Adv. Appl. Prob. **24,** 302–321 (1992)
Printed in N. Ireland
© *Applied Probability Trust* 1992

GENERALISED EIGENPROBLEMS ARISING IN AGGREGATED MARKOV PROCESSES ALLOWING FOR TIME INTERVAL OMISSION

ASSAD JALALI AND
ALAN G. HAWKES, *University of Wales, Swansea*

Abstract

We consider a continuous-time Markov chain in which one cannot observe individual states but only which of two sets of states is occupied at any time. Furthermore, we suppose that the resolution of the recording apparatus is such that small sojourns, of duration less than a constant deadtime, cannot be observed. We obtain some results concerning the poles of the Laplace transform of the probability density function of apparent occupancy times, which correspond to a problem about generalised eigenvalues and eigenvectors. These results provide useful asymptotic approximations to the probability density of occupancy times. A numerical example modelling a calcium-activated potassium channel is given. Some generalisations to the case of random deadtimes complete the paper.

EIGENVALUES; OCCUPANCY TIMES; ION CHANNELS

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60K15
SECONDARY 60J27, 15A18, 15AH2, 92C05, 92C30

1. Introduction

Our basic model is a continuous-time finite-state Markov process, $X(t)$, in which the rate constants for transitions between states i and j $(i \neq j)$ are the elements, q_{ij}, of the transition rate matrix Q. The diagonal elements, q_{ii}, are defined so that the rows sum to 0, so $-1/q_{ii}$ is the mean lifetime of a sojourn in state i.

We suppose the states are divided into just two mutually exclusive subsets, denoted $\mathscr{A}$ and $\mathscr{F}$, and that we are unable to observe the occupancy of individual states but, at any time, all we can say is that the system is in the set $\mathscr{A}$ or in the set $\mathscr{F}$. In the context of modelling ion channels in biological membranes, see Colquhoun and Hawkes (1982), subset $\mathscr{A}$ corresponds to the channel being open, and a current flow being observed, while subset $\mathscr{F}$ corresponds to the channel being shut. We suppose that $k_{\mathscr{A}}$ is the number of $\mathscr{A}$ states and $k_{\mathscr{F}}$ the number of $\mathscr{F}$ states, so $k_{\mathscr{A}} + k_{\mathscr{F}} = k$ is the total number of states, then the Q-matrix may be partitioned as

$$Q = \begin{bmatrix} Q_{\mathscr{A}\mathscr{A}} & Q_{\mathscr{A}\mathscr{F}} \\ Q_{\mathscr{F}\mathscr{A}} & Q_{\mathscr{F}\mathscr{F}} \end{bmatrix}.$$

Received 1 March 1991; revision received 22 April 1991.

Postal address for both authors: Statistics and Operational Research Group, European Business Management School, University of Wales, Swansea, Singleton Park, Swansea SA2 8PP, UK.

In practice this may be complicated by the resolution of the recording equipment, leading to an inability to detect very small intervals. We shall model this by supposing a constant critical gap or deadtime, τ, such that occupancies of either subset $\mathscr{A}$ or subset $\mathscr{F}$ of duration less than this time are missed. We suppose, after Colquhoun and Sigworth (1983), that an observable $\mathscr{A}$-occupancy begins with a sojourn in the $\mathscr{A}$ states of duration at least τ and ends at the beginning of the next sojourn in $\mathscr{F}$ of duration greater than τ. Thus, the observed occupancy may consist of r sojourns in the set of $\mathscr{F}$ states, each of duration less than τ, and $r+1$ sojourns in the set of $\mathscr{A}$ states, of which the first must exceed τ. Observed $\mathscr{F}$-occupancies may be defined similarly.

Following Ball and Sansom (1988), we consider a semi-Markov process whose events occur at time τ after the start of observed occupancies. An event type is the state of the underlying Markov process which is occupied at that time. The durations of the intervals between events, which we call $e\mathscr{A}$ and $e\mathscr{F}$ intervals, are identical to the durations of the observed $\mathscr{A}$-occupancies and $\mathscr{F}$-occupancies. Any such duration, T say, is necessarily greater than τ, so it will often be convenient to consider the excess duration $U = T - \tau$. The definitions are illustrated in Figure 1.1.

Intervals of this process will be alternately of type $e\mathscr{A}$ and $e\mathscr{F}$, so the semi-Markov transition densities will be given by a matrix of the form

$$ {}^eG(t) = \begin{bmatrix} 0 & {}^eG_{\mathscr{A}\mathscr{F}}(t) \\ {}^eG_{\mathscr{F}\mathscr{A}}(t) & 0 \end{bmatrix} $$

with Laplace transform

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

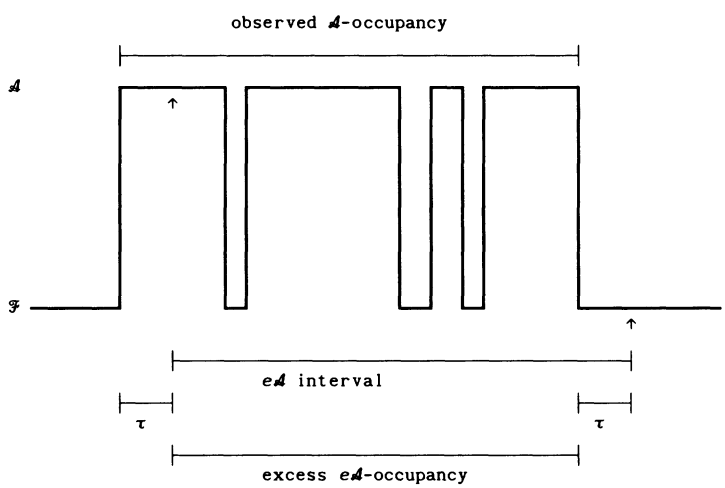


Figure 1.1. Illustration of the definition of an observed $\mathscr{A}$-occupancy and excess $\mathscr{A}$-occupancy. The $e\mathscr{A}$ interval is equivalent to an observed $\mathscr{A}$-occupancy shifted by an amount τ. The events of the semi-Markov process defined in the text occur at points marked ↑

The Markov chain embedded at the event points has transition matrix

$$ {}^eG = \begin{bmatrix} 0 & {}^eG_{\mathcal{AF}} \\ {}^eG_{\mathcal{FA}} & 0 \end{bmatrix}. $$

Here we simplify the notation when setting $s = 0$ in a Laplace transform by omitting the asterisk and the argument: for example, in this case ${}^eG^*_{\mathcal{AF}}(0)$ is written as ${}^eG_{\mathcal{AF}}$.

By looking only at alternate events, and ignoring the interval durations, we have a Markov chain on the $\mathcal{A}$ states with transition matrix ${}^eG_{\mathcal{AF}}{}^eG_{\mathcal{FA}}$ and equilibrium probability vector $\phi_{\mathcal{A}}$, satisfying

$$ \phi_{\mathcal{A}} = \phi_{\mathcal{A}}{}^eG_{\mathcal{AF}}{}^eG_{\mathcal{FA}}, \qquad \phi_{\mathcal{A}} u_{\mathcal{A}} = 1, $$

where $u_{\mathcal{A}}$ is a column vector whose elements are all equal to 1.

We also discuss the probability density of observed $\mathcal{A}$-occupancies; the distribution of observed $\mathcal{F}$-occupancies can be obtained simply by interchanging $\mathcal{A}$ and $\mathcal{F}$ in the notation. Let ${}^{\mathcal{A}}R(t)$ be a matrix whose ijth element ($i, j \in \mathcal{A}$) is

$$ {}^{\mathcal{A}}R_{ij}(t) = P[X(t) = j \text{ and no } \mathcal{F}\text{-sojourn is detected over } (0, t) \mid X(0) = i], $$

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

$$ {}^eG_{\mathcal{AF}}(t) = {}^{\mathcal{A}}R(t - \tau)Q_{\mathcal{AF}} \exp(Q_{\mathcal{FF}}\tau). \tag{1.1} $$

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

${}^eG_{\mathcal{AF}}(t)$ is a most important function as it enables one to write down a likelihood for an observed record, see Section 6 of Hawkes et al. (1990), and because the probability density of observed $\mathcal{A}$-occupancies is given by

$$ {}^ef_T(t) = \phi_{\mathcal{A}}{}^eG_{\mathcal{AF}}(t)u_{\mathcal{F}}. \tag{1.2} $$

The excess $\mathcal{A}$-occupancy $U = T - \tau$ and so the p.d.f. ${}^ef_T(t) = f_U(t - \tau)$ and thus

$$ f_U(t) = \phi_{\mathcal{A}}{}^eG_{\mathcal{AF}}(t + \tau)u_{\mathcal{F}} = \phi_{\mathcal{A}}{}^{\mathcal{A}}R(t)Q_{\mathcal{AF}} \exp(Q_{\mathcal{FF}}\tau)u_{\mathcal{F}}. \tag{1.3} $$

It follows that ${}^{\mathcal{A}}R(t)$ is the key to the problem. Hawkes et al. (1990) show that its Laplace transform can be written as

$$ {}^{\mathcal{A}}R^*(s) = \{I - G^*_{\mathcal{AF}}(s)S^*_{\mathcal{FF}}(s)G^*_{\mathcal{FA}}(s)\}^{-1}(sI - Q_{\mathcal{AA}})^{-1} \tag{1.4} $$

where $G^*_{\mathcal{AF}}(s) = (sI - Q_{\mathcal{AA}})^{-1}Q_{\mathcal{AF}}$, $G^*_{\mathcal{FA}}(s) = (sI - Q_{\mathcal{FF}})^{-1}Q_{\mathcal{FA}}$ and $S^*_{\mathcal{FF}}(s)$ is defined by

$$ \int_0^\tau e^{-st} \exp(Q_{\mathcal{FF}}t)\, dt = \{I - \exp(-(sI - Q_{\mathcal{FF}})\tau)\}(sI - Q_{\mathcal{FF}})^{-1} $$

$$ = S^*_{\mathcal{FF}}(s)(sI - Q_{\mathcal{FF}})^{-1}. $$

Substituting these into (1.4) yields the alternative expression

$$(1.5) \qquad {}^{\mathscr{A}}R^*(s) = \left[sI - Q_{\mathscr{A}\mathscr{A}} - Q_{\mathscr{A}\mathscr{F}}\left(\int_0^\tau \exp(-st)\exp(Q_{\mathscr{F}\mathscr{F}}t)\,dt\right)Q_{\mathscr{F}\mathscr{A}}\right]^{-1}.$$

These results were given, using different notation, by Ball and Sansom (1988), generalised to allow different critical intervals, $\tau_{\mathscr{A}}$ and $\tau_{\mathscr{F}}$ which could be random variables.

Hawkes et al. (1990) obtained the inverse of this transform, and hence the p.d.f. $f_U(t)$ in a form such that, for t in the interval $I_{n+1} = (n\tau, (n+1)\tau)$, $f_U(t) = \sum_{i=1}^{k} \theta_{in}(t)\exp(-\lambda_i t)$, where $\theta_{in}(t)$ is a matrix polynomial of degree n in t and $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the eigenvalues of $-Q$. Formulae for the recursive computation of the coefficient matrices of the polynomials $\theta_{in}(t)$ were also given. 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, increase as n increases.

However, we may hope for some good approximation by a simple form for large t. In this paper we show that the density $f_U(t)$ is asymptotically exponential in form. We also show that, if the transition matrix Q is reversible, ${}^{\mathscr{A}}R^*(s)$ has exactly k real poles which lead to an approximate expression for $f_U(t)$ as a linear combination of exponentials which is, in practice, very good for all except small values of t. For those values of t, the exact density mentioned above is simple and accurate to compute.

2. Roots of a determinantal equation

From Equation (1.5) we see that the asymptotic behaviour of ${}^{\mathscr{A}}R(t)$, and hence of $f_U(t)$, depends on the values of s which render singular the matrix defined as $W(s) = sI - H(s)$, where

$$(2.1) \qquad H(s) = Q_{\mathscr{A}\mathscr{A}} + Q_{\mathscr{A}\mathscr{F}}\left(\int_0^\tau \exp(-st)\exp(Q_{\mathscr{F}\mathscr{F}}t)\,dt\right)Q_{\mathscr{F}\mathscr{A}},$$

or, if s is not an eigenvalue of $Q_{\mathscr{F}\mathscr{F}}$,

$$(2.1a) \qquad H(s) = Q_{\mathscr{A}\mathscr{A}} + Q_{\mathscr{A}\mathscr{F}}(sI - Q_{\mathscr{F}\mathscr{F}})^{-1}(I - \exp(-(sI - Q_{\mathscr{F}\mathscr{F}})\tau))Q_{\mathscr{F}\mathscr{A}}.$$

In other words, we are interested in the roots of the determinantal equation

$$(2.2) \qquad \det W(s) = 0.$$

In this section we prove two results. First we prove an analogue of the Perron–Frobenius theorem. Next we show that, if Q corresponds to a reversible process, the above equation has exactly $k_{\mathscr{A}}$ real roots, where $k_{\mathscr{A}}$ is the number of states in $\mathscr{A}$.

Let $\lambda_i(s)$, $i = 1, \cdots, k_{\mathscr{A}}$, be the eigenvalues of $H(s)$. These are continuous functions of s; for complex s they are even analytic functions of $s^{1/m}$ for some

$m \leqq k_{\mathscr{A}}$, see Lancaster (1969), p. 237. Clearly, s_n is a root of (2.2) if and only if $s_n I - H(s_n)$ is singular, i.e. if and only if s_n is an eigenvalue of $H(s_n)$, or $s_n = \lambda_i(s_n)$ for some i. Thus the real roots of Equation (2.2) are exactly the points of intersection of the graphs $\lambda_i(s)$, as functions of real s when the function is also real, with the diagonal straight-line function s through the origin. Corresponding to any such point s_n, real or complex, there will be column and row vectors such that $vW(s_n) = W(s_n)u = 0$, which are obviously eigenvectors of $H(s_n)$. We call such vectors p-eigenvectors (p for pseudo).

It is useful here to establish a couple of results on maximal eigenvalues and ML matrices.

Lemma 2.1. Denote by A^+ the matrix whose elements are the moduli of the elements of the square matrix A and let $\lambda_{\max}(A)$ be the eigenvalue of A with largest absolute value. Then for any non-negative matrix P with $A^+ \leqq P$ elementwise, $\lambda_{\max}(P)$ is real and non-negative (positive if all elements of P are strictly positive) and $|\lambda_{\max}(A)| \leqq \lambda_{\max}(P)$. If A is irreducible and there is at least one pair of indices i, j such that $|a_{ij}| < P_{ij}$, then $|\lambda_{\max}(A)| < \lambda_{\max}(P)$.

These results are due to Frobenius; see Ostrowski (1964), pp. 81–83 and Taussky (1964), pp. 127–128.

Definition. An ML (Metzler–Leontief) matrix is a matrix, M say, such that all the elements of $\mu I + M$ are positive for some constant μ.

Lemma 2.2. Let M be an ML matrix and let $\lambda_1(M)$ be the real eigenvalue of M which is greater than the real part of every other eigenvalue of M. Then there exists a real number a_M such that, $\forall b > a_M$, $b + \lambda_1(M)$ is the eigenvalue with maximum modulus among all the eigenvalues of $bI + M$.

Corollary. If M_1 and M_2 are ML matrices with $M_1 \leqq M_2$ elementwise, then $\lambda_1(M_1) \leqq \lambda_1(M_2)$. The eigenvalue inequality is strict if M_1 is irreducible and the matrix inequality is strict for at least one element.

Proof. The existence of a real eigenvalue with the stated property follows from Theorem 2.5 of Seneta (1973). Let $\lambda = x + iy$ be any other eigenvalue. Then $\lambda_1(M) > x$ and, for any real b, $b + \lambda = (b + x) + iy$ is an eigenvalue of $bI + M$ and has smaller modulus than $b + \lambda_1(M)$ if $(b + \lambda_1(M))^2 > (b + x)^2 + y^2$, that is if $b > (x^2 + y^2 - \lambda_1^2(M))/2(\lambda_1(M) - x)$. Thus $b + \lambda_1(M)$ is the eigenvalue of maximum modulus of $bI + M$ for all $b > a_M = \max_{\lambda \neq \lambda_1(M)} \{(|\lambda|^2 - \lambda_1^2(M))/2(\lambda_1(M) - \operatorname{Re} \lambda)\}$.

To prove the corollary we take b large enough so that all the elements of both $bI + M_1$ and $bI + M_2$ are positive and also $b > \max\{a_{M_1}, a_{M_2}\}$. Then $b + \lambda_1(M_i)$ are the eigenvalues of $bI + M_i$ of maximal modulus and so, by Lemma 2.1, they satisfy $|b + \lambda_1(M_1)| \leqq |b + \lambda_1(M_2)|$ and are also real and positive: hence $b + \lambda_1(M_1) \leqq b + \lambda_1(M_2)$ and, therefore, also $\lambda_1(M_1) \leqq \lambda_1(M_2)$.

We are now ready to state the first of our theorems.

Theorem 2.1. If $H(s)$ is irreducible, $\det W(s) = 0$ always has a simple real root $s_1 < 0$ which is greater than the real part of any other root. This is the point of intersection of the strictly decreasing function $\lambda_1(s)$, defined as the largest real root of $H(s)$ for real s, with the diagonal line s through the origin. Corresponding to this there are unique left and right p-eigenvectors which are non-negative.

Proof. Obviously, $H(s)$ is an ML matrix for real s. If it is also irreducible, then it follows, from Theorem 2.5 of Seneta (1973), that there exists a real eigenvalue of $H(s)$, namely $\lambda_1(s)$, which is simple and is greater than the real part of every other eigenvalue. To this eigenvalue there correspond non-negative left and right eigenvectors which are unique up to a scalar multiplier. Note that the irreducibility condition is not a serious restriction, because the results can still be obtained by a continuity argument.

As s varies from $-\infty$ to ∞, $\lambda_1(s)$ varies from ∞ to ρ_1, the eigenvalue of $Q_{\mathcal{A}\mathcal{A}}$ with largest real part. Moreover, $\lambda_1(s)$ is strictly decreasing: this follows from the corollary to Lemma 2.1 because, for real $s_1 < s_2$, it is easy to see that $H(s_2) \leqq H(s_1)$ with strict inequality for at least one element. Consequently $\lambda_1(s)$ intersects the diagonal line only once, see Figure 2.1, at the point $s_1 > \rho_1$, which is therefore a simple root of our equation, and the corresponding left and right p-eigenvectors can be chosen to be non-negative. As all other real $\lambda_k(s)$, $k \neq 1$, lie below $\lambda_1(s)$, s_1 is the largest real root of $\det W(s) = 0$.

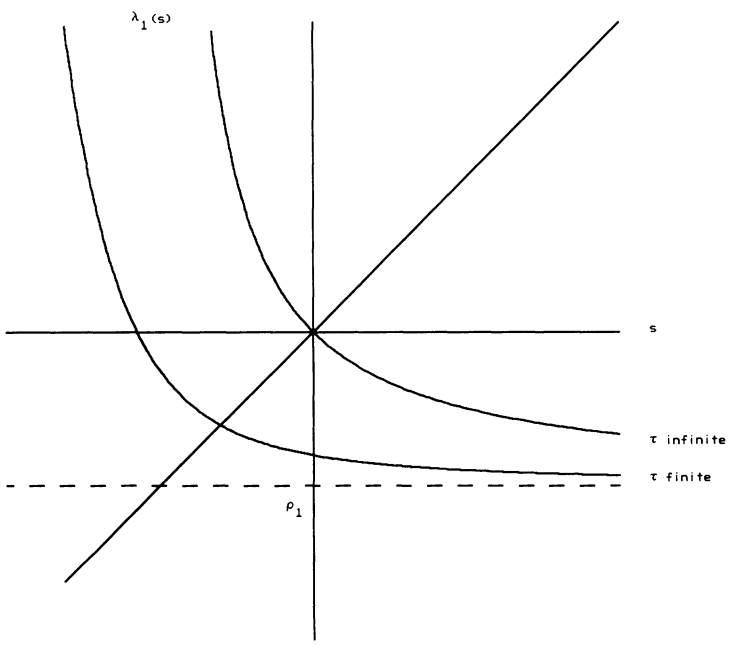


Figure 2.1. Graphs of $\lambda_1(s)$ for finite τ and for $\tau = \infty$

Now suppose s is a non-real root of $\det W(s) = 0$, with $\operatorname{Re} s = u$, then s is an eigenvalue of $H(s)$ and therefore $b + s$ is an eigenvalue of $bI + H(s)$. Choose b large enough so that $bI + Q_{\mathcal{A}\mathcal{A}}$ is non-negative and, see Lemma 2.2, so that $b + \lambda_1(u) = \lambda_{\max}(bI + H(u))$. Then, elementwise,

$$\begin{aligned}(bI + H(s))^+ &\leqq bI + Q_{\mathcal{A}\mathcal{A}} + Q_{\mathcal{A}\mathcal{F}} \int_0^\tau [\exp(-(sI - Q_{\mathcal{F}\mathcal{F}})t)]^+ \, dt Q_{\mathcal{F}\mathcal{A}} \\ &= bI + Q_{\mathcal{A}\mathcal{A}} + Q_{\mathcal{A}\mathcal{F}} \int_0^\tau \exp(-(uI - Q_{\mathcal{F}\mathcal{F}})t) \, dt Q_{\mathcal{F}\mathcal{A}} \\ &= bI + H(u).\end{aligned}$$

Then, by Lemma 2.1, $b + u < |b + s| \leqq |\lambda_{\max}(bI + H(s))| \leqq \lambda_{\max}(bI + H(u)) = b + \lambda_1(u)$ and so $u < \lambda_1(u)$. But this implies that $\operatorname{Re} s = u < s_1$, see Figure 2.1.

Now consider different values of the deadtime $\tau_1 < \tau_2$. It is easy to see that, for any real s, $H(s, \tau_1) \leqq H(s, \tau_2)$ elementwise, with strict inequality for at least one element. Then the same arguments as used above imply that $\lambda_1(s, \tau_1) < \lambda_1(s, \tau_2)$, and so $s_1(\tau_1) < s_1(\tau_2)$ and, therefore, $s_1(\tau) < s_1(\infty)$ for all finite τ. But $H(0, \infty) = Q_{\mathcal{A}\mathcal{A}} - Q_{\mathcal{A}\mathcal{F}} Q_{\mathcal{F}\mathcal{F}}^{-1} Q_{\mathcal{F}\mathcal{A}}$ is an ML matrix whose rows sum to 0, being the Q-matrix of a Markov process obtained from the original Markov process by stopping the clock whenever the system is not in the set of states $\mathcal{A}$, so it is well known that the root with maximum real part is $\lambda = 0$, see for example Cox and Miller (1965), p. 184. Thus $s_1(\infty) = 0$ and so $s_1 < 0$ for all finite τ. This completes the proof of the theorem.

Our second theorem concerns reversible processes, which are of great interest in many physical problems, including the modelling of ion channel currents (Colquhoun and Hawkes (1982)).

Theorem 2.2. When Q is reversible, $\det W(s) = 0$ has exactly $k_{\mathcal{A}}$ real roots (taking the multiplicity into account). Order the roots so that $s_1 > s_2 \geqq \cdots \geqq s_{k_{\mathcal{A}}}$ and let $\rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_{k_{\mathcal{A}}}$ be the eigenvalues of $Q_{\mathcal{A}\mathcal{A}}$ and $\mu_1 = 0 > \mu_2 \geqq \cdots \geqq \mu_k$ the eigenvalues of Q, then $\rho_i \leqq s_i \leqq \mu_i$ for $i = 1, \cdots, k_{\mathcal{A}}$.

Proof. As Q is reversible there is a symmetric matrix $\tilde{Q} = \Pi^{\frac{1}{2}} Q \Pi^{-\frac{1}{2}}$, where the diagonal matrix $\Pi = \operatorname{diag}(\pi_1, \cdots, \pi_k)$ contains the equilibrium probabilities, Fredkin et al. (1985). Clearly $\tilde{Q}$ and Q have the same eigenvalues. Let the matrices $\tilde{Q}$ and W be partitioned the same way as Q and define $\tilde{W}(s) = sI - \tilde{H}(s)$, where $\tilde{H}(s) = \tilde{Q}_{\mathcal{A}\mathcal{A}} + \tilde{Q}_{\mathcal{A}\mathcal{F}} \int_0^\tau \exp(-(sI - \tilde{Q}_{\mathcal{F}\mathcal{F}})t) \, dt \tilde{Q}_{\mathcal{F}\mathcal{A}}$. Clearly, $\tilde{W}(s) = \Pi_{\mathcal{A}\mathcal{A}}^{\frac{1}{2}} W(s) \Pi_{\mathcal{A}\mathcal{A}}^{-\frac{1}{2}}$ and $\tilde{H}(s) = \Pi_{\mathcal{A}\mathcal{A}}^{\frac{1}{2}} H(s) \Pi_{\mathcal{A}\mathcal{A}}^{-\frac{1}{2}}$ so $\tilde{H}(s)$ and $H(s)$ have the same eigenvalues and the equations $\det \tilde{W}(s) = 0$ and $\det W(s) = 0$ have the same roots. Now $\tilde{H}(s)$ is symmetric so that, for real s, $\lambda_1(s) > \lambda_2(s) \geqq \cdots \geqq \lambda_{k_{\mathcal{A}}}(s)$ are real graphs, see Bellman (1970), p. 35. Similarly, the eigenvalues of $\tilde{Q}_{\mathcal{F}\mathcal{F}}$ are real and so the eigenvalues of $\exp(\tilde{Q}_{\mathcal{F}\mathcal{F}} t)$, which are exponentials of t times the eigenvalues of $\tilde{Q}_{\mathcal{F}\mathcal{F}}$, are positive: thus $\exp(\tilde{Q}_{\mathcal{F}\mathcal{F}} t)$ is positive definite for all t, see Bellman (1970), p. 54.

Now let $s_1 < s_2$ and consider

$$\tilde{H}(s_1) - \tilde{H}(s_2) = \tilde{Q}_{\mathscr{A}\mathscr{F}} \int_0^\tau \exp(\tilde{Q}_{\mathscr{F}\mathscr{F}}t)(\exp(-s_1 t) - \exp(-s_2 t)\, dt \tilde{Q}_{\mathscr{F}\mathscr{A}}.$$

As $\exp(-s_1 t) - \exp(-s_2 t) > 0$, the above integral, which we denote by $\mathscr{D}$, is positive definite. Thus for any vector $v \neq 0$ let $w = \tilde{Q}_{\mathscr{F}\mathscr{A}} v$, and note that $\tilde{Q}_{\mathscr{A}\mathscr{F}} = \tilde{Q}^{\mathrm{T}}_{\mathscr{F}\mathscr{A}}$, then we have $v^{\mathrm{T}}\tilde{Q}_{\mathscr{A}\mathscr{F}}\mathscr{D}\tilde{Q}_{\mathscr{F}\mathscr{A}}v = w^{\mathrm{T}}\mathscr{D}w \geqq 0$. Thus $\tilde{H}(s_1) = \tilde{H}(s_2) + \tilde{Q}_{\mathscr{A}\mathscr{F}}\mathscr{D}\tilde{Q}_{\mathscr{F}\mathscr{A}}$, where the final term is non-negative definite. Hence a well-known result in matrix theory, see e.g. Bellman (1970), p. 117, Theorem 3, tells us that $\lambda_i(s_1) \geqq \lambda_i(s_2)$ for all $i = 1, \cdots, k_{\mathscr{A}}$. This means that each graph $\lambda_i(s)$, $i = 1, \cdots, k_{\mathscr{A}}$, is a decreasing function of s and therefore cuts the diagonal graph s exactly once. Hence $\det W(s) = 0$ has exactly $k_{\mathscr{A}}$ roots. This establishes the main part of the theorem.

Now if we consider different deadtimes $\tau_1 < \tau_2$, essentially the same argument as that used above implies that $\lambda_i(s, \tau_1) \leqq \lambda_i(s, \tau_2)$ and therefore $\lambda_i(s, \tau) \leqq \lambda_i(s, \infty)$ for any finite τ. But, as we have seen, the root $s_i(\tau)$ of $\det W(s, \tau) = 0$ is the intersection of the graph $\lambda_i(s, \tau)$ with the diagonal line s, so $s_i = s_i(\tau) \leqq s_i(\infty)$. But $W(s, \infty) = (sI - Q_{\mathscr{A}\mathscr{A}}) - Q_{\mathscr{A}\mathscr{F}}(sI - Q_{\mathscr{F}\mathscr{F}})^{-1}Q_{\mathscr{F}\mathscr{A}}$ and thus $\det(sI - Q) = \det(SI - Q_{\mathscr{F}\mathscr{F}}) \times \det W(s, \infty)$, see for example Morrison (1967), Section 2.11. Therefore, the roots of $\det W(s, \infty) = 0$ are among the eigenvalues of Q so that, for any finite τ, $s_i = s_i(\tau) \leqq \mu_i$. When $s \to \infty$, $H(s) \to Q_{\mathscr{A}\mathscr{A}}$ and so the eigenvalues $\lambda_i(s) \to \rho_i$. As $\lambda_i(s)$ is decreasing, the root $s_i \geqq \rho_i$. This completes the proof of the theorem.

Irreversible processes. As we have already said, the case where Q corresponds to a reversible process is important in ion channel modelling. However, it is of interest to see what happens to our results if this does not hold. Below we give an example of an irreversible process which may have more or less than $k_{\mathscr{A}}$ roots.

Consider a Markov process in which $\mathscr{A}$ and $\mathscr{F}$ each consist of two states. The Q-matrix is partitioned as

$$\begin{array}{c} \\ \mathscr{A} \\ \mathscr{F} \end{array} \begin{array}{c} \begin{array}{cc} \mathscr{A} & \mathscr{F} \end{array} \\ \left(\begin{array}{cc:cc} -\lambda & 0 & \lambda & 0 \\ 0 & -\lambda & 0 & \lambda \\ \hdashline 0 & \mu & -\mu & 0 \\ \mu & 0 & 0 & -\mu \end{array}\right) \end{array} \tag{2.3}$$

It is clear from the graphical representation, Figure 2.2, that this is irreversible. Then

$$H(s) = \begin{pmatrix} -\lambda & 0 \\ 0 & -\lambda \end{pmatrix} + f(s) \begin{pmatrix} 0 & \lambda\mu \\ \lambda\mu & 0 \end{pmatrix}$$

where

$$f(s) = \int_0^\tau \exp(-(s+\mu)t)\, dt = (1 - \exp(-(s+\mu)\tau))/(s+\mu). \tag{2.4}$$

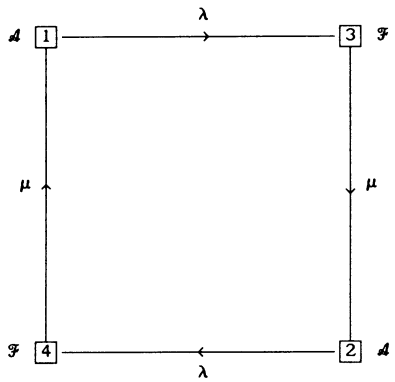


Figure 2.2. State diagram of an irreversible process, with two states in each of the subsets $\mathscr{A}$ and $\mathscr{F}$, whose Q-matrix is given by Equation (2.3)

Then

$$\lambda_1(s) = -\lambda + \lambda\mu f(s), \qquad \lambda_2(s) = -\lambda - \lambda\mu f(s), \tag{2.5}$$

and so

$$\lambda_1'(s) = -\lambda_2'(s) = -\lambda\mu \int_0^\tau t \exp(-(s+\mu)t)\, dt,$$

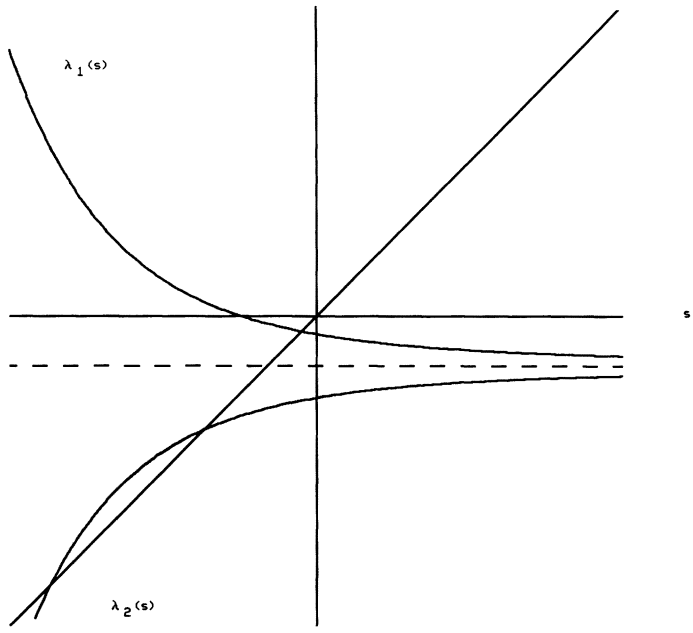
$$\lambda_1''(s) = -\lambda_2''(s) = \lambda\mu \int_0^\tau t^2 \exp(-(s+\mu)t)\, dt.$$

Hence, $\lambda'(s)$ is decreasing, as expected from Theorem 2.1, convex and meets the diagonal graph s in one point. $\lambda_2(s)$ is increasing and concave, so it may intersect the diagonal in two points or in none, see Figure 2.3. Therefore, $\det W(s) = 0$ has three real roots in the case of Figure 2.3a and only one real root in the case of Figure 2.3b. Our example shows that when Q is not reversible, the equation $\det W(s) = 0$ may have more or less than $k_{\mathscr{A}}$ real roots.

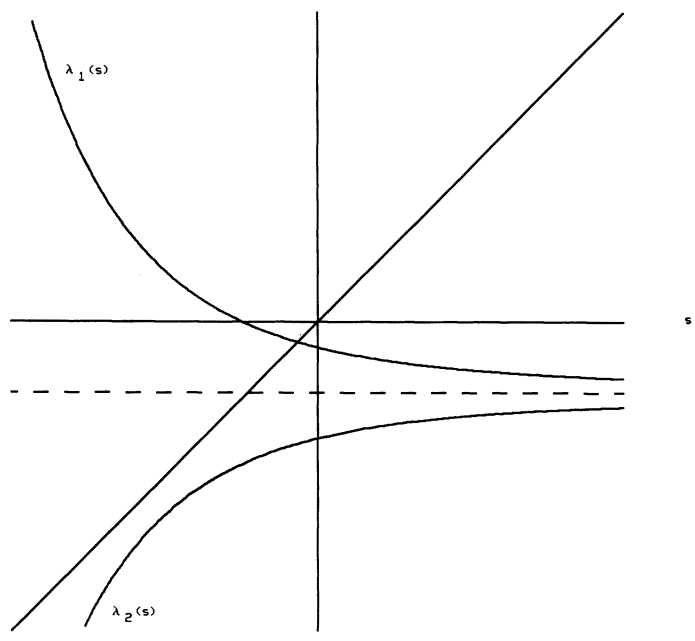
Complex roots. In this paper we deal exclusively with real roots. For the two-state case, with one $\mathscr{A}$ state and one $\mathscr{F}$ state, Jalali and Hawkes (1992) found that, in addition to the one real root expected, there are infinitely many complex conjugate pairs of roots. It is almost certainly also the case in the more general situation studied here. However, it is much more difficult to deal with them in general and the results to be presented in the next section make use of the real roots only. They appear to be adequate for practical purposes.

3. Asymptotic distributions

Our main concern in this paper is to obtain good approximations for ${}^{\mathscr{A}}R(t)$ for large t, and hence obtain similar approximations for the transition density matrix



(a)



(b)

Figure 2.3. Graphs of $\lambda_1(s)$ and $\lambda_2(s)$ given by Equations (2.4), (2.5) for the irreversible model of Equation (2.3), with $\tau = 0{\cdot}2$ and $\mu = 5$. In (a), when $\lambda = 8{\cdot}75/e$, there are three roots; in (b), with $\lambda = 12{\cdot}5/e$, there is just one

${}^{e}G_{\mathcal{AF}}(t)$ from Equation (1.1) and for the density $f_U(t)$ from Equation (1.3). It follows from Theorem 2.1 that, as $t \to \infty$, ${}^{\mathcal{A}}R(t) \sim$ the residue of $\exp(st)\,{}^{\mathcal{A}}R^*(s)$ at $s = s_1$, the real root which exceeds the real part of any other root of $\det W(s) = 0$, see for instance Smith (1966), Chapter 10. As s_1 is a simple root, we have a simple pole and so asymptotically ${}^{\mathcal{A}}R(t)$ has an exponential form $\exp(s_1 t)M$ where the matrix M is the residue of ${}^{\mathcal{A}}R^*(s)$ at s_1. The form of this matrix is given in the following theorem.

Theorem 3.1. If $H(s)$ is irreducible, then asymptotically ${}^{\mathcal{A}}R(t) \sim \exp(s_1 t)c_1 r_1 / r_1 W'(s_1) c_1$ where c_1, r_1 are the right (column) and left (row) eigenvectors of $H(s_1)$ corresponding to eigenvalue s_1.

Proof. First note that

$$W(s_1)c_1 = (s_1 I - H(s_1))c_1 = 0 = r_1(s_1 I - H(s_1)) = r_1 W(s_1), \tag{3.1}$$

so that c_1, r_1 are also right and left eigenvectors of $W(s_1)$ corresponding to the zero eigenvalue. We may normalise in the usual way, so that

$$r_1 c_1 = 1. \tag{3.2}$$

Let $\tilde{W}(s)$ be the adjoint matrix of $W(s)$, then

$$W(s)\tilde{W}(s) = \tilde{W}(s)W(s) = I \det W(s) \tag{3.3}$$

and

$$W^{-1}(s) = \tilde{W}(s)/\det W(s).$$

As s_1 is a simple root of $\det W(s_1) = 0$, $W(s_1)$ has rank $n-1$ and $\tilde{W}(s_1)$ has rank 1, so that we have

$$\tilde{W}(s_1) = C c_1 r_1$$

for some scalar constant C. Then the residue of ${}^{\mathcal{A}}R(s)$ at $s = s_1$ is given by

$$M = \lim_{s \to s_1} (s - s_1)\,{}^{\mathcal{A}}R^*(s) = \lim_{s \to s_1} (s - s_1) W^{-1}(s) = C c_1 r_1 \Big/ \left(\frac{d}{ds} \det W(s)\right)_{s = s_1}. \tag{3.4}$$

But differentiating Equation (3.3) yields

$$I \frac{d}{ds} \det W(s) = W'(s)\tilde{W}(s) + W(s)\tilde{W}'(s).$$

If we multiply this by r_1 on the left and c_1 on the right, substitute $s = s_1$, and make use of Equations (3.1) and (3.2), we get

$$\left(\frac{d}{ds} \det W(s)\right)_{s = s_1} = r_1 W'(s_1) C c_1.$$

Substitution of this into Equation (3.4) gives the required result.

This result is related also to the fact that, if $\eta_1(s)$ is an eigenvalue of $W(s)$, then $\eta_1'(s_1) = r_1 W'(s_1) c_1$, see Lancaster (1969), p. 224.

Note that the result is independent of the scaling of the eigenvectors.

If Q is reversible we can obtain an improved approximation for ${}^{\mathscr{A}}R(t)$, as follows.

Theorem 3.2. If Q is irreducible and reversible and the $k_{\mathscr{A}}$ roots of $\det W(s) = 0$ are distinct, then, as $t \to \infty$,

$$ {}^{\mathscr{A}}R(t) \sim \sum_{i=1}^{k_{\mathscr{A}}} \exp(s_i t) c_i r_i / r_i W'(s_i) c_i $$

where c_i, r_i are the right (column) and left (row) eigenvectors of $H(s_i)$ corresponding to eigenvalue s_i.

Proof. If Q is irreducible and reversible, then Theorem 2.2 guarantees that there are exactly $k_{\mathscr{A}}$ real roots s_i of $\det W(s) = 0$. If these are all distinct they will all correspond to simple poles and the residues can be found in the same manner as for s_1, as outlined in Theorem 3.1. As s_1 is the largest root, the expression given above is asymptotically equivalent to that in Theorem 2.1: this theorem is therefore trivially true.

However, our claim that the result given in Theorem 3.2 is a much better approximation to ${}^{\mathscr{A}}R(t)$ than the simple form of Theorem 3.1 is based on the fact that under certain conditions ${}^{\mathscr{A}}R(t)$ is equal to the sum of the residues of all the poles of $\exp(st)\,{}^{\mathscr{A}}R^*(s)$, see Smith (1966), Chapter 4. The result given here contains just the real ones.

Jalali and Hawkes (1992) show that this is certainly the case for the two-state model with just one $\mathscr{A}$-state and one $\mathscr{F}$-state, and that there are infinitely many complex-conjugate pairs of poles, apart from the single real root s_1. This leads to a series of damped oscillations in addition to the asymptotic exponential so that, in the two-state case,

$$ {}^{\mathscr{A}}R(t) = w_1 \exp(s_1 t) + 2 \sum_{n=2}^{\infty} c_n \exp(\sigma_n t) \cos(\omega_n t + \vartheta_n), $$

where $s_n = \sigma_n + i\omega_n$.

We conjecture that the same form holds for the general reversible case but with $k_{\mathscr{A}}$ real exponentials, corresponding to the form in Theorem 3.2, in addition to infinitely many complex ones. The result, however, is rather more difficult to prove and the complex roots more difficult to find. We have found that in practice, in numerical examples, Theorem 3.2 does indeed give us an improved approximation to ${}^{\mathscr{A}}R(t)$, compared to Theorem 3.1, over a wide range of t.

Finding the roots and eigenvectors. In order to implement the above results in practice we need to be able to find the real roots and the corresponding eigenvectors. We have used two methods which, between them, have proved satisfactory.

Newton–Raphson. If $F(x)$ is a vector of n functions $F^{\mathrm{T}}(x) = (f_1(x), f_2(x), \cdots, f_n(x))$, where x is itself a vector of n variables $x^{\mathrm{T}} = (x_1, x_2, \cdots, x_n)$, then we have the iteration $x_{r+1} = x_r - F'(x_r)^{-1}F(x_r)$ for solving the set of equations $F(x) = 0$, where the Jacobian is $F'(x) = (\partial f_i/\partial x_j)$. In our case we take $x^{\mathrm{T}} = (c^{\mathrm{T}}, s)$ and let $F(x) = 0$ be the set of equations $W(s)c = 0$ together with the normalisation $u^{\mathrm{T}}c - 1 = 0$, where u is a vector of ones. Then

$$F'(s) = \begin{pmatrix} W(s) & W'(s)c \\ u^{\mathrm{T}} & 0 \end{pmatrix}$$

and, after some algebra, we get

$$c_{r+1} = (u^{\mathrm{T}}W^{-1}(s_r)W'(s_r)c_r)^{-1}W^{-1}(s_r)W'(s_r)c_r$$
$$s_{r+1} = s_r - (u^{\mathrm{T}}W^{-1}(s_r)W'(s_r)c_r)^{-1}.$$

Once a root s has been found, the left eigenvalue can be found as a solution to $rW(s) = 0$, $ru = 1$. This is just like the equations for finding an equilibrium vector of a Q-matrix and can be solved in a similar manner, see for example Hawkes and Sykes (1990).

As usual, the Newton–Raphson method either works very well or it fails, for instance if the required solution is a repulsion point. If that happens we use instead the following.

Bisection method. One can make a rough plot of $\det W(s)$ as a function of real s and identify the roots approximately. Each of these can then be located precisely by a simple bisection method. Once an s has been found the left eigenvector can be found as above; the right eigenvector can be found in the same way, as its transpose satisfies $c^{\mathrm{T}}W^{\mathrm{T}}(s) = 0$, $c^{\mathrm{T}}u = 1$.

4. Numerical example

For illustration we apply the foregoing theory to the following example first introduced by Magleby and Pallota (1983) as a model of a calcium-activated potassium channel, and further considered by Blatz and Magleby (1986), Crouzy

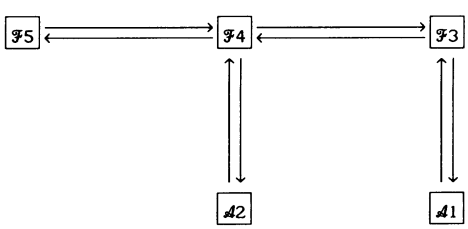


Figure 4.1. State diagram of a model of a calcium-activated potassium channel with two $\mathscr{A}$ states and three $\mathscr{F}$ states, corresponding to a channel being open or shut. The Q-matrix of transition rates is given in Equation (4.1)

and Sigworth (1990) and Hawkes et al. (1990). There are five states, of which $\mathscr{A} = (1, 2)$ correspond to the channel being open and the remaining three states, $\mathscr{F} = (3, 4, 5)$, to the channel being closed. The model is illustrated in Figure 4.1. The deadtime is $\tau = 0{\cdot}15$ ms and the Q-matrix is

$$Q = \left(\begin{array}{ccc:cc} -0{\cdot}322 & 0 & 0{\cdot}322 & 0 & 0 \\ 0 & -2{\cdot}86 & 0 & 2{\cdot}86 & 0 \\ \hdashline 3{\cdot}95 & 0 & -4{\cdot}55 & 0{\cdot}6 & 0 \\ 0 & 0{\cdot}12 & 0{\cdot}285 & -0{\cdot}585 & 0{\cdot}18 \\ 0 & 0 & 0 & 0{\cdot}034 & -0{\cdot}034 \end{array}\right) \tag{4.1}$$

partitioned into the two sets. The rates are given as events per ms.

We have expressed the roots s_i as time constants $\mu_i = -1/s_i$. Then the asymptotic p.d.f. of excess $e\mathscr{A}$-occupancies can be calculated approximately from Equation (1.3) and Theorem 3.2 as

$$f_U(t) \approx \sum_i (a_i/\mu_i) \exp(-t/\mu_i). \tag{4.2}$$

a_i can be thought of as the 'area' contributed by the component with time constant μ_i, although it is not necessarily positive. Results for excess $\mathscr{F}$-occupancies are obtained by interchanging the roles of $\mathscr{A}$ and $\mathscr{F}$. The time constants and areas are given in Table 4.1. Note that we are interested in the asymptotic behaviour of the probability densities, not in an approximation to the whole distribution. There is, therefore, no reason why the areas should sum to 1 although in this example, with the deadtime relatively small compared to the mean occupancy times, they very nearly do. It is interesting to observe that in this example the rate constants come out very close to those obtained in approximations by Blatz and Magleby and also those of Crouzy and Sigworth (1990), although the areas are somewhat different.

The single exponential asymptotic p.d.f. and the double exponential asymptotic p.d.f. of excess $e\mathscr{A}$-occupancies are shown in Figure 4.2. The latter is visually indistinguishable from the exact p.d.f. as calculated by the method described in Section 1. Compared to the exact p.d.f., this asymptotic p.d.f. is in error by only 0·72% at $t = 0$ and by less than 0·02% for all $t > \tau = 0{\cdot}15$. The single exponential

TABLE 4.1

Time constants μ_i and areas a_i for asymptotic probability densities of excess $e\mathscr{A}$-occupancies and excess $e\mathscr{F}$-occupancies, as expressed in Equation (4.2), for the model whose Q-matrix is given by Equation (4.1) and with deadtime $\tau = 0{\cdot}15$ ms.

	mean (ms)	area	mean (ms)	area	mean (ms)	area
$e\mathscr{A}$-occupancies	5·4961	0·9322	0·3573	0·0676		
$e\mathscr{F}$-occupancies	46·9198	0·1135	1·9607	0·1988	0·2308	0·6849

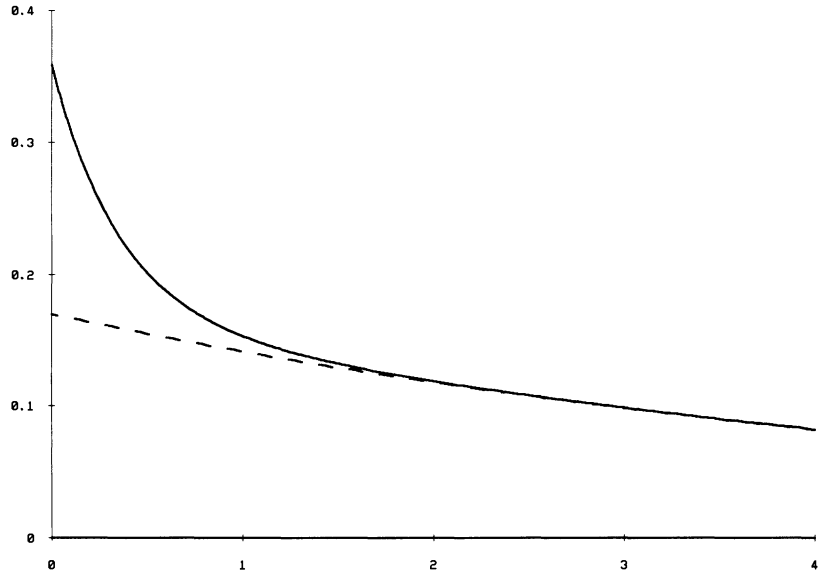


Figure 4.2. Probability density of excess $e\mathcal{A}$-occupancies above $\tau = 0{\cdot}15$ ms for the model with transition rates given in Equation (4.1). The second asymptotic p.d.f. with two exponential components is shown as (——) and the asymptotic single exponential dashed (– – –). The second asymptotic p.d.f. with two exponential components is visually indistinguishable from the exact p.d.f. calculated by the methods of Section 1

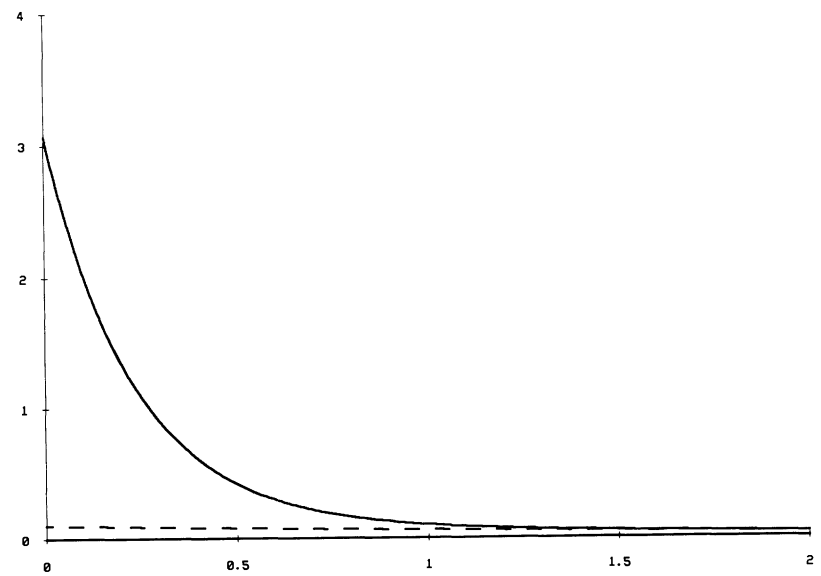


Figure 4.3. Probability density of excess $e\mathcal{F}$-occupancies above $\tau = 0{\cdot}15$ ms for the model with transition rates given in Equation (4.1). The asymptotic p.d.f. with three exponential components, shown as (——), is indistinguishable by eye from the exact p.d.f. The asymptotic p.d.f. with only two components, given by the two largest time constants in Table 4.1, is well below the exact curve up to $t = 2$, after which it is accurate to within 2%: it is shown dashed (– – –)

component, corresponding to the largest real root s_1, is too low initially, but is in error by less than 0·6% for all $t > 2$.

The double exponential asymptotic p.d.f. and the triple exponential asymptotic p.d.f. of excess $e\mathcal{F}$-occupancies are shown in Figure 4.3. The latter is visually indistinguishable from the exact p.d.f. as calculated by the method described in Section 1. Compared to the exact p.d.f. this asymptotic p.d.f. is in error by 2·09% at $t = 0$ and by less than 0·008% for all $t > \tau = 0{\cdot}15$. The double exponential asymptotic p.d.f. using the largest two real roots, is much too low initially, but is in error by less than 2% for all $t > 2$. The single exponential component, corresponding to the largest real root s_1, does not become effective in this case until way out in the tail of the distribution.

This example shows that, if τ is small compared to the mean occupancy times, the asymptotic p.d.f. may be very close to the exact p.d.f. for all t. For larger values of τ this is not usually the case, but the asymptotic p.d.f. is very accurate for all t greater than two or three multiples of τ in all realistic cases we have looked at. Our recommended procedure, therefore, is to compute the exact p.d.f. by the method of Section 1 for two or three multiples of τ, in which range it is quite easy to calculate (and numerically stable). If the asymptotic p.d.f. agrees closely with this at the end of this range, then use the asymptotic form for all larger t.

Further numerical examples are given in Hawkes et al. (1992).

5. Non-constant deadtimes

Ball and Sansom (1988) considered a more general situation in which the critical deadtime τ, instead of being constant, is random. A sojourn in the $\mathcal{F}$ states is not detected if its duration is less than $\tau_{\mathcal{F}}$, a random variable distributed independently of the sojourn time and of the critical deadtime corresponding to any other sojourn. If this has a probability density whose Laplace transform is denoted by ${}^{\mathcal{F}}f_{\tau}^{*}(s)$, then the basic theory outlined in Section 1 remains except that we replace Equation (1.1) by a more general result which is better expressed in terms of the Laplace transform

$$ {}^{e}G_{\mathcal{A}\mathcal{F}}^{*}(s) = {}^{\mathcal{A}}R^{*}(s)Q_{\mathcal{A}\mathcal{F}}\,{}^{\mathcal{F}}f_{\tau}^{*}(sI - Q_{\mathcal{F}\mathcal{F}}) $$

where, to generalise Equation (1.5),

$$ {}^{\mathcal{A}}R^{*}(s) = [sI - Q_{\mathcal{A}\mathcal{A}} - Q_{\mathcal{A}\mathcal{F}}(sI - Q_{\mathcal{F}\mathcal{F}})^{-1}\{I - {}^{\mathcal{F}}f_{\tau}^{*}(sI - Q_{\mathcal{F}\mathcal{F}})\}Q_{\mathcal{F}\mathcal{A}}]^{-1}. $$

This is equivalent to Theorem 3.1 of Ball and Sansom (1988). The previous expressions, corresponding to constant τ, are obtained if we set the transform ${}^{\mathcal{F}}f_{\tau}^{*}(s)$ equal to $\exp(-s\tau)$. Thus, as before, we are looking for the roots of $\det W(s) = 0$, where $W(s) = sI - H(s)$, but now Equations (2.1) and (2.1a) are replaced by the

two alternative forms

$$H(s) = Q_{\mathcal{A}\mathcal{A}} + Q_{\mathcal{A}\mathcal{F}}\left(\int_0^\infty \exp\left(-(sI - Q_{\mathcal{F}\mathcal{F}})t\right)(1 - {}^{\mathcal{F}}F_\tau(t))\,dt\right)Q_{\mathcal{F}\mathcal{A}}, \tag{5.1}$$

$$H(s) = Q_{\mathcal{A}\mathcal{A}} + Q_{\mathcal{A}\mathcal{F}}(sI - Q_{\mathcal{F}\mathcal{F}})^{-1}\{I - {}^{\mathcal{F}}f_\tau^*(sI - Q_{\mathcal{F}\mathcal{F}})\}Q_{\mathcal{F}\mathcal{A}}, \tag{5.1a}$$

where ${}^{\mathcal{F}}F_\tau(t)$ is the cumulative distribution of the critical $\mathcal{F}$-deadtimes.

What we have to be careful about is the fact that the integral in Equation (5.1) may not converge for all values of s, although the alternative form given by Equation (5.1a) may do so. However, if we restrict ourselves to values of s with sufficiently large real part for the integral to converge, the main arguments used in Theorem 2.1 still hold and a modified result can be stated as follows.

Theorem 5.1. If $H(s)$ is irreducible, $\det W(s) = 0$ has always a simple real root s_1 which is greater than the real part of any other root. This is the largest point of intersection of $\lambda_1(s)$, defined as the largest real root of $H(s)$ for real s, with the diagonal line s through the origin. Corresponding to this there are unique left and right p-eigenvectors which are non-negative.

Implicit in the statement of this theorem is the possibility that $\lambda_1(s)$ may intersect the diagonal line s more than once. This possibility is confirmed in the following example.

Example. Consider a two-state model with $\mathcal{A}$ and $\mathcal{F}$ containing just one state each. Let $Q_{\mathcal{A}\mathcal{F}} = -Q_{\mathcal{A}\mathcal{A}} = \alpha$, $Q_{\mathcal{F}\mathcal{A}} = -Q_{\mathcal{F}\mathcal{F}} = \beta$ and suppose that $\tau_{\mathcal{F}}$ has a negative exponential distribution with rate parameter γ, so that ${}^{\mathcal{F}}f_\tau^*(s) = \gamma/(s+\gamma)$. Now the integral in Equation (5.1a) has the form $\int_0^\infty \exp(-(s+\beta+\gamma)t)\,dt$, which is convergent if and only if $\operatorname{Re} s > -(\alpha+\beta)$. However, the form given by Equation (5.1a) may still be written down as $H(s) = -\alpha + \alpha\beta/(s+\beta+\gamma) = \lambda_1(s)$. For $s > -(\beta+\gamma)$ this is a decreasing function which tends to $+\infty$ as $s \to -(\beta+\gamma)$ from the right. For $s < -(\beta+\gamma)$ it is again decreasing, but this time from $-\alpha$ to $-\infty$ as s goes from $-\infty$ to $-(\beta+\gamma)$: this part, therefore, has a second intersection with the diagonal line s, see Figure 5.1. This is because the equation $W(s) = s - H(s) = 0$ is equivalent to the quadratic $s^2 + (\alpha+\beta+\gamma)s + \alpha\gamma = 0$, which clearly has two negative roots.

The above problems can be avoided if $W(s)$ has no poles, i.e. if the integral converges for the whole s-plane. Thus we shall assume that ${}^{\mathcal{F}}f_\tau^*(s)$ is an entire function, which means that the p.d.f. of $\tau_{\mathcal{F}}$ should fall off faster than any exponential. This would happen if the p.d.f. had finite support, which is quite natural to assume in the physical context postulated, as there must surely be an upper bound to the size of an interval which may be missed. We now state a generalisation of Theorem 2.2.

Theorem 5.2. If the matrix Q is reversible, and if ${}^{\mathcal{F}}f_\tau^*(s)$ is entire, then the real curves $\lambda_i(s)$, for $i = 1$ to $k_{\mathcal{A}}$ and real s, are all decreasing functions tending to ρ_i as

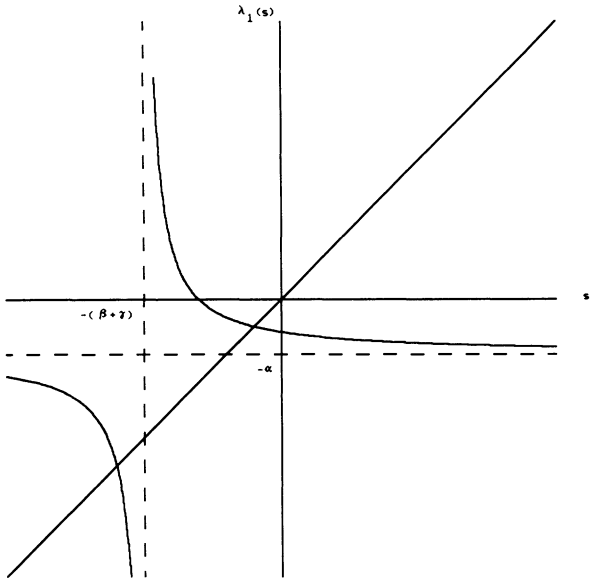


Figure 5.1. Graph of $\lambda_1(s) = -\alpha[1 - \beta/(s + \beta + \gamma)]$ for two-state model with transition rates rates α, β and exponentially distributed deadtimes with mean $1/\gamma$. There are two roots, one each side of the asymptote at $s = -(\beta + \gamma)$

$s \to \infty$. Hence each has a unique intersection with the diagonal line s and thus $\det W(s) = 0$ has exactly $k_{\mathscr{A}}$ real roots, allowing for possible multiplicity.

The proof follows exactly that of the main part of Theorem 2.2.

In the proof of Theorem 2.2 we argued that, for different constant deadtimes $\tau_1 < \tau_2$, $\lambda_i(s, \tau_1) \leqq \lambda_i(s, \tau_2)$. Here the deadtimes are random, but essentially the same argument leads to the following theorem.

Theorem 5.3. If $\tau_{1\mathscr{F}}$ is stochastically less than $\tau_{2\mathscr{F}}$, i.e. the c.d.f. ${}^{\mathscr{F}}F_{\tau_1}(t) \geqq {}^{\mathscr{F}}F_{\tau_2}(t)$ for all t, then $\lambda_i(s, \tau_{1\mathscr{F}}) \leqq \lambda_i(s, \tau_{2\mathscr{F}})$ for $i = 1$. If Q is reversible and ${}^{\mathscr{F}}f^*_{\tau_1}(s)$ and ${}^{\mathscr{F}}f^*_{\tau_2}(s)$ are both entire, the result holds for all $i = 1, \cdots, k_{\mathscr{A}}$.

Asymptotic densities. Under the conditions of Theorem 5.1 or Theorem 5.2, the results of Section 3 remain valid, so that ${}^{\mathscr{A}}R(t)$ and $f_{\bar{T}}(t)$ can be represented asymptotically by one or more exponential terms.

Another example. Consider a general model with Q-matrix partitioned as in Section 1 and suppose that the critical deadtimes are exponentially distributed with parameters $\gamma_{\mathscr{A}}$ and $\gamma_{\mathscr{F}}$, respectively, for omitted $\mathscr{A}$-intervals and omitted $\mathscr{F}$-intervals. Then the process with omitted intervals is equivalent to a continuous-time Markov chain with an extra set of $k_{\mathscr{F}}$ $\mathscr{A}$-states which are really $\mathscr{F}$-states which have not been recognised as such (denote this set by $\mathscr{A}^{\mathrm{P}}$ for pseudo), and an extra set of $k_{\mathscr{A}}$ $\mathscr{F}$-states which are really $\mathscr{A}$-states which have not been recognised as such

(denote the set by $\mathscr{F}^{\mathrm{P}}$). Then the modified sets are $\mathscr{A}' = \mathscr{A} \cup \mathscr{A}^{\mathrm{P}}$ and $\mathscr{F}' = \mathscr{F} \cup \mathscr{F}^{\mathrm{P}}$ and the modified Q-matrix is

$$Q' = \begin{array}{c} \\ \mathscr{A}' \\ \mathscr{F}' \end{array} \begin{array}{c} \begin{array}{cc} \mathscr{A}' & \mathscr{F}' \end{array} \\ \left(\begin{array}{cc:cc} Q_{\mathscr{A}\mathscr{A}} & Q_{\mathscr{A}\mathscr{F}} & 0 & 0 \\ Q_{\mathscr{F}\mathscr{A}} & Q_{\mathscr{F}\mathscr{F}} - \gamma_{\mathscr{F}} I & 0 & \gamma_{\mathscr{F}} I \\ \hdashline \gamma_{\mathscr{A}} I & 0 & Q_{\mathscr{A}\mathscr{A}} - \gamma_{\mathscr{A}} I & Q_{\mathscr{A}\mathscr{F}} \\ 0 & 0 & Q_{\mathscr{F}\mathscr{A}} & Q_{\mathscr{F}\mathscr{F}} \end{array}\right) \end{array},$$

see Ball (1990). Note that the new process is not reversible even if the original one is.

Then the standard theory of Colquhoun and Hawkes (1982) implies that

$$^{\mathrm{e}}G_{\mathscr{A}\mathscr{F}}(t) = [\exp(Q'_{\mathscr{A}'\mathscr{A}'}t)]_{\mathscr{A}\mathscr{A}^{\mathrm{P}}}\gamma_{\mathscr{F}},$$

i.e. the $\mathscr{A}\mathscr{A}^{\mathrm{P}}$ partition of $\exp(Q'_{\mathscr{A}'\mathscr{A}'}t)$ where

$$Q'_{\mathscr{A}'\mathscr{A}'} = \begin{pmatrix} Q_{\mathscr{A}\mathscr{A}} & Q_{\mathscr{A}\mathscr{F}} \\ Q_{\mathscr{F}\mathscr{A}} & Q_{\mathscr{F}\mathscr{F}} - \gamma_{\mathscr{F}} I \end{pmatrix}.$$

Hence it is a linear combination of k negative exponential functions corresponding to the eigenvalues of $Q'_{\mathscr{A}'\mathscr{A}'}$, which are all real and negative if Q is reversible. The previous example in this section is a special case of this model with $k_{\mathscr{A}} = k_{\mathscr{F}} = 1$, which explains the two roots obtained in that case.

This example illustrates that there may be more than $k_{\mathscr{A}}$ real exponential components in the probability density of apparent $\mathscr{A}$-sojourns if the assumption of entire functions does not hold.

References

BALL, F. G. AND SANSOM, M. S. P. (1988) Aggregated Markov processes incorporating time interval omission. *Adv. Appl. Prob.* **20,** 546–572.

BALL, F. G. AND 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.

BALL, F. G. (1990) Aggregated Markov processes with negative exponential time interval omission. *Adv. Appl. Prob.* **22,** 802–830.

BELLMAN, R. (1970) *Introduction to Matrix Analysis,* 2nd edn, McGraw-Hill, New York.

BLATZ, A. L. AND MAGLEBY, K. L. (1986) Correcting single channel data for missed events. *Biophys. J.* **49,** 967–980.

COLQUHOUN, D. AND 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.

COLQUHOUN, D. AND SIGWORTH, F. J. (1983) Fitting and statistical analysis of single channel records. In *Single Channel Recording,* ed. B. Sakmann and E. Neher, Plenum, New York, 191–263.

COX, D. R. AND MILLER, H. D. (1965) *The Theory of Stochastic Processes.* Methuen, London.

CROUZY, S. C. AND SIGWORTH, F. J. (1990) Yet another approach to the dwell time omission problem of single channel analysis. *Biophys. J.* **58,** 731–743.

FREDKIN, D. R., MONTAL, M. AND 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 and R. A. Ohlsen, Belmont, Wadsworth, 269–289.

HAWKES, A. G., JALALI, A. AND COLQUHOUN, D. (1990) The distributions of the apparent open times and shut times in a single channel record when brief events can not be detected. *Phil. Trans R. Soc. Lond.* A **332,** 511–538.

HAWKES, A. G., JALALI, A. AND COLQUHOUN, D. (1992) Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil. Trans. R. Soc. London B.* To appear.

HAWKES, A. G. AND SYKES, A. M. (1990) Equilibrium distributions of finite-state Markov processes. *IEEE Trans. Reliability* **39,** 592–595.

JALALI, A. AND HAWKES, A. G. (1992) The distribution of apparent occupancy times in a two-state Markov process in which brief events can not be detected. *Adv. Appl. Prob.* **24,** 288–301.

LANCASTER, P. (1969) *Theory of Matrices.* Academic Press, New York.

ROUX, B. AND SAUVÉ, R. (1985) A general solution of the time interval omission problem applied to single channel analysis. *Biophys. J.* **48,** 149–158.

MAGLEBY, K. S. AND PALLOTA, B. L. (1983) Calcium dependence of open and shut interval distributions from calcium-activated potassium channels in cultured rat muscle. *J. Physiol. (Lond.)* **344,** 585–604.

MORRISON, D. F. (1967) *Multivariate Statistical Methods.* McGraw-Hill, New York.

OSTROWSKI, A. M. (1964) Positive matrices and functional analysis. In *Recent Advances in Matrix Theory,* ed. H. Schneider, University of Wisconsin Press, Madison, 125–138.

SENETA, E. (1973) *Non-negative Matrices.* George Allen and Unwin, London.

SMITH, M. G. (1966) *Laplace Transform Theory.* Van Nostrand, London.

TAUSSKY, O. (1964) On the variation of the characteristic roots of a finite matrix under various changes of its elements. In *Recent Advances in Matrix Theory,* ed. H. Schneider, University of Wisconsin Press, Madison, 125–138.