Model properties underlying non-identifiability in single channel inference

R. O. EDESON¹, F. G. BALL², G. F. YEO³, R. K. MILNE¹ and S. S. DAVIES²

¹ Department of Anaesthesia, Sir Charles Gairdner Hospital, Nedlands 6009, Western Australia, Australia
² Department of Mathematics, University of Nottingham, Nottingham, NG7 2RD, U.K.
³ School of Mathematical and Physical Sciences, Murdoch University, Murdoch 6150, Western Australia, Australia
⁴ Department of Mathematics, University of Western Australia, Nedlands 6009, Western Australia, Australia

SUMMARY
Models of ion channel kinetics subserve inferential methods applied to patch clamp data. For Markov models the density function of a sojourn time in a class of states is a mixture of exponentials. Determination of kinetic parameters from density functions may be complicated by non-uniqueness of solutions. This non-identifiability is investigated analytically for a class of two states, assuming detailed balance; relations between model properties, observable density parameters, and non-uniqueness are presented. The results are further developed in terms of similarity transform methods. Additional information provided by joint distributions is discussed. An example is given where identifiability of a model can be demonstrated explicitly. Attention is drawn to instances where the number of components in a density function may be misleading when used to infer the number of underlying states.

1. INTRODUCTION
The electrophysiological technique of patch clamp recording allows direct observation of transitions between conductance states in a single ion channel. Recorded sojourn times in experimentally distinguishable states are used to make inferences about underlying kinetic states, their connectivity, and the chemical rate constants which govern transitions between them. Because the observed channel behaviour is apparently random, methods of inference are widely based upon stochastic process (usually Markov) models of the ion channel kinetics. Thus it is common practice to fit sojourn time data with a (marginal) probability density function of a form implied by a proposed model. The parameters of the fitted density are then used in some way to characterize the model, for example by providing estimates of kinetic constants. In some instances, such as the simple linear four-state model of nicotinic receptor activation, this can be done explicitly and uniquely (Edeson et al. 1990a, equation 78), but there are many other cases in which multiple sets of model parameter values give rise to the same marginal density function, or even the same joint density of open and closed sojourn times. It then follows that possible non-uniqueness of solutions must be considered in inference based on these distributions.

We examine here some of the model structural properties which underlie the above non-identifiability. The ideas are illustrated by reference to a widely studied scheme of receptor kinetics involving two open and three closed states (Colquhoun & Hawkes 1982), see scheme 1 below; we use the term scheme to define a set of discrete states together with allowable transitions between them, and their respective conductances). A Markov model based on this five-state scheme is specified by a set of transition rates. We consider what information about these transition rates is given by the density of sojourn times in the open class, together with detailed balance. The solution for transition rates, parametrized by the entry probability vector, is studied in some detail, and conditions for uniqueness obtained. An analysis based on a similarity transform (Kienker 1989) for the open class is then presented, along with properties of the similarity matrix which entail positive transition rates and detailed balance. We show that a Markov model based on scheme 1 is identifiable, in the sense that transition rates can be uniquely determined, given the joint distribution of open and closed sojourn times.

2. MARKOV MODELS
In the kinetic scheme

\[
\begin{align*}
\text{(scheme 1)} & \quad 5 \rightarrow 4 \leftrightarrow 3 \\
& \quad 1 \leftarrow 2
\end{align*}
\]

\(\mathcal{A} = \{1, 2\}\) is a class of open states and \(\mathcal{B} = \{3, 4, 5\}\) a class of closed states. Much of this paper is concerned with sojourn times in the open class, for which purpose scheme 1 may be considered a special case of

\[
\begin{align*}
\text{(scheme 2)} & \quad 1 \rightarrow 2 \leftrightarrow \\
& \quad 1 \leftrightarrow 2
\end{align*}
\]

where the connections between closed states (\(\kappa_g\) in number) are not specified. Transitions between states are assumed to follow a continuous-time homogeneous
Markov chain in equilibrium. This is represented by a transition rate matrix \( Q = [q_{ij}] \) where for \( i \neq j \) the transition rate \( q_{ij} \) from state \( i \) to state \( j \) is positive if and only if the connection \( i \to j \) is present in the scheme, and \( q_{ii} = -\sum_{j \neq i} q_{ij} \). Further, let \( d_i = -q_{ii} \), so \( d_i > 0 \) is the finite and positive mean sojourn time in state \( i \). As usual, \( Q \) is partitioned into submatrices

\[
Q = \begin{bmatrix}
Q_{dd} & Q_{ds} \\
Q_{sd} & Q_{ss}
\end{bmatrix},
\]

(1)

where, for example,

\[
Q_{dd} = \begin{bmatrix}
-d_1 & q_{12} \\
q_{21} & -d_2
\end{bmatrix}
\]

(2)

contains transition rates within the open class, and so on.

3. SIMILARITY TRANSFORMATIONS

Experimentally it is possible to observe only whether the channel is open or closed, rather than the precise state it is in. Kienker (1989) has discussed the occurrence of distinct Markov models of the underlying channel gating process having probabilistically indistinguishable observable processes. Let \( Q = [q_{ij}] \) and \( Q' = [q'_{ij}] \) be the transition rate matrices of two processes each based on scheme 2. Then, subject to mild conditions discussed below, Kienker (1989) has shown that the two processes are equivalent (that is, have probabilistically indistinguishable observable processes) if and only if there exists a nonsingular block-diagonal matrix \( S \)

\[
S = \begin{bmatrix}
S_{dd} & 0 \\
0 & S_{ss}
\end{bmatrix},
\]

where \( S \) is partitioned as in (1) and 0 denotes a matrix with all elements zero, such that

\[
Q' = S^{-1}QS.
\]

(3)

and \( S \) has all row sums equal to one, that is, \( S1 = 1 \), where 1 denotes a column vector of ones whose dimension (5 for scheme 1) is apparent from the context. Note that there is no requirement that all elements of \( S \) be non-negative; furthermore equation (3) can also be expressed as

\[
SQ' = QS.
\]

(4)

Let \( p = (p_1, p_2, \ldots, p_n)^T \) (where \( T \) denotes transpose and \( n \) the total number of states), where \( p_i \) is the equilibrium (steady state) probability that the process with transition rate matrix \( Q \) is in the state \( i \), and define \( p' \) similarly. Now \( p \) satisfies \( p^TQ = 0 \), so it follows from equation (4) and uniqueness of equilibrium distributions that \((p')^T = p^TS\).

In what follows we refer to the process with transition rate matrix \( Q \) as the original process and that with \( Q' \) as the transformed process. Suppose the two have probabilistically indistinguishable observable processes. Then, expanding (4) in partitioned form gives

\[
S_{dd}Q_{dd}' = Q_{dd}S_{dd}, \quad S_{dd}Q_{ds}' = Q_{ds}S_{dd}, \quad S_{ss}Q_{ss}' = Q_{ss}S_{ss}, \quad (5a, b)
\]

\[
S_{dd}Q_{sd}' = Q_{sd}S_{dd}, \quad S_{dd}Q_{ss}' = Q_{ss}S_{dd}, \quad S_{ss}Q_{sd}' = Q_{sd}S_{ss}, \quad (5c, d)
\]

The marginal open time density function for the original process is \((\text{Colquhoun & Hawkes 1977; Fredkin et al. 1985})\)

\[
f_o(t) = \pi_1^o e^{Q_{dd}t}(-Q_{dd}) 1 \quad (t > 0),
\]

where \( \pi_1^o = [\pi_1, \pi_s] \) gives the probabilities that a sojourn in the open class begins in state 1 or 2 (\( \pi_1 + \pi_s = 1 \)) and \exp \left( Q_{dd}t \right) = \sum \pi_i e^{Q_{dd}t} / k! \) is the usual matrix exponential (for example, see Bellman 1960, p. 165). Because \( S_{ss}1 = 1 \) and \( (\pi_1)^T = \pi_1^o S_{ss} \) (Kienker 1989, p. 294), and using the matrix decomposition \( cQ_{dd}' = S_{dd}' cQ_{dd} S_{dd} \),

\[
(7)
\]

the equality of the open time densities for the transformed and original processes can be shown by (see, for example, Kienker 1989)

\[
f_o(t) = \pi_1^o e^{Q_{dd}t}(-Q_{dd}) 1,
\]

\[
= \pi_1^o e^{Q_{ss}t}(-Q_{ss}) 1,
\]

\[
= f_o(t).
\]

(8)

The marginal open time density is completely specified by \( Q_{dd}' \) and \( \pi_o' \). In contrast, computation of the joint density of a sequence of open and closed times (see equation (29) below) requires all submatrices in the partition of \( Q \). Even if for two models \( Q_{dd}' \) and \( Q_{ss}' \) are identical (that is, \( S_{ss} = I \) under the similarity transformation of \( Q \) ), \( Q_{dd}' \) and \( Q_{ss}' \) must be related in each case in accordance with equations \((5b, c)\) for the joint densities to be identical. This last condition is not necessary for the marginal sojourn time densities to be identical. Thus it is possible for the transformed and original processes to have identical marginal open and closed time distributions, but have probabilistically distinguishable observable processes.

There are two rather mild regularity conditions required in the proof of Kienker's theorem. First, each of the submatrices \( Q_{dd}', Q_{ss}', Q_{dd} \), and \( Q_{ss} \) must have distinct eigenvalues; then it follows that \( f_o(t) \) (equation (6)) and the corresponding marginal closed time density function \( f_c(t) \) can be written as linear combinations of exponential densities, whose parameters are the eigenvalues of \( -Q_{dd}' \) and \( -Q_{ss}' \), respectively. Secondly, all the coefficients in the latter must be non-zero. These conditions will be satisfied in many practical situations but may be violated in particular cases, for example where certain symmetries are present.

4. MARGINAL DENSITY

Using the spectral expansion of \( Q_{dd}' \) the density function \((6)\) of a sojourn time in \( S \) can be written (Colquhoun & Hawkes 1982)

\[
f_o(t) = A_1 x_1 e^{-x_1t} + A_2 x_2 e^{-x_2t} \quad (t > 0),
\]

where \( x_1, x_2 \) are the eigenvalues (assumed distinct) of \( -Q_{dd}' \) and are given by

\[
x_1, x_2 = 0.5[d_1 + d_2 \pm \sqrt{(d_1 + d_2)^2 - 4d_1d_2}];
\]

(10)

the subscripts chosen so that \( x_1 < x_2 \). Here

\[
s = 1 - p_{12} p_{21},
\]

(11)

where \( p_{ij} = q_{ij} / d_i \) is the probability that a
transition from state $i$ to state $j$. In fact, $s$ is the success probability in the geometric distribution of the number of visits to one of the states in $\mathcal{A}$ during a sojourn in $\mathcal{A}$ which begins and ends in the other state. Edeson et al. (1990b) considered the use of the eigenvalues $\lambda_s$ as estimates of the reciprocal mean lifetimes $d_s$ and, by developing equation (10), the dependence of error in this approximation on the parameter $s$.

Let $L(\omega) = \int_0^\infty e^{-\omega t} f_\mathcal{A}(t) \, dt$ be the Laplace transform of the density function $f_\mathcal{A}(t)$. It follows from (9) that

$$L(\omega) = \sum_{i=1}^{2} A_i \lambda_s / (\omega + \lambda_s).$$

Thus $L(\omega)$ can also be expressed by a ratio of polynomials as

$$L(\omega) = \frac{u_0 + u_1 \omega}{u_0 + v_1 \omega + \omega^2},$$

where $u_0$, $u_1$, and $v_1$ are explicit functions of the $q_{ij}$. Equating coefficients of powers of $\omega$ in (12a) and (12b), and using (10) gives

$$u_0 = \pi_1 \lambda_s = d_1 d_2 s,$$

$$u_1 = A_1 \lambda_s + A_2 \lambda_s = \pi_1 q_{1s} + \pi_2 q_{2s},$$

$$v_1 = \pi_1 + \pi_2 = d_1 + d_2,$$

where $q_{1s} = d_1 - q_{12}$, $q_{2s} = d_2 - q_{31}$ (so for scheme 1, $q_{1s} = q_{12}$ and $q_{2s} = q_{23}$). The coefficients in (9) are readily obtained from (13b) as

$$A_1 = 1 - A_2 = [\lambda_s - \pi_1 q_{1s} - \pi_2 q_{2s}] / (\lambda_s - \lambda_s),$$

and are non-negative under detailed balance conditions (Kijima & Kijima 1987).

The three ‘observable’ parameters $A_1, \pi_1, \pi_2$ can be obtained from the marginal density function (9) of open times. However, our objective is to determine the elements of $Q_{\mathcal{A},\mathcal{A}'}$ of which there are four. For the purpose of studying possible solutions for the $q_{ij}$ it is convenient to consider $\pi_1$ as a fourth given parameter. Now from, for example, Ball & Rice (1989, equation 7), $\pi_1 = \rho_1 q_{11}/(\rho_1 q_{11} + \rho_2 q_{21})$ under detailed balance; also, $\rho_1 q_{12} = \rho_2 q_{21}$, so it follows immediately that

$$\pi_1 = q_{21} q_{11}/(q_{21} q_{11} + q_{23} q_{23})$$

(see Edeson et al. 1990a, equation 74b). In addition, because $\pi_1$ can also be expressed in terms of $Q_{\mathcal{A},\mathcal{A}'}$ and $Q_{\mathcal{A},\mathcal{A}'}$, $\pi_1$ is constant for all solutions $Q_{\mathcal{A},\mathcal{A}'}$ for which those two submatrices are fixed. Moreover, in some cases the value of $\pi_1$ may be inferred; for example, in scheme 1 with states 3 and 4 not adjacent, if marginal closed time distributions at two agonist concentrations are known ($q_{12}$ and $q_{34}$ being concentration-dependent). Note that equations (15) and (6) together imply that, for scheme 1 under detailed balance, the marginal open time density is fully determined by $Q_{\mathcal{A},\mathcal{A}'}$. Indeed, this is true for any time reversible model in which the open states form a communicating class. (A class of states is communicating if for every pair of states, $(i,j)$ say, in the class there exists a pathway joining $i$ to $j$ that remains within the class.)

Using equations (13a), (13b), (11) and (15) we can express the $q_{ij}$ in terms of $u_0, u_1, v_1, \pi_1$ and either $q_{1s}$ or

<table>
<thead>
<tr>
<th>$\pi_1$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$\pi_1$</th>
<th>$\pi_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1. Behaviour of $-Q_{\mathcal{A},\mathcal{A}'}$ and $-Q_{\mathcal{A},\mathcal{A}'}^*$. As functions of $A_1$ and $\pi_1$.
Figure 1. Solutions of $d_i$ as functions of $\pi_i$ for the open class (1, 2) of the Colquhoun & Hawkes (1982) five-state model of nicotinic receptor kinetics, where $d_i = q_{12} + q_{14}$ (see text) is the reciprocal mean lifetime of sojourns in state $i$ (monoligated open) and $\pi_i$ is the probability that an opening begins in state $i$. Transition rates $q_{12}$, $q_{14}$ were calculated, using equation (17), from the open time density given by (9) with $\omega_A = 500.654$, $\omega_2 = 3050.013$ and $A_1 = 1 - A_2 = 0.9276$. In the range $A < \pi_i < A_1$ there are two proper solutions for $d_i$, denoted $d_{i1}$ and $d_{i2}$, where ‘proper’ signifies that $q_{12}$ and $q_{14}$ are both positive. Values of $d_i$ as $\pi_i \to 0$ and as $\pi_i \to 1$ were found using equation (14) of Edeson et al. (1990b). The unit of time is seconds.

For a given density (9) (and therefore given values of $u_0$, $u_1$ and $n_1$ it can be shown (starting from (13b)) that there are restrictions on the value of $\pi_1$ for which a proper solution for $-Q_{1d}$ exists, where a proper solution is defined as one for which $q_{12} > 0$, $q_{21} > 0$, and $q_{12} = d_i < \infty$, $q_{21} < d_i < \infty$. Thus: (i) for $A_1 < \pi_1 < A_2$ there are no proper solutions; (ii) for $A_2 < \pi_1 < A_1$ there are two proper solutions (both $-Q_{1d}$ and $-Q_{2d}$); and (iii) for $\pi_1 > \max(A_1, A_2)$ or $\pi_1 < \min(A_1, A_2)$ there is only one proper solution.

These results are illustrated numerically in table 1 which presents $-Q_{1d}$ as a function of $A_1$ and $\pi_1$ for the case $\omega_1 = 1$, $\omega_2 = 19$ (arbitrary time units). It can be seen that both the number and type of solutions satisfy the above limits. Note the symmetry between $-Q_{1d}$ and $-Q_{2d}$ which corresponds to a simple relabelling of states. Let $R$ be a 2 x 2 matrix with diagonal elements 0 and off-diagonals 1; then $R^1 = R$ and $R^2 = I$. Pre- and post-multiplying $Q_{1d}$ by $R$ corresponds to such a relabelling of the two open states, $Q_{1d}^{(r)}(A_1, \pi_i) = RQ_{1d}(A_1, \pi_i) R$.

Now using the similarity transformation (adapted from (5a))

$$Q_{1d}^{(r)} = S_{1d}^{(r)}Q_{1d}^{(r)}S_{1d}^{-1},$$

we have

$$Q_{1d}^{(r)}(A_1, \pi_i) = S_{1d}^{(r)}Q_{1d}^{(r)}(A_1, \pi_i) S_{1d}^{-1},$$

$$= TQ_{1d}^{(r)}(A_1, \pi_i)T^{-1},$$

where $T = S_{1d}^{(r)}R$, $T^{-1} = R S_{1d}^{(r)}$. Other symmetries exist between diagonal elements in $-Q_{1d}$ and $-Q_{2d}^{(r)}$, and between those in $-Q_{1d}$ and $-Q_{2d}^{(r)}$. It is clear from the above that a particular set of kinetic parameter values may give rise to a density function from which, in turn, a second set of parameter values can be found yielding the same density, and also the same entry probability vector $n_i$. In the model (scheme 1) of nicotinic receptor kinetics studied by Colquhoun & Hawkes (1982) the calculated density of sojourn times in the open class, using their transition rate matrix

$$Q = \begin{bmatrix}
-3050 & 50 & 0 & 3000 & 0 \\
0.667 & -500.667 & 500 & 0 & 0 \\
0 & 15000 & -19000 & 4000 & 0 \\
15 & 0 & 50 & -2065 & 2000 \\
0 & 0 & 0 & 10 & -10
\end{bmatrix}$$

with unit of time second (where for exact detailed balance, 0.667 should be read as 2/3), is given by (9) with $\omega_1 = 500.654$, $\omega_2 = 3050.013$, and $A_1 = 0.9276$. Because $A_1 > A_2$ we expect two solutions for $-Q_{1d}$ for any value of $\pi_1$, satisfying $A_2 < \pi_1 < A_1$. In figure 1 values of $d_{i1}$, $d_{i2}$ are plotted against $\pi_1$ in the range for which (17) has a proper solution. At any value of $\pi_1$ for which two such solutions exist these afford plausible yet
significantly different alternatives. For the particular parameter values of Colquhoun & Hawkes (1982), \( \pi_1 = 0.0741 \) and the two solutions for \( -Q_{ab} \) are

\[
\begin{align*}
-Q_{a1} &= \begin{bmatrix}
3050.00 & -50.00 \\
-0.667 & 500.667
\end{bmatrix} \\
-Q_{a2} &= \begin{bmatrix}
1945.15 & -1823.44 \\
-875.25 & 1605.52
\end{bmatrix}
\end{align*}
\]

(19)

the first being that specified originally. These are related through (18), with \( S_{a1} = \pi_1^T S_{a2} \pi_1^T = [0.0741, 0.9259] \).

The five-state model, but with different parameter values, was also used by Ball & Sansom (1989) to test their method of parameter estimation based upon a likelihood algorithm. They used a range of agonist concentration \( (a) \) in producing simulated data from which they attempted to recover the assumed parameter values. For \( a = 10^{-4} \) and \( 10^{-5} \) m these assumed values predict an open time density from which, in turn, only one proper solution (the (+) one) to (17) exists. For \( a = 10^{-4} \) and \( 10^{-3} \) m, there are two such solutions. For example, when \( a = 10^{-4} \) m, the theoretical open time density is given by equation (9) with \( \tau_j = 1040.09, \tau_2 = 1977.00 \) and \( A_p = 0.941741 \), with \( \pi_1 = 0.13333 \) (where their states Oα and Oα2 are relabelled 1 and 2 respectively). Because \( A_2 < \pi_1 < A_1 \) we anticipate two proper solutions for \( -Q_{a1,a2} \), these being

\[
\begin{align*}
-Q_{a1,a2}^1 &= \begin{bmatrix}
19687.50 & -10000 \\
-153.85 & 1122.60
\end{bmatrix} \\
-Q_{a1,a2}^2 &= \begin{bmatrix}
7390.63 & -7148.44 \\
-10997.60 & 13419.47
\end{bmatrix}
\end{align*}
\]

(21)

where the first corresponds to the assumed parameter set.

5. PROPERTIES OF THE SIMILARITY MATRIX \( S_{a1,a2} \)

In the transformation (18) above, \( \pi_1 \) is not free to vary. We now relax this restriction and return to the more general transformation (5a). Consider a (proper) submatrix \( Q_{a1,a2} \) defined by equation (2) with \( d_1 > q_{12} > 0 \) and \( d_2 > q_{21} > 0 \), and the similarity matrix \( S_{a1,a2} \), where the first corresponds to the assumed parameter set.

Recall that \( (\pi')^T = \pi_2^T S_{a1,a2} \). Thus

\[
(\pi')^T = [\pi_1^T, \pi_2^T] = [(\pi_1 s_{11} + \pi_2 s_{21}), (\pi_1 s_{12} + \pi_2 s_{22})],
\]

(22)

which yields

\[
-r s_{21} < s_{22} < -r s_{11} + 1/\pi_2,
\]

(23)

where \( r = \pi_1/\pi_2 \), and (i) \( s_{11} > s_{22} \) implies \( s_{11} > 0, s_{22} < 1 \) and \( s_{21} > 0 \), and (ii) \( s_{11} < s_{22} \) implies \( s_{21} > 0, s_{11} < 1 \) and \( s_{22} > 0 \). In the special case that \( \pi_1 = \pi_2 \) we obtain from (22) that \( s_{21} = r s_{22} \).

The transition rates \( q'_{ij} \) for the transformed process can be expressed in terms of the \( q_{ij} \) and \( s_{ij} \) by developing (5a),

\[
\begin{align*}
q'_{12} &= s_{11}(s_{22} q_{12} - s_{21} d_1) + s_{12}(s_{22} d_1 - s_{12} q_{12}), \\
q'_{21} &= s_{21}(s_{11} q_{21} - s_{12} d_2) + s_{22}(s_{11} d_2 - s_{22} q_{21}),
\end{align*}
\]

(24a)

\[
\begin{align*}
q'_{11} &= s_{22} q_{12} - s_{21} d_1, \\
q'_{22} &= s_{11} q_{21} - s_{12} d_2.
\end{align*}
\]

(24b)

(24c)

(24d)

By examining in turn the cases \( s_{11} > s_{21} \) and \( s_{11} < s_{21} \), equations (24) can be used to obtain inequalities for which the \( q'_{ij} \) are positive. For example, first suppose that \( s_{11} > s_{21} \); then \( q'_{21} > 0 \) in (24b) requires

\[
s_{11}(s_{11} q_{21} - s_{12} d_2) + s_{21}(s_{11} d_2 - s_{22} q_{12}) > 0.
\]

Dividing by \( s_{11} \) gives a quadratic in \( s_{21}/s_{11} \) which, by use of (10), yields

\[
s_{11} > s_{21} < s_{11} d_2 - q_{12}.
\]

(25)

The equations of the lines defining such inequalities, and those of (23), are given in table 2 (labelled \( A \) to \( H \)), along with the direction of the inequality ensuring positivity of the \( q'_{ij} \) in each of the cases \( s_{11} > s_{21} \). These four pairs of parallel lines (one line of each pair through \( 0,0 \) and one through \( 1,1 \)) enclose two distinct, congruent (quadrilateral) regions within which the inequalities are satisfied, one in each of the half planes \( s_{11} > s_{21} \) and \( s_{11} < s_{21} \).

Assuming detailed balance in the transformed process we can express the entry probabilities in a form analogous to (15), obtaining \( \pi_1^T/\pi_2^T = q'_{12}/(q'_{12} q'_{22}) \). Substituting from (22) and (24) this simplifies after considerable algebra to

\[
s_{11}(1-s_{11})q_{21} + s_{21}(1-s_{21})q_{12} = 0.
\]

(26)

On completing the squares this can be written as

\[
\frac{(s_{11} - 0.5)^2}{a^2} + \frac{(s_{21} - 0.5)^2}{b^2} = 1,
\]

(27)

where \( a = 0.5 \sqrt{(1+q_{12}/q_{21})}, b = 0.5 \sqrt{(1+q_{21}/q_{12})} \). Thus equation (27) describes an ellipse having centre \((0.5,0.5)\) and passing through the points \((0,0), (1,0), (1,1)\) and \((0,1)\). For \( s_{12} > 0 \) (respectively \( q_{12} > q_{21} \)) the semimajor axis has length \( a \) (respectively \( b \)) and the eccentricity is \( \sqrt{1-q_{12}/q_{21}} \) (respectively \( \sqrt{1-q_{21}/q_{12}} \)). When \( q_{12} = q_{21} \) the curve is a circle of radius \( \sqrt{0.5} \).

The points of intersection of the lines \( A-H \) with the ellipse are given in table 2. Because there are only four distinct \( Q_{a1,a2} \)-dependent points \( P_{i1}, P_{i2}, P_{b1} \) and \( P_{b2} \),
Table 2. Properties of the similarity matrix $S_{d,d}$ for which the transformed submatrix $Q_{d,d}$ and entry probability vector $\pi'$ are proper.

(For example, if $t_{11} > t_{21}$ then $t_{21} < t_{11}(d_1 - \alpha_1)/q_{12}$, line E, see equation (25) and also figure 2.)

<table>
<thead>
<tr>
<th>line</th>
<th>$t_{11} &gt; t_{21}$</th>
<th>$t_{11} &lt; t_{21}$</th>
<th>intersection with ellipse (equation (27))</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$s_{21} = -s_{51}$</td>
<td>$&gt;$</td>
<td>$(0, 0)$ $(u_1 - q_{12}u_2, u_1 - q_{12}u_2)$</td>
</tr>
<tr>
<td>B</td>
<td>$s_{21} = -s_{51} - \frac{1}{\pi_2}$</td>
<td>$&lt;$</td>
<td>$(1, 1)$ $(q_{12}, q_{12})$</td>
</tr>
<tr>
<td>C</td>
<td>$s_{21} = s_{51} q_{12} - q_{12} q_{12} = q_{12}$</td>
<td>$&lt;$</td>
<td>$(0, 0)$ $(q_{12}, q_{12})$</td>
</tr>
<tr>
<td>D</td>
<td>$s_{21} = s_{51} q_{12} + q_{12} q_{12} = q_{12}$</td>
<td>$&lt;$</td>
<td>$(1, 1)$ $(u_1 - q_{12}u_2, u_1 - q_{12}u_2)$</td>
</tr>
<tr>
<td>E</td>
<td>$s_{21} = s_{51} q_{12} - q_{12} q_{12} = q_{12}$</td>
<td>$&lt;$</td>
<td>$(0, 0)$ $(q_{12}, q_{12})$</td>
</tr>
<tr>
<td>F</td>
<td>$s_{21} = s_{51} q_{12} + q_{12} q_{12} = q_{12}$</td>
<td>$&lt;$</td>
<td>$(1, 1)$ $(u_1 - q_{12}u_2, u_1 - q_{12}u_2)$</td>
</tr>
<tr>
<td>G</td>
<td>$s_{21} = s_{51} q_{12} - q_{12} q_{12} = q_{12}$</td>
<td>$&lt;$</td>
<td>$(0, 0)$ $(q_{12}, q_{12})$</td>
</tr>
<tr>
<td>H</td>
<td>$s_{21} = s_{51} q_{12} + q_{12} q_{12} = q_{12}$</td>
<td>$&lt;$</td>
<td>$(1, 1)$ $(u_1 - q_{12}u_2, u_1 - q_{12}u_2)$</td>
</tr>
</tbody>
</table>

\[ m = \frac{\alpha_2 u_1 - q_{22}(\alpha_2 - \alpha_1 + u_1)}{\alpha_2 u_1 - q_{12}(\alpha_2 - \alpha_1 + u_1)} \]

Note that $m < 0$ and $\epsilon_1 > \epsilon_2 \geq 0$.

These results are also illustrated in figure 2, in which a particular $S_{d,d}$ can be viewed as characterized by a point in the plane. The points $(0, 0)$ and $(1, 1)$ are excluded as $S_{d,d}$ is invertible; points $(1, 0)$ and $(0, 1)$ correspond to $S_{d,d} = I$ and $S_{d,d} = R$, respectively. Any two points (say, $(1)$ and $(2)$) on the ellipse connected by a diameter determine similarity matrices related by $S_{d,d}^{(1)} = S_{d,d}^{(2)} R$, that is, $s_{ij}^{(1)} = 1 - s_{ij}^{(2)}$. Given $Q_{d,d}^{(1)}$ and letting $Q_{d,d}^{(0)} = Q_{d,d}^{(1)} S_{d,d}^{(0)}$, $i = 1, 2$, we have $Q_{d,d}^{(2)} = R Q_{d,d}^{(1)} R$, with $\pi_{d,d}^{(0)} = \pi_{d,d}^{(1)} R$. Now let $Q_{d,d}^{(2)} = R Q_{d,d}^{(1)} R$ and $S_{d,d}^{(1)} Q_{d,d}^{(1)} = Q_{d,d}^{(2)} S_{d,d}$. Then $R Q_{d,d}^{(1)} = Q_{d,d}^{(2)} R$, where $R = S_{d,d}^{(1)} R S_{d,d}^{(1)}$. Note that similar symmetries hold for corresponding pairs of points within the congruent quadrilaterals without necessarily satisfying detailed balance.

For the transformed submatrix $Q_{d,d}$ to be proper and preserve detailed balance $S_{d,d}$ must lie on the ellipse and between the lines (equations (28)) labelled [1] and [2] in the figure. As is clear from (26), the form of the ellipse is dependent on $q_{12}$ and $q_{22}$ only; the lines [1] and [2], however, also depend on $q_{12}$ and $q_{22}$ and their behaviour in certain limiting cases is of interest. When $q_{12} > 0$, line [1] approaches $(0, 1)$ from above ($\epsilon_1 < 1$) and line [2] approaches $(1, 0)$ from below ($\epsilon_2 > m = (\alpha_1 - \alpha_2)/\alpha_2$). When $q_{22} > 0$, line [1]
approaches are suggested by the (1,0) from above \( (q_1 \rightarrow m = \alpha_3/(u_1 - x_1)) \) and line [2] approaches (0,1) from below \( (q_2 \rightarrow 1) \). When \( q_{1k} \rightarrow q_{22} > 0 \), \( c_2 \rightarrow 0 \) and line [1] approaches the tangent to the ellipse (1,1) having slope \( -q_{21}/q_{12} \).

Given a proper \( Q_{d.w} \), the ellipse and the lines \([1],[2]\) are, of course, determined. When \( S_{d.w} \) (strictly, \( (s_{11}, s_{21}) \)) lies on the ellipse between the lines \([1],[2]\) (the ‘proper region’), \( Q_{d.w} \) will be proper. It is now of interest to examine the behaviour of \( Q_{d.w} \) at the boundaries of this region. As \( (s_{11}-s_{21}) \rightarrow P_{1w} \), \( q_{1w} \rightarrow 0 \); as \( (s_{11}+s_{21}) \rightarrow P_{2w} \), \( q_{2w} \rightarrow 0 \); and as \( (s_{11}, s_{21}) \rightarrow P_{vw} \) or \( P_{wv} \), \( q_{12} \) and \( q_{21} \rightarrow 0 \).

The intersection of the ellipse and the line \( s_{21} = r(1-s_{11}) \) (see after equation (23)) signifies values of \( S_{d.w} \) satisfying (18). These are the points \((0,0),(0,1),(1,0)\), and \( (1,1) \), corresponding respectively to the identity matrix and, for example, the matrix of (20). It is not difficult to show that the second point lies in the proper region (that for which \( s_{11} < s_{21} \)) only when \( s_{11} < A_1 \), in the case \( q_{1w} > q_{2w} \) and when \( s_{11} > A_2 \), in the case \( q_{1w} < q_{2w} \), consistent with results given in §4.

6. PATHOLOGICAL DENSITIES

When the two transition rates \( q_{1w} \) and \( q_{2w} \) out of the open states are equal (to \( q_w \), say) we have from (13b) \( u_1 = q_w \) and from (10) \( x_1 = q_w \). Substituting in (14) yields \( A_1 = 1, A_2 = 0 \). On the other hand, when \( A_1 = 1 \) it follows from (14) that \( x_1 = u_1 \) and from (16b) that

\[
b_1 = b_2 = \frac{(\pi_2 - \pi_1)(u_1^2 - u_1 v_1 + u_0)}{(\pi_1 \pi_2 (v_1 - u_1))};
\]

but from (13) \( u_0 = u_1 v_1 + u_0 = 0 \), so that \( b_1 = b_2 \) and (from (16a)) \( q_{1w} = q_{2w} \). Thus whenever \( q_{1w} = q_{2w} \) we have \( A_w = 0 \) and the marginal distribution of open times will be single exponential with parameter \( q_w \). This is intuitively reasonable, as transitions between the two open states have no effect on the rate \( q_w = x_1 \) out of the class. Under a similarity transformation of \( Q_{d.w} \) we would expect (because of identical distributions) this pathology to be conserved. That this is so can readily be ascertained by inspection of (24c, d) where \( q_{1w} = q_{2w} \) implies \( q_1 = q_{1w} \). Then also \( q_{11} = q_{12} = q_{12}, q_{21} = q_{22}, q_{21} = q_{21} \). The existence of such special cases is important from the inference point of view, at least theoretically, as the number of distinct exponential components in an experimentally determined sojourn time distribution is generally used to infer the (supposed equivalent) number of kinetically distinct states.

7. JOINT DENSITY

The foregoing has shown that two distinct sets of parameter values can result in the same marginal sojourn time distributions. This may cause problems if inference about underlying parameters of the channel mechanism is based solely on fitted marginal distributions. In some cases these problems can be overcome by basing inference on the joint densities of a sequence of open and closed times, and it is preferable to use the latter whenever feasible. Some practical demonstrations that the two-dimensional sojourn time distributions (open-closed and closed-open) can distinguish between models that have the same marginal open and closed time distributions have been given by Magleby and co-workers (see, for example, Magleby & Weiss 1990a).

We illustrate by considering a model based on scheme 1. The joint density function \( f(t,s) \) for a series of open and closed times \( (t, s) = (t_1,s_1, t_2, s_2, \ldots, t_m,s_m) \), where \( t_j \) is the \( j \)th open time and \( s_j \) the \( j \)th closed time, is given by (Fredkin et al. 1985; Ball & Sansom 1989)

\[
f(t,s) = \pi_0^m \prod \{e^{Q_{11}t}Q_{12}e^{Q_{21}t}Q_{22}e^{Q_{33}t}}. \tag{29}
\]

Suppose \( Q \) is such that two distinct choices, \( Q_{11}^{(+)}, Q_{12}^{(+)}, Q_{21}^{(+)}, Q_{22}^{(+)}, Q_{33}^{(+)}, \) of \( Q_{d.w} \) give rise to the same marginal open time distribution. Suppose also that \( Q_{11}^{(-)}, Q_{12}^{(-)}, Q_{21}^{(-)}, Q_{22}^{(-)}, Q_{33}^{(-)} \), of \( Q_{d.w} \) give rise to the same marginal open time distribution. Suppose also that \( Q_{11}^{(+)} = Q_{11}^{(-)} \) and let \( Q^{(a)} = Q \). A second transition rate matrix \( Q^{(a)} \) can be calculated from equation (5) with \( S_{d.w} \) calculated as in, for example (19) and (20), and \( S_{d.w} = I \). (If \( Q_{d.w}^{(a)} \) is strictly, then \( Q^{(a)} = Q \) and \( Q^{(a)} \) is obtained by a similarity transformation.) By using numerical values for \( Q \) from Colquhoun & Hawkes (1982) and Ball & Sansom (1989) we simulated data sets of length up to 10000 open and closed time pairs with \( Q = Q^{(a)} \), and calculated \( f(t,s) \) with \( Q = Q^{(a)} \) and \( Q = Q^{(a)} \). In every case \( f(t,s) \) was appreciably larger under \( Q^{(a)} \) than under \( Q^{(a)} \). We then repeated the exercise using data simulated under \( Q = Q^{(a)} \) and in that case \( f(t,s) \) was larger under \( Q^{(a)} \). Thus, for the present model, the joint density function enables us to distinguish between \( Q^{(a)} \) and \( Q^{(a)} \).

This example emphasizes the importance of basing inference on the complete channel record rather than on just the fitted open and closed time distributions. In practice, of course, the underlying transition rate matrix \( Q \) is unknown but can in principle be estimated by maximizing the joint density function (likelihood) over the parameter values. Even then good estimation may be difficult (for example, see Fredkin & Rice 1992a).

8. IDENTIFIABILITY OF THE FIVE-STATE MODEL

Although parameter estimation with respect to scheme 1 based on the marginal open time density may give rise to multiple solutions for \( Q_{d.w} \), it was noted above that numerical evaluation of the likelihood function identified the correct parameter values (at least in the case of \( Q_{d.w}^{(1)} \) and \( Q_{d.w}^{(1)} \)). We now prove that this method is always identifiable by showing that there exists no similarity transform in the sense of Kienker (1989) other than the identity transform, provided the mild regularity conditions for his theorem are satisfied.

Let \( Q \) and \( Q \) be the transition rate matrices of equivalent processes based on scheme 1, each partitioned as in equation (1), and \( S = [s_{ij}] \). From (5b) we obtain, by equating terms,

\[
s_{33} = (1-\epsilon_{31})q_{23}/q_{23}, s_{34} = q_{23}/q_{23}, s_{35} = 0; \tag{30}
\]

\[
s_{43} = (1-\epsilon_{41})q_{23}/q_{23}, s_{44} = q_{23}/q_{23}, s_{45} = 0. \tag{30}
\]
28. R. O. Edeson and others Non-identifiability in single channel inference

As \( \sum_{i,j} s_{ij} = 1 \) for \( i = 3, 4, 5 \), we have \( s_{34} = 1 - s_{33} \), so

\[
(1 - s_{33}) q_{33} + s_{31} q_{31} = q_{33}, \quad (1 - s_{33}) q_{33} + s_{31} q_{31} = q_{34}. \tag{31}
\]

Now equating terms in (5c) gives \( s_{33} = s_{34} = 0 \), \( s_{55} = 1 \), and

\[
(1 - s_{30}) q_{41} = s_{42} q_{42}, \quad s_{23} q_{23} = (1 - s_{21}) q_{22}. \tag{32}
\]

By division in (30) and (32) to cancel \( q_{23} \) and \( q_{22} \), respectively, we obtain

\[
q_{11}/q_{23} = q_{32}/q_{41} = (1 - s_{31}) s_{23}/(1 - s_{31}),
\]

and similarly \( q_{14}^2/q_{23} = q_{32}^2/q_{41} = (1 - s_{31}) (1 - s_{34})/(1 - s_{31}) \); these are already very restrictive conditions on a possible \( Q \) matrix. Next, (5d) gives from the off-diagonal corner elements (for example, \( s_{35} q_{35} + s_{34} q_{34} = q_{35} q_{35} + s_{34} q_{34} \) with \( q_{35} = q_{34} = s_{35} = 0 \) that \( s_{35} = 0 \) and \( s_{34} = 0 \); hence \( S_{35} = I \).

Returning to (30) we now have that \( s_{21} q_{14}/q_{32} = 0 \) and \( 1 - s_{31} q_{14}/q_{31} = 0, \) so \( s_{21} = 0, s_{31} = 1 \) and hence \( S_{21} = I \).

Therefore, given a model based on scheme 1, the only similarity matrix is the identity matrix; that is, the model is identifiable. If \( q_{12} = q_{21} = 0 \) the same argument holds, and the model remains identifiable. If state 5 is deleted \( (q_{53} = q_{43} = 0) \), the model is identifiable unless the very restrictive condition \( q_{14}/q_{23} = q_{32}/q_{41} \) is satisfied. The fact that the full five-state model is identifiable explains why consideration of the joint density function enables us to distinguish between \( Q^+ \) and \( Q^- \) in the previous section.

9. DISCUSSION

Although there has been extensive development of ion channel models over the past decade, statistical inference for such models is considerably less advanced. However, progress has been made, for example by Magleby & Weiss (1990a, b) using simulation-based fitting of the two-dimensional sojourn time distributions, and by Fredkin & Rice (1992b) using hidden Markov models. The challenges of inference with respect to model selection and parameter estimation are not trivial; even simple models may have surprising properties which advise caution, particularly in attempting to estimate kinetic parameters from empirical sojourn time densities. Where practicable, an approach based on joint density information is preferable to use of marginal distributions. For a given model, constraints on the similarity matrix which ensure that the transformed \( Q \) matrix is proper (all \( q_{ij} \) non-negative) are not known (Kienker 1989). Some progress has been made here in elucidating these constraints analytically by focusing on a two-state submatrix; extension to larger classes or the full \( Q \) matrix has not been attempted, and the problem in general remains open. For a given kinetic scheme, it may be possible to demonstrate identifiability of the corresponding Markov model explicitly, as was the case for the five-state example considered above.

Finally, the basic model-determined non-uniqueness examined here may be compounded in practice by other difficulties related, for example, to limited time resolution (Milne et al. 1989; Ball et al. 1993), numerical fitting methods, or lack of information about the number of active channels contributing to the record. The latter problem is illustrated by the case of two independent, identical channels having the kinetics of scheme 2 with two closed states. If their probability of opening is low, concurrent events may not be observed; then the single level openings, whose density is a mixture of four exponentials (cf. Yeo et al. 1989), could be interpreted as arising from a single channel with four open states (see Colquhoun & Hawkes (1990) for a discussion of this phenomenon).

Supported by the Faculty of Anaesthetists of the Royal Australasian College of Surgeons through a Harry Daly Research Fellowship (R.O.E.), the U.K. Science and Engineering Research Council (F.G.B.), the Wellcome Trust (F.G.B., S.S.D.), and by the Australian Research Council for a visit by F.G.B. and S.S.D. to the University of Western Australia and Murdoch University. We thank the referees for their detailed reading of the manuscript.

APPENDIX

The derivation of the quadratics in (16), although mathematically straightforward, is somewhat lengthy; an outline of the method is therefore given for the case in \( q_{1*} \).

From equation (15) we have

\[
q_{2*} q_{1*} = q_{12} q_{2*} \pi_1/\pi_2, \tag{A 1}
\]

which after substitution in \( u_0 = q_{12} q_{2*} + q_{21} q_{1*} + q_{1*} q_{2*} \) (from \( u_0 = d_0 d_2 s \), (13a)) yields

\[
u_0 = q_{2*}(q_{1*} + q_{12}/\pi_2), \tag{A 2}
\]

thence by rearrangement giving (17c) (and (17d)). In addition, from (13b) we have \( q_{2*} = (u_1 - \pi_1 q_{1*})/\pi_2 \), and substituting into (A 2) gives

\[
q_{12} = \pi_2 \left[ \frac{u_0 \pi_2}{(u_1 - \pi_1 q_{1*}) - q_{1*}} - q_{1*} \right]. \tag{A 3}
\]

This leads by (A 1) to

\[
q_{2*} = \pi_1 \left[ (u_0/q_{1*}) - (u_1 - \pi_1 q_{1*})/\pi_2 \right]. \tag{A 4}
\]

Substituting into (15c) gives

\[
u_0 \pi_2^2/(u_1 - \pi_1 q_{1*}) + u_0 \pi_1/q_{1*} + u_1 = \pi_1, \tag{A 5}
\]

which by rearrangement results in the quadratic (16a) in \( q_{1*} \).

REFERENCES


Ball, F. G., Davies, S. S. & Sansom, M. S. P. 1993 Ion channel gating and time interval omission: statistical inference for a two-state Markov model. (Submitted.)


Colquhoun, D. & Hawkes, A. G. 1990 Stochastic properties of ion channel openings and bursts in a membrane patch that contains two channels: evidence concerning the number of channels present when a record containing only single openings is observed. *Proc. R. Soc. Lond.* B 240, 453–477.


Received 16 August 1993; accepted 30 September 1993