}$ states of duration at least τ and ends at the start of the next sojourn in $\mathscr{F}$ which is greater than τ. Thus, the observed open time may consist of r shut times, each less than τ, and $r+1$ open times, of which the first must exceed τ. Observed shut times may be defined similarly.

A theory describing these observed open and shut times may be developed. This is slightly awkward theoretically because in 'real time' an open period is not detected until it has lasted for a time τ, although this is no problem analysing recorded data. Therefore, as in Ball & Sansom (1988*a*), we consider a semi-Markov process whose events occur at time τ after the start of observed open or closed periods; see figure 3. An event type is the state of the underlying Markov process occupied at that time. The durations of the intervals between events, which we call e-open and e-closed intervals (not the same terminology as Milne *et al.* (1988)), are identical to those of the observed open and closed intervals, being simply shifted by an amount τ. Intervals of this process will be alternately e-open and e-closed, so the transition densities are given by a matrix whose Laplace transform has the form

$${}^{e}G^{*}(s) = \begin{bmatrix} 0 & {}^{e}G^{*}_{\mathscr{A}\mathscr{F}}(s) \\ {}^{e}G^{*}_{\mathscr{F}\mathscr{A}}(s) & 0 \end{bmatrix}. \tag{2.12}$$

Downloaded from rsta.royalsocietypublishing.org

The Markov chain embedded at the event points has transition matrix

$$ {}^{\mathrm{e}}G = \begin{bmatrix} 0 & {}^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}} \\ {}^{\mathrm{e}}G_{\mathscr{F}\mathscr{A}} & 0 \end{bmatrix}. \tag{2.13} $$

Here, as elsewhere, we simplify the notation when $s = 0$ in a Laplace transform: we omit the '*' and the argument. For example, in this case ${}^{\mathrm{e}}G^*_{\mathscr{A}\mathscr{F}}(0)$ is written as ${}^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}}$.

Looking only at alternate events, ignoring the interval durations, we have a Markov chain on the $\mathscr{A}$ states with transition matrix ${}^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}}\,{}^{\mathrm{e}}G_{\mathscr{F}\mathscr{A}}$ and equilibrium probability vector, $\phi_{\mathscr{A}}$, satisfying

$$ \phi_{\mathscr{A}} = \phi_{\mathscr{A}}\,{}^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}}\,{}^{\mathrm{e}}G_{\mathscr{F}\mathscr{A}}, \quad \phi_{\mathscr{A}}\,u_{\mathscr{A}} = 1. \tag{2.14} $$

See Colquhoun & Hawkes (1983) and Hawkes & Sykes (1990) for simple solutions to such equations. A Markov chain at the closed events has transition matrix ${}^{\mathrm{e}}G_{\mathscr{F}\mathscr{A}}\,{}^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}}$ with equilibrium vector

$$ \phi_{\mathscr{F}} = \phi_{\mathscr{A}}\,{}^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}}. \tag{2.15} $$

To find ${}^{\mathrm{e}}G^*_{\mathscr{A}\mathscr{F}}(s)$ we need to distinguish short (less than τ) and long (greater than τ) sojourns in $\mathscr{F}$ and so, in the manner of Hawkes (1970), we break up the Laplace transform into parts.

$$ \int_0^\tau \mathrm{e}^{-st} \exp(Q_{\mathscr{F}\mathscr{F}}t)\,\mathrm{d}t = \{I - \exp(-(sI - Q_{\mathscr{F}\mathscr{F}})\tau)\}(sI - Q_{\mathscr{F}\mathscr{F}})^{-1} = S^*_{\mathscr{F}\mathscr{F}}(s)(sI - Q_{\mathscr{F}\mathscr{F}})^{-1}, \tag{2.16} $$

$$ \int_\tau^\infty \mathrm{e}^{-st} \exp(Q_{\mathscr{F}\mathscr{F}}t)\,\mathrm{d}t = \exp(-(sI - Q_{\mathscr{F}\mathscr{F}})\tau)(sI - Q_{\mathscr{F}\mathscr{F}})^{-1} = L^*_{\mathscr{F}\mathscr{F}}(s)(sI - Q_{\mathscr{F}\mathscr{F}})^{-1}. \tag{2.17} $$

These expressions define $S^*_{\mathscr{F}\mathscr{F}}$ and $L^*_{\mathscr{F}\mathscr{F}}$, such that

$$ S^*_{\mathscr{F}\mathscr{F}}(s) + L^*_{\mathscr{F}\mathscr{F}}(s) = I. \tag{2.18} $$

Now $L_{\mathscr{F}\mathscr{F}} = L^*_{\mathscr{F}\mathscr{F}}(0) = \exp(Q_{\mathscr{F}\mathscr{F}}\tau)$ gives the probabilities of a long closed time and the state occupied after the dead-time τ, conditional on the initial state. Similar results hold for the open states $\mathscr{A}$. Then it is easy to see that

$$ \left.\begin{aligned} {}^{\mathrm{e}}G^*_{\mathscr{A}\mathscr{F}}(s) &= \sum_{r=0}^{\infty} \{G^*_{\mathscr{A}\mathscr{F}}(s)\,S^*_{\mathscr{F}\mathscr{F}}(s)\,G^*_{\mathscr{F}\mathscr{A}}(s)\}^r G^*_{\mathscr{A}\mathscr{F}}(s)\,L^*_{\mathscr{F}\mathscr{F}}(s), \\ {}^{\mathrm{e}}G^*_{\mathscr{A}\mathscr{F}}(s) &= \{I - G^*_{\mathscr{A}\mathscr{F}}(s)\,S^*_{\mathscr{F}\mathscr{F}}(s)\,G^*_{\mathscr{F}\mathscr{A}}(s)\}^{-1} G^*_{\mathscr{A}\mathscr{F}}(s)\,L^*_{\mathscr{F}\mathscr{F}}(s). \end{aligned}\right\} \tag{2.19} $$

This allows for the possibility of r cycles, each consisting of a sojourn in $\mathscr{A}$ and a short sojourn in $\mathscr{F}$ before returning to $\mathscr{A}$, followed by a final sojourn in $\mathscr{A}$ and then a long sojourn in $\mathscr{F}$: the first period τ of this closed interval is counted as part of the e-open interval (see figure 3). Similarly, we have

$$ {}^{\mathrm{e}}G^*_{\mathscr{F}\mathscr{A}}(s) = \{I - G^*_{\mathscr{F}\mathscr{A}}(s)\,S^*_{\mathscr{A}\mathscr{A}}(s)\,G^*_{\mathscr{A}\mathscr{F}}(s)\}^{-1} G^*_{\mathscr{F}\mathscr{A}}(s)\,L^*_{\mathscr{A}\mathscr{A}}(s). \tag{2.20} $$

The distributions of e-open times and e-closed times have Laplace transforms

$$ {}^{\mathrm{e}}f^*_{\mathscr{A}}(s) = \phi_{\mathscr{A}}\,{}^{\mathrm{e}}G^*_{\mathscr{A}\mathscr{F}}(s)\,u_{\mathscr{F}}, \quad {}^{\mathrm{e}}f^*_{\mathscr{F}}(s) = \phi_{\mathscr{F}}\,{}^{\mathrm{e}}G^*_{\mathscr{F}\mathscr{A}}(s)\,u_{\mathscr{A}}. \tag{2.21} $$

These results were given, using different notation, by Ball & Sansom (1988*a*), generalized to allow different dead-times, $\tau_{\mathscr{A}}$ and $\tau_{\mathscr{F}}$, for open and closed intervals

Downloaded from rsta.royalsocietypublishing.org

and, furthermore, they could be random. They point out that the results (2.19) and (2.20) are implicit in the work of Roux & Sauvé (1985) who, however, use the wrong equilibrium vectors in attempting to obtain the results in (2.21). They also give expressions for moments, which are easily obtained from (2.19) to (2.21).

3. Probability densities of e-intervals

Laplace transforms of the distributions

Laplace transforms are useful for obtaining moments. However, one would like to know probability densities, particularly in view of their role in identifying the numbers of states. Several authors have noted that one may invert Laplace transforms numerically. This is certainly useful but rather prohibitive if one wants to use the results for maximum likelihood inference (Ball & Sansom 1989), and one would really like to know something about the functional form of these densities. They are clearly not mixtures of exponentials, but may perhaps be approximated as such.

We will obtain the probability density of e-open times; the distribution of e-closed times can be obtained simply by interchanging $\mathscr{A}$ and $\mathscr{F}$ in the notation. Let ${}^{\mathscr{A}}R(t)$ be a matrix whose ijth element $(i,j\in\mathscr{A})$ is

$$ {}^{\mathscr{A}}R_{ij}(t) = \text{Prob}\,[X(t)=j \quad \text{and no shut time is detected over} \quad (0,t)\,|\,X(0)=i], \tag{3.1}$$

where a detectable shut time is a sojourn in $\mathscr{F}$ of duration greater than τ. This is a kind of reliability or survivor function: it gives the probability that an e-open time, starting in state i, has not yet finished and is currently in state j. Then the transition density is given by

$$ {}^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}}(t) = {}^{\mathscr{A}}R(t-\tau)\,Q_{\mathscr{A}\mathscr{F}}\exp(Q_{\mathscr{F}\mathscr{F}}\tau). \tag{3.2}$$

This is because, for the e-open interval to end at time t, there must be a transition from $\mathscr{A}$ to $\mathscr{F}$ at time $t-\tau$ (there being no detectable sojourn in $\mathscr{F}$ up to that time) followed by a sojourn of at least τ in $\mathscr{F}$. The corresponding Laplace transform is

$$ {}^{\mathrm{e}}G^{*}_{\mathscr{A}\mathscr{F}}(s) = {}^{\mathscr{A}}R^{*}(s)\,\mathrm{e}^{-s\tau}Q_{\mathscr{A}\mathscr{F}}\exp(Q_{\mathscr{F}\mathscr{F}}\tau). \tag{3.3}$$

Equations (3.2) and (2.21) imply that ${}^{\mathscr{A}}R(t)$ is the key to the required density.

Equations (3.3), (2.19), (2.17) and (2.10) lead to

$$ {}^{\mathscr{A}}R^{*}(s) = \{I-G^{*}_{\mathscr{A}\mathscr{F}}(s)\,S^{*}_{\mathscr{F}\mathscr{F}}(s)\,G^{*}_{\mathscr{F}\mathscr{A}}(s)\}^{-1}(sI-Q_{\mathscr{A}\mathscr{A}})^{-1}$$

and further, using (2.16) to (2.18), this can be written as

$$ {}^{\mathscr{A}}R^{*}(s) = \{sI-Q_{\mathscr{A}\mathscr{A}}-Q_{\mathscr{A}\mathscr{F}}(sI-Q_{\mathscr{F}\mathscr{F}})^{-1}Q_{\mathscr{F}\mathscr{A}}+Q_{\mathscr{A}\mathscr{F}}\,\mathrm{e}^{-s\tau}\exp(Q_{\mathscr{F}\mathscr{F}}\tau)\,(sI-Q_{\mathscr{F}\mathscr{F}})^{-1}Q_{\mathscr{F}\mathscr{A}}\}^{-1},$$

and finally, using (2.6) and (2.7),

$$ {}^{\mathscr{A}}R^{*}(s) = \{I+\mathrm{e}^{-s\tau}T^{*}_{\mathscr{A}\mathscr{F}}(s)\exp(Q_{\mathscr{F}\mathscr{F}}\tau)\,Q_{\mathscr{F}\mathscr{A}}\}^{-1}T^{*}_{\mathscr{A}\mathscr{A}}(s). \tag{3.4}$$

This may also be obtained by taking the Laplace transform of the integral equation:

$$ T_{\mathscr{A}\mathscr{A}}(t) = {}^{\mathscr{A}}R(t)+\int_{0}^{t-\tau} T_{\mathscr{A}\mathscr{F}}(u)\exp(Q_{\mathscr{F}\mathscr{F}}\tau)\,Q_{\mathscr{F}\mathscr{A}}\,{}^{\mathscr{A}}R(t-u-\tau)\,\mathrm{d}u. \tag{3.5}$$

This holds because, to be in $\mathscr{A}$ at t, either there has been no detectable sojourn in $\mathscr{F}$ or there has been at least one such sojourn: then let the last such sojourn terminate

Downloaded from rsta.royalsocietypublishing.org

at time $u+\tau$ with a transition into $\mathscr{A}$ and be followed by a period of duration $t-u-\tau$ with no detectable sojourn in $\mathscr{F}$. A different equation can be obtained by considering instead the beginning of the first detectable sojourn in $\mathscr{F}$ to occur at time u. Then

$$T_{\mathscr{A}\mathscr{A}}(t) = {}^{\mathscr{A}}R(t) + \int_0^{t-\tau} {}^{\mathscr{A}}R(u)\, Q_{\mathscr{A}\mathscr{F}} \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, T_{\mathscr{F}\mathscr{A}}(t-u-\tau)\, \mathrm{d}u, \tag{3.6}$$

which, on taking the Laplace transform, leads to

$${}^{\mathscr{A}}R^*(s) = T^*_{\mathscr{A}\mathscr{A}}(s)\{I + \mathrm{e}^{-s\tau} Q_{\mathscr{A}\mathscr{F}} \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, T^*_{\mathscr{F}\mathscr{A}}(s)\}^{-1}. \tag{3.7}$$

Inversion of the Laplace transforms

If we make series expansions of the expressions in (3.4) and (3.7), we get

$$\begin{aligned} {}^{\mathscr{A}}R^*(s) &= \sum_{m=0}^{\infty} (-1)^m \mathrm{e}^{-ms\tau} \{T^*_{\mathscr{A}\mathscr{F}}(s) \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, Q_{\mathscr{F}\mathscr{A}}\}^m T^*_{\mathscr{A}\mathscr{A}}(s) \\ &= \sum_{m=0}^{\infty} (-1)^m \mathrm{e}^{-ms\tau} T^*_{\mathscr{A}\mathscr{A}}(s) \{Q_{\mathscr{A}\mathscr{F}} \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, T^*_{\mathscr{F}\mathscr{A}}(s)\}^m. \end{aligned} \tag{3.8}$$

Now let

$$H_{\mathscr{A}\mathscr{A}}(t) = T_{\mathscr{A}\mathscr{F}}(t) \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, Q_{\mathscr{F}\mathscr{A}}, \quad t>0; \quad H_{\mathscr{A}\mathscr{A}}(t) = 0, \quad t<0. \tag{3.9}$$

$$K_{\mathscr{A}\mathscr{A}}(t) = Q_{\mathscr{A}\mathscr{F}} \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, T_{\mathscr{F}\mathscr{A}}(t), \quad t>0; \quad K_{\mathscr{A}\mathscr{A}}(t) = 0, \quad t<0. \tag{3.10}$$

Then the inverses of (3.8) can be expressed as

$${}^{\mathscr{A}}R(t) = \sum_{m=0}^{\infty} (-1)^m (H^{\oplus m}_{\mathscr{A}\mathscr{A}} \oplus T_{\mathscr{A}\mathscr{A}})(t-m\tau) = \sum_{m=0}^{\infty} (-1)^m (T_{\mathscr{A}\mathscr{A}} \oplus K^{\oplus m}_{\mathscr{A}\mathscr{A}})(t-m\tau), \tag{3.11}$$

where $\oplus$ denotes convolution and $\oplus m$ denotes m-fold convolution. We will work exclusively with the first of these expressions, but the other would do as well.

Note that, in view of (3.9), the series in (3.11) has only a finite number of terms for any fixed t, namely $n+1$ terms for any t in the interval $I_{n+1} = (n\tau, (n+1)\tau)$. Thus there is no simple functional form, but a different form over each of the intervals I_n. Unfortunately, the number of terms, and their complexity, increases with n. However, we may hope for a good approximation by a simple form for large t.

It is convenient to write

$${}^{\mathscr{A}}R(t) = \sum_{m=0}^{\infty} (-1)^m M_m(t-m\tau). \tag{3.12}$$

We shall prove the following theorem.

Theorem. *If $-Q$ has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, with $\lambda_1 = 0$, then*

$$M_m(t) = \sum_{i=1}^{k} B_{im}(t) \exp(-\lambda_i t), \quad t>0,$$

where $B_{im}(t)$ is a polynomial of degree m in t with matrix-valued coefficients, so

$$B_{im}(t) = \sum_{r=0}^{m} C_{imr} t^r.$$

Downloaded from rsta.royalsocietypublishing.org

Before proving the theorem, let us note the following lemma.

Lemma 3.1. *The convolution of* $t^r \exp(-\lambda_i t)$ *and* $\exp(-\lambda_j t)$ *equals*

$$\begin{cases} t^{r+1}\exp(-\lambda_i t)/(r+1) & (\lambda_i = \lambda_j) \\ \dfrac{r!}{(\lambda_i-\lambda_j)^{r+1}}\left\{\exp(-\lambda_j t)-\exp(-\lambda_i t)\sum\limits_{l=0}^{r}(\lambda_i-\lambda_j)^l t^l/l!\right\} & (\lambda_i \neq \lambda_j). \end{cases}$$

The proof is by integration by parts.

Proof of the theorem. From equations (3.11) and (2.3)

$$M_0(t) = T_{\mathscr{A}\mathscr{A}}(t) = [\exp(Qt)]_{\mathscr{A}\mathscr{A}} = \sum_{i=1}^{k} A_{i\mathscr{A}\mathscr{A}}\exp(-\lambda_i t). \tag{3.13}$$

Thus the theorem is true for $m = 0$, since

$$B_{i0}(t) = C_{i00} = A_{i\mathscr{A}\mathscr{A}} \tag{3.14}$$

is a polynomial of degree zero, i.e. a constant matrix. Also, from (3.9) and (2.3)

$$H_{\mathscr{A}\mathscr{A}}(t) = \sum_{j=1}^{k} D_j \exp(-\lambda_j t), \tag{3.15}$$

where

$$D_j = A_{i\mathscr{A}\mathscr{F}}\exp(Q_{\mathscr{F}\mathscr{F}}\tau)\,Q_{\mathscr{F}\mathscr{A}}. \tag{3.16}$$

Now $M_m(t) = (H_{\mathscr{A}\mathscr{A}}^{\oplus m} \oplus T_{\mathscr{A}\mathscr{A}})(t) = (H_{\mathscr{A}\mathscr{A}} \oplus M_{m-1})(t)$. Therefore, if we assume the theorem is true for $m-1$, we have

$$\begin{aligned} M_m(t) &= \int_0^t \left\{\sum_{j=1}^{k} D_j \exp(-\lambda_j(t-u))\right\}\left\{\sum_{i=1}^{k}\left(\sum_{r=0}^{m-1} C_{i(m-1)r}u^r\right)\exp(-\lambda_i u)\right\}\mathrm{d}u \\ &= \sum_{j=1}^{k}\sum_{i=1}^{k}\sum_{r=0}^{m-1} D_j C_{i(m-1)r}\int_0^t u^r \exp[-\lambda_i u-\lambda_j(t-u)]\,\mathrm{d}u, \end{aligned}$$

and thus, from the lemma,

$$\begin{aligned} M_m(t) &= \sum_{i=1}^{k}\sum_{r=0}^{m-1} D_i C_{i(m-1)r}t^{r+1}\exp(-\lambda_i t)/(r+1) \\ &\quad + \sum\sum_{i\neq j}\sum_{r=0}^{m-1} D_j C_{i(m-1)r}\frac{r!}{(\lambda_i-\lambda_j)^{r+1}}\left\{\exp(-\lambda_j t)-\exp(-\lambda_i t)\sum_{l=0}^{r}(\lambda_i-\lambda_j)^l t^l/l!\right\}. \end{aligned} \tag{3.17}$$

Therefore $M_m(t)$ has the required form with

$$\left.\begin{aligned} C_{imm} &= D_i C_{i(m-1)(m-1)}/m, \\ C_{iml} &= D_i C_{i(m-1)(l-1)}/l - \sum_{j\neq i}\sum_{r=l}^{m-1} D_j C_{i(m-1)r}\,r!/[l!(\lambda_i-\lambda_j)^{r-l+1}],\ l = 1,\ldots,m-1, \\ C_{im0} &= \sum_{j\neq i}\sum_{r=0}^{m-1}\{D_i C_{j(m-1)r}\,r!/(\lambda_j-\lambda_i)^{r+1} - D_j C_{i(m-1)r}\,r!/(\lambda_i-\lambda_j)^{r+1}\}. \end{aligned}\right\} \tag{3.18}$$

The theorem is true for $m = 0$, and so it must be true, by induction, for all $m \geqslant 0$. Moreover, (3.18) gives a recursive algorithm for computing the necessary coefficients.

Downloaded from rsta.royalsocietypublishing.org

The density of e*-open times*

By definition, any e-open time, T say, must exceed τ in duration and so it is convenient to consider the excess time $U = T-\tau$.

Then the density ${}^{\mathrm{e}}f_T(t) = f_U(t-\tau)$ and so the results of the theorem, together with equations (3.12), (3.2) and (2.14), imply that the probability density of U is

$$f_U(t) = \sum_{m=0}^{\infty} (-1)^m f_m(t-m\tau), \tag{3.19}$$

where
$$f_m(t) = \phi_{\mathscr{A}} M_m(t)\, Q_{\mathscr{A}\mathscr{F}} \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, u_{\mathscr{F}} = \sum_{i=1}^{k} \beta_{im}(t) \exp(-\lambda_i t) \tag{3.20}$$

and $\beta_{im}(t)$ is a polynomial of degree m in t with real constant coefficients. Then

$$\beta_{im}(t) = \sum_{r=0}^{m} \gamma_{imr} t^r, \tag{3.21}$$

where
$$\gamma_{imr} = \phi_{\mathscr{A}} C_{imr} Q_{\mathscr{A}\mathscr{F}} \exp(Q_{\mathscr{F}\mathscr{F}}\tau)\, u_{\mathscr{F}}. \tag{3.22}$$

It follows that, for t in the interval I_{n+1},

$$f_U(t) = \sum_{m=0}^{n} (-1)^m f_m(t-m\tau) = \sum_{i=1}^{k} \theta_{in}(t) \exp(-\lambda_i t), \tag{3.23}$$

where
$$\theta_{in}(t) = \sum_{m=0}^{n} (-1)^m \sum_{r=0}^{m} \gamma_{imr}(t-m\tau)^r \tag{3.24}$$

is a polynomial of degree n in t. Note that the λ_i that appear here are the k eigenvalues of the whole matrix $-Q$, rather than the $k_{\mathscr{A}}$ eigenvalues of $-Q_{\mathscr{A}\mathscr{A}}$ which appear in the solution when no intervals are missed.

These results, containing exponentials and piecewise polynomials over intervals of length τ, are similar to results of Garwood (1940) on the distribution of waiting time at a vehicle actuated traffic light: an arriving vehicle waits for a gap of at least τ in a Poisson stream of traffic before crossing. Garwood's results, obtained by combinatorial arguments, are easily derived by methods similar to those used here (see Jalali & Hawkes 1990). Equation (4.2) below is a generalization of Garwood's results and a special case of equation (10) of Hawkes (1965). Results for more general semi-Markov processes, including multilevel processes, have also been obtained (Jalali, personal communication 1989).

4. The two-state model

In this section we discuss the simplest special case of the model, comprising just one open state and one closed state, and compare it with a number of approximations which have appeared in the literature. The Q-matrix can be written

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$

Downloaded from rsta.royalsocietypublishing.org

so $-Q$ has eigenvalues $\lambda_1 = 0, \lambda_2 = \alpha+\beta$. In this case $\phi_{\mathscr{A}}$ and $u_{\mathscr{F}}$ are simply unit scalars, so that (2.21), (3.3) and (3.4) imply

$$ {}^{e}f^{*}_{\mathscr{A}}(s) = {}^{e}G^{*}_{\mathscr{A}\mathscr{F}}(s) = \{1+\mathrm{e}^{-s\tau}\alpha\beta\,\mathrm{e}^{-\beta\tau}/[s(s+\alpha+\beta)]\}^{-1}(s+\beta)\,\alpha\,\mathrm{e}^{-(s+\beta)\tau}/[s(s+\alpha+\beta)]. \tag{4.1}$$

If, as before, we consider only the excess of an e-open time over τ we have

$$ f^{*}_{U}(s) = \{s(s+\alpha+\beta)+\mathrm{e}^{-s\tau}\alpha\beta\,\mathrm{e}^{-\beta\tau}\}^{-1}(s+\beta)\,\alpha\,\mathrm{e}^{-\beta\tau}. \tag{4.2}$$

It is possible to use the recursive solution given in §3. However, in this case we can obtain a more explicit result by expanding equation (4.2) as

$$ f^{*}_{U}(s) = \frac{1}{\beta}\sum_{m=0}^{\infty}(-1)^m\,\mathrm{e}^{-sm\tau}\rho^{m+1}\{1/[s^m(s+\lambda_2)^{m+1}]+\beta/[s^{m+1}(s+\lambda_2)^{m+1}]\}, \tag{4.3}$$

where

$$ \rho = \alpha\beta\,\mathrm{e}^{-\beta\tau}. \tag{4.4}$$

Using partial fractions, one can show that, for positive integers p, q

$$ 1/[s^p(s+\lambda_2)^q] = \sum_{r=0}^{p-1} a_{qr}/s^{p-r} + \sum_{r=0}^{q-1} b_{pr}/(s+\lambda_2)^{q-r}, \tag{4.5}$$

where for $r = 0, 1, \ldots$

$$ a_{qr} = (-1)^r\binom{q+r-1}{r}\Big/\lambda_2^{q+r}, \quad b_{pr} = (-1)^p\binom{p+r-1}{r}\Big/\lambda_2^{p+r}. \tag{4.6}$$

Equation (4.5) is the Laplace transform of

$$ \sum_{r=0}^{p-1} a_{qr}\,t^{p-r-1}/(p-r-1)! + \sum_{r=0}^{q-1} b_{pr}\,t^{q-r-1}\exp(-\lambda_2 t)/(q-r-1)!. \tag{4.7}$$

It follows that

$$ f_U(t) = \sum_{m=0}^{\infty}(-1)^m f_m(t-m\tau), \tag{4.8}$$

where

$$ f_m(t) = \beta_{1m}(t)+\beta_{2m}(t)\exp(-\lambda_2 t). \tag{4.9}$$

Here $\beta_{1m}(t)$ and $\beta_{2m}(t)$ are polynomials of degree m in t, as indicated in equation (3.21) with, for $m > 0$,

$$ \gamma_{1mr} = \begin{cases} \rho^{m+1}\{a_{(m+1)(m-r-1)}+\beta a_{(m+1)(m-r)}\}/r!\beta & (r = 0, 1, \ldots, m-1), \\ \rho^{m+1}a_{(m+1)0}/m! & (r = m), \end{cases} \tag{4.10}$$

$$ \gamma_{2mr} = \rho^{m+1}\{b_{m(m-r)}+\beta b_{(m+1)(m-r)}\}/r!\beta \quad (r = 0, 1, \ldots, m). \tag{4.11}$$

We treat $m = 0$ as a special case, because (4.6) does not hold when $p = 0$. It is easy to see, by letting $m = 0$ in (4.3), that

$$ f_0(t) = \rho(\beta+\alpha\exp(-\lambda_2 t))/(\lambda_2\beta), \tag{4.12}$$

so that over $(0, \tau)$ $f_U(t) = f_0(t)$ is a constant plus an exponential.

Downloaded from rsta.royalsocietypublishing.org

Approximate densities

It can be seen that even the simple two-state case is not all that simple. Previously only approximations, or a numerically inverted transform, have been available. It would be useful to see how good these approximations are. Yeo *et al.* (1989) considered a number of simple exponential approximations

$$f_U(t) \approx \mathrm{e}^{-t/\nu}/\nu \quad (t > 0) \tag{4.13}$$

of which the most satisfactory uses the correct mean

$$\nu = \mathrm{e}^{\beta\tau}/\alpha - (1-\mathrm{e}^{\beta\tau})/\beta - \tau. \tag{4.14}$$

Yeo *et al.* (1989) also obtain a bi-exponential density by using the second-order approximation $\mathrm{e}^{-s\tau} \approx 1 - s\tau + \frac{1}{2}s^2\tau^2$ in equation (4.2). Then

$$f^*_U(s) \approx (s+\beta)\,\rho/\beta[\rho + s(\lambda_2 - \rho\tau) + s^2(1+\tfrac{1}{2}\tau^2\rho)].$$

The inverse of this is

$$f_U(t) \approx (\theta/\nu_3)\exp(-t/\nu_3) + ((1-\theta)/\nu_4)\exp(-t/\nu_4), \tag{4.15}$$

where

$$\theta = (\nu_3 - 1/\beta)/(\nu_3 - \nu_4), \quad \nu_3, \nu_4 = \tfrac{1}{2}(\alpha_1 \mp (\alpha_1^2 - 4\alpha_2)^{\frac{1}{2}}), \quad \nu_3 > \nu_4,$$
$$\alpha_1 = \lambda_2/\rho - \tau, \quad \alpha_2 = 1/\rho + \tfrac{1}{2}\tau^2. \tag{4.16}$$

Crouzy & Sigworth (1990) also obtain a bi-exponential approximation using a method of 'virtual states'. This means introducing an extra state corresponding to missed closed times, which are therefore assumed to be exponentially distributed (although in fact the distribution is truncated) with rate constant β', chosen to make the mean missed closed time correct, so $1/\beta' = 1/\beta - \tau\,\mathrm{e}^{-\beta\tau}/(1-\mathrm{e}^{-\beta\tau})$. Also a fraction $(1-\mathrm{e}^{-\beta\tau})$ or shut times are missed, so consider a modified Q-matrix

$$Q' = \left[\begin{array}{cc:c} -\alpha & \alpha(1-\mathrm{e}^{-\beta\tau}) & \alpha\,\mathrm{e}^{-\beta\tau} \\ \beta' & -\beta' & 0 \\ \hdashline \beta & 0 & -\beta \end{array}\right]. \tag{4.17}$$

The modified set of open states $\mathscr{A}'$, comprising the open state and the 'missed' closed state, is indicated in the above matrix. A different, but similar, virtual state and modified matrix can be introduced when finding an approximate distribution for e-closed times. An observed open time comprises a sojourn in the open state, state 1, of duration τ, followed by a time U which is a sojourn in $\mathscr{A}'$ starting in state 1 and ending with a transition into the single closed state. With these assumptions, the excess observed open time, U, has probability density given from equation (2.11)

$$f_U(t) \approx (1,0)\exp(Q'_{\mathscr{A}'\mathscr{A}'}t)\,Q'_{\mathscr{A}'\mathscr{F}}. \tag{4.18}$$

After some algebra, this can be written as

$$f_U(t) \approx A\{[l_1 + l_2(1-c)^{\frac{1}{2}}]\exp(-\eta_1 t) + [-l_1 + l_2(1-c)^{\frac{1}{2}}]\exp(-\eta_2 t)\}, \tag{4.19}$$

where $A = \alpha\,\mathrm{e}^{-\beta\tau}/\{2l_2(1-c)^{\frac{1}{2}}\}$, $l_1 = \frac{1}{2}(\beta' - \alpha)$, $l_2 = \frac{1}{2}(\beta' + \alpha)$, $c = \alpha\beta'\,\mathrm{e}^{-\beta\tau}/l_2^2$, (4.20)

and η_1, η_2 are the eigenvalues of $-Q'_{\mathscr{A}'\mathscr{A}'}$

$$\eta_1 = l_2\{1-(1-c)^{\frac{1}{2}}\}, \quad \eta_2 = l_2\{1+(1-c)^{\frac{1}{2}}\}. \tag{4.21}$$

Downloaded from rsta.royalsocietypublishing.org

A third bi-exponential approximation can be obtained by assuming that the critical gaps are, for each interval, independent identically distributed observations from an exponential distribution with mean Δ. Then the generalization of equation (2.19) given by Ball & Sansom (1988*a*) gives

$$f_T^*(s) = \alpha/[\Delta s^2 + s(1+\Delta(\alpha+\beta))+\alpha]. \tag{4.22}$$

If we choose

$$\Delta = (e^{\beta\tau}-1)/\beta, \tag{4.23}$$

then both the probability of detecting a closed time and the mean e-open time are the same as for the case of a constant critical gap of duration τ. Note, however, that the bi-exponential obtained by inverting (4.22) is the distribution of the full e-open time T, not the excess time $U = T-\tau$; note also that one of the weights is negative, leading to a density which rises from close to zero to a maximum before decaying. Considered as an approximation to the constant case it is poor, see figure 4.

The approximation of Yeo *et al.* (1988) is not good at the origin because as $s \to \infty$ $\lim sf_U^*(s)$ for the approximation (4.15) is not the same as for the exact distribution, given by equation (4.2). This is because the approximation $e^{-s\tau} \approx 1-s\tau-\frac{1}{2}s^2\tau^2$ is poor for large s. If, instead, we substitute the approximation $e^{-s\tau} \approx (1-\frac{1}{2}s\tau)/(1+\frac{1}{2}s\tau)$ into equation (4.2), we get a rational approximation for $f_U^*(s)$ which corresponds to a mixture of three exponentials for $f_U(t)$. This has the correct mean and variance, as does the Yeo approximation; however, $f_U(t)$ and its derivative at the origin, $t = 0$, are also correct. In examples we have tried, we have found that one of the areas is negative, so it is not a proper mixture, while overall it is only marginally better than the Crouzy–Sigworth approximation for most t. We do not, therefore, pursue this line here. This is one of a series of approximations, related to Bessel polynomials (see Burchnall & Chaundy 1931; Grosswald 1978), leading to further exponential components. These have interesting properties outside the scope of this paper.

Asymptotic behaviour of exact density

The behaviour of $f_U(t)$ for large t is governed by the roots of the denominator of $f_U^*(s)$. It can be shown that this denominator, given from equation (4.2) by

$$D(s) = s(s+\alpha+\beta)+\alpha\beta\, e^{-(s+\beta)\tau}, \tag{4.24}$$

has two real negative roots. One of these is clearly $s = -\beta$, which cancels with the factor $(s+\beta)$ in the numerator, so the behaviour is governed by the other root $s = -\eta$, say. The value of η, which lies in the interval $(0, \alpha e^{-\beta\tau})$, must be found numerically, e.g. by a bisection method. The derivative of $D(s)$ is given by

$$D'(s) = 2s+\alpha+\beta-\tau\alpha\beta\, e^{-(s+\beta)\tau}. \tag{4.25}$$

We approximate $D(s)$ by $c(s+\beta)(s+\eta)$, where we choose c to equate the derivative of this approximation to that of $D(s)$ at $s = -\eta$. Then

$$c = D'(-\eta)/(\beta-\eta) \tag{4.26}$$

and so $f_U^*(s) \approx \alpha e^{-\beta\tau}/\{c(s+\eta)\}$, which implies that

$$f_U(t) \approx [\alpha e^{-\beta\tau}/c\eta]\,\eta\, e^{-\eta t}, \tag{4.27}$$

a single exponential density with an area which is not unity. This is not a problem because this is the asymptotic behaviour of the exact density: it is not an approximation for the whole distribution. In practice we have found that it gives an

Downloaded from rsta.royalsocietypublishing.org

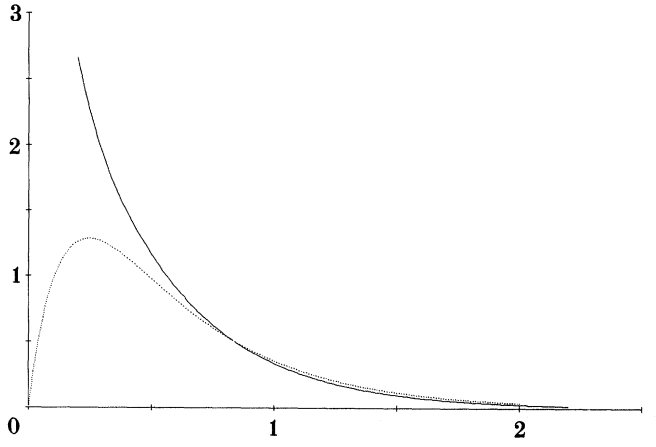

Figure 4. Exact (——) and approximate (....) densities for e-open times for the 'slow' model with $\tau = 0.2$ ms, $\mu_o = 0.2990$ ms, $\mu_c = 0.8787$ ms. The approximation assumes exponentially distributed deadtimes, see equations (4.22) and (4.23).

Table 1. *Comparison of mixed exponential approximations for densities of* e-*open and* e-*closed times in excess of* τ *for 'fast' two-state model with* $\tau = 0.2$ ms, $\mu_o = 0.1063$ ms, $\mu_c = 0.2148$ ms; *the asymptotic exponential expressions for the exact densities are also given*

		mean/ms	area	mean/ms	area	mean/ms	area
open density	C–S	0.4317	0.9161	0.0529	0.0839	—	—
	Yeo	0.4360	0.8599	0.1788	0.1401	—	—
	exp	0.4	1	—	—	—	—
	tri-exp	0.4075	1.0756	0.0568	0.1006	0.2504	−0.1761
	asymp.	0.4212	0.9464	—	—	—	—
closed density	C–S	1.8169	0.9911	0.0546	0.0089	—	—
	Yeo	1.8138	0.9926	0.0936	0.0074	—	—
	exp	1.8	1	—	—	—	—
	tri-exp	1.8131	0.9934	0.0629	0.0134	0.1314	−0.0069
	asymp.	1.8133	0.9931	—	—	—	—

extremely accurate approximation to the exact density for t larger than a few multiples of τ.

Numerical examples

We consider two examples from Colquhoun & Sigworth (1983); a 'slow' model with dead-time $\tau = 0.2$ ms and mean open and closed occupancies $\mu_o = 1/\alpha = 0.2990$ ms, $\mu_c = 1/\beta = 0.8787$ ms; a 'fast' model with the same τ but with $\mu_0 = 0.1063$ ms, $\mu_c = 0.2148$ ms. In the latter case the mean open time is less than the dead-time, while the mean closed time only just exceeds it. These two models both have mean observed open and closed times which are 0.6 ms and 2.0 ms respectively; see (4.28) below.

The density of e-open times for the slow model is shown in figure 4, with the bi-exponential approximation given by (4.22) and (4.23). The latter has two components with means 0.451 ms and 0.149 ms with corresponding areas 1.492 and −0.492. These sum to unity but it is not a proper mixture because one of them is negative. It is clearly a poor approximation to the exact density and will not be considered further.

Downloaded from rsta.royalsocietypublishing.org

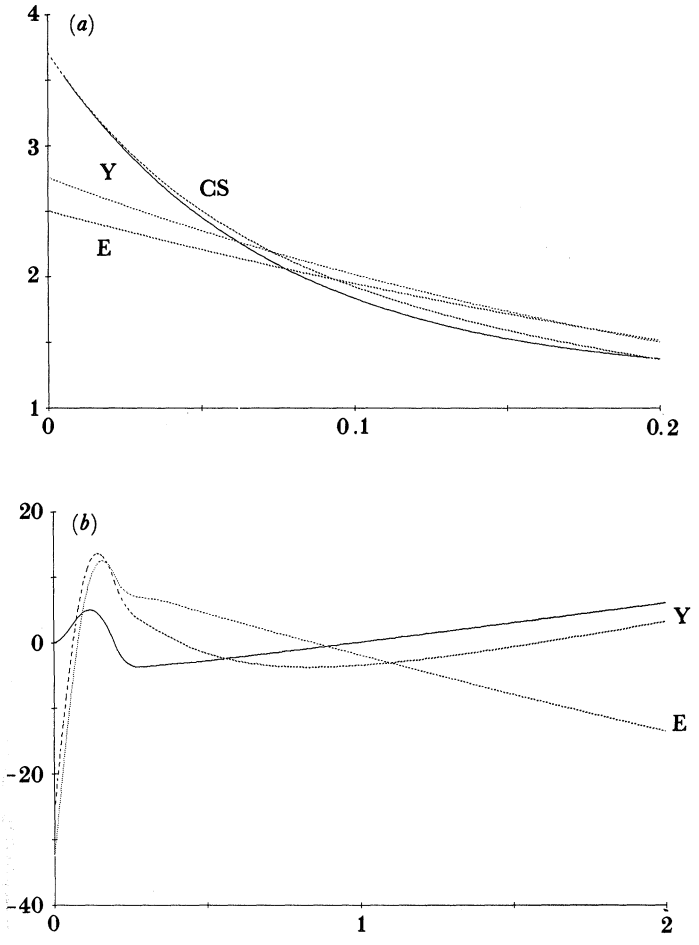

Figure 5. Exact and approximate densities of e-open times in excess of τ for the 'fast' model $\tau = 0.2$ ms, $\mu_o = 0.1063$ ms, $\mu_c = 0.2148$ ms. In (*a*) the densities are shown for an interval near the origin where they are discernably different. The exact density is shown as a solid curve (——). The Crouzy–Sigworth, Yeo *et al.* and exponential approximations are shown dotted (....), and identified by CS, Y and E, respectively. In (*b*) we show percentage errors of the three approximations relative to the exact density. Crouzy–Sigworth is shown as solid (——); Yeo and the exponential approximations are shown as dotted (....).

Turning now to the fast model, we consider the density of excess e-open and e-closed times, $U = T - \tau$, obtained by subtracting the dead-time τ. The parameters of the two bi-exponential approximations (equations (4.15) and (4.16) and (4.19)–(4.21)), the single exponential approximation, and the asymptotic exponential (equations (4.24)–(4.27)) are given in table 1. We also give the tri-exponential approximation (note the negative weight), but do not consider it further.

The densities for e-open times are shown in figure 5 and for e-closed times in figure 6. In each case they only differ visibly for a range of values near the origin. Therefore, in part (*a*) of each figure, we have shown only a small range of time values and a portion of the vertical scale, so as to magnify the differences; also we have shown percentage errors of each approximation relative to the exact curve over a larger interval in part (*b*) of each figure.

Downloaded from rsta.royalsocietypublishing.org

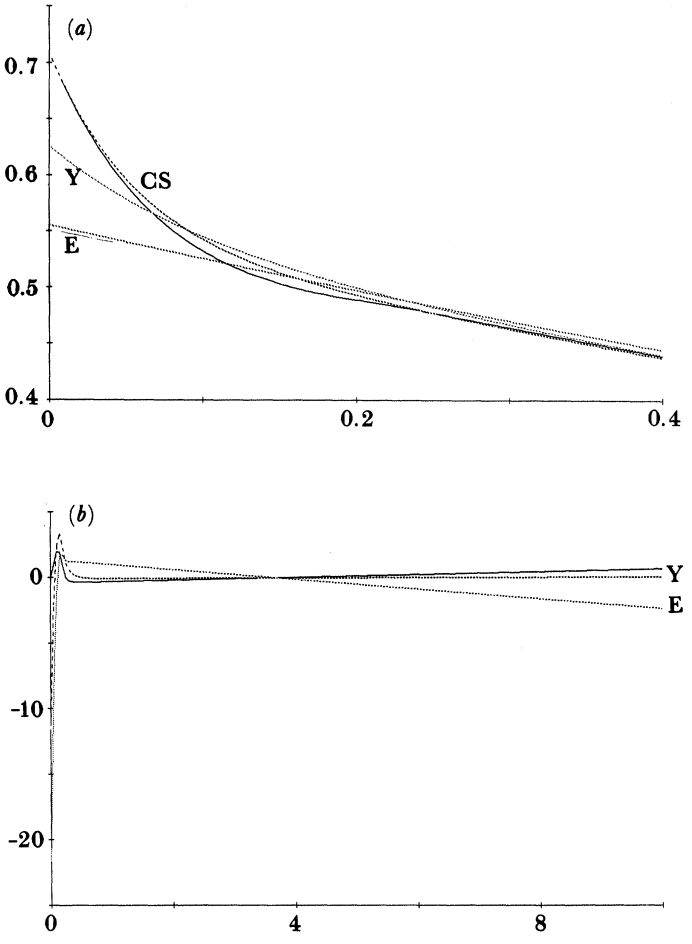

Figure 6. Exact and approximate densities of e-closed times in excess of τ for the 'fast' model $\tau = 0.2$ ms, $\mu_o = 0.1063$ ms, $\mu_c = 0.2148$ ms. In (*a*) the densities are shown for an interval near the origin where they are discernably different. The exact density is shown as a solid curve (——). The Crouzy–Sigworth, Yeo *et al.* and exponential approximations are shown dotted (....), and identified by CS, Y and E, respectively. In (*b*) we show percentage errors of the three approximations relative to the exact density. Crouzy–Sigworth is shown as solid (——); Yeo and the exponential approximations are shown as dotted (....).

It can be seen that the Crouzy–Sigworth approximation is very much better than the others near the origin, while for larger values of t there is little to choose between them. For both open and closed distributions the asymptotic exponential agreed with the exact density to within a 10^{-6} percentage error for all t in the interval $5\tau < t < 20\tau$: a very close agreement indeed!

The conclusions drawn from these, and other examples we have considered, are the following.

(i) If a mean open or closed occupancy is moderately larger than the dead-time τ, then the corresponding e-open or e-closed time distribution is adequately represented by a single exponential distribution with effective mean given by equation (4.14).

(ii) If a mean open or closed occupancy is of comparable duration to τ, then all three approximations are reasonable for large t but differ near the origin. In all cases

Downloaded from rsta.royalsocietypublishing.org

studied the Crouzy–Sigworth approximation has been quite good. The single exponential is too low near the origin and a little too high for medium values of t; the Yeo *et al.* approximation behaves similarly and is not all that much better.

(iii) The exact distribution can be easily calculated for t up to about 20 multiples of τ. Long before that it may be approximated very accurately by the single exponential asymptotic expression.

The examples we have illustrated here are relatively bad. For larger means the approximations differ much less, but show the same general features. If a mixed exponential approximation is needed, we recommend the single exponential, if that is adequate, or the Crouzy–Sigworth approximation. These have the added advantage that they can be generalized to more complex models, as illustrated in §5.

Inference

From these results we conclude that densities of observed open or closed times will often be adequately represented by a single exponential; then we would infer that there was probably no more than one open or closed state. If, however, a mean is not much bigger than τ, one should be cautious about inferring the existence of a second state from the presence of a second component in the observed distribution.

This is a useful qualitative conclusion: we now consider parameter estimation. Colquhoun & Sigworth (1983) used a method-of-moments approach, equating the theoretical mean e-open and e-closed times to their observed values $\bar{x}_o$ and $\bar{x}_c$. Setting $\mu_o = 1/\alpha, \mu_c = 1/\beta$, equation (4.14) leads to

$$\left.\begin{aligned} {}^{e}\mu_o &= (\mu_o+\mu_c)\exp(\tau/\mu_c)-\mu_c = \bar{x}_o,\\ {}^{e}\mu_c &= (\mu_o+\mu_c)\exp(\tau/\mu_o)-\mu_o = \bar{x}_c. \end{aligned}\right\} \tag{4.28}$$

They point out that there are often two solutions for (μ_o, μ_c), called fast and slow solutions. The examples considered in the previous subsection were of this form, being the two solutions when $\bar{x}_o = 0.6$ ms and $\bar{x}_c = 2.0$ ms with $\tau = 0.2$ ms. They arise from a kind of aliasing effect caused by the finite resolution of the recording equipment.

Yeo *et al.* (1988) discuss the estimation problem in detail. They point out that the moment estimators are maximum likelihood estimators if the single exponential approximation is used, the likelihood having two peaks of equal height. Using their bi-exponential approximation, they find that there is usually a unique maximum likelihood estimate but that there may be another local maximum of very nearly the same height as the main peak. They also discuss the possibility of discriminating between the two solutions by repeating the analysis with different dead-times.

Figure 7 shows the exact densities of e-open times for both fast and slow solutions. They do not differ much so, while in theory they are capable of being discriminated, one would need quite a lot of data to do so. Notice that the fast distribution, has a distinct change in form at $\tau = 0.2$ and, to a lesser extent, at $2\tau = 0.4$. It is now feasible to compute likelihood surfaces from the exact model. We expect them to show similar features to those based on the bi-exponential approximation, possibly with a slightly bigger difference between the heights of the two peaks.

As a simple illustration of the aliasing effect we simplify to a single parameter by considering a symmetric model with $\mu_o = \mu_c = \mu$, say, taking the timescale so that $\tau = 1$. Figure 8 shows the likelihood $l(\mu)$, calculated from the exact density, based on a single observation. This is done for each of four different values of an observed e-

Downloaded from rsta.royalsocietypublishing.org

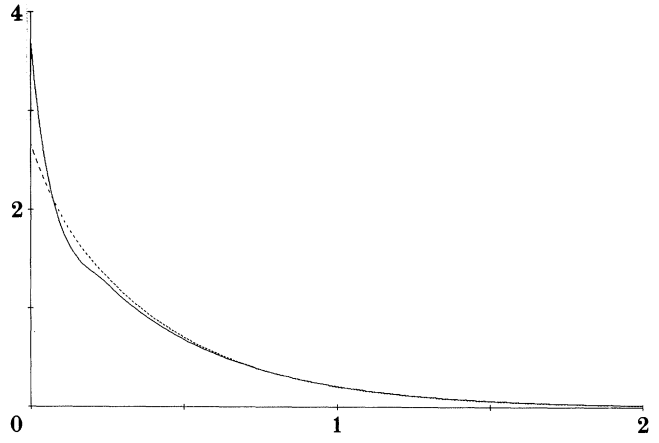

Figure 7. Exact densities of e-open times in excess of τ for both fast (——) and slow (....) models discussed in the text.

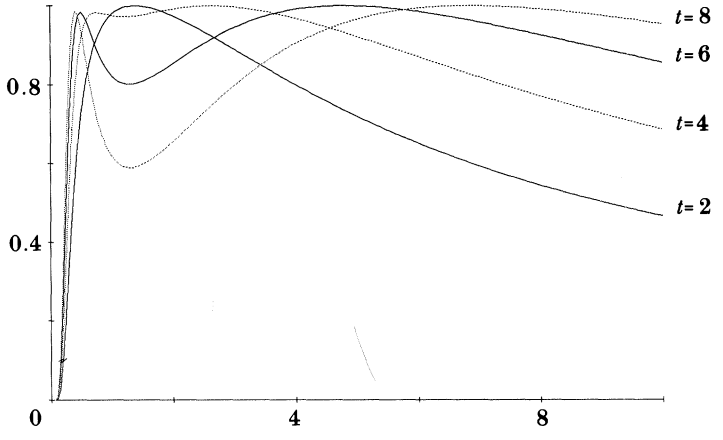

Figure 8. Relative likelihood curves for symmetric two-state model with mean occupancies both equal to μ and dead-time 1. Each likelihood is a function of μ arising from a single observation t on an observed e-interval; each is scaled to have maximum of 1. Likelihoods corresponding to $t = 2, 6$ are shown as (——) and those for $4, 8$ as (....).

open interval $t = 2, 4, 6, 8$. As t increases, the likelihood moves from a unimodal to a flat-topped function to a two-peaked function. The peaks are of about the same height, a sharp one for the fast solution and a flat one for the slow solution.

5. More complex models

There are two approximations to the density of e-open times in the general case, both leading to a mixture of exponentials. The B–M approximation, of Blatz & Magleby (1986), alters the elements of the Q-matrix to a set of 'effective rates'. The numbers of open and closed states are not altered, so the number of components in the mixture is the same as for the original model, without time interval omission, but with different parameters. The C–S approximation of Crouzy & Sigworth (1990) leads to a mixture with $k_{\mathscr{A}} + k_{\mathscr{F}}$ components instead of $k_{\mathscr{A}}$: however, some components

Downloaded from rsta.royalsocietypublishing.org

usually have areas too small to be identified in practice. In this section we describe these approximations and suggest improvements. We try them out on a numerical example.

The B–M approximation

The B–M method of approximating the density of e-open times assumes that any sojourn in $\mathscr{F}$ which is too short to be observed consists only of a sojourn in a single state. The relatively unlikely possibility of several moves among the $\mathscr{F}$ states in a short time is ignored. It takes into account the probability of such a sojourn being short, and its average duration, to obtain modified effective rates. The detailed argument is given by Blatz & Magleby (1986); we summarize the algorithm below.

For each $i \in \mathscr{A}$ and $j \in \mathscr{F}$ for which the transition rate $q_{ij} > 0$, calculate

$$\left.\begin{aligned} F_j^{\mathrm{m}} &= 1-\exp(q_{jj}\tau),\\ 1-F_j^{\mathrm{cap}} &= F_j^{\mathrm{miss}} = F_j^{\mathrm{m}}\{q_{ji}/(-q_{jj})\},\\ T_j^{\mathrm{miss}} &= -[1/q_{jj}+\tau(1-F_j^{\mathrm{m}})/F_j^{\mathrm{m}}],\\ F_{ij}^{\mathrm{r}} &= F_j^{\mathrm{cap}}/\{1-F_j^{\mathrm{miss}}T_j^{\mathrm{miss}}q_{ii}\}. \end{aligned}\right\} \tag{5.1}$$

The effective rates are then given by

$$^{\mathrm{e}}q_{ij} = F_{ij}^{\mathrm{r}}\,q_{ij}, \quad ^{\mathrm{e}}q_{ji} = F_{ij}^{\mathrm{r}}\,q_{ji}. \tag{5.2}$$

All rates within the open states or within the closed states are unchanged. After these modifications to all the relevant q_{ij}, the diagonal elements q_{ii} are adjusted so that the rows of the new matrix of effective rates, $^{\mathrm{e}}Q$ say, sum to zero. The density of open times is then found using the standard theory, applying equation (2.11) to the matrix $^{\mathrm{e}}Q$ instead of to Q. The distribution thus obtained allows e-open times to be less than τ. To adjust for this, Blatz & Magleby suggest truncating the distribution, dividing by Prob$(T > \tau)$ for $t > \tau$. As this is a mixture of exponentials,

$$f_T(t) = \Sigma\, a_i \exp(-t/v_i)/v_i, \tag{5.3}$$

the modified distribution of excess e-open times, $U = T-\tau$, is given from Colquhoun & Sigworth (1983), equation (60), as

$$f_U(t) = \Sigma\, a_i' \exp(-t/v_i)/v_i, \quad a_i' = a_i \exp(-\tau/v_i)/\{\Sigma\, a_j \exp(-\tau/v_j)\}. \tag{5.4}$$

The single exponential approximation for the two-state case discussed in §4 is a special case of the B–M approximation.

The approximate distribution of e-closed times can be found in a similar way, using a separate effective Q-matrix obtained by interchanging the roles of $\mathscr{A}$ and $\mathscr{F}$ in the above algorithm.

The C–S approximation

For their approximation to the density of e-open times, Crouzy & Sigworth (1990) first, if necessary, transform the Q-matrix to a dynamically similar one by the method of Kienker (1989). Thus Q is replaced by the similar matrix

$$Q' = S^{-1}QS, \tag{5.5}$$

where S is a block diagonal matrix, whose rows sum to unity, of the form

$$S = \begin{bmatrix} I_{\mathscr{A}\mathscr{A}} & 0 \\ 0 & S_{\mathscr{F}\mathscr{F}} \end{bmatrix}. \tag{5.6}$$

Downloaded from rsta.royalsocietypublishing.org

Optionally, one may replace the identity matrix $I_{\mathscr{A}\mathscr{A}}$ by a more general matrix. If $S_{\mathscr{F}\mathscr{F}}$ is chosen to diagonalize $Q_{\mathscr{F}\mathscr{F}}$, the closed states are said to be uncoupled and transitions between these modified closed states are not possible. This may be done by taking the columns of $S_{\mathscr{F}\mathscr{F}}$ to be eigenvectors of $Q_{\mathscr{F}\mathscr{F}}$ scaled so that the rows sum to unity. If the closed states are already uncoupled, this step may be omitted.

Next introduce some 'virtual' states; these are shut states which are missed because sojourns in them are too short, and so they count as virtual open states when finding the density of e-open times. For each $j \in \mathscr{F}$ create a virtual open state j', so there will be $k_{\mathscr{F}}$ virtual states $\mathscr{V}$. Calculate F_j^{m} and T_j^{miss} from (5.1) then, denoting the transition rates of the virtual scheme by ${}^{v}q_{ij}$, for each $i \in \mathscr{A}$ we calculate

$$\left.\begin{aligned} {}^{v}q_{ij} &= q_{ij}(1-F_j^{\mathrm{m}}), \quad {}^{v}q_{ij'} = q_{ij}F_j^{\mathrm{m}}, \\ \beta_j &= 1/\{T_j^{\mathrm{miss}} \sum_{i\in\mathscr{A}} q_{ji}\}, \quad {}^{v}q_{j'i} = \beta_j q_{ji}. \end{aligned}\right\} \tag{5.7}$$

This deals with all the transition rates in the partitions ${}^{v}Q_{\mathscr{A}\mathscr{F}}$, ${}^{v}Q_{\mathscr{A}\mathscr{V}}$ and ${}^{v}Q_{\mathscr{V}\mathscr{A}}$. The partitions $Q_{\mathscr{A}\mathscr{A}}$, $Q_{\mathscr{F}\mathscr{F}}$ and $Q_{\mathscr{F}\mathscr{A}}$ are left unaltered. ${}^{v}Q_{\mathscr{F}\mathscr{V}}$ and ${}^{v}Q_{\mathscr{V}\mathscr{F}}$ are set to zero as there is no direct transition between the shut states and the virtual states. Finally, ${}^{v}Q_{\mathscr{V}\mathscr{V}}$ is a diagonal matrix whose diagonal elements are chosen so that the rows of the virtual matrix ${}^{v}Q$ sum to zero as usual.

The approximate density of e-open times is obtained from equation (2.11), using the matrix ${}^{v}Q$ instead of Q and the set of virtual open states $\mathscr{A}' = \mathscr{A} \cup \mathscr{V}$ instead of $\mathscr{A}$. As there are $k_{\mathscr{A}} + k_{\mathscr{F}}$ of these, that will be the number of exponential components in the mixture. The distribution of e-closed times can be obtained similarly, creating a separate virtual process by the above method with the roles of $\mathscr{A}$ and $\mathscr{F}$ interchanged.

Crouzy & Sigworth do not discuss the problem of what to do about the fact that an e-open time must exceed τ. There seem to be two possibilities.

C–S1. As the density is a mixture of exponentials, one could apply the adjustment of equation (5.4).

C–S2. Use equation (2.11*b*), applied to the original Q-matrix, to obtain the initial vector, ϕ_{o}, for entry to all open times, visible or not. Our observed open times must begin with a period of duration τ in the set of open states, $\mathscr{A}$, so that the initial probability vector of the state in which the excess e-open time begins is given by

$${}^{\mathrm{e}}\phi_{\mathrm{o}} = \phi_{\mathrm{o}} \exp(Q_{\mathscr{A}\mathscr{A}}\tau)/\phi_{\mathrm{o}} \exp(Q_{\mathscr{A}\mathscr{A}}\tau)\, u_{\mathscr{A}}. \tag{5.8}$$

Add to this a set of zeros corresponding to the virtual states $\mathscr{V}$, which together with $\mathscr{A}$ makes up the virtual open set $\mathscr{A}'$, to obtain an initial vector ϕ_{o}'. Now apply equation (2.11*a*) using this initial vector, the set $\mathscr{A}'$ and the matrix ${}^{v}Q$. The result is an approximation to the density of the excess of e-open times over τ. The adjustment (5.8) applied to ${}^{v}Q$ instead of Q, and $\mathscr{A}'$ instead of $\mathscr{A}$, is equivalent to method C–S1.

If there is only one open (closed) state the adjustment C–S2 gives exactly the same e-open (e-closed) density as the unadjusted Crouzy–Sigworth method. In particular, this is true for the two-state case, for which it proved successful in §4. C–S1, in contrast, performs poorly. We therefore recommend the modified method C–S2.

Modified B–M method

The B–M method ignores short sojourns *in* $\mathscr{F}$ which involve occupying more than one state. We suggest that the closed states be uncoupled, if necessary, as described

Downloaded from rsta.royalsocietypublishing.org

Table 2. *Comparison of Crouzy–Sigworth and Blatz–Magleby mixed exponential approximations for densities of* e-*open and* e-*closed times in excess of* τ *for the scheme whose transition rates are given by the matrix in equation* (5.11), *and with critical deadtime* $\tau = 0.2$ ms

		mean/ms	area	mean/ms	area	mean/ms	area
distribution of e-open times							
C–S	unadjusted	0.9687	0.9183	0.0283	0.0011	0.0647	0.0175
	C–S1	0.9687	0.9990	0.0283	1.20E-6	0.0647	9.99E-4
	C–S2	0.9687	0.9813	0.0283	0.0011	0.0647	0.0175
B–M	standard	0.4374	1	—	—	—	—
	uncoupled	0.9666	1	—	—	—	—
distribution of e-closed times							
C–S	unadjusted	0.5879	0.9339	0.0279	0.0532	0.0692	0.0129
	C–S1	0.5879	0.9989	0.0279	6.19E-5	0.0692	0.0011
	C–S2	0.5879	0.9698	0.0279	0.0035	0.0692	0.0267
B–M	unadjusted	0.8561	0.9789	0.0327	0.0211	—	—
	adjusted	0.8561	0.9999	0.0327	0.0001	—	—

above, to obtain an equivalent system in which any sojourn in the modified $\mathscr{F}$ can only involve one state. The B–M method may then be applied to the modified system.

A numerical example

Consider the simple model with three states discussed in Colquhoun & Hawkes (1977, equations 83–86). We call this the KM model, after Katz & Miledi (1972).

$$\left.\begin{array}{c} \mathrm{A}+\mathrm{T}\underset{k_2}{\overset{k_1}{\rightleftharpoons}}\mathrm{AT}\underset{\alpha}{\overset{\beta}{\rightleftharpoons}}\mathrm{AR} \\ \text{state}: \mathscr{F}(3) \quad \mathscr{F}(2) \quad \mathscr{A}(1) \end{array}\right\} \tag{5.9}$$

The transition-rate matrix, Q, is shown partitioned according to the open state, $\mathscr{A} = (1)$, and the closed states, $\mathscr{F} = (2.3)$.

$$Q = \left[\begin{array}{c:cc} -\alpha & \alpha & 0 \\ \hdashline \beta & -(\beta+k_2) & k_2 \\ 0 & k_2 c & -k_2 c \end{array}\right]. \tag{5.10}$$

We have taken c in this matrix to be a standardized drug concentration $c = xk_1/k_2$, where x is the concentration of agonist A. As a particular case we take $\alpha = 5$, $\beta = k_2 = 10$, $c = 2$, making the transition-rate matrix

$$Q = \left[\begin{array}{c:cc} -5 & 5 & 0 \\ \hdashline 10 & -20 & 10 \\ 0 & 20 & -20 \end{array}\right]. \tag{5.11}$$

If these are rates per millisecond, the mean occupancies are 0.2 ms for the open state and 0.05 ms for each shut state. If $\tau = 0.2$ ms, we are testing the approximations severely as the means are short compared with the dead-time, particularly for the shut states.

Downloaded from rsta.royalsocietypublishing.org

The distribution of e*-open times*

For the C–S method we first uncouple the shut states to get a dynamically equivalent scheme with Q-matrix

$$Q = \left[\begin{array}{c|cc} -5 & 4.268 & 0.732 \\ \hline 5.858 & -5.858 & 0 \\ 34.142 & 0 & -34.142 \end{array}\right]. \tag{5.12}$$

We then form the virtual scheme with two extra states shown in (5.13) and with virtual matrix ${}^{v}Q$ shown in (5.14).

$$\begin{array}{ccccc} & & \mathscr{F}(3) & & \\ & & \Big\uparrow\Big\downarrow & & \\ \mathscr{F}(2) & \rightleftharpoons & \mathscr{A}(1) & \rightleftharpoons & \mathscr{V}(5) \\ & & \Big\uparrow\Big\downarrow & & \\ & & \mathscr{V}(4) & & \end{array} \tag{5.13}$$

$$ {}^{v}Q = \left[\begin{array}{c|cc|cc} -5 & 1.322 & 7.9\times10^{-4} & 2.945 & 0.731 \\ \hline 5.858 & -5.858 & 0 & 0 & 0 \\ 34.142 & 0 & -34.142 & 0 & 0 \\ \hline 12.360 & 0 & 0 & -12.360 & 0 \\ 34.397 & 0 & 0 & 0 & -34.397 \end{array}\right] \begin{array}{l} \mathscr{A} \\ \\ \mathscr{F} \\ \\ \mathscr{V} \end{array} \tag{5.14}$$

$\mathscr{A}$ $\mathscr{F}$ $\mathscr{V}$

The set $\mathscr{A}' = \mathscr{A} \cup \mathscr{V}$ has three states, so there are three components in the approximate distribution derived from ${}^{v}Q$ in the manner described earlier.

The matrix of effective rates for the B–M method is

$${}^{e}Q = \left[\begin{array}{c|cc} -2.286 & 2.286 & 0 \\ \hline 4.572 & -14.572 & 10 \\ 0 & 20 & -20 \end{array}\right] \tag{5.15}$$

and the effective matrix obtained from the matrix (5.12), if the shut states are first uncoupled, is

$${}^{e}Q = \left[\begin{array}{c|cc} -1.035 & 1.034 & 6.9\times10^{-4} \\ \hline 1.419 & -1.419 & 0 \\ 0.032 & 0 & -0.032 \end{array}\right]. \tag{5.16}$$

As there is only one open state, this approximation is a single exponential distribution. The distribution for excess e-open times, obtained from equation (5.4) will therefore be the same as the unadjusted distribution.

Table 2 shows the components of the mixtures obtained from the various B–M and C–S approximations. In this case we chose transitions between the shut states to be

Downloaded from rsta.royalsocietypublishing.org

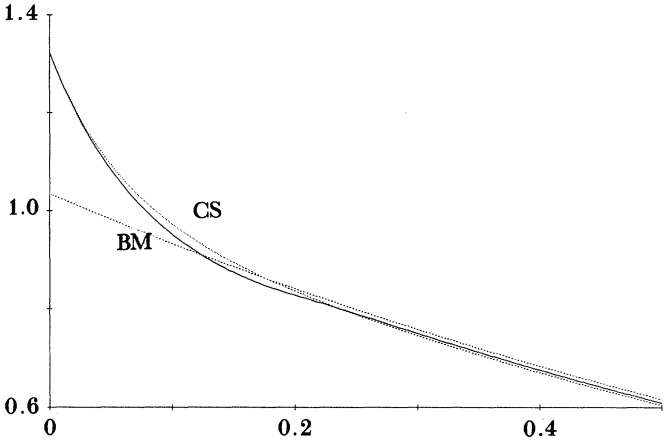

Figure 9. Behaviour near the origin of exact (——) and approximate (....) densities of excess of e-open times above $\tau = 0.2$ ms for the KM model with transition rates given in equation (5.11). The Crouzy–Sigworth approximation, C–S2, is marked as CS; the Blatz–Magleby approximation, using the uncoupled scheme (5.12) is marked as BM.

fast. Consequently, we see that uncoupling the closed state has made a big difference to the results of the B–M method, bringing it much closer to those for the C–S method. We find that C–S2 is better than C–S1. This can be illustrated, for example, by comparing the densities of e-open time in excess of τ at the origin, $t = 0$. These are respectively 1.047, 1.323 and 1.323 for C–S1, C–S2 and the exact distribution. The initial parts of the exact, uncoupled B–M and C–S2 densities are shown in figure 9. Both approximations do quite well, especially C–S2. They remain good up to $t = 4$.

The distribution of e*-closed times*

The virtual scheme for the C–S approximation to the density of e-closed times is

$$\begin{array}{ccccc} \mathscr{F}(3) & \rightleftharpoons & \mathscr{F}(2) & \rightleftharpoons & \mathscr{A}(1)\ . \\ & & \| & & \\ & & \mathscr{V}(4) & & \end{array} \tag{5.17}$$

This time we are interested in the set of virtual shut states $\mathscr{F}' = \mathscr{F} \cup \mathscr{V}$, which consists of three states. The matrix of virtual rates is

$${}^{v}Q = \begin{array}{c} \\ \mathscr{A} \\ \mathscr{F} \\ \mathscr{V} \end{array} \begin{array}{c} \begin{array}{cccc} \mathscr{A} & \mathscr{F} & & \mathscr{V} \end{array} \\ \begin{bmatrix} -5 & 5 & 0 & 0 \\ 3.679 & -20 & 10 & 6.321 \\ 0 & 20 & -20 & 0 \\ 0 & 11.961 & 0 & -11.961 \end{bmatrix} \end{array} \tag{5.18}$$

and the matrix of effective rates for the B–M method is

$${}^{e}Q = \begin{bmatrix} -0.894 & 0.894 & 0 \\ 1.788 & -11.788 & 10 \\ 0 & 20 & -20 \end{bmatrix}. \tag{5.19}$$

Downloaded from rsta.royalsocietypublishing.org

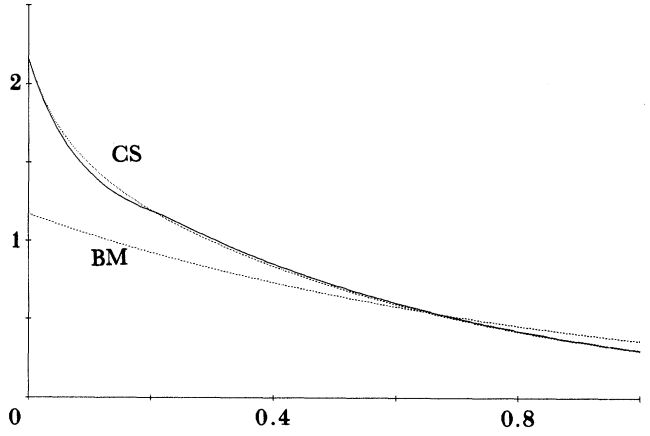

Figure 10. Behaviour near the origin of exact (——) and approximate (....) densities of excess of e-closed times above $\tau = 0.2$ ms for the KM model with transition rates given in equation (5.11). The Crouzy–Sigworth approximation, C–S2, is marked as CS; the Blatz–Magleby approximation is marked as BM.

The exponential components derived from these are again shown in table 2. The C–S2 approximation is good and better than C–S1 : for example the values at the origin are respectively 1.717, 2.161 and 2.161 for the C–S1, C–S2 and exact densities. The initial parts of the densities for the exact, C–S2 and B–M, adjusted by equation (5.4), methods are shown in figure 10. As there is only one open state, there is no need for uncoupling. Usually the B–M approximation is good away from the origin. In this case it is fairly rough, as can be inferred from the fact that the dominant time constants differ somewhat between the C–S and B–M solutions, also the curve in figure 10 is beginning to pull away from the exact curve from about $t = 0.7$ onwards.

6. Discussion

For examples presented here, and others, we found the modified Crouzy–Sigworth approximation generally to be quite good. The Blatz–Magleby approximation is usually good if the processes are slow relative to the dead-time, but may not be if they are fast. In that case one should first uncouple the complementary set of states ($\mathscr{F}$ for the open time distribution, $\mathscr{A}$ for the closed time distribution) before using the method. In very fast situations this may not be enough.

For moderate speed processes, the B–M approximation may not be adequate near the origin. The C–S approximation, with more components, fits well. Thus inspection of an observed histogram may indicate the presence of extra components and so give the impression of the existence of more states than there really are. The effect of time interval omission on inference about the number of pathways between $\mathscr{A}$ and $\mathscr{F}$ based on autocorrelation is discussed by Ball & Sansom (1988*b*), and based on cross-correlation by Ball *et al.* (1988).

We have found that computation of the exact density is generally quite feasible, both in time and accuracy, for up to 20 times the dead-time. For larger values of t the series becomes complicated and numerically unstable. For many purposes the C–S2 approximation will be adequate for all t. If greater accuracy is required near the origin, our current recommendation is to calculate the exact distribution for the first

Downloaded from rsta.royalsocietypublishing.org

few multiples of τ then switch to the C–S2 approximation for larger t; for the two-state case the asymptotic approximation should be used for large t. In the future, we hope to get better results for large t by finding an asymptotic expression in the general case; we know that asymptotically it has the form of a mixture of exponentials and damped oscillations, in which one exponential always dominates, but full computational details have still to be worked out.

In particular, it should be perfectly feasible to study the likelihood surface of the overall distribution of e-open times, or e-closed times, using this method. For that purpose the Crouzy–Sigworth approximation is almost certainly good enough for most data-sets available in practice. However, a more exciting possibility is that it may be computationally feasible to compute the likelihood for the complete observed process, consisting of an alternating sequence of e-open and e-closed times $\{t_o, t_c\} = (t_{o1}, t_{c1}, \ldots, t_{om}, t_{cm})$, where t_{oj} is the jth e-open time and t_{cj} is the jth e-closed time. From the underlying semi-Markov process, following the approach of Fredkin *et al.* (1985), we can write the likelihood as

$$l(\{t_o, t_c\} \mid \theta) = \phi_{\mathscr{A}} \prod_{j=1}^{m} \{{}^{e}G_{\mathscr{A}\mathscr{F}}(t_{oj})\, {}^{e}G_{\mathscr{F}\mathscr{A}}(t_{cj})\}\, u_{\mathscr{A}}, \tag{6.1}$$

where ${}^{e}G_{\mathscr{A}\mathscr{F}}(t)$ is given by equation (3.2), and with a similar expression for ${}^{e}G_{\mathscr{F}\mathscr{A}}(t)$. This involves ${}^{\mathscr{A}}R(t-\tau)$, and a similar ${}^{\mathscr{F}}R(t-\tau)$, which can be computed from equations (3.12) to (3.18). The matrices C_{iml} involved in this need only be computed once for a given set of parameters, θ.

Equation (6.1) is a generalization of eq. (18) of Ball & Sansom (1989), who successfully applied this approach to the case of no time interval omission, $\tau = 0$. Although the computational problems in this case are more severe, we believe that, with adequate computing resources and careful attention to numerical analysis, it is quite feasible to attempt such a study.

References

Ball, F. G. & Sansom, M. S. P. 1987 Temporal clustering of ion channel openings incorporating time interval omission. *IMA J. Math. appl. Med. Biol.* **4**, 333–361.

Ball, F. G. & Sansom, M. S. P. 1988*a* Aggregated Markov processes incorporating time interval omission. *Adv. appl. Prob.* **20**, 546–572.

Ball, F. G. & Sansom, M. S. P. 1988*b* Single channel autocorrelation functions – the effects of time interval omission. *Biophys. J.* **53**, 819–832.

Ball, F. G., Kerry, C. J., Ramsey, R. L., Sansom, M. S. P. & Usherwood, P. N. R. 1988 The use of dwell time cross-correlation functions to study single ion channel gating kinetics. *Biophys. J.* **54**, 309–320.

Ball, F. G. & Sansom, M. S. P. 1989 Ion channel gating mechanisms: model identification and parameter estimation from single channel recordings. *Proc. R. Soc. Lond.* B **236**, 385–416.

Blatz, A. L. & Magleby, K. L. 1986 Correcting single channel data for missed events. *Biophys. J.* **49**, 967–980.

Burchnall, J. & Chaundy, T. W. 1931 Commutative ordinary differential operators. II. The identity $P^n = Q^m$. *Proc. R. Soc. Lond.* A **134**, 471–485.

Colquhoun, D. & Hawkes, A. G. 1977 Relaxation and fluctuations of membrane currents that flow through drug-operated channels. *Proc. R. Soc. Lond.* B **199**, 205–235.

Colquhoun, D. & Hawkes, A. G. 1981 On the stochastic properties of single ion channels. *Proc. R. Soc. Lond.* B **211**, 205–235.

Colquhoun, D. & Hawkes, A. G. 1982 On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond.* B **300**, 1–59.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Downloaded from rsta.royalsocietypublishing.org

Colquhoun, D. & Hawkes, A. G. 1983 The principles of the stochastic interpretation of ion channel records. In *Single channel recording* (ed. B. Sakmann & E. Neher), pp. 135–175. New York: Plenum.

Colquhoun, D. & Hawkes, A. G. 1987 A note on correlations in single ion channel records. *Proc. R. Soc. Lond.* B **230**, 15–52.

Colquhoun, D. & Sakmann, B. 1981 Fluctuations in the microsecond time range of the current through single acetylcholine receptor ionchannels. *Nature, Lond.* **294**, 464–466.

Colquhoun, D. & Sakmann, B. 1985 Fast events in single-channel currents activated by acetylcholine and its analogues at the frog muscle end-plate. *J. Physiol., Lond.* **369**, 501–557.

Colquhoun, D. & Sigworth, F. J. 1983 Fitting and statistical analysis of single channel records. In *Single channel recording* (ed. B. Sakmann & E. Neher), pp. 191–263. New York: Plenum.

Crouzy, S. C. & Sigworth, F. J. 1990 A practical approximate solution to the dwell time omission problem. (Submitted.)

Fredkin, D. R., Montal, M. & Rice, J. A. 1985 Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. In *Proceedings of the Berkeley Conference in honour of Jerzy Neyman and Jack Kiefer* (ed. L. M. LeCam & R. A. Ohlsen), pp. 269–289. Belmont: Wadsworth.

Garwood, F. 1940 An application of the theory of probability to the operation of vehicular-controlled traffic signals. *Jl R. statist. Soc.* **7**, 65–77.

Grosswald, E. 1978 *Bessel polynomials*. Lecture Notes in Mathematics, vol. 698. Berlin, Heidelburg and New York: Springer-Verlag.

Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. 1981 Improved patch clamp techniques for high resolution current recording from cells and cell-free membrane patches. *Pflügers Arch. ges. Physiol.* **391**, 85–100.

Hawkes, A. G. 1965 Queueing for gaps in traffic. *Biometrika* **52**, 79–85.

Hawkes, A. G. 1970 Bunching in a semi-Markov process. *J. appl. Prob.* **7**, 175–182.

Hawkes, A. G. & Sykes, A. M. 1990 Equilibrium distributions of finite-state Markov processes. *IEEE Trans. Reliability*. (In the press.)

Howe, J. R., Cull-Candy, S. G. & Colquhoun, D. 1990 Currents through single glutamate-receptor channels in outside-out patches from rat cerebellar granule cells. *J. Physiol., Lond.* (In the press.)

Jalali, A. & Hawkes, A. G. 1990 Waiting for a gap in Poisson traffic. University of Wales, Swansea, working paper EBMS/1990/11.

Katz, B. & Miledi, R. 1972 The statistical nature of the acetylcholine potential and its molecular components. *J. Physiol., Lond.* **224**, 665–699.

Kienker, P. 1989 Equivalence of aggregated Markov models of ion channel gating. *Proc. R. Soc. Lond.* B **236**, 269–309.

Marshall, C. G., Ogden, D. C. & Colquhoun, D. 1990 The actions of suxamethonium (succinyldicholine) as an agonist and channel blocker at the nicotinic receptor of frog muscle. *J. Physiol., Lond.* (In the press.)

McCammon, J. A. & Harvey, S. C. 1987 *Dynamics of proteins and nucleic acids*. Cambridge University Press.

Milne, R. K., Yeo, G. F., Edeson, R. O. & Madsen, B. W. 1988 Stochastic modelling of a single ion channel: an alternating renewal approach with application to limited time resolution. *Proc. R. Soc. Lond.* B **233**, 247–292.

Neher, E. & Sakmann, B. 1976 Single channel currents recorded from membrane of denervated frog muscle fibres. *Nature, Lond.* **260**, 799–802.

Neher, E. & Steinbach, J. H. 1978 Local anaesthetics transiently block current through single acetylcholine receptor channels. *J. Physiol., Lond.* **277**, 153–176.

Neher, E. 1983 The charge carried by single channel currents of rat cultured muscle cell in the presence of local anaesthetics. *J. Physiol., Lond.* **339**, 663–678.

Pyke, R. 1961 Markov renewal processes: Definitions and preliminary properties. *Ann. math. Statist.* **32**, 1231–1242.

Rickard, J. T. 1977 The zero-crossing interval statistics of the smoothed random telegraph signal. *Info. Sci.* **13**, 253–268.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Downloaded from rsta.royalsocietypublishing.org

Roux, B. & Sauvé, R. 1985 A general solution of the time interval omission problem applied to single channel analysis. *Biophys. J.* **48**, 149–158.

Sachs, F. & Auerbach, A. 1983 Single channel electrophysiology : use of the patch clamp. *Methods Enzymol.* **103**, 147–176.

Sachs, F., Neil, J. & Barkakati, N. 1982 The automated analysis of data from single ionic channels. *Pflügers Arch. Eur. J. Physiol.* **395**, 331–340.

Sine, S. M., Claudio, T. & Sigworth, F. J. 1990 Activation of Torpedo acetylcholine receptors expressed in mouse fibroblasts : single channel current kinetics reveal distinct agonist binding affinities. *J. gen. Physiol.* (In the press.)

Wilson, D. L. & Brown, A. M. 1985 Effect of limited interval resolution on single channel measurements with application to Ca channels. *IEEE Trans. Biomed. Engng* **BME-32**, 786–797.

Yeo, G. F., Milne, R. K., Edeson, R. O. & Madsen, B. W. 1988 Statistical inference from single channel records : two-state Markov model with limited time resolution. *Proc. R. Soc. Lond.* B **235**, 63–94.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES