Robust inference and modelling for the single ion channel

BRENTON R. CLARKE* and PETER L. McKINNON

Mathematics and Statistics, Division of Science and Engineering, Murdoch University, Murdoch, WA 6150, Australia

(Revised 10 December 2003; in final form 17 June 2004)

Statistical modelling and inference for a single-ion channel have principally been carried out using finite-state space continuous-time Markov chains. Statistical inferences for the closed and open dwell times and the kinetic rate constants between states have then been arrived at via maximum likelihood methods, including the use of the EM algorithm. The fundamental assumption behind this theory is that one has the correct number of closed and open states, something which may not be easily determined by the use of current methods for modelling, say, the number of components in a mixture of exponential distributions used to fit, say, the ‘closed’ dwell times. Here, we show that the use of a robust $L_2$ estimator can outperform the EM algorithm both when the correct number of states is apparent and also when there are small deviations from the supposed models. After describing the statistical models used to demonstrate these results and how they lead to particular mixtures of exponential distributions the comparison is then made between the performances of the estimators (robust $L_2$ and maximum likelihood (via the EM algorithm)). The resulting performances in terms of means and standard errors of the estimated kinetic rate constants are then assessed. The estimating equations derived from minimizing the $L_2$ distance are given explicitly in the appendices.

Keywords: M-estimator; Minimum distance estimation; Ion channels; Mixtures of exponential distributions; EM algorithm; Markov chain

1. Introduction

The single-ion channel involves a gating mechanism in the biological cell membrane. An electrical signal marked by the passage of ions through the cell membrane indicates that the ion channel is open. When the signal in terms of electric current stops the ion channel is essentially closed and so no ions are passing through. The duration with which the ion channel remains closed(open) can be modelled by assuming there are one or more closed(open) states linked in terms of a finite state continuous time Markov chain. The duration with which the ion channel remains closed(open) relies on the time spent in successive closed(open) states of the Markov chain. It is a feature of this modelling that the resultant closed(open) dwell times have as their distribution mixtures of exponential distributions.

References to the fact that gating mechanisms of a single-ion channel are usually modelled by a finite-state space continuous time Markov chain include Colquhoun and Hawkes [1],

*Corresponding author. Email: clarke@prodigal.murdoch.edu.au
Ball et al. [2], see Sakmann and Neher [3] for an overview. Ball and Rice [4] give a bibliography for the Markovian models widely used for ion channel kinetics. For a time-reversible Markov model of single ion channel gating that is in equilibrium, the length of a typical sojourn in, say, the closed states follows a distribution given by a finite mixture of exponential distributions (A1). A similar result holds for open sojourns, see Kijima and Kijima [5]. Fredkin and Rice [6] discuss maximum likelihood estimation and identification from single-channel recordings; see also Edeson et al. [7]. Maximum likelihood estimation can be approached several ways, either by writing down the likelihood directly and using maximum likelihood routines, as in Fredkin and Rice [6], or in the cases where the assumed models lead to independent closed times (or independent open times) using the EM algorithm (starting from good initial values) using just the closed times (open times). Theoretically, the EM algorithm leads to the maximum likelihood solution which is asymptotically the most efficient if the model has been identified correctly. However, there are other ways of estimating parameters in mixture models such as in equation (A1) and we put forward in this paper the $L_2$-minimum distance estimator. This estimator has proved highly efficient and robust in estimating parameters of related distributions as discussed in Heathcote and Silvapulle [8], Clarke [9], Clarke and Heathcote [10] and in Hettmansperger et al. [11], and Clarke [12]. The estimator, given as a solution of M-estimating equations is in fact consistent, asymptotically normal, and can have high efficiency at the parametric model.

In our study, as opposed to that of Fredkin and Rice [6] we do not add noise to the signal before doing the estimation, for we argue that there is no information in practical ion channel modelling about what form the real noise actually takes. In fact this only confirms that using robust estimation methods, such as in the use of the $L_2$ estimator, is most appropriate in ion channel research. Theoretically implementing the modelling and estimation as we do here without the addition of noise and consequent reconstruction of the signal should advantage the maximum likelihood estimator (MLE) as carried out by the EM algorithm. However, even with large sample sizes of say 1000 observations from a mixture, the EM algorithm fails to assert its theoretical asymptotic efficiency at the parametric model. Moreover, in small departures from the asserted Markovian model the $L_2$-estimator does better. Robust estimators are likely to do better in resolving the transition rates either with or without noise added to the signal, since if we model the noise inappropriately as we have no information about what true form it might take, this will inevitably lead to a departure from the real model, something for which maximum likelihood techniques are known to be sensitive [13, 14]. Despite this there appears to have been a spurt in discussion of the EM and related methods under various regimes of noise as in Michalek and Timmer [15], Michalek et al. [16], Qin et al. [17], Venkataramanan et al. [18, 19]. The purpose of this paper is to highlight a valid alternative approach for some statistical models which does perform better than straightforward use of the EM algorithm. A recent paper of De Gunst et al. [20] considers attempts at modelling various forms of noise, but again no real knowledge of noise involved in actual nature is offered.

Letting the parametric model for a mixture distribution be $F_\theta$ where $\theta \in \Theta$ is the parameter space, an $L_2$-estimator can be obtained by solving the minimizing equations of $L_2$ distance

$$J_n(\theta) = \int_0^\infty (F_n(x) - F_\theta(x))^2 \, dx$$

Here $F_n$ represents the empirical distribution function formed from a sample $X_1, X_2, \ldots, X_n$ of independent identically distributed random variables. The estimating equations are of the M-estimator form

$$n^{-1} \sum_{i=1}^n \psi(X_i, \theta) = 0$$
See the appendices for explicit forms of the vector $\psi$ when $F_\theta$ represents the finite mixture of two or more exponential distributions. It can be noted here that robustness of the $L_2$ estimator follows from the bounded and smooth nature of the corresponding $\psi$-function. See Clarke [12, 21, 22] for the relevant details. In contrast the MLE satisfies equations which are associated with an unbounded efficient score function. Consequently the MLE, for example, as arrived at by use of the EM algorithm is non-robust. We concentrate here on the EM algorithm used in mixture studies as in maximum likelihood estimation ever since the early papers of Hasselblad [23] and Dempster et al. [24]. Background study of direct solution of the maximum likelihood equations showed that the EM algorithm performed better and so we do not report that study here.

In this paper, we compare the solution of the nonlinear equations (2), for the $L_2$ distance method solving the equations using Matlab 6.0 equation solver ‘FSOLVE’, to the solution of the EM algorithm to estimate parameters in a mixture of two component distributions, that is, we compare the performance of the estimators when the model is correctly specified. The parameters of the underlying mixture are chosen to reflect the dwell times in a class of closed states for ion channel models similar to those discussed in Fredkin and Rice [6]. Following this we compare the performance of the estimators when in fact there are two extra states: one closed and one open, for which there is an underlying three component mixture for the equilibrium distribution of closed time sojourns, though the third component has a small proportion but a large mean. In behaving as being unaware of the extra states, the $L_2$ minimum distance estimator outperforms the solution obtained from the EM algorithm. This also has relevance to the actual transition rates derived from the estimated mixtures.

It is readily admitted in Fredkin and Rice [6] that there are many situations where it may be difficult or practically impossible to identify the underlying kinetic scheme and estimate its parameters, even when using the MLE when the model is correctly identified. Partly this can be explained by model distributions for the closed state sojourns having parameters too near the boundary of the parameter space, whereupon unusually large samples would be needed to gain information about the models generating the data. On the other hand, while some models appear to be resolvable using EM, we show the $L_2$ estimator to be quite competitive in terms of standard errors when applied to such models, the EM algorithm tending to stray from the underlying population parameters.

Seidel et al. [25] note the sensitivity of the EM algorithm to starting values in the context of hypothesis testing with two component mixtures of exponential distributions. This is related to determining the number of component distributions in the mixture. Given the failure of the EM algorithm even in such simple situations it is clear that robust modelling and inference has a part to play in the discussion of ion channels. Further work in this direction is clearly necessary.

In section 2, we describe the particular models associated with ion channels and how they lead to the mixture distributions in equation (A1). Comparisons of estimation and fitting of the two methods, EM and $L_2$ follow.

2. Application of modelling in connection to mixtures in ion channels

2.1 Estimation of rate constants in 3-state and 5-state models

In this section, we consider the kinetic rate constants associated with two models, a 5-state model, that leads to an equilibrium distribution for closed dwell times of a three component exponential mixture distribution, and a more simple 3-state model for which the equilibrium distribution of closed times is a mixture of two exponentials. Figures 1 and 2 depict the models,
respectively. For instance, the model in figure 2 is a special case of the five-state model, putting transition rates \( q_{21} = q_{24} = 0 \). Here \( q_{ij} \) are transition rates (per second) from state \( i \) to state \( j \).

States C1–C3 form the closed class of states and O4–O5 represent the open states. In the models of figures 1 and 2 there are no cycles, so the system is reversible. Consequently, it is known [5] that the equilibrium density function of the closed sojourn times \( T \) is a mixture of at most \( n_c \) exponential densities, where \( n_c \) is the number of closed states. Random sum methods have been used by, for example Milne et al. [26], Edeson et al. [27] to obtain Laplace transforms \( \chi(s) = E(e^{-sT}) \) and hence corresponding density functions.

The aim of this section is to see how accurately we can estimate a three component mixture by a two component mixture when one (or more) \( q_{ij} \) in figure 1 is (are) close to 0 (which results in a small proportion for one of the \( \epsilon_i \) in equation (3) corresponding to a component distribution with a large mean). Using the estimated \( \hat{\epsilon}_1, \hat{\lambda}_1 \) and \( \hat{\lambda}_2 \) we can then determine the \( q_{ij} \) of the three state model. The respective distributions for the dwell times in the closed class of states is for the 5-state model of figure 1 and the 3-state model of figure 2 given by densities

\[
f(t) = \epsilon_1 \lambda_1 e^{-\lambda_1 t} + \epsilon_2 \lambda_2 e^{-\lambda_2 t} + \epsilon_3 \lambda_3 e^{-\lambda_3 t} \tag{3}
\]

\[
f(t) = \epsilon_1 \lambda_1 e^{-\lambda_1 t} + \epsilon_2 \lambda_2 e^{-\lambda_2 t} \tag{4}
\]

with

\[
\sum_i \epsilon_i = 1
\]

and

\[
\lambda_i > 0, \quad \epsilon_i > 0 \quad \text{for all} \ i \tag{5}
\]

Here \( q_{ij} \) are the transition rates from state \( i \) to state \( j \), where \( d_i = \sum_{j \neq i} q_{ij} \) is the transition rate out of state \( i \). Let \( p_{ij} = q_{ij}/d_i \) which implies \( \sum_{j \neq i} p_{ij} = 1 \). A single sojourn in state \( i \) has an exponential distribution with Laplace transform \( \varphi_i(s) = d_i/(d_i + s) \ (s \geq 0) \). This is the Laplace transform of an exponential random variable with mean \( 1/d_i \).

### 2.2 Some preliminary calculations

We first look at the 3-state model as given in figure 2. The sojourn time \( (T) \) in class C is a random sum, visualize

\[
T = X_{30} + \sum_{i=1}^{N-1} (X_{3i} + X_{2i})
\]

Figure 2. 3-State model.
Robust inference and modelling for the single ion channel

with

\[ P(N = n) = p^{n-1}(1 - p) \quad \text{for } n = 1, 2, \ldots \]

where \( p = q_{32}/d_3 \). Therefore, letting \( \chi \) denote the Laplace transform so that

\[
\chi(s) = E[e^{-sT}] \\
= \sum_{n=1}^{\infty} p^{n-1}(1 - p)\varphi_2(s)\varphi_3(s)^n \\
= \frac{(1 - p)\varphi_3(s)}{1 - p\varphi_2(s)\varphi_3(s)} \\
= L\left(\sum_{i=1}^{2} \epsilon_i \lambda_i e^{-\lambda_i t}\right) \tag{7}
\]

This then leads to

\[
\frac{(q_{23} + s)q_{35}}{s^2 + (d_3 + q_{23})s + q_{23}q_{35}} = \frac{s(\epsilon_1 \lambda_1 + \lambda_2 - \lambda_2 \epsilon_1) + \lambda_1 \lambda_2}{s^2 + (\lambda_1 + \lambda_2)s + \lambda_1 \lambda_2}
\]

which gives

\[
\epsilon_1 = \frac{q_{35} - \lambda_2}{\lambda_1 - \lambda_2} \tag{8}
\]

and the \( \lambda_i \) are the negative of the roots of

\[
s^2 + (d_3 + q_{23})s + q_{23}q_{35} = 0 \tag{9}
\]

This demonstrates how we find \( \epsilon_1, \lambda_1 \) and \( \lambda_2 \) for the distribution given in equation (4). On the other hand if we are given \( \epsilon_1, \lambda_1 \) and \( \lambda_2 \) we can find the \( q_{ij} \) in figure 2. Equation (8) implies

\[
q_{35} = \epsilon_1(\lambda_1 - \lambda_2) + \lambda_2 \tag{10}
\]

and from equation (9)

\[
q_{32} = \frac{(\lambda_1 - q_{35})(q_{35} - \lambda_2)}{q_{35}} \tag{11}
\]

\[
q_{23} = \frac{\lambda_1 \lambda_2}{q_{35}} \tag{12}
\]

2.3 Estimating a correctly specified 3-state model

In this section, we consider 3-state models of figure 2 where the distribution of closed dwell times is defined by a two component mixture—we will set the \( q_{ij} \) in figure 2, determine the relevant \( \epsilon_1, \lambda_1 \) and \( \lambda_2 \) and use these to generate 100 samples of sample size 1000. Using these samples the \( L_2 \) and EM algorithms are used to find the average and standard errors (SE) of the estimates and then from these estimates we determine the \( q_{ij} \) using equations (10) to (12).
First, we must define the stopping criteria for the algorithms. For the $L_2$, we use the MATLAB default of absolute tolerance of $10^{-5}$, and for EM a relative tolerance of $10^{-5}$ (suggested in conversation with Prof. Krishnan) defined for example when

$$\frac{|\hat{e}_1^{(k+1)} - \hat{e}_1^{(k)}|}{|\hat{e}_1^{(k+1)}|} < 10^{-5}$$

(13)

and similar constraints for $\lambda_1$ and $\lambda_2$ are also satisfied. It is important to note that the EM algorithm is implemented with the suggested tolerance from the book by McLachlan and Krishnan [28]. The algorithm FSOLVE in MATLAB is different in character from the EM algorithm and the absolute tolerance of $10^{-5}$ is easily implemented. No real advantage is gained in using an absolute tolerance of $10^{-5}$ for the EM algorithm, in fact computing times become longer. To ensure FSOLVE does not stop too early in simulations the $\psi$-functions should be suitably scaled so they are all of the same order. This is easily achieved by computing each function with the approximate parameters checking the output and scaling accordingly.

For example, suppose the ‘generating parameters’ are $\theta_0$ and the value of the vector function on the left hand side of equation (2) for the first sample in the simulation we denote as $K_n(\theta_0) = n^{-1} \sum \psi(X_i, \theta_0)$. Here $(K_{n1}(\theta_0), K_{n2}(\theta_0), K_{n3}(\theta_0))$ is the vector $K_n(\theta_0)$. Suppose now for the first sample $K_n(\theta_0)$ is of the order of magnitude $10^{-5}$, $K_{n2}(\theta_0)$ is of order of magnitude $10^{-7}$ and $K_{n3}(\theta_0)$ is of order of magnitude $10^{-9}$. From then on the equation solver substituted the function $\psi = (10^3 \psi_1, 10^5 \psi_2, 10^7 \psi_3)$ for the rest of the simulation. This procedure ensured that the equation solver never dropped out too early. A simple method to find initial values is to seek solutions which minimize the Cramér von Mises distance, $\int_{\mathbb{R}} (F_n - F_0)^2(x)dF_0(x)$, which has the computational formula

$$\frac{1}{n} \sum_{i=1}^{n} \left[ F_0(X(i)) - \frac{i - 1/2}{n} \right]^2 + \frac{1}{12n^2}$$

(14)

Here $X(i)$ is the $i$th order statistic. Equation (14) is simple to compute, provides reasonable estimates and is reasonably robust to starting values. While the Cramér von Mises distance yields reliable and easily determined initial estimates, the $L_2$-estimator given as an iteratively derived solution of equation (2) proved more efficient or less variable at least on empirical analysis. In the simulations, the same initial estimates were used for the EM algorithm as were used for the solution of the $L_2$ estimating equations, these being the ‘generating parameters’. This should not advantage one or the other of either the EM algorithm or the solution to the $L_2$ estimating equations.

First, we consider transition rate parameters as given in Fredkin and Rice [6] where even the MLE has some difficulty in resolving the parameters; let $q_{23} = 91$, $q_{32} = 4$ and $q_{35} = 50$. The underlying distribution for the closed dwell times is given by parameters $\epsilon_1 = 0.9233$, $\lambda_1 = 45.9246$ and $\lambda_2 = 99.0754$. The estimates along with their standard errors for each method are shown in table 1.

<table>
<thead>
<tr>
<th>$\epsilon_1$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9233</td>
<td>45.9246</td>
<td>99.0754</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\epsilon}_1$</th>
<th>$\hat{\lambda}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9753</td>
<td>47.2034</td>
</tr>
</tbody>
</table>

Overall, the $L_2$ estimates are superior. This is a positive result, since while the asymptotically efficient estimator is the MLE, the $L_2$-estimator proves competitive when the model is correctly specified, suggesting high relative efficiency. The EM estimate for $\lambda_2$ is quite different with a large standard error explained by the fact that estimates range from 51.4451 to 409.0734 (out of interest this estimate is (0.9753, 47.2034, 409.0734) which is within one standard deviation for $\hat{\epsilon}_1$ and $\hat{\lambda}_1$ but not for $\hat{\lambda}_2$). We also output the average number of iterations for the EM estimates, 400.76, and find that the number of iterations taken ranges from 1 to 10,025. While it is understandable that the number of iterations could be as low as 1 since we are starting from good initial estimates, it is frequently the case that the EM algorithm needs a larger number of iterations to arrive at an answer.
Table 1. Estimates for $\epsilon_i$, $\lambda_1$ and $\lambda_2$ when $q_{23} = 91$, $q_{32} = 4$ and $q_{35} = 50$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\epsilon}$ Average</th>
<th>$\hat{\lambda}_1$ Average</th>
<th>$\hat{\lambda}_2$ Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_2$</td>
<td>0.9232</td>
<td>45.98</td>
<td>99.13</td>
</tr>
<tr>
<td>EM</td>
<td>0.9403</td>
<td>47.80</td>
<td>108.78</td>
</tr>
<tr>
<td>True</td>
<td>0.9233</td>
<td>45.9246</td>
<td>99.0754</td>
</tr>
</tbody>
</table>

Given the estimates from the 100 samples generated for each algorithm, we can use equations (10)–(12) to find the $q_{ij}$ and their standard errors, as given in table 2.

Again, the $L_2$ estimates and the true values are comparable, and once again the EM estimates and standard errors are poor. Most notable is the average of the estimates of $q_{32}$, 15.37, which almost quadruple the true $q_{32}$, and the very large standard error for the estimates of $q_{23}$ – explained by estimates ranging from 45.32 to 331.21. These results confirm the observation of Fredkin and Rice [6] of the model not being easy to resolve.

A model which is better to estimate from the sequence of closed dwell times is if we alternate the values for $q_{23}$ and $q_{32}$ such that $q_{23} = 4$, $q_{32} = 91$ and $q_{35} = 50$, we find $\epsilon_1 = 0.6582$, $\lambda_1 = 1.3927$ and $\lambda_2 = 143.6073$. The estimates from $L_2$ and EM are shown in table 3.

This time the EM algorithm performs more credibly with the averages and standard errors being comparable to $L_2$ except for the large standard error for $\hat{\lambda}_2$. Again, we find the average number of iterations for the EM algorithm is 8.32, and we also find the $q_{ij}$ from the estimates, as given in table 4.

The results for $L_2$ and EM are all comparable with both giving accurate estimates of the original $q_{ij}$.

Finally, an example is presented which will be of relevance in later sections. In figure 2, we let $q_{23} = 300$, $q_{32} = 600$ and $q_{35} = 4000$, and find that $\lambda_1 = 258.5393$, $\lambda_2 = 4641.4607$ and $\epsilon_1 = 0.1464$. Entering these parameters into MATLAB we find the following:

The output in table 5 again illustrates that EM performs poorly compared to $L_2$. The very large standard error for $\hat{\lambda}_2$ (from the EM estimates) can again be explained by the estimates

Table 2. Estimates for $q_{23}$, $q_{32}$ and $q_{35}$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{q}_{23}$ Average</th>
<th>$\hat{q}_{32}$ Average</th>
<th>$\hat{q}_{35}$ Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_2$</td>
<td>91.73</td>
<td>3.298</td>
<td>50.83</td>
</tr>
<tr>
<td>EM</td>
<td>90.71</td>
<td>15.37</td>
<td>50.48</td>
</tr>
<tr>
<td>True</td>
<td>91</td>
<td>4</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 3. Estimates for $\epsilon_i$, $\lambda_1$ and $\lambda_2$ when $q_{23} = 4$, $q_{32} = 91$ and $q_{35} = 50$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\epsilon}$ Average</th>
<th>$\hat{\lambda}_1$ Average</th>
<th>$\hat{\lambda}_2$ Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_2$</td>
<td>0.6564</td>
<td>1.3940</td>
<td>143.64</td>
</tr>
<tr>
<td>EM</td>
<td>0.6572</td>
<td>1.3941</td>
<td>143.86</td>
</tr>
<tr>
<td>True</td>
<td>0.6582</td>
<td>1.3927</td>
<td>143.6073</td>
</tr>
</tbody>
</table>
which range from 4233.99 to 5168.09. Finally, we can go back and find the \( q_{ij} \) of the original model which are given in table 6 where we see that once again the EM does quite poorly in estimating the parameters of the distribution generating the data. The very large standard error for \( q_{35} \) is partly a result of the large standard error in \( \hat{\lambda}_2 \).

We can now be confident that the \( L_2 \) method performs relatively well with respect to the solutions of the maximum likelihood equations and when the 3-state model corresponding to a two component mixture for the closed dwell times is correct.

### 2.4 Fitting three component models with a two component distribution

The 5-state model of Colquhoun and Hawkes [1, 29], which provides a model for the gating behaviour of the nicotinic acetylcholine receptor, has been widely used as an example in the literature, and includes as special cases several simpler models. The model is illustrated in figure 1. Now, it is a feature of many robustness studies to consider, for example, the behaviour of an estimation procedure in some small departure from the assumed model. We do this in this paper by considering parametric estimation of the closed dwell time distributions that would arise in the model as described by figure 2. However, to see how robust is the estimation procedure, we consider what happens if in fact the underlying model is given in figure 1 in which case there are two extra states \( C_1 \) and \( O_4 \). For the relevant change in the underlying distribution we, therefore, consider modelling the closed time distribution in the 5-state model. These distributions mimic only slight departures from the original distributions of the entertained 3-state model in figure 2, for instance when \( q_{21}, q_{24} \ll q_{23}, q_{32} \). For example,
a transition to the state $C_1$ when the ion channel is spending a sojourn in the class of closed states occurs infrequently, however, when it occurs (at least for the transition rates entertained in this paper) the typical stay is on average longer than for a visit to any other of the other two closed states. The effect noted here is to yield an underlying closed dwell time distribution which is a three component mixture of exponentials but with one component having a small proportion with a large mean. While we simulate the data under the resultant three component distribution, one component of which is normally undetectable by classical methods for the sample sizes given because of its small proportion, the estimation proceeds assuming only a two component mixture distribution. Due to the large mean of the extra component in the underlying distribution the simulation occasionally throws up a large observation which has an adverse effect on the EM algorithm method of estimation though not so much on the $L_2$ method of estimation.

We can similarly find the sojourn times for the 5-state model by considering Laplace transforms.

\[ \chi_3(s) = \varphi_3(s)p_{35} + \varphi_3(s)p_{32}\chi_2(s) \]  
\[ \chi_2(s) = \varphi_2(s)(p_{24} + p_{21}\varphi_1(s)\chi_2(s) + p_{23}\chi_3(s)) \]

We then solve for $\chi_2(s)$ by substituting equation (15) in equation (16), which leads to

\[ \frac{(d_1 + s)d_2(p_{24}(d_3 + s) + p_{23}p_{35}d_3)}{(d_1 + s)(d_2 + s)(d_3 + s) - d_2(p_{21}d_1(d_3 + s) + p_{32}p_{23}d_3(d_1 + s))} = L \left( \sum_{i=1}^{3} \epsilon_i \lambda_i e^{-\lambda_i s} \right) \]

This implies that $\lambda_i$ are the negative roots of (with some rearranging)

\[ s^3 + s^2(d_1 + d_2 + d_3) + s(d_1d_2 + d_1d_3 + d_2d_3 - d_2p_{21}d_1 - d_2p_{32}p_{23}d_3) + d_1d_2d_3(p_{24} + p_{23}p_{35}) = (\lambda_1 + s)(\lambda_2 + s)(\lambda_3 + s) = 0 \]

We can similarly solve equations (15) and (16) for $\chi_3(s)$ to get

\[ \frac{p_{35}(d_1 + s)(d_2 + s)d_3 - d_1d_2d_3p_{21}p_{35} + d_3d_2(d_1 + s)p_{32}p_{24}}{(d_1 + s)(d_2 + s)(d_3 + s) - d_2(p_{21}d_1(d_3 + s) + p_{32}p_{23}d_3(d_1 + s))} = L \left( \sum_{i=1}^{3} \epsilon_i \lambda_i e^{-\lambda_i s} \right) \]

We notice that the denominator is the same as in equation (17). This means that the $\lambda_i$ when solving for $\chi_3$ are again the negative roots of equation (18).

Given that we know $\lambda_i$ (for $i = 1, 2, 3$) we can then solve equations (17) and (19) to find their corresponding $\epsilon_i$. An interesting point is that when equating the numerator of either equation we find

\[ \lambda_1\lambda_2\lambda_3 = d_1d_2d_3(p_{24} + p_{23}p_{35}) \]

which is equation (18) for $s = 0$ which is already satisfied. We can replace this equation with the other condition on the $\epsilon_i$, equation (5).
Therefore, from equation (17) we find
\[
\begin{pmatrix}
\lambda_1 \lambda_2 + \lambda_1 \lambda_3 & \lambda_1 \lambda_2 + \lambda_2 \lambda_3 & \lambda_1 \lambda_3 + \lambda_2 \lambda_3 \\
\lambda_1 & \lambda_2 & \lambda_3 \\
1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
\epsilon_1 \\
\epsilon_2 \\
\epsilon_3
\end{pmatrix}
= 
\begin{pmatrix}
d_1 d_2 p_{24} + d_2 d_3 (p_{24} + p_{23} p_{35}) \\
p_{24} d_2 \\
p_{24} d_2
\end{pmatrix}
\] (20)

and from equation (19) we find
\[
\begin{pmatrix}
\lambda_1 \lambda_2 + \lambda_1 \lambda_3 & \lambda_1 \lambda_2 + \lambda_2 \lambda_3 & \lambda_1 \lambda_3 + \lambda_2 \lambda_3 \\
\lambda_1 & \lambda_2 & \lambda_3 \\
1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
\epsilon_1 \\
\epsilon_2 \\
\epsilon_3
\end{pmatrix}
= 
\begin{pmatrix}
d_4 d_3 p_{35} + d_2 d_3 (p_{35} + p_{32} p_{24}) \\
p_{35} d_3 \\
p_{35} d_3
\end{pmatrix}
\] (21)

It is obvious that the \( \epsilon_i \) found when solving the system of equations given in equation (20)
will be different from those when solving the system in equation (21), therefore denote \( \epsilon_{ik} \)
as being \( \epsilon_i \) from \( \chi_k \). It is now easy to write a MATLAB program to first determine \( \lambda_i \)
and then their corresponding \( \epsilon_{ik} \).

### 2.5 Finding \( p_1 \) and \( p_2 \)

Now that we have found the \( \lambda_i \) and their corresponding \( \epsilon_{ik} \), we now have to find the probability
of starting in state 2 (\( p_1 \)) and the probability of starting in state 3 (\( p_2 \)) which are found as follows.

First, let \( P \) denote the probability transition matrix such that
\[
P = 
\begin{pmatrix}
0 & p_{12} & 0 & 0 & 0 \\
p_{21} & 0 & p_{23} & 0 & 0 \\
0 & p_{32} & 0 & 0 & p_{35} \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{pmatrix}
\] (22)

As expected the rows of \( P \) sum to 1. We now define \( p_1 \) and \( p_2 \) as
\[
p_1 = \frac{\pi_4}{\pi_4 + \pi_5} \quad \text{and} \quad p_2 = \frac{\pi_5}{\pi_4 + \pi_5}
\] (23)

where \( \pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5) \), and \( \pi \) comes from \( \pi P = \pi \), i.e. the left eigenvector of \( P \)
corresponding to the eigenvalue of 1. Therefore, \( \pi_4 \) and \( \pi_5 \) can be computed either analytically
or numerically. If computed analytically, we find that
\[
p_1 = \frac{p_{24} p_{32}}{p_{24} p_{32} + p_{35} p_{23}} \quad \text{and} \quad p_2 = \frac{p_{35} p_{23}}{p_{24} p_{32} + p_{35} p_{23}}
\]

A problem arises when computing \( \pi \) numerically as MATLAB finds the right eigenvector
of \( P \). Therefore,
\[
\pi P = \pi \implies (\pi P)' = \pi' \implies P' \pi' = \pi'
\]

which is simply the right eigenvector of \( P' \).

Therefore, the three component mixture for the 5-state model in figure 1 is
\[
f(t) = \epsilon_1 \lambda_1 e^{-\lambda_1 t} + \epsilon_2 \lambda_2 e^{-\lambda_2 t} + \epsilon_3 \lambda_3 e^{-\lambda_3 t}
\] (24)

where
\[
\epsilon_i = p_1 \epsilon_{i2} + p_2 \epsilon_{i3}
\] (25)
2.6 Computation results

The aim of this section is to investigate the relative performance of the $L_2$ estimator when estimation is carried out with a two component mixture, but the underlying model is given by figure 1 with transition parameters that lead to the distribution of closed dwell times being a mixture of three component distributions, but where one component has a large mean but is present with a small proportion. This is when one or more of the $q_{ij}$ in figure 1 are close to 0. If we ignored these low transition rates we would be back to the model in figure 2.

We now consider a distribution for the 5-state model as given in equation (24), and we can now go ahead and numerically calculate the $\lambda_i$ and their corresponding $\epsilon_{ik}$ and then fit a 3-state model to the data.

First, let $q_{12} = 30, q_{21} = 4, q_{23} = 300, q_{32} = 600, q_{24} = 4$ and $q_{35} = 4000$. The resultant $\epsilon_{ik}$ and $\lambda_i$ are shown in table 7. As the system of figure 1 is reversible it follows that the $\epsilon_i \geq 0, i = 1, 2, 3$. However, it is not necessary that the coefficients $\epsilon_{ij}$ ($i = 2, 3$) be all non-negative.

We find from equation (23) that $p_1 = 0.0020$ and $p_2 = 0.9980$, and therefore $\epsilon_1 = 0.0025, \epsilon_2 = 0.8517$ and $\epsilon_3 = 0.1458$ (where $\epsilon_i$ is defined in equation (25)). We can then enter these $\epsilon_i$ and $\lambda_i$ into MATLAB and fit a biexponential model to the data from which we can then find the $q_{ij}$ of the 3-state model by using equations (10)–(12). Table 8 gives the estimates of $\epsilon_1, \lambda_1$ and $\lambda_2$ for the 3-state model and table 9 gives the resultant $q_{ij}$. Please note that the marginal distribution of the aggregated closed times does not depend on the parameters $q_{42}$ and $q_{53}$. However, the generated aggregated closed times are in fact dependent in this 5-state model and depend on the latter values. See Colquhoun and Hawkes [29]. For completeness, we generated these using $q_{42} = 1000$ and $q_{53} = 100$. An estimated correlation from these data is in the order of $10^{-4}$, that is the dependence is not strong.

It is clear from table 8 that $L_2$ performs better than the EM which again stands out with its large standard error for $\lambda_2$. Again this is explained by the estimates which range from 4113 to 5411. It is interesting also that the $L_2$ estimates for $q_{ij}$ are very close to the true $q_{ij}$

\begin{table}[h]
\centering
\begin{tabular}{cccccccc}
\hline
$\lambda_1$ & $\lambda_2$ & $\lambda_3$ & $\epsilon_{12}$ & $\epsilon_{13}$ & $\epsilon_{22}$ & $\epsilon_{23}$ & $\epsilon_{32}$ & $\epsilon_{33}$ \\
\hline
29.4982 & 4641.5367 & 266.9651 & 0.0189 & 0.0025 & -0.0591 & 0.8535 & 1.0402 & 0.1440 \\
\hline
\end{tabular}
\caption{Estimates for $\lambda_i$ and $\epsilon_{ik}$ for $q_{12} = 30, q_{21} = 4, q_{23} = 300, q_{32} = 600, q_{24} = 4$ and $q_{35} = 4000$.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{llllll}
\hline
Method & $\hat{\epsilon}$ & SE & $\hat{\lambda}_1$ & SE & $\hat{\lambda}_2$ & SE \\
\hline
$L_2$ & 0.1456 & 0.014 & 234.15 & 35.75 & 4651.3 & 15.20 \\
EM & 0.1438 & 0.015 & 231.78 & 36.88 & 4579.8 & 207.00 \\
\hline
\end{tabular}
\caption{Estimates for $\epsilon_i$ and $\lambda_i$ when $q_{12} = 30, q_{21} = 4, q_{23} = 300, q_{32} = 600, q_{24} = 4, q_{42} = 1000, q_{35} = 4000$ and $q_{53} = 100$.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{llllll}
\hline
Method & $\hat{q}_{23}$ & SE & $\hat{q}_{32}$ & SE & $\hat{q}_{35}$ & SE \\
\hline
$L_2$ & 301.10 & 29.358 & 609.734 & 55.11 & 3982.70 & 86.11 \\
EM & 274.024 & 39.61 & 592.76 & 65.18 & 3953.69 & 169.32 \\
\hline
\end{tabular}
\caption{Estimates for $q_{23}, q_{32}$ and $q_{35}$ when $q_{12} = 30, q_{21} = 4, q_{23} = 300, q_{32} = 600, q_{24} = 4, q_{42} = 1000, q_{35} = 4000$ and $q_{53} = 100$.}
\end{table}
from the 5-state model. Essentially, the robustness in terms of weak continuity and Fréchet
differentiability in neighbourhoods of the parametric model indicated the asymptotic theory
in Clarke [12, 21, 22] and Clarke and Heathcote [10] is reflected in the stability of the $L_2$
estimates.

Finally, we consider the case where $q_{24} = 0$ and $q_{21} = 10, 4$ and $1$. We have essentially
eliminated state $O_4$ in figure 1, which leads to $p_1 = 0$ and $p_2 = 1$ which implies that $\epsilon_i = \epsilon_3$.
Table 10 gives the $\epsilon_i$ and $\lambda_i$ for the various $q_{21}$.

If we have $q_{21} = 0$ then we have the last model mentioned in section 2.2. Using the above
$\epsilon_i$ and $\lambda_i$, and performing the same operations as above we get the results given in table 11
(included is table 5 along with the true $\epsilon_1, \lambda_1$ and $\lambda_2$ for comparison). We notice from table 11
the performance of $L_2$ compared to EM. As $q_{21}$ changes from 10 to 4 to 1, the $L_2$ estimates
change very little compared to the MLE estimates, and we also notice how close the $L_2$
estimates are to the true $\epsilon_1, \lambda_1$ and $\lambda_2$ when $q_{21} = q_{24} = 0$.

Table 11. Estimates of $\epsilon_1, \lambda_1$ and $\lambda_2$ for different $q_{21}$.

<table>
<thead>
<tr>
<th>$q_{21}$</th>
<th>$\hat{\epsilon}_1$</th>
<th>$\hat{\epsilon}_2$</th>
<th>$\hat{\epsilon}_3$</th>
<th>$\hat{\lambda}_1$</th>
<th>$\hat{\lambda}_2$</th>
<th>$\hat{\lambda}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.138</td>
<td>0.015</td>
<td>232.1</td>
<td>30.44</td>
<td>4569.9</td>
<td>210.65</td>
</tr>
<tr>
<td></td>
<td>0.133</td>
<td>0.015</td>
<td>189.3</td>
<td>39.70</td>
<td>4517.2</td>
<td>212.52</td>
</tr>
<tr>
<td>4</td>
<td>0.144</td>
<td>0.014</td>
<td>249.2</td>
<td>26.91</td>
<td>4626.8</td>
<td>201.19</td>
</tr>
<tr>
<td></td>
<td>0.140</td>
<td>0.014</td>
<td>227.2</td>
<td>35.85</td>
<td>4591.3</td>
<td>198.11</td>
</tr>
<tr>
<td>1</td>
<td>0.146</td>
<td>0.014</td>
<td>258.4</td>
<td>27.49</td>
<td>4651.8</td>
<td>203.60</td>
</tr>
<tr>
<td></td>
<td>0.144</td>
<td>0.014</td>
<td>253.1</td>
<td>30.03</td>
<td>4637.3</td>
<td>201.63</td>
</tr>
<tr>
<td>0</td>
<td>0.147</td>
<td>0.015</td>
<td>259.5</td>
<td>25.87</td>
<td>4649.0</td>
<td>195.26</td>
</tr>
<tr>
<td></td>
<td>0.146</td>
<td>0.015</td>
<td>260.3</td>
<td>25.28</td>
<td>4644.2</td>
<td>186.94</td>
</tr>
<tr>
<td>0</td>
<td>True</td>
<td>–</td>
<td>258.5</td>
<td>–</td>
<td>4641.5</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 12. Estimates for $q_{23}$, $q_{32}$ and $q_{35}$ given the estimates for $\epsilon_1$, $\lambda_1$ and $\lambda_2$ as given in table 11.

<table>
<thead>
<tr>
<th>$q_{21}$</th>
<th>$\hat{q}_{23}$</th>
<th>$\hat{q}_{32}$</th>
<th>$\hat{q}_{35}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>267.5</td>
<td>36.90</td>
<td>565.8</td>
</tr>
<tr>
<td></td>
<td>217.4</td>
<td>47.12</td>
<td>550.0</td>
</tr>
<tr>
<td>4</td>
<td>288.6</td>
<td>33.11</td>
<td>590.3</td>
</tr>
<tr>
<td></td>
<td>262.4</td>
<td>43.12</td>
<td>577.4</td>
</tr>
<tr>
<td>1</td>
<td>299.5</td>
<td>33.5</td>
<td>598.9</td>
</tr>
<tr>
<td></td>
<td>293.3</td>
<td>36.5</td>
<td>593.8</td>
</tr>
<tr>
<td>0</td>
<td>301.44</td>
<td>32.13</td>
<td>602.0</td>
</tr>
<tr>
<td></td>
<td>302.26</td>
<td>31.26</td>
<td>600.74</td>
</tr>
<tr>
<td>0</td>
<td>True</td>
<td>300</td>
<td>–</td>
</tr>
</tbody>
</table>
We conclude this discussion with a look at the $q_{ij}$. Now that we have $\hat{\epsilon}_1$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ we can also determine the $\hat{q}_{ij}$ for the 3-state model, as given in table 12. The standard errors of the $\hat{q}_{ij}$ for the $L_2$ and EM estimates are comparable, however, we can observe that as $q_{21}$ increases the $L_2$ provides estimates with smaller bias. We can conclude from these analyses that $L_2$ out-performs EM when we try to fit a 3-state model to data from a 5-state model where for the closed dwell time distribution one of the $\epsilon_i$ is a small proportion, corresponding to a component distribution with a large mean.

3. Conclusion

We have introduced into the ion channel area the use of a robust estimator defined by the minimizing equations of the $L_2$ distance given by equation (2). We have shown how the resulting $L_2$ estimator is competitive or in fact improves on the maximum likelihood estimates when a 3-state model is correctly specified. In addition, it is able to capture the dominant structure in terms of transition rates of a 3-state model when in fact the underlying kinetic model is a 5-state model, but where two of the transition rates are near 0. Essentially, as is indicated in the study of Clarke and Heathcote [10] this is because the $L_2$ estimator has a bounded influence function, whereas the maximum likelihood estimator is not robust and has an unbounded influence function. The $\psi$ function for the equation (2) which minimise $J_n(\theta)$ is a bounded function of the observation space variable, whereas the efficient score function is unbounded in this regard. Thus, while the $L_2$ estimator can perform in a stable manner under slight misspecification of the underlying model, the MLE in comparison is quite sensitive to even slight departures from the proposed model. The $L_2$-estimating functional derived from equation (2) is weakly continuous and Fréchet differentiable yielding a consistent and asymptotically normal estimator [10, 12, 22 for related discussions of the theory].

Acknowledgements

The authors have benefited from correspondence, e-mail and conversations from a number of people through the preparation of this paper. Particular thanks are owing to Prof. F. G. Ball, Prof. C. R. Heathcote, Prof. T. Krishnan, Dr. R. K. Milne, Dr. N. Kahn, Dr. G. F. Yeo, and Mr. C. De Haer. The first author thanks the Department of Mathematics and Statistics at the University of Western Australia where he spent a profitable sabbatical in preparing some of this paper.

References

Appendix A  Estimating equations for \( k \)-component mixtures of exponentials

Here, we consider general expressions for estimating a mixture of \( k \)-component exponentials.

\[
F_{\theta}(x) = \sum_{j=1}^{k} \epsilon_j G(x; \lambda_j) = \sum_{j=1}^{k} \epsilon_j (1 - e^{-\lambda_j x})
\]

\[
= 1 - \epsilon_1 e^{-\lambda_1 x} - \cdots - \epsilon_{k-1} e^{-\lambda_{k-1} x} - (1 - \epsilon_1 - \cdots - \epsilon_{k-1}) e^{-\lambda_k x} \quad (A1)
\]

\[
= (1 - e^{-\lambda_k x}) + \epsilon_1 (e^{-\lambda_1 x} - e^{-\lambda_k x}) + \cdots + \epsilon_{k-1} (e^{-\lambda_{k-1} x} - e^{-\lambda_k x})
\]

where we have the constraints \( \sum_{j=1}^{k} \epsilon_j = 1 \), and \((\epsilon_j, \lambda_j)\) are both positive, \( j = 1, \ldots, k \); \( \lambda_i \neq \lambda_j \) for \( i \neq j \).
Again consider minimizing the distance $J_n(\theta)$ to give equations in the form of equation (2). See equations (3) and (4) of Clarke and Heathcote [10] for example. Consider now the first $k - 1$ derivatives with respect to the mixing proportions

$$
\frac{\partial}{\partial \epsilon_i} F_\theta(x) = e^{-\lambda_i x} - e^{-\lambda_k x}, \quad i = 1, \ldots, k - 1
$$

Denote

$$
A_i(x) = \int_0^x \frac{\partial}{\partial \epsilon_i} F_\theta(y) \, dy
$$

$$
= \frac{1}{\lambda_i} e^{-\lambda_i x} - \frac{1}{\lambda_k} e^{-\lambda_k x} - \frac{1}{\lambda_i} + \frac{1}{\lambda_k}
$$

$$
E_\theta[A_i(X)] = \int_0^\infty \left( \frac{1}{\lambda_i} e^{-\lambda_i x} - \frac{1}{\lambda_k} e^{-\lambda_k x} \right) [\epsilon_1 \lambda_1 e^{-\lambda_1 x} + \cdots + \epsilon_{k-1} \lambda_{k-1} e^{-\lambda_{k-1} x}]
$$

$$
+ (1 - \epsilon_1 - \cdots - \epsilon_{k-1}) \lambda_k e^{-\lambda_k x} \right) \, dx - \frac{1}{\lambda_i} + \frac{1}{\lambda_k}
$$

$$
= \frac{1}{\lambda_i} B_{i,k}(\theta) - \frac{1}{\lambda_k} B_{k,k}(\theta) - \frac{1}{\lambda_i} + \frac{1}{\lambda_k}
$$

where

$$
B_{i,k}(\theta) = \frac{\lambda_k}{\lambda_i + \lambda_k} + \epsilon_1 \left[ \frac{\lambda_1}{\lambda_1 + \lambda_i} - \frac{\lambda_k}{\lambda_k + \lambda_i} \right] + \cdots + \epsilon_{k-1} \left[ \frac{\lambda_{k-1}}{\lambda_{k-1} + \lambda_i} - \frac{\lambda_k}{\lambda_k + \lambda_i} \right]
$$

The first $k - 1$ estimating equations involve functions

$$
\psi_i(x; 0) = \frac{1}{\lambda_i} e^{-\lambda_i x} - \frac{1}{\lambda_k} e^{-\lambda_k x} - \frac{1}{\lambda_i} B_{i,k}(\theta) + \frac{1}{\lambda_k} B_{k,k}(\theta)
$$

for $i = 1, \ldots, k - 1$. Now consider

$$
C_i(x) = \int_0^x \frac{\partial}{\partial \lambda_i} F_\theta(y) \, dy
$$

$$
= \frac{\epsilon_i}{\lambda_i} C_i^*(x)
$$

where

$$
C_i^*(x) = \frac{1}{\lambda_i} - (xe^{-\lambda_i x} + \frac{1}{\lambda_i} e^{-\lambda_i x})
$$

Also note

$$
C_i(\infty) = \frac{\epsilon_i}{\lambda_i^2}
$$

and

$$
C_i^*(\infty) = \frac{1}{\lambda_i}
$$

Now

$$
E_\theta[C_i^*(X)] = \frac{1}{\lambda_i} - D_{i,k}(\theta)
$$
where

\[ D_{i,k} = \frac{\lambda_k}{(\lambda_i + \lambda_k)^2} + \left[ \epsilon_1 \left( \frac{\lambda_1}{(\lambda_1 + \lambda_i)^2} - \frac{\lambda_k}{(\lambda_k + \lambda_i)^2} \right) + \cdots + \epsilon_{k-1} \left( \frac{\lambda_{k-1}}{(\lambda_{k-1} + \lambda_i)^2} - \frac{\lambda_k}{(\lambda_k + \lambda_i)^2} \right) \right] + \frac{\lambda_k}{\lambda_i(\lambda_i + \lambda_k)} + \frac{\lambda_k}{\lambda_i(\lambda_i + \lambda_k)} \]

Therefore, the last \( k \) equations can be represented as

\[ \psi_i'(x; \theta) = -xe^{-\lambda_i x} - \frac{1}{\lambda_i}e^{-\lambda_i x} + D_{i,k}(\theta) \]

\[ i = i' - k + 1, \ i' = k, \ldots, 2k - 1 \]

It can be checked that these formulae are consistent with the equations derived for a mixture of two exponentials. See the next section for details.

**Appendix B  Restriction to the case \( k = 2 \)**

Consider

\[ \psi_1(x; \theta) = \frac{1}{\lambda_1}e^{-\lambda_1 x} - \frac{1}{\lambda_2}e^{-\lambda_2 x} - \frac{1}{\lambda_1}B_{1,2}(\theta) + \frac{1}{\lambda_2}B_{2,2}(\theta) \]

Here

\[ B_{1,2}(\theta) = \frac{\lambda_2}{\lambda_1 + \lambda_2} + \epsilon_1 \left[ \frac{\lambda_1}{2\lambda_1} - \frac{\lambda_2}{\lambda_1 + \lambda_2} \right] \]

and

\[ B_{2,2}(\theta) = \frac{\lambda_2}{\lambda_2 + \lambda_2} + \epsilon_1 \left[ \frac{\lambda_1}{\lambda_1 + \lambda_2} - \frac{\lambda_2}{\lambda_2 + \lambda_2} \right] \]

\[ = \frac{1}{2} + \epsilon_1 \left[ \frac{\lambda_1}{\lambda_1 + \lambda_2} - \frac{1}{2} \right] \]

Therefore

\[ \psi_1(x; \theta) = \frac{1}{\lambda_1}e^{-\lambda_1 x} - \frac{1}{\lambda_2}e^{-\lambda_2 x} - \frac{\lambda_2}{\lambda_1(\lambda_1 + \lambda_2)} - \epsilon_1 + \epsilon_1 \frac{\lambda_2}{\lambda_1(\lambda_1 + \lambda_2)} + \frac{2\lambda_1}{\lambda_1(\lambda_1 + \lambda_2)} + \frac{2\lambda_1}{\lambda_1(\lambda_1 + \lambda_2)} \]

\[ + \frac{\epsilon_1 \lambda_1}{\lambda_2(\lambda_1 + \lambda_2)} - \frac{\epsilon_1}{2\lambda_2} \]

It is also easy to see that

\[ \psi_2(x; \theta) = -xe^{-\lambda_1 x} - \frac{1}{\lambda_1}e^{-\lambda_1 x} + D_{1,2}(\theta) \]

and

\[ \psi_3(x; \theta) = -xe^{-\lambda_2 x} - \frac{1}{\lambda_2}e^{-\lambda_2 x} + D_{2,2}(\theta) \]
In fact as can be verified by direct minimisation of $J_n(\theta)$

$$\psi_2(x; \theta) = \left\{ \frac{1}{\lambda_1} - xe^{-\lambda_1 x} - \frac{1}{\lambda_1} e^{-\lambda_1 x} \right\} - \int_0^\infty \left( \frac{1}{\lambda_1} - xe^{-\lambda_1 x} - \frac{1}{\lambda_1} e^{-\lambda_1 x} \right) dF_\theta(x)$$

$$= -xe^{-\lambda_1 x} - \frac{1}{\lambda_1} e^{-\lambda_1 x} + \frac{3\epsilon}{4\lambda_1} + \frac{(1 - \epsilon) \lambda_2}{(\lambda_1 + \lambda_2)^2} + \frac{(1 - \epsilon) \lambda_2}{\lambda_1 (\lambda_1 + \lambda_2)}$$

Similarly,

$$\psi_3(x; \theta) = -xe^{-\lambda_2 x} - \frac{1}{\lambda_2} e^{-\lambda_2 x} + \frac{\epsilon \lambda_1}{(\lambda_1 + \lambda_2)^2} + \frac{3(1 - \epsilon)}{4\lambda_2} + \frac{\epsilon \lambda_1}{\lambda_2 (\lambda_1 + \lambda_2)}$$