Investigating Markov model discrimination for ion channels

BY MATTHEW E. A. HODGSON AND PETER J. GREEN
Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK

First revision: 21st August 1998

Hodgson (1998) formulates a Bayesian method of ion channel analysis, in which he uses an alternating renewal model of channel kinetics and an autoregressive process to model patch clamp noise. He computes the resulting posterior by a reversible jump Markov chain Monte Carlo method (Green 1995). Our focus here is on extending this analysis to the more physically realistic hidden Markov models of channel dynamics. Here, the states of the channel are partitioned into classes of states having the same conductance, so that patch clamp recording is informative only about the aggregated process. This loss of information increases the challenge of making structural inference about the underlying Markov chain. The study we present, involving discrimination between a few simple models, serves to give some indication as to the feasibility of more general Markov model determination for ion channels.

1 Hidden Markov models for ion channels

Ion channels, large proteins spanning cell membranes, are thought to exist in a finite number of physicochemically distinct states, some of which allow the passage of current in the form of selected ions. Their kinetics are most frequently modelled by a continuous-time Markov chain in which the state space is partitioned into classes of states having the same conductance. Since it is the current through the channel which is measured (with noise), states having the same conductance are indistinguishable. A thorough exposition of the relevant theoretical properties of hidden Markov models is given by Colquhoun and Hawkes (1982). In theory the technique of patch clamp recording (see, for example, Sakmann and Neher 1995) reveals which class the channel state belongs to, so that inference about the postulated 'hidden' Markov chain may be based on this so-called 'aggregated' process. Such inference is the subject of Fredkin and Rice (1986). In practice, degradation of patch clamp records by additive noise and filtering prior to digitisation further complicates the statistician's task. We restrict attention to the usual case of just two conductance classes, termed 'open' and 'closed'.

A conventional algorithm to restore the true channel current from a patch clamp record involves thresholding followed by the deletion of apparent sojourns in either class of duration less than some critical value, a practice known as "time interval omission". The use of the thresholded signal as data, with and without time interval omission, in making inference about hidden Markov models has received much attention; see, for example, Ball *et al.* (1993) and Edeson *et al.* (1994). An unwieldy likelihood hampers frequentist inference in the case of time interval omission. Means of overcoming this difficulty, in the case of the simplest two-state Markov model, are presented by Ball, Chen, and Sansom (1992) and Ball and Davies (1995). From the Bayesian viewpoint, Ball *et al.* (1996) analyse discrete ion channel data simulated from the two-state model, using Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution of the transition matrix.

Some authors base inference on raw rather than thresholded patch clamp data; see Magleby and Weiss (1990) and Fredkin and Rice (1992b) for frequentist approaches. Fredkin and Rice (1992a) use recursive algorithms to compute marginal and global Bayes restorations of ion channel signals based on a two-state Markov chain prior and a Gaussian 'white noise' likelihood. Hodgson (1998) treats the opening and closing of the channel as an alternating renewal process with gamma distributed sojourn times, without incorporating a more fundamental hidden stochastic process to drive the aggregated process. He draws inference by generating a sample from the joint posterior distribution of the sojourn time parameters, the channel conductance levels and the binary step function representing the true channel current. The unknown number of discontinuities ('switches') of this function induces variable dimensionality in the joint posterior, the situation for which Green's (1995) reversible jump protocol was designed. The approach of Ball *et al.* (1997) is similar, except that they treat the latent process as a Markov chain and incorporate it explicitly into their model. The realised path of this Markov chain, rather than the derived aggregated process, is the variable-dimension object updated by reversible jump moves in their sampler. In addition to the two-state and three-state linear models (models *11* and *21* in the notation we introduce in the next section), they consider three-state cyclic and five-state models for the channel kinetics.

This paper extends Hodgson's methods to the more familiar and physically realistic class of hidden Markov models. Kienker (1989) studies classes of *equivalent* models; that is distinct Markov chains yielding probabilistically identical aggregated processes. A strategy for discriminating between alternative models, based on the Schwarz criterion, is presented by Ball and Sansom (1989). We shall attempt more fully Bayesian model selection from a set of four simple candidate models. The numbers of open and closed states and the transition rates of the Markov chain will all be unknown. Hence our MCMC sampler requires reversible jump moves to update the model in addition to those used by Hodgson to update the step function. Although our models are mathematically identifiable, the comprehensive inference we attempt may be infeasible owing to the low information content of the data. Our study should reveal whether the idea of Markov model determination for ion channels has the potential to be generalised.

In section 2 we introduce the set of candidate models and derive the implied sojourn time densities that are non-trivial, we derive it. We also describe both the Bayesian model for ion channel model determination and the reversible jump MCMC algorithm to sample from the posterior distribution. Section 3 illustrates the performance of our method on data simulated from one of the candidate models and on the same synthetic data used by Hodgson (1998). Section 4 addresses the issues of MCMC mixing and convergence, and section 5 details an investigation into the sensitivity of results to various factors. A discussion concludes the paper.

2 Discrimination between four simple Markov models

2.1 Introduction

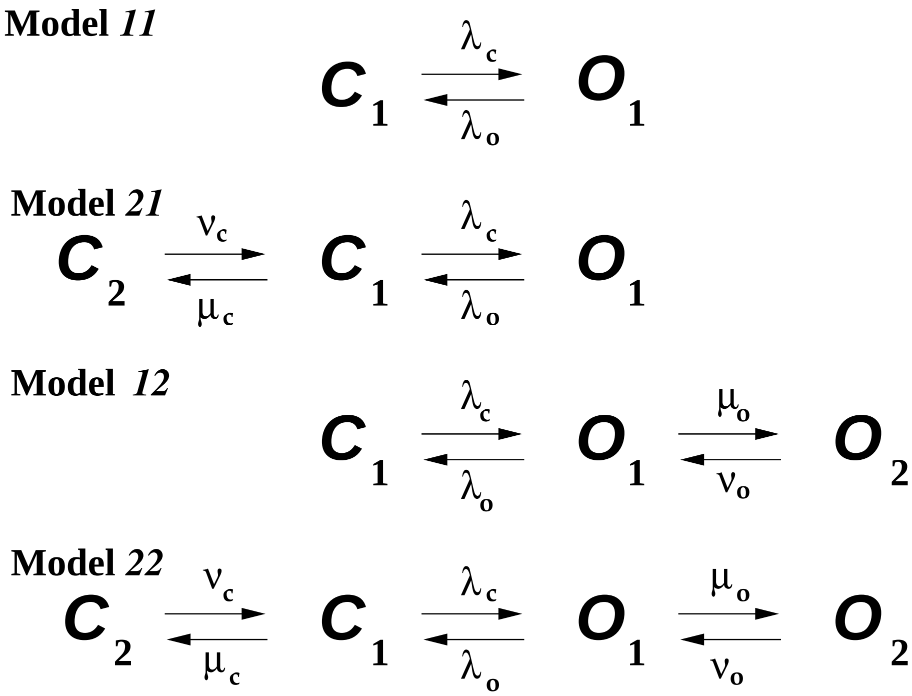

Figure 1: The candidate Markov models

The four models of ion channel kinetics considered in this paper are shown in figure 1. As these models have at most two states in each class and a single gateway between the classes, deriving the probability density function of the sojourn time in a two-state class is a simple exercise. This enables us to integrate the latent Markov process out analytically and include only the aggregated process in our ion channel model. We will simulate segments of raw data from the different Markov models for analysis and then sample from the joint posterior distribution of the model indicator, transition rates and aggregated process. The chief interest will be the extent to which the data 'select' the true model if we are indifferent between them *a priori*.

2.2 The density of the sojourn time in a two-state class

Consider Markov model *21* illustrated in figure 1. Closed sojourns of course begin and end in state $\mathcal{C}_1$; let $N = 0, 1, 2, \ldots$ be the number of visits to state $\mathcal{C}_2$ during a particular closed sojourn. Since the probability that the jump immediately after a sojourn in $\mathcal{C}_1$ is to $\mathcal{C}_2$ is $\mu_c/(\mu_c+\lambda_c)$, irrespective of the holding time in $\mathcal{C}_1$, then $N+1$ has the geometric distribution, parameter $\lambda_c/(\mu_c+\lambda_c)$. Conditional on N, the closed sojourn time T is the sum of $N+1$ independent exponentials, rate $\mu_c+\lambda_c$, and N independent exponentials, rate ν_c. Denote by $f_c(t)$ the density of the closed sojourn times. Then the Laplace transform of the density is

$$\begin{aligned}
\widehat{f}_c(\theta) &= E[\exp(-\theta T)] \\
&= E\left[E[\exp(-\theta T)|N]\right] \\
&= E\left[E[\exp(-\theta(X_N+Y_N))]\right], \ X_N \sim \Gamma(N+1, \mu_c+\lambda_c), \ Y_N \sim \Gamma(N, \nu_c) \\
&\quad \text{independently} \\
&= E\left[\left(\frac{\mu_c+\lambda_c}{\mu_c+\lambda_c+\theta}\right)^{N+1} \left(\frac{\nu_c}{\nu_c+\theta}\right)^{N}\right] \\
&= \frac{\lambda_c}{\mu_c+\lambda_c+\theta} \sum_{n=0}^{\infty} \left(\frac{\mu_c+\lambda_c}{\mu_c+\lambda_c+\theta}\right)^{n} \left(\frac{\nu_c}{\nu_c+\theta}\right)^{n} \left(\frac{\mu_c}{\mu_c+\lambda_c}\right)^{n} \\
&= \frac{\lambda_c(\nu_c+\theta)}{\theta^2+\theta(\nu_c+\mu_c+\lambda_c)+\nu_c\lambda_c}.
\end{aligned}$$

This may be inverted to give

$$f_c(t) = (1/2-c)(a-b)\ \exp(-(a-b)t) + (1/2+c)(a+b)\ \exp(-(a+b)t), \ t > 0, \quad (1)$$

where $a = (\nu_c + \mu_c + \lambda_c)/2$, $b = \sqrt{a^2 - \nu_c\lambda_c}$ and $c = (\lambda_c - a)/2b$. It is easily shown that c lies in the interval $(-1/2, 1/2)$. Thus T has the surprisingly simple form of a mixture of two exponential distributions (with positive weights). This is a special case of a result, known for many years (see, for example, Colquhoun and Hawkes 1981), that the sojourn time in a size n subset of the state space of a finite continuous-time Markov chain has in general an exponential mixture distribution with n components.

2.3 The Bayesian model

Figure 2 illustrates the joint distribution of the Markov model indicator, transition rates, aggregated process and data by its directed acyclic graph (DAG). As usual circles contain random quantities and squares contain fixed or observed quantities. The four-dimensional vector m gives the prior probabilities assigned to models *11*—*22* shown in section 2.1, and $k \in \mathcal{K} = \{11, 21, 12, 22\}$ is the model indicator. Given k, Λ is the vector of transition rates, of dimension 2 if $k = 11$, 4 if $k = 21$ or *12* and 6 if $k = 22$. If there are two closed states ($k = 21$ or *22*), the three transition rates determining the density of the closed sojourn times are given independent gamma priors:

$$\nu_c,\ \mu_c,\ \lambda_c\ \sim \Gamma(\alpha, \beta/2).$$

In the limit as μ_c tends to 0, the sojourn time distribution collapses to the $\mathrm{Exp}(\lambda_c)$ distribution. This is of course the sojourn time distribution when there is a single closed state, in which case λ_c has a gamma distribution but with a different inverse scale parameter:

$$\lambda_c \sim \Gamma(\alpha, \beta).$$

Transition rates governing open sojourn times are assigned priors in the same manner. When there are two closed states, the expected closed sojourn time is

$$E[T] = \frac{\mu_c + \nu_c}{\nu_c\ \lambda_c}.$$

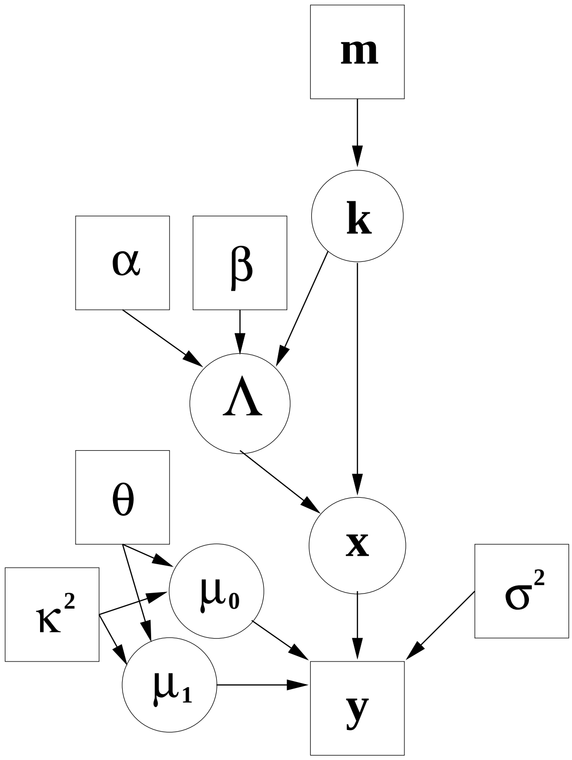

Figure 2: The DAG corresponding to the ion channel model

Therefore halving the inverse scale parameters of the priors on the transition rates, and hence doubling their prior means, has the effect of stabilising the location of the closed sojourn time distribution in moving from one to two closed states; this is its motivation.

The rest of the symbols appearing in figure 2 have exactly the same interpretation as in Hodgson (1998). In summary, $\mathbf{x}$ is the indicator function of channel openness on the continuous interval $[0, T]$ over which the patch clamp recording is taken and μ_1 and μ_0 are the channel conductance levels in the open and closed classes respectively. The vector $\mathbf{y}$ of data points is a digitised filtered noisy version of the binary channel current. The reader is referred to Hodgson for further details and for a justification of the treatment of the noise variance σ^2 as known.

2.4 Design of the reversible jump sampler

We require to simulate from the joint posterior

$$p(k, \Lambda, \mathbf{x}, \mu_0, \mu_1 \mid \mathbf{y}) \propto p(k)p(\Lambda|k)p(\mathbf{x}|\Lambda, k)p(\mu_0)p(\mu_1)p(\mathbf{y}|\mathbf{x}, \mu_0, \mu_1).$$

First, consider the case where the channel is closed at time $t = 0$ and there are an even number of switches between the closed and open classes in the interval of observation. Let $s = 2j$ be the number of switches and $0 < t_1 < \ldots < t_{2j} < T$ be the switch times. Then there are j open sojourns on the intervals $[t_1, t_2]$, $\ldots$, $[t_{2j-1}, t_{2j}]$, $(j-1)$ closed sojourns on $[t_2, t_3]$, $\ldots$, $[t_{2j-2}, t_{2j-1}]$ and two 'censored' closed sojourns on $[0, t_1]$ and $[t_{2j}, T]$. The conditional distribution of $\mathbf{x}$ given (Λ, k) is

$$
\begin{aligned}
p(\mathbf{x}|\Lambda, k) &= \frac{m_c}{m_c + m_o} \times \frac{1 - F_c(t_1|\Lambda, k)}{m_c} \times \prod_{i=1}^{j} f_o(t_{2i} - t_{2i-1}|\Lambda, k) \\
&\times \prod_{i=1}^{j-1} f_c(t_{2i+1} - t_{2i}|\Lambda, k) \times (1 - F_c(T - t_{2j}|\Lambda, k)). \qquad (2)
\end{aligned}
$$

Here, $f_c(.|\Lambda, k)$ and $f_o(.|\Lambda, k)$ are the density functions of closed and open sojourn times implied by the Markov model k and transition rates Λ, and m_c and m_o are the mean closed and open sojourn times. The modifications to the above expression when $\mathbf{x}$ has an odd number of switches, when the channel is open initially or when both occur are obvious.

The moves implemented in our sampler are:

1. updating the Markov model indicator k, necessarily altering Λ simultaneously,
2. updating the transition rates Λ for a given k,
3. updating the indicator function $\mathbf{x}$ for given (k, Λ), and
4. updating the channel conductance levels (μ_0, μ_1).

Note that the reversible jump technique (Green 1995) is in general needed in the first and third of the above moves, as Λ and $\mathbf{x}$ have variable dimensionality. We describe only the first move in detail. Suppose the current model is the two-state one, $k = \mathit{11}$. With probability 0.5, we propose to create a second closed state; otherwise we propose a second open state. In either case, the dimensionality of the corresponding transition rate vector Λ increases from 2 to 4. Following Green's recipe, we generate a two-dimensional random vector $\mathbf{z} = (z_1, z_2)$ independently of the existing rates and set

$$\Lambda' = \Lambda'(\Lambda, \mathbf{z});$$

that is, the new transition rates Λ' are a deterministic function of the existing rates Λ and the random vector $\mathbf{z}$. Suppose that a second closed state is proposed. We shall preserve the mean sojourn times, so that the new transition rates $(\lambda_c', \mu_c', \nu_c', \lambda_o')$ are related to the existing rates (λ_c, λ_o) by:

$$
\begin{aligned}
\frac{\mu_c' + \nu_c'}{\nu_c' \lambda_c'} &= \frac{1}{\lambda_c} \\
\lambda_o' &= \lambda_o. \qquad (3)
\end{aligned}
$$

A necessary condition to satisfy the first equation is $\lambda_c' > \lambda_c$. We draw $z_1 \sim \Gamma(\rho_1, \rho_2)$ and set

$$\lambda_c' = \lambda_c \ \exp(z_1).$$

As both are transition rates out of the inner closed state $\mathcal{C}_1$, we relate λ_c' and μ_c' by a multiplicative perturbation. We generate $z_2 \sim N(0, \tau^2)$ independently of z_1 and set

$$\mu_c' = \lambda_c \ \exp(z_1 + z_2).$$

Here, ρ_1, ρ_2 and τ are simulation parameters. The final rate $\nu_c^{'}$ is determined by (3):

$$\nu_c^{'} = \frac{\lambda_c \exp(z_1 + z_2)}{\exp(z_1) - 1}.$$

In the reverse move jumping from Markov model *21* to model *11*, which is chosen with probability 0.5 when is model is currently *21*, the inverse of the transformation specified above must be used for the new transition rates, namely:

$$\begin{aligned}
\lambda_c^{'} &= \frac{\nu_c \, \lambda_c}{\nu_c + \mu_c} \\
\lambda_o^{'} &= \lambda_o \\
z_1 &= \log\left(\frac{\nu_c + \mu_c}{\nu_c}\right) \\
z_2 &= \log\left(\frac{\mu_c}{\lambda_c}\right).
\end{aligned}$$

From Green (1995), the appropriate acceptance probability for the move creating a second closed state is

$$\min\left\{1, \frac{p(21)p(\Lambda^{'}|k = 21)p(\mathbf{x}|\Lambda^{'}, k = 21)}{p(11)p(\Lambda|k = 11)p(\mathbf{x}|\Lambda, k = 11)q(\mathbf{z})} \left|\frac{\partial \Lambda^{'}}{\partial(\Lambda, \mathbf{z})}\right|\right\}, \tag{4}$$

where $q(.)$ is the density of $\mathbf{z}$. The term $p(21)/p(11)$ is read directly from the vector m, $p(\Lambda^{'}|21)/p(\Lambda|11)$ is

$$\frac{\beta^{2\alpha}}{2^{3\alpha}(\Gamma(\alpha))^2}\left(\frac{\nu_c^{'}\mu_c^{'}\lambda_c^{'}}{\lambda_c}\right)^{\alpha-1} \exp(-\beta/2(\nu_c^{'} + \mu_c^{'} + \lambda_c^{'} - 2\lambda_c)),$$

the density of $\mathbf{z}$ is

$$q(\mathbf{z}) = \frac{\rho_2^{\rho_1}}{\Gamma(\rho_1)\tau\sqrt{2\pi}} z_1^{\rho_1 - 1} \exp(-\rho_2 z_1 - z_2^2/2\tau^2),$$

and the transformation of variables from $(\lambda_c, \lambda_o, z_1, z_2)$ to $(\nu_c^{'}, \mu_c^{'}, \lambda_c^{'}, \lambda_o^{'})$ has Jacobian

$$\left|\frac{\partial \Lambda^{'}}{\partial(\Lambda, \mathbf{z})}\right| = \frac{\nu_c^{'2} \lambda_c^{'2}}{\lambda_c^2}.$$

The term $p(\mathbf{x}|\Lambda^{'}, k = 21)/p(\mathbf{x}|\Lambda, k = 11)$ follows from (2); note that the open sojourn density and the mean closed sojourn time are unchanged by the proposals. The acceptance probability of the reverse move has the same form as (4) with the ratio terms inverted and some obvious relabelling of variables.

Proposals to create or delete the second open state $\mathcal{O}_2$ are generated in exactly the same way as for the second closed state $\mathcal{C}_2$. Given the current model k, it is proposed to change the number of closed states (from 1 to 2 or from 2 to 1) with probability 0.5 and change the number of open states otherwise. Hence k may not change from *11* to *22* or from *21* to *12* in a single move.

Turning now to the second move type, the proposed updates to the transition rates for a given model are multiplicative perturbations of their current values. If $k = 22$ for example, we draw $z_i \sim N(0, \tau^2)$, $i = 1, \ldots, 6$ independently and set

$$\nu_c^{'} = \nu_c \exp(z_1), \ \mu_c^{'} = \mu_c \exp(z_2), \ \ldots, \nu_o^{'} = \nu_o \exp(z_6).$$

The parameter τ takes the same value as in the model-changing move type. The acceptance probability for this Metropolis-Hastings joint update is $\min\{1, R\}$, where

$$\begin{aligned} R &= \text{full conditional ratio} \times \text{ proposal ratio} \\ &= \left(\frac{\nu_c{}'\mu_c{}'\lambda_c{}'\lambda_o{}'\mu_o{}'\nu_o{}'}{\nu_c\mu_c\lambda_c\lambda_o\mu_o\nu_o}\right)^{\alpha} \times \frac{p(\mathbf{x}|\Lambda', k = \mathit{22})}{p(\mathbf{x}|\Lambda, k = \mathit{22})} \\ &\times \; e^{-\beta(\nu_c{}'+\mu_c{}'+\lambda_c{}'+\lambda_o{}'+\mu_o{}'+\nu_o{}'-\nu_c-\mu_c-\lambda_c-\lambda_o-\mu_o-\nu_o)/2}. \end{aligned}$$

If $k = \mathit{11}$, only 2 normal variates need be drawn; if $k = \mathit{21}$ or $\mathit{12}$, 4 need be drawn.

The design of the remainder of the moves and the sampling scheme follows that of Hodgson (1998), the only difference being that the distribution of the function $\mathbf{x}$ given the channel's kinetic parameters is now given by (2), causing changes in the expressions for the acceptance probabilities of the moves updating $\mathbf{x}$.

3 Performance of the model discrimination methodology

3.1 Data simulated from the maximal model

A useful and easily implemented means of partially validating any MCMC sampler intended to simulate from a posterior distribution is to disconnect the data from the model and simulate from the prior. The output can then be compared with the chosen prior. We simulated from the prior

$$p(k, \Lambda, \mathbf{x}) = p(k)p(\Lambda|k)p(\mathbf{x}|\Lambda, k),$$

checking that the Markov model distribution $p(k)$, the priors on the transition rates $p(\Lambda|k)$ and the indicator function distribution $p(\mathbf{x}|\Lambda, k)$ were satisfactorily approximated by the empirical distributions. Having satisfied ourselves of the correctness of the code, we proceeded with simulating from the posterior.

The 'truth' underlying the first data set we simulated was $k = \mathit{22}$, $\nu_c = \mu_c = 0.05, \lambda_c = 0.025, \lambda_o = 0.04$ and $\mu_o = \nu_o = 0.02$. The choice of transition rates is arbitrary save for it yielding expected mean closed and open times (80 and 50 respectively) comparable to those in the data set analysed in Hodgson. We added noise from an AR(15) process identical to that fitted to the noise in Hodgson, and then applied the same filter. Our choice of model parameters is $m = (0.25, 0.25, 0.25, 0.25), \alpha = 1, \beta = 40$, corresponding to prior indifference between the four Markov models and exponential priors on the transition rates. With these model parameters, we found from experimental runs that $\rho_1 = \rho_2 = 2, \tau = 0.3$ was a sensible choice of simulation parameters. These are the settings of model and simulation parameters in all subsequent runs.

We ran the sampler for 3 million sweeps with no burn-in, recording realisations every thirtieth sweep purely due to storage constraints. This took 79 minutes on a 167 MHz Sun UltraSparc 1 workstation. Pilot runs had indicated that using aggressive optimisation options on compiling the code cut the resulting computation time by around 60% without affecting numerical results. The

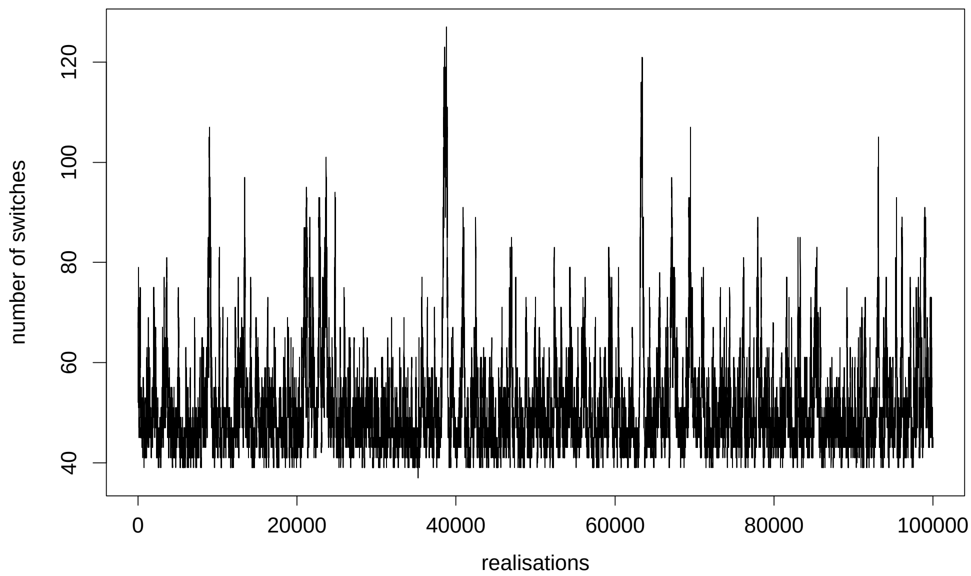

Figure 3: Trace of the number of switches

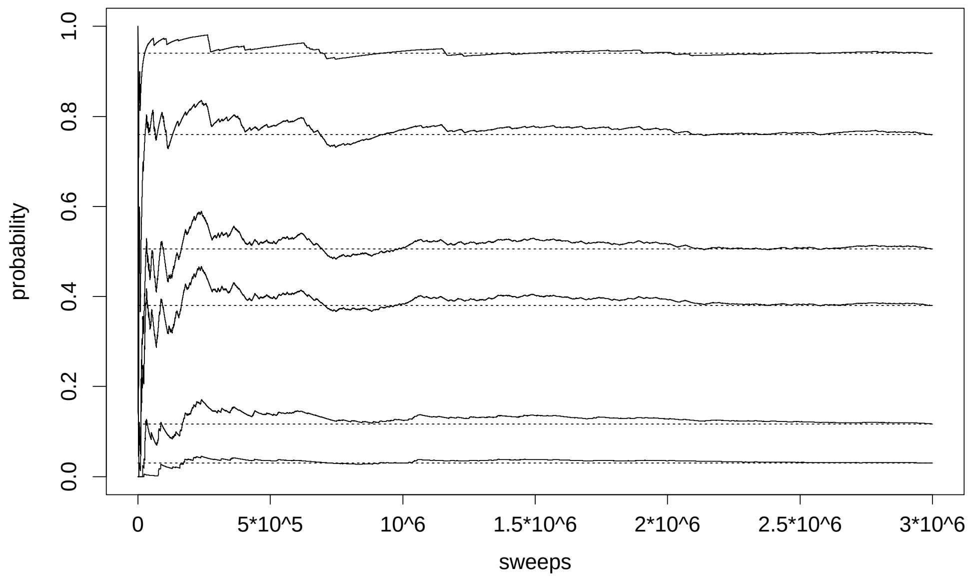

Figure 4: Quantiles diagnostic for the number of switches

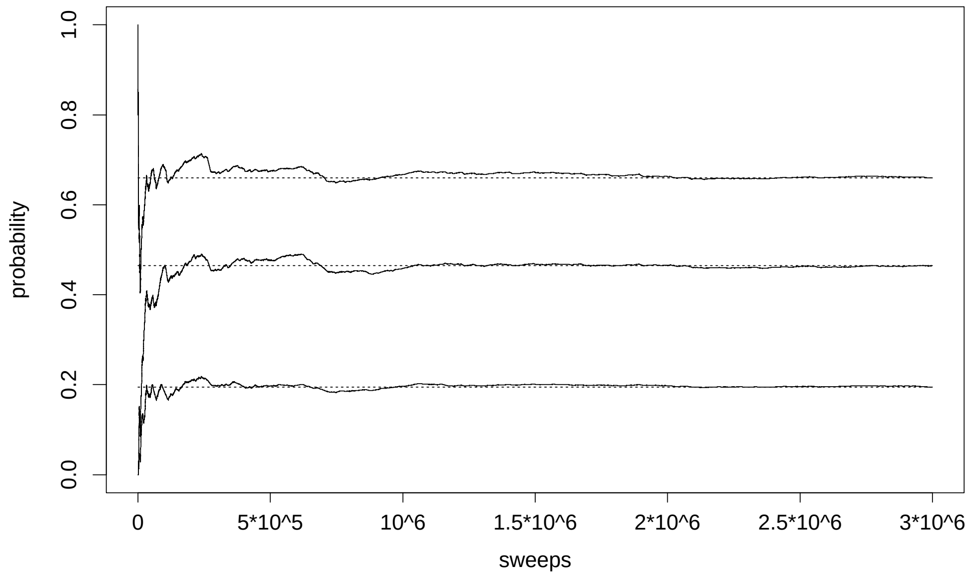

Figure 5: Quantiles diagnostic for the Markov model indicator

trace of the number of switches of the indicator function $\mathbf{x}$ is shown in figure 3. The broken lines of figure 4 represent quantiles of the empirical distribution of the number of switches, while the solid lines trace the cumulative occupancy fractions below these quantiles. The equivalent convergence diagnostic for the Markov model indicator $k \in \{11,21,12,22\}$ is shown in figure 5. Mixing over the number of switches appears to be satisfactory, and if we take the stabilisation of the occupancy fractions as an indication of convergence then the quantiles plots suggest discarding as burn-in the initial 40% of the realisations, basing posterior inference on the remainder.

It would be desirable to make convincing inference about hidden Markov models for ion channels based on weakly informative prior information. Completely uninformative priors on the transition rates given the model, achieved by setting $\alpha = \beta = 0$ in our construction, are not possible: all the posterior probability would then be assigned to $k = 11$, the minimal model. This phenomenon is known as 'Lindley's paradox' (Lindley 1957, Bartlett 1957). Section 5 will present a sensitivity analysis with results which we believe justify our decision not to impose more diffuse priors on the transition rates (taking values of α less than 1 also increased time to convergence as estimated by the quantile plots).

The empirical distribution of the model variable k, based on the final 60 000 recorded realisations, is displayed in figure 6. Encouragingly, the strongest support is for the true Markov model, $k = 22$. The estimated posterior probabilities for models *11, 21, 12* and *22* are 0.193, 0.269, 0.191 and 0.347 respectively. The standard errors of these estimates were assessed by the batched means method and estimated to be 0.007, 0.009, 0.007 and 0.012 respectively. Quantiles plots for the transition rates conditional on k suggested that the sampler mixes faster within than between k, and also faster within $k = 11$ than within more complicated models. An explanation, supported by the evidence of section 4, is that when there are two open or closed states, the three-dimensional joint posterior for the relevant rates is diffuse and the dimension-preserving moves updating Λ have a low acceptance rate.

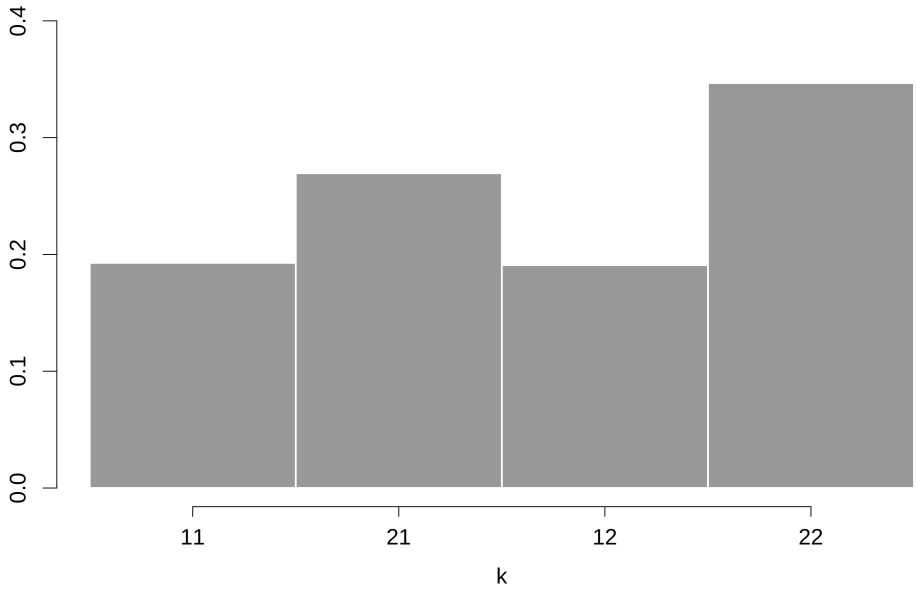

Figure 6: Posterior distribution of the model indicator k

3.2 A seemingly paradoxical phenomenon

We stated in the previous section that the choice of the six transition rates given $k = \mathit{22}$ was constrained only by the loose requirement that they give realistic expected sojourn times. It is known (section 2.2) that sojourn times in a two-state class have two-component exponential mixture distributions. Given that $k = \mathit{22}$ is the most favoured model *a posteriori*, it is natural to ask how far these distributions 'depart' from exponentiality, and hence how remarkable this finding is.

Note that for both sojourn times, we set the true outer transition rates, that is ν_c and μ_c for closed times, to be equal. Denote by ξ the ratio of these outer rates to the 'gateway' rate (λ_c for closed times). Then $\xi = 2$ for closed times and $\xi = 1/2$ for open times. Figure 7 illustrates the degree of departure from exponentiality for various values of ξ by Q–Q plots of the pure exponential against the two-component exponential mixture distributions. All distributions are scaled to have mean 1. It is clear that the mixture distributions assign higher probability density to high and low values and lower density to intermediate ones, this trend being less marked with increasing ξ. Under this set-up, it is easily shown that the sojourn time distribution has coefficient of variation

$$\sqrt{1 + \frac{1}{2\xi}}.$$

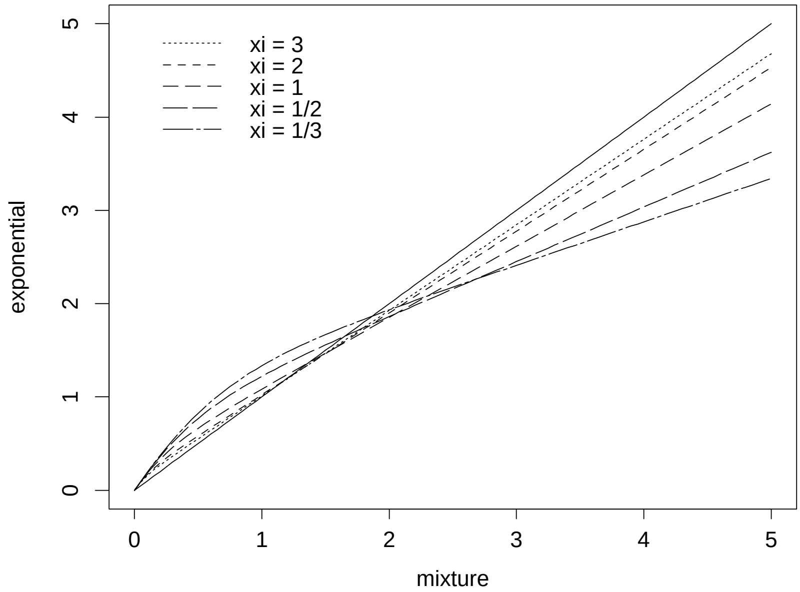

Figure 7: Departure from exponentiality for a range of values of ξ

Given our choice of transition rates, the theoretical open sojourn time distribution is markedly further from the exponential than that of the closed times, the respective coefficients of variation being $\sqrt{2}$ and $\sqrt{5}/2$. However the previous section shows that $k = 21$, the model having two closed and one open state, is favoured *a posteriori* over $k = 12$, the model having a single closed and two open states. The obvious explanation is that sampling variation leads to a simulated channel signal having open times fitting a single exponential better than closed times do.

In an effort to verify or discredit this theory, we conducted off-line model discrimination experiments for the sample of completed realised closed times and then separately for the completed realised open times from the true signal underlying the data set of the previous section. There were 27 in each sample. The associated DAG is in figure 8. In this reduced model, $\tilde{k} \in \{1, 2\}$ is the number of exponential components in the mixture distribution fitted to the sojourn times and p_1 is the prior probability of there being just one component. The vector $\tilde{\Lambda}$ contains the associated transition rates, and hence is one-dimensional if $\tilde{k} = 1$ and three-dimensional otherwise. The data $\mathbf{t}$ are the sojourn times. The priors on the rates given α and β, and their default values, are as before. We set $p_1 = 0.5$. It was trivial to write a program to compute the posterior under this simplified model as the necessary code is an extract from that already written. The results of these experiments were conclusive. The posterior probability $p(\tilde{k} = 2 \mid \mathbf{t})$ was 0.237 for the closed time sample but as high as 0.987 for the open time sample. The simple explanation of the surprising finding of the previous paragraph can be discounted.

For each $t \in \{0, 1, \ldots, T = 2^{12} - 1\}$, we may estimate the posterior probability of the channel being open at time t, and hence obtain a pointwise estimate of the

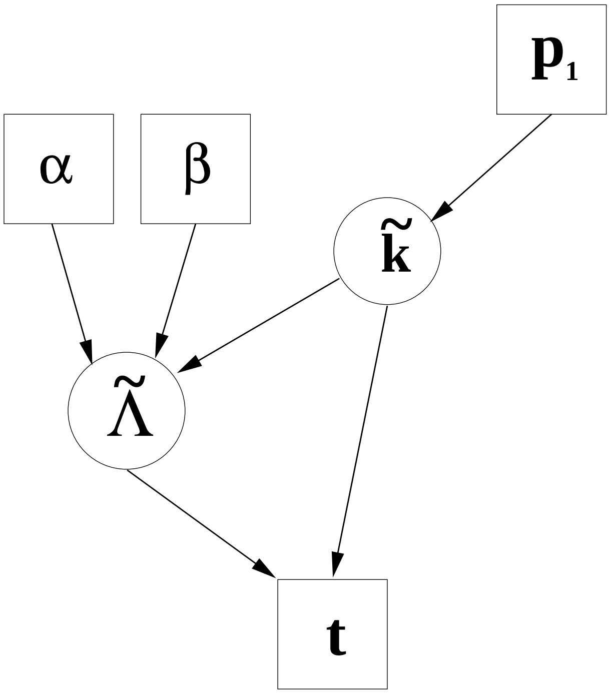

Figure 8: The DAG corresponding to the simplified model

posterior mean of $\mathbf{x}$, the indicator function of channel openness. Thresholding the mean function at the 50% level yields an estimate $\widehat{\mathbf{x}}$ of the true indicator function. We estimated the posterior mean functions conditional on the Markov model k and thresholded them to obtain estimates of the true signal conditional on k. We found negligible dependence on k, the largest number of points of disagreement between any two estimates being 5 out of 4096. However, the threshold estimate is insensitive, suppressing much relevant information in the sample. The differences in the distributions of $\mathbf{x}$ under models *21* and *12* are illustrated in figure 9. For the first half of the channel record, we show as a solid curve the difference $\mathbf{d}$ between the posterior mean functions conditional on $k = \mathit{21}$ and on $k = \mathit{12}$. The dotted line represents the function $(2\widehat{\mathbf{x}} - 1) \sup |\mathbf{d}_i|$. The unconditional estimate $\widehat{\mathbf{x}}$ of the true indicator function and those conditional on k are almost identical. The positive peaks of $\mathbf{d}$ are all in the neighbourhood of a 'switch' of $\widehat{\mathbf{x}}$. In contrast, some negative peaks are located within lengthy closed and open sojourns. Negative peaks within closed sojourns correspond to times at which the evidence for a brief opening is stronger under model *12* than model *21*. As model *12* induces an open time distribution with coefficient of variation greater than 1, this is to be expected. Similarly, negative peaks within open sojourns correspond to times at which model *21* supports a brief closing more strongly than model *12*.

In summary, the small differences in the distributions of $\mathbf{x}$ under the competing models indicate some association between closed and open sojourn times. This association probably accounts for the disparity between the posterior model probabilities given the noisy data and those one would obtain given the noiseless signal.

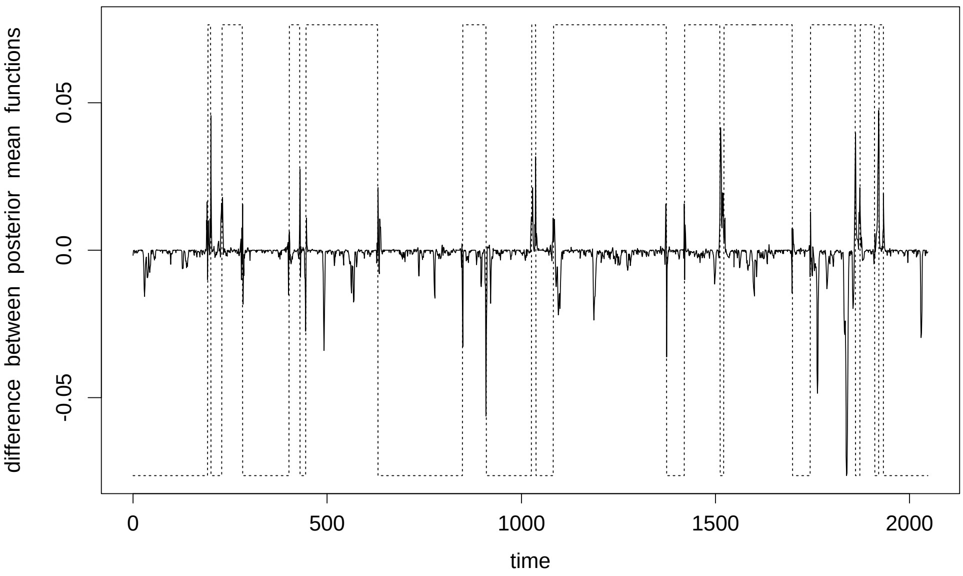

Figure 9: The difference between the posterior mean functions under models *21* and *12*

3.3 The Eisenberg and Levis data

The data analysed by Hodgson (1998) and previously by Johnstone and Silverman (1997) is a segment of length $2^{12} = 4096$ from a time series of 10^5 data points simulated by physiologists Eisenberg and Levis. The time series represents the recorded current through a single ion channel with two conductance levels, and is intended to stimulate the development of new signal processing techniques for this kind of data (Eisenberg 1994). Hodgson assigned an alternating renewal model to the channel's aggregated process. It would be interesting to see which of our candidate Markov models are well supported by these data, and to compare our results with those of Hodgson.

The estimated posterior probabilities for models *11*-*22* given the Eisenberg and Levis data are 0.405, 0.119, 0.369 and 0.107 respectively. As in section 3.1 these estimates are based on a run of 3 million sweeps of the reversible jump MCMC sampler, subsampling every thirtieth sweep and discarding the first 40% of the output. It is $k = 11$, the minimal two-state model, and $k = 12$, the model with one closed and two open states, which together receive the bulk of the posterior support. It is unsurprising that there is more support for a second open state than a second closed state. Hodgson (1998) placed gamma distributions on the channel sojourn times. With independent Exp(1) priors on the closed and open time 'shape' parameters s_0 and s_1, their posterior medians were 1.88 and 0.72 respectively. Since the coefficient of variation of the gamma distribution with shape s is $1/\sqrt{s}$, there is clear evidence that the open times have a higher coefficient of variation than the closed times. One would therefore anticipate the posterior distribution placing greater probability on the open times having the two-component exponential mixture distribution, which always has coefficient of variation greater than 1.

3.4 Predictive distributions

After every sampler sweep, the realised values of (k, Λ) determine a realised closed sojourn time density $f_c(.|k, \Lambda)$ and an open time density $f_o(.|k, \Lambda)$. In the same manner as for other functionals of interest, averaging these densities across the MCMC output yields estimates of their posterior expectations. Averaging conditionally on a given value of k gives estimates of $E\{f_c(.|k, \Lambda) \mid k, \mathbf{y}\}$ and $E\{f_o(.|k, \Lambda) \mid k, \mathbf{y}\}$, Bayesian predictive density estimates of the sojourn times under the given Markov model. Further averaging across k gives estimates of $E\{f_c(.|k, \Lambda) \mid \mathbf{y}\}$ and $E\{f_o(.|k, \Lambda) \mid \mathbf{y}\}$, the unconditional predictive density estimates of the sojourn times.

We display both conditional and unconditional open time density estimates for the Eisenberg and Levis data set in figure 10. The estimates conditional on Markov models *11* and *21*, in which there is just one open state, are very similar to each other, as are those conditional on models *12* and *22*, in which there are two open states. This suggests, contrary to the findings of section 3.2, that given the number of states in one class, the association between the transition rate(s) governing the distribution of sojourns in that class and the number of states in the other class is weak. The density estimates conditional on $k = \mathit{12}$ and $k = \mathit{22}$ have higher coefficient of variation than the estimates given $k = \mathit{11}$ and $k = \mathit{21}$. The overall (unconditional) predictive density estimate, given by the solid line in figure 10, is a compromise between the two classes of similar estimates. All the above remarks apply equally to the closed time density estimates (not shown), only with models *21* and *12* interchanged.

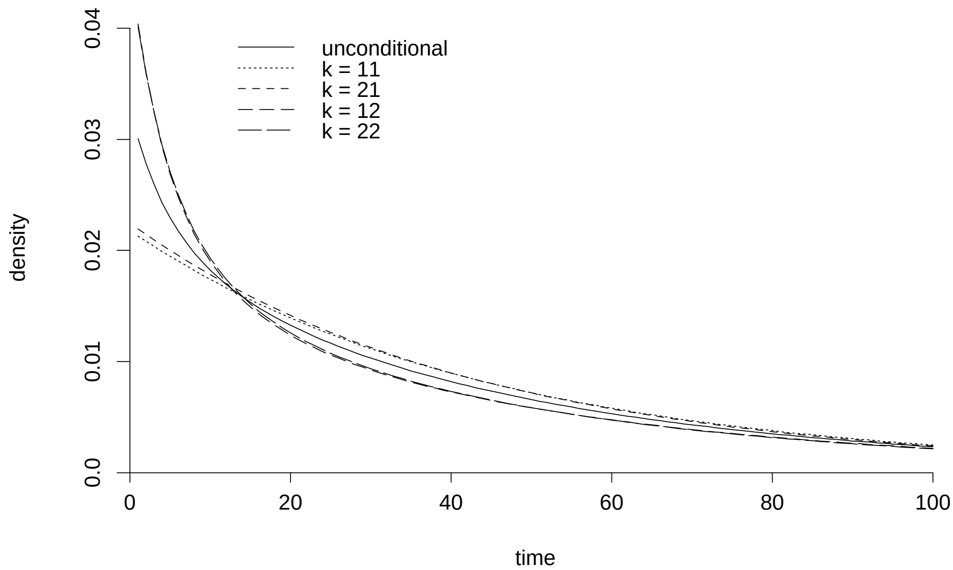

Figure 10: Bayesian predictive density estimates for open sojourn times

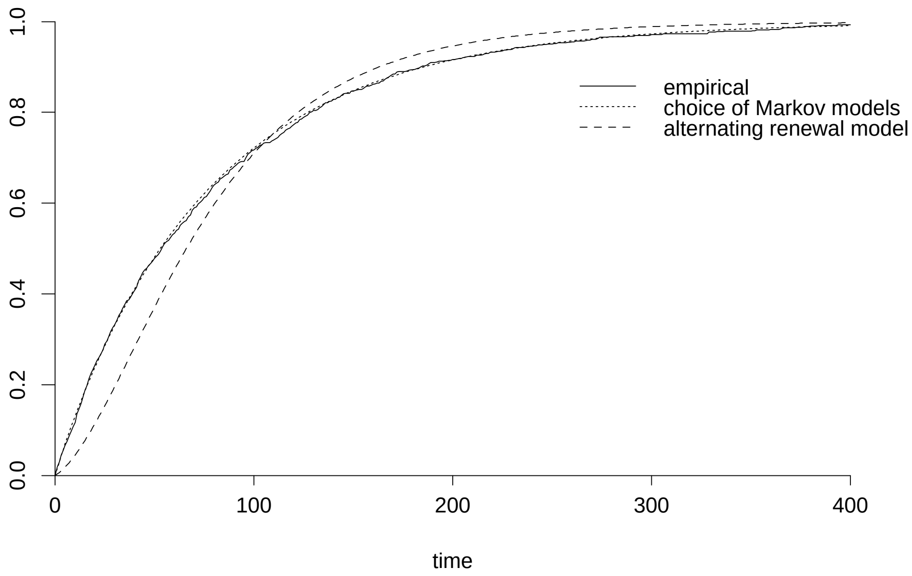

Figure 11: Predictive distribution function estimates for closed times

We may also compare predictive sojourn time distributions under the current set-up of a choice between hidden Markov models and under the alternating renewal model, with gamma distributed sojourn times, of Hodgson (1998). Figure 11 shows the unconditional closed time distribution function estimate under the current model (dotted line) and the estimate under Hodgson's model (dashed line). Also shown as a solid line is the empirical distribution function of the 811 completed closed times from the entire 'true signal' record of Eisenberg and Levis. Clearly the current model provides much the better fit. Under our model, the predictive distribution function (and its estimate) must be concave as it is a continuous mixture of exponential distribution functions. Since Hodgson places no prior constraint on the closed time shape parameter s_0, this need not be the case under his model. His predictive distribution has lower coefficient of variation than ours, illustrating the fact that the bulk of the posterior support in his analysis was for values of s_0 greater than 1.

4 Assessing sampler mixing and convergence

In section 3.1 our decision to discard the first 40% of the MCMC output as burn-in was based on figures 3 to 5. None of these informal diagnostics give cause for grave concern, but it would be unwise to rely solely on them as evidence of satisfactory mixing and convergence to stationarity. There is a danger that such plots fail to spot serious difficulties, especially in the current context where there are two degrees of variable dimensionality. Richardson and Green (1997) evaluate the performance of their reversible jump algorithm by firstly assessing mixing over the dimension variable and then examining mixing within the subspaces corresponding to a fixed dimension, for those dimensions with significant posterior support. We shall examine mixing over both the Markov model indicator k and

Table 1: Summary of the posterior output from runs 1 and 2

Parameter	Run 1	Run 2
$I\{k = 11\}$	.409 (.013)	.413 (.011)
$I\{k = 21\}$	.116 (.005)	.125 (.004)
$I\{k = 12\}$	.373 (.014)	.355 (.009)
$I\{k = 22\}$	.102 (.004)	.107 (.005)
s	68.1 (0.8)	67.5 (0.5)
$\lambda_c \mid k = 11$	.0125 (.000048)	.0125 (.000047)
$\lambda_o \mid k = 11$	.0217 (.000077)	.0218 (.000087)
$\lambda_c \mid k = 12$	.0139 (.00028)	.0136 (.00013)
$\lambda_o \mid k = 12$	.051 (.004)	.046 (.002)
$\mu_o \mid k = 12$	.055 (.003)	.052 (.002)
$\nu_o \mid k = 12$	.068 (.001)	.068 (.001)

the number of switches of the indicator function, and then condition on the more likely values of k.

In initialising the variables for the MCMC run on the Eisenberg and Levis data (section 3.3), we firstly thresholded the data at their median value. We then made a forward pass through the thresholded data, applying time interval omission by sequentially removing all apparent sojourns in either class of duration less than 5 time units. The resulting estimates of the signal and channel conductance levels were the starting values of $(\mathbf{x}, \mu_0, \mu_1)$. The starting values of (k, Λ) were set to be $k = 11$, $\lambda_c = 0.0125$ and $\lambda_o = 0.02$. The initial number of switches may be changed dramatically by varying the detection limit in the second stage of the algorithm. With limits of 3, 5 or 9 time units, the estimates have 254, 44 and 16 switches respectively; the true number is 55. This provides a means of generating over-dispersed initial values from which to run the sampler, as advocated by Gelman and Rubin (1992).

Initialising $\mathbf{x}$ to be the highly 'switchy' signal generated by setting the detection limit to be 3, we ran the sampler for 50 000 sweeps *whilst holding* $\mathbf{x}$ *fixed*, released $\mathbf{x}$ and then performed a further 3 million sweeps. This formula was repeated after initialising $\mathbf{x}$ by setting the detection limit to be 9. Keeping $\mathbf{x}$ constant whilst sampling all other parameters allows time for the other parameters to 'equilibrate' before allowing $\mathbf{x}$ to vary. This should ensure that the two runs to be compared start from genuinely well separated regions of the joint posterior. Table 1 displays the posterior means and standard error estimates (parenthesised) of some key parameters. They are based on subsampling the entire 3 million sweeps of both runs after every thirtieth sweep. Here, $I\{E\}$ is the indicator variable associated with the event E and s is the number of switches of $\mathbf{x}$. Results from the two runs are reassuringly similar.

In figure 12 we compare the traces of the first 30 000 recorded realisations of s from runs 1 (solid line) and 2 (dotted line). This suggests that the influence of the starting state does not extend beyond the first 5000 realisations (150 000 sweeps). Gelman and Rubin's test diagnoses convergence of both the s and the k series. Other tests give less encouraging results, however. The simple quantiles plots used in section 3.1 (not shown) suggest discarding as much as 50% of the realisations. The tests of Geweke (1992) and Heidelberger and Welch (1983) are failed, except for the k series arising from run 1, which passes the latter. Of course, none of these diagnostics were designed with reversible jump output in mind, and their applicability to series of discrete variables is questionable. Acceptance rates are much higher for moves updating k (around 24%) than for moves updating s (around 1%). This partly explains the much greater autocorrelation displayed by the s series.

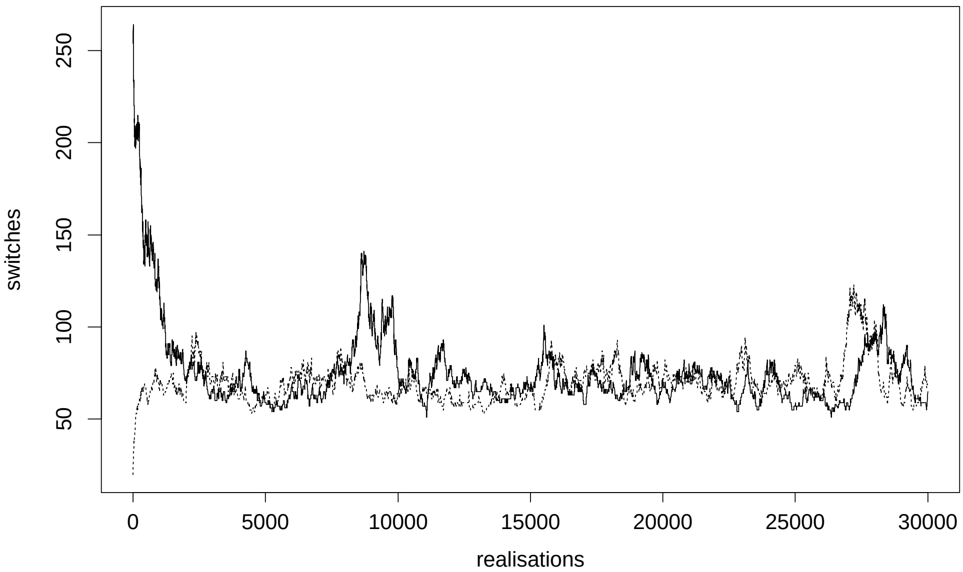

Figure 12: Traces of the number of switches for runs 1 and 2

We confine our examination of within-k mixing and convergence to the cases $k = 11$ and $k = 12$, attracting between them over 75% of the posterior support. Given $k = 11$, the posteriors for the transition rates λ_c and λ_o are fairly tight: both have coefficient of variation less than 0.2. This indicates that the data are quite informative about the parameters of the two-state Markov model. The quantiles plots and the tests of Gelman and Rubin (1992) and Raftery and Lewis (1992) suggest rapid convergence. The results of Heidelberger and Welch's test are positive except when applied to the λ_o series arising from run 2. Geweke's diagnostic advises caution, however, as a high proportion of the Z-scores for both parameters on both runs lie outside the 95% confidence interval of the standard normal distribution (figure 13).

Estimates of the posterior densities of the transition rates given $k = 12$ are shown in figure 14. These are based on combining the second half of the output conditional on $k = 12$ for the two runs. Clearly the rates governing open sojourn times (namely λ_o, μ_o and ν_o) have considerably more diffuse posteriors than λ_c, the single transition rate governing closed sojourn times. The data carry relatively little information about these three transition rates given this model. Although both Gelman and Rubin's and Raftery and Lewis' tests point to convergence of the marginals for all rates, the cumulative occupancy fractions traced by the quantiles plots do not stabilise until around half the sweeps for

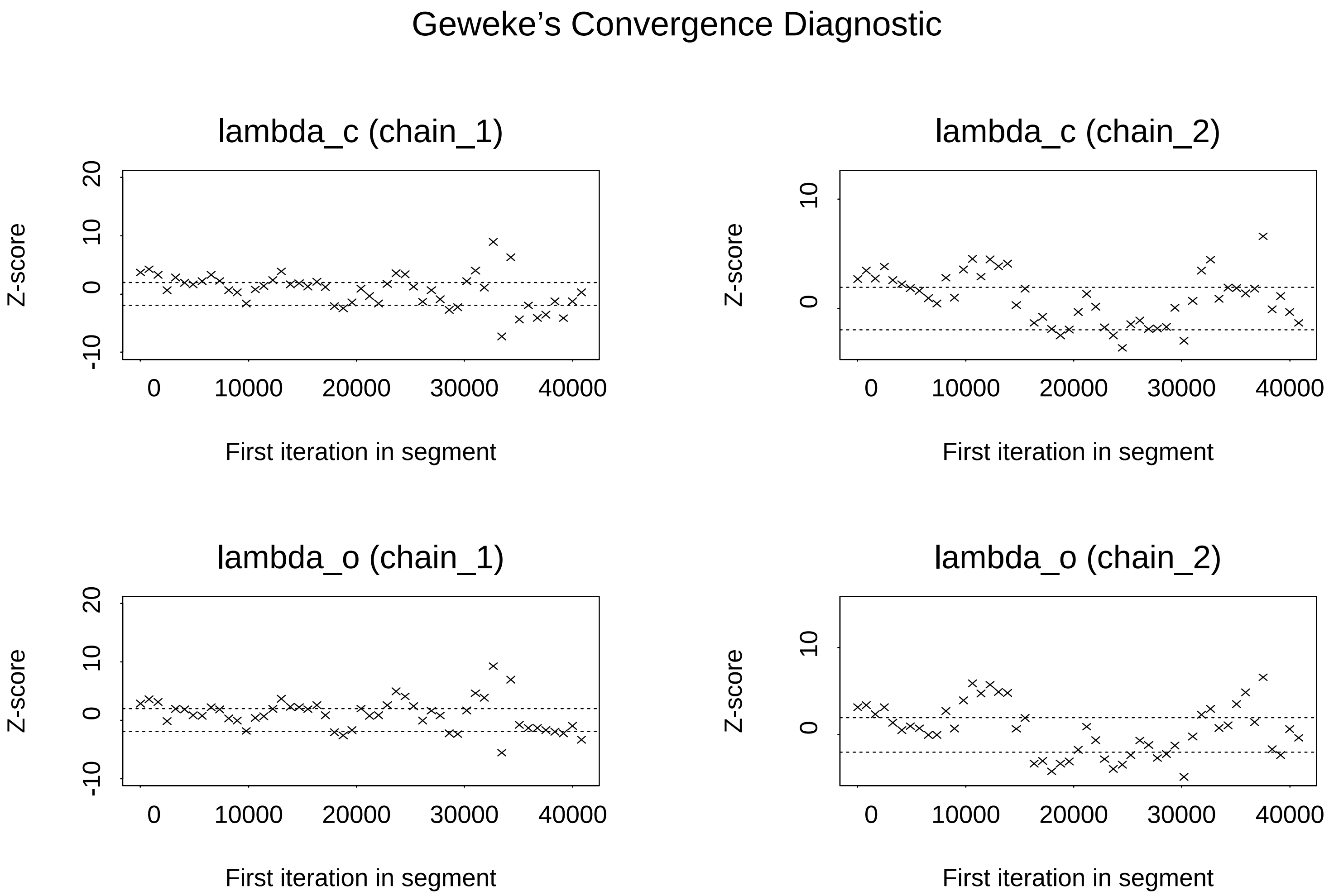

Figure 13: Geweke’s diagnostic for transition rates given $k = 11$

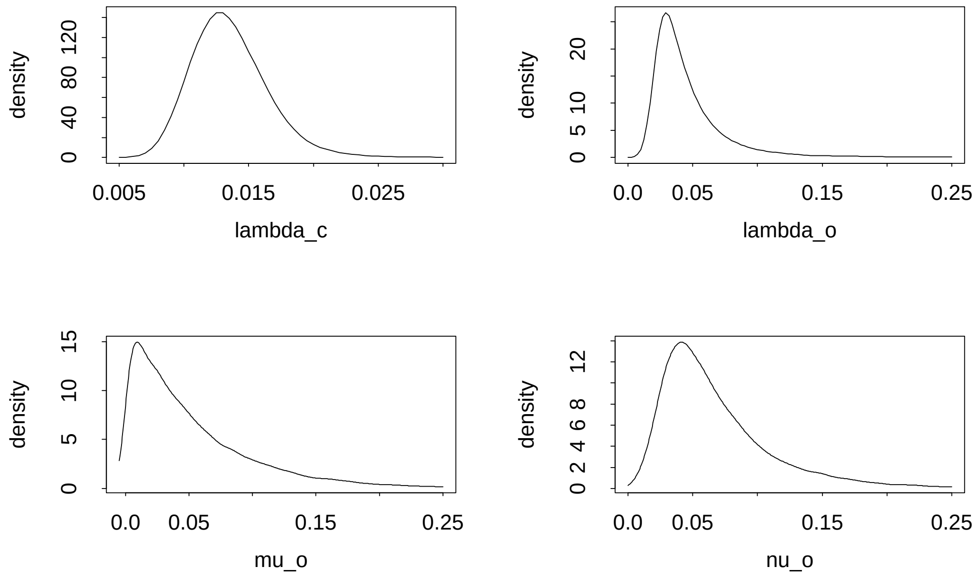

Figure 14: Posteriors for transition rates conditional on $k = 12$

which $k = 12$ have been completed. Heidelberger and Welch's and Geweke's tests are failed comprehensively by all rates for both runs. All series display higher autocorrelation than the transition rates series for $k = 11$, presumably due to the lower acceptance rates of the dimension-preserving moves updating Λ when $k = 12$ than when $k = 11$.

5 A sensitivity analysis

5.1 Factors to be investigated

In section 3.1 we simulated a data set based on $k = 22$. Given the settings we chose for m, α and β, the true model receives the strongest, though by no means overwhelming, posterior support (figure 6). However this result has little significance without a qualitative understanding of some of the key factors influencing the posterior probabilities attached to the competing models. We alluded to Lindley's paradox: namely that $p(k = 11 \mid \mathbf{y})$ will tend to 1 as the priors on the transition rates become less and less informative. With only mildly informative priors ($\alpha = 1, \beta = 40$) one might anticipate that the simplest model will be 'easier' to identify than the others. The level of noise in the data and the length of the data segment are obvious factors of possible importance. Intuitively we would anticipate that lowering the noise level and lengthening the channel recording would result in a higher posterior probability on the correct value of k, but at the price of slower sampler mixing.

The fourth and final factor in the forthcoming study may be expressed informally as the ease with which the true model may be identified. In section 3.2 we defined the ratio ξ of the 'outer' rates to the 'gateway' rate in the case of the two-state class with equal outer rates. We saw (figure 7) that the exponential mixtures implied for the sojourn time distributions by the transition rates showed greater departure from exponentiality the lower the value of ξ. In this context, we would anticipate low values of ξ for the open and closed times to facilitate identification of the true model $k = 22$. We cannot however calibrate this factor by a quantity whose definition is model-dependent as the model is another factor. Calibrating this rather abstract factor is indeed problematic, and is discussed in the next section.

5.2 Calibrating the final factor and the experimental set-up

We shall study the effects on the posterior distribution of the model indicator k of the four factors identified above. It is straightforward to define and set 'high' and 'low' levels for three of the factors. Factor A is the true model itself, $k = 11$ being the high level and $k = 22$ being the low level. Factor B is the length of the data set, with high level $n = 2^{13}$ and low level $n = 2^{12}$. Factor C is described loosely as a measure of how easily the true model is identified, with high level corresponding to cases in which the truth is easily identified. Finally factor D is the noise precision σ^{-2}, with high level equal to the precision of Eisenberg and Levis' data (Hodgson 1998) and low level equal to half the high level. By virtue of the definitions of high and low levels, each factor is expected *a priori* to have a positive main effect. Our first attempt to provide a satisfactory definition of factor C was as follows: given the true model and the transition rate(s) determining the (open or closed) sojourn time distribution, we generated a random sample of 100 such times. With sensible choices of rates, this sample size is of the same order as the number of switches in the indicator functions underlying the data sets to be analysed. We measured the infidelity of this sample to its underlying model by the posterior probability allocated to the 'incorrect' number of exponential components under the choice between pure exponential and two-component exponential mixture distributions, given prior indifference between the two (section 3.2 and figure 8). The posterior probability on the wrong number of components shall henceforth be called the *error probability*.

We generated 50 samples of 100 times from the closed sojourn time distribution implied by $k = 22, \nu_c = \mu_c = 0.0533, \lambda_c = 0.0267$ (so that $\xi = 2$). The resulting error probabilities are displayed in figure 15; their mean is 0.27. Clearly the sampling variation in error probabilities for samples of this size is considerable. Similar degrees of dispersion were observed for the error probabilities derived from samples of 100 times from both the single exponential of the same mean (75) and exponential mixtures of the same mean but having different values of ξ. In the mixture case, mean error probabilities increase with ξ as expected: the mean rises from 0.15 to 0.44 in going from $\xi = 1$ to $\xi = 3$. For a single sample of sojourn times, however, this effect is overwhelmed by the sampling variation. Hence it is wholly unsatisfactory to calibrate factor C by mean error probabilities from samples of size 100 from the theoretical sojourn time distributions implied by the true model (level of factor A) and the transition rates. However, these findings suggest a related but superior way to assign levels to factor C. Rather than simulating samples of size 100 from the sojourn time distributions implied by the choice of Markov model and transition rates, suppose that for each combination of levels of factors A (true model) and B (record length) we simulated several noiseless channel records. We could then extract the completed sojourn times from these

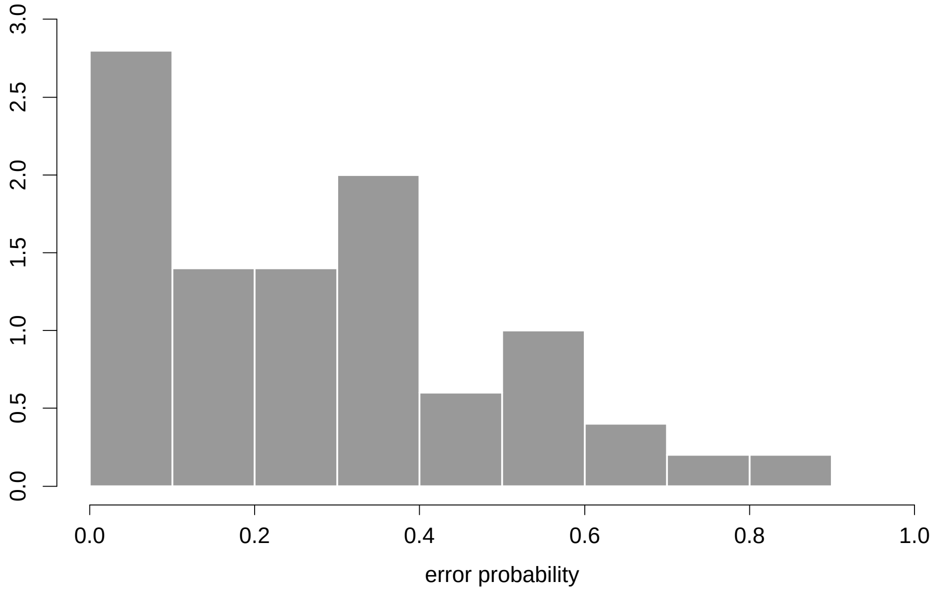

Figure 15: Histogram of error probabilities for 50 samples of 100 sojourn times

records and compute the error probabilities associated with these randomly-sized samples. Choosing transition rates to give the same distribution for open and closed times, so that realised sojourn times need not be partitioned, simplifies matters. For each combination of A and B levels, we hand-pick 2 records from the set of those simulated, one yielding a 'high' error probability and the other a 'low' error probability. Having done this, we would have 8 records, 4 yielding approximately the same 'high' error probability and 4 yielding approximately the same 'low' error probability. Given the high sampling variation of error probabilities associated with samples of size of order 100, it should be easy to establish two well separated bands of error probability and simulate signals yielding error probabilities falling within one or other of these bands. The 4 records giving 'high' error probability will be used to generate noisy data sets corresponding to the low level of factor C (true model 'hard to identify'), and those giving 'low' error probability will generate data sets for which factor C takes its high level (true model 'easily identified'). When $k = \mathit{11}$, we chose $\lambda_c = \lambda_o = 0.0133$ and when $k = \mathit{22}$, we chose $\nu_c = \mu_c = \mu_o = \nu_o = 0.0533,\ \lambda_c = \lambda_o = 0.0267$. Table 2 summarises our efforts to calibrate factor C by means of error probabilities. We follow the convention that $+$ denotes the high level of a two-level factor and $-$ the low level. We have not endeavoured to match error probabilities corresponding to the low and high levels of factor C too closely as the forthcoming experiments are essentially qualitative.

We conduct a single replicate of the 2^4 factorial design. To each of the 8 ion channel signals corresponding to all combinations of factors A, B and C, we added the autoregressive noise processes of variance corresponding to both levels of factor D. This yielded the 16 data sets to be analysed by our Markov model selection methodology.

Table 2: Calibrating factor C by error probabilities

Factor A	Factor B	Error probability	Factor C
$-$	$-$	0.375	$-$
$-$	$+$	0.346	$-$
$+$	$-$	0.278	$-$
$+$	$+$	0.317	$-$
$-$	$-$	0.143	$+$
$-$	$+$	0.157	$+$
$+$	$-$	0.141	$+$
$+$	$+$	0.176	$+$

Table 3: Results of the factorial experiment

Run	A	B	C	D	Treatment combination	$2 \log B_{A,-A}$
1	$-$	$-$	$-$	$-$	(1)	0.67
2	$+$	$-$	$-$	$-$	a	-1.26
3	$-$	$+$	$-$	$-$	b	-5.05
4	$-$	$-$	$+$	$-$	c	0.67
5	$-$	$-$	$-$	$+$	d	0.14
6	$+$	$+$	$-$	$-$	ab	5.81
7	$+$	$-$	$+$	$-$	ac	2.42
8	$+$	$-$	$-$	$+$	ad	-2.54
9	$-$	$+$	$+$	$-$	bc	-3.50
10	$-$	$+$	$-$	$+$	bd	-5.10
11	$-$	$-$	$+$	$+$	cd	1.45
12	$+$	$+$	$+$	$-$	abc	5.75
13	$+$	$+$	$-$	$+$	abd	5.93
14	$+$	$-$	$+$	$+$	acd	3.70
15	$-$	$+$	$+$	$+$	bcd	-4.31
16	$+$	$+$	$+$	$+$	abcd	5.54

5.3 Results

Summarised in table 3 are the results of the factorial experiment. As it is restricted to $[0, 1]$, we do not take the posterior probability assigned to the true model as our response variable. Instead we measure the evidence for the true model against the model corresponding to the 'incorrect' level of factor A by twice the logarithm of the Bayes factor $B_{A,-A}$. This statistic has the same scale as the classical likelihood ratio test statistic, and is given a rough calibration by Kass and Raftery (1995). Their guidelines are that values of $2 \log B_{12}$ greater than 2 provide 'positive' evidence for model 1 over model 2, and values greater than 6 provide 'strong' evidence.

The estimates of the main effects derived from table 3 are 5.05, -0.02, 1.64 and -0.09 for A, B, C and D respectively. The interactions largest in magnitude are 5.20 for AB and -1.43 for ABC. Given that Bayes factors tend to favour parsimonious models when prior information on model parameters is weak, the strong positive effect of A is unsurprising. The positive main effect of factor C is also as expected, but D has neither a significant main effect nor any significant interactions. The noise precision presumably must be altered by a factor greater than 2 before inferences about the underlying channel signals given the low and high precision data generated from them are significantly different. As precision tends to zero data sets become increasingly uninformative and so posterior beliefs increasingly reflect prior beliefs. In particular, the posterior odds in favour of one model over another will approach the prior odds, and so Bayes factors will tend to 1. Since data sets are also less informative as they become shorter, this phenomenon of *prior shrinkage* would explain the AB interaction. When factor A is at its low level (so that the true model is $k = \mathit{22}$), the Bayes factor $B_{A,-A}$ is typically less than 1, so that a shorter data set increases $B_{A,-A}$ towards 1. Conversely, at its high level, A typically raises $B_{A,-A}$ above 1, so that a shorter data set reduces $B_{A,-A}$ towards 1. Hence AB has a positive effect.

In summary, table 3 reveals that when B takes its high level, it is always far better to have A at its high level than at its low level. When B takes its low level, the difference in performance caused by changing the level of A is smaller, and it is better to have A at the same level as C.

6 Discussion and further work

This paper has described a study into the feasibility of carrying out general Markov model selection for ion channel kinetics. It extends the work of Hodgson (1998) from an alternating renewal model to a choice of four simple Markov models. We do not however accomplish this by explicitly incorporating the latent Markov process into our Bayesian model (figure 2), but rather by integrating it out analytically (sections 2.1 and 2.2). This approach is convenient only for the simplest extensions of the basic two-state model, $k = \mathit{11}$. Although our extensions are generally simpler than those appearing in the literature, model *21* was proposed by Castillo and Katz (1957) as the simplest mechanism for non-cooperative agonist action.

Of greatest interest in this study is the marginal posterior for k. General model selection for ion channels will be worth pursuing only if there is clear evidence of the ability to discriminate between these four models. In this respect, the factorial experiment of section 5 gives mixed results. Model *11* is correctly favoured *a posteriori* over model *22* in six cases out of eight, but incorrectly favoured in four cases out of eight. Lindley's paradox manifests itself in the strong positive effect of factor A. Even with mildly informative exponential priors on the transition rates within each model, the bias towards the simpler model is clearly evident. Ion channel data, at least in the quantities and subject to the noise levels of this paper, are insufficiently informative to make the ideal of Bayesian model determination with 'uninformative' priors viable. Factor C measures the practical identifiability of the true model given the noiseless channel signal. All our models are mathematically identifiable, in that distinct choices of transition rates will induce distinct distributions on the closed and open sojourn times. However, some choices of transition rates governing sojourn times in a two-state class, for example choosing ξ to be large in the set-up of section 3.2, result in a sojourn time distribution virtually indistinguishable from a single exponential, so that the two-state class tends to degenerate into a single state. Since the parameter of a single exponential is just a scale factor, the evidence of departure from exponentiality in a sample of a given size from an exponential distribution is determined solely by sampling variation, and it is the prior on the transition rates given $k = 11$ which determines the distribution of the error probabilities in section 5.2. Hence factor C is rather one-sided.

To summarise, in any future work with real channel data, we will require the collaboration of neurophysiologists to establish plausible Markov models and appropriate priors on the transition rates within each model, which may incorporate dependence between the rates. Of course, real patch clamp records are orders of magnitude longer than the records we simulate. Although it is not reflected in the main effect of factor B in our experiment, there is no doubt that increasing the amount of data will aid identification of the true model. However, implementation of the Bayesian method is then likely to be hampered by poorer mixing of the MCMC sampler. Perhaps by allowing different Markov models to apply to different subsections of the data set, along the lines of the simulated tempering idea in Hodgson (1998), these difficulties could be overcome.

References

Ball, F. G., Cai, Y., Kadane, J. B., and O'Hagan, A. (1996). MCMC methods for discrete sojourn time ion channel data. Technical report, Statistics Group, University of Nottingham.

Ball, F. G., Cai, Y., Kadane, J. B., and O'Hagan, A. (1997). Bayesian inference for ion channel gating mechanisms directly from single channel recordings, using Markov chain Monte Carlo. Technical report, Statistics Group, University of Nottingham.

Ball, F. G., Chen, A., and Sansom, M. S. P. (1992). Poisson sampling-based inference for single ion channel data with time interval omission. *Proceedings of the Royal Society of London B*, **250**, 263–269.

Ball, F. G. and Davies, S. S. (1995). Statistical inference for a two-state Markov model of a single-ion channel, incorporating time-interval omission. *Journal of the Royal Statistical Society B*, **57**, 269–287.

Ball, F. G. and Sansom, M. S. P. (1989). Ion-channel gating mechanisms: model identification and parameter estimation from single channel recordings. *Proceedings of the Royal Society of London B*, **236**, 385–416.

Ball, F. G., Yeo, G. F., Milne, R. K., Edeson, R. O., Madsen, B. W., and Sansom, M. S. P. (1993). Single ion channel models incorporating aggregation and time interval omission. *Biophysical Journal*, **64**, 357–374.

Bartlett, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika*, **44**, 533–534.

Castillo, J. and Katz, B. (1957). Interaction at end-plate receptors between different choline derivatives. *Proceedings of the Royal Society of London B*, **146**, 369–381.

Colquhoun, D. and Hawkes, A. G. (1981). On the stochastic properties of single ion channels. *Proceedings of the Royal Society of London B*, **211**, 205–235.

Colquhoun, D. and Hawkes, A. G. (1982). On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Philosophical Transactions of the Royal Society of London B*, **300**, 1–59.

Edeson, R. O., Ball, F. G., Yeo, G. F., Milne, R. K., and Davies, S. S. (1994). Model properties underlying non-identifiability in single channel inference. *Proceedings of the Royal Society of London B*, **255**, 21–29.

Eisenberg, R. (1994). Biological signals that need detection: Currents through single membrane channels. In J. Norman and F. Sheppard (Eds.), *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 32a–33a.

Fredkin, D. R. and Rice, J. A. (1986). On aggregated Markov processes. *Journal of Applied Probability*, **23**, 208–214.

Fredkin, D. R. and Rice, J. A. (1992a). Bayesian restoration of single-channel patch clamp recordings. *Biometrics*, **48**, 427–448.

Fredkin, D. R. and Rice, J. A. (1992b). Maximum likelihood estimation and identification directly from single-channel recordings. *Proceedings of the Royal Society of London B*, **249**, 125–132.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511. With discussion.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger (Eds.), *Bayesian Statistics 4*, pp. 169–193. New York: Oxford University Press.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, **31**, 1109–1144.

Hodgson, M. E. A. (1998). A Bayesian restoration of an ion channel signal. *Journal of the Royal Statistical Society B*, **60**. To appear.

Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society B*, **59**, 319–351.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kienker, P. (1989). Equivalence of aggregated Markov models of ion-channel gating. *Proceedings of the Royal Society of London B*, **236**, 269–309.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, **44**, 187–192.

Magleby, K. L. and Weiss, D. S. (1990). Estimating kinetic parameters for single channels with simulation: a general method that resolves the missed events problem and accounts for noise. *Biophysical Journal*, **58**, 1411–1426.

Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger (Eds.), *Bayesian Statistics 4*, pp. 763–773. Oxford University Press.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B*, **59**, 731–792. With discussion.

Sakmann, B. and Neher, E. (Eds.) (1995). *Single-Channel Recording* (Second ed.)., New York. Plenum Press.