PROCEEDINGS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Bayesian inference for ion–channel gating mechanisms directly from single –channel recordings, using Markov chain Monte Carlo

F. G. Ball, Y. Cai, J. B. Kadane and A. O'Hagan

Proc. R. Soc. Lond. A 1999 **455**, 2879-2932
doi: 10.1098/rspa.1999.0432

THE ROYAL SOCIETY

Bayesian inference for ion-channel gating mechanisms directly from single-channel recordings, using Markov chain Monte Carlo

By F. G. Ball[1], Y. Cai[1]†, J. B. Kadane[2] and A. O'Hagan[1]‡

[1]*School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

[2]*Department of Statistics, Carnegie-Mellon University, 232 Baker Hall, PA 15213, USA*

Received 11 November 1997; revised 1 October 1998; accepted 5 January 1999

The gating mechanism of a single-ion channel is usually modelled by a finite-state-space continuous-time Markov chain. The patch-clamp technique enables the experimenter to record the current flowing across a single-ion channel. In practice, the current is corrupted by noise and low-pass filtering, and is sampled with a typically very short sampling interval. We present a method for performing Bayesian inference about parameters governing the underlying single-channel gating mechanism and the recording process, directly from such single-channel recordings. Our procedure uses a technique known as Markov chain Monte Carlo, which involves constructing a Markov chain whose equilibrium distribution is given by the posterior distribution of the unknown parameters given the observed data. Simulation of that Markov chain then enables the investigator to estimate the required posterior distribution. As well as providing a method of estimating the transition rates of the underlying Markov chain used to model the single-channel gating mechanism and the means and variances of open and closed conductance levels, the output from our Markov chain Monte Carlo simulations can also be used to estimate single-channel properties, such as the mean lengths of open and closed sojourn times, and to reconstruct the unobserved quantal signal which indicates whether the channel is open or closed. The theory is illustrated by several numerical examples taken mainly from the ion-channel literature.

Keywords: Bayesian inference; continuous-time Markov chain; hidden Markov chain; Markov chain Monte Carlo; single ion channel; time reversibility

1. Introduction

Ion channels are protein molecules that form pores in cell membranes. Movement of ions through these pores determines, in part, the electrical properties of nerve and muscle cells, and is responsible, in part, for transmission of information through a nervous system. An understanding of the process by which ion channels open and close will provide important insights into the modes of actions of drugs (for example,

† Present address: Department of Statistics, The University of Warwick, Coventry, Warwickshire CV4 7AL, UK.

‡ Present address: Department of Probability and Statistics, University of Sheffield, Hicks Building, Sheffield S3 7RH, UK.

local anaesthetics) which perturb the functioning of these proteins, and into the rational design of novel drugs acting on a nervous system.

The patch-clamp technique enables the experimenter to record the current flowing across a single-ion channel (see, for example, Sakmann & Neher 1995). The gating mechanism of a single-ion channel is usually modelled by a finite-state-space continuous-time Markov chain (see, for example, Colquhoun & Hawkes 1977, 1981, 1982). The state space is usually partitioned into two classes, termed open and closed, corresponding to the receptor channel being open or closed. In principle, it is possible to observe only which class, rather than which state, the channel is in. In practice, the current is corrupted by noise and low-pass filtering, and is sampled, with the sampling interval being typically of the order of 100 μs, prior to analysis. The sequence of open and closed sojourns of the channel is often then reconstructed using some kind of 'threshold' algorithm. This reconstruction results in the loss of very short sojourns of the channel, in either the open or closed classes of states, a phenomenon known as time-interval omission. A principal objective of single-channel analysis is to make inferences concerning the structure and parameter values of the continuous-time Markov chain used to model the channel-gating mechanism. Such inferences can be made either from the reconstructed sequence of open and closed sojourns, or directly from the current record.

Maximum-likelihood parameter estimation, based on sojourn-time data, was first considered by Horn & Lange (1983). Ball & Sansom (1989) used a form of the likelihood of a sequence of sojourn times given by Fredkin *et al*. (1985) to develop a numerically more efficient algorithm; see also Chay (1988). All of these studies ignored the effect of time-interval omission. Ball *et al*. (1991, 1993) developed a semi-Markov framework for analysing single-channel sojourn-time data, both with and without time-interval omission. The key component of that framework, which is required for any likelihood-based inference, is its associated semi-Markov kernel. When time-interval omission is ignored, an explicit expression exists for the kernel and is essentially what was used by Ball & Sansom (1989). However, there is no simple expression for the kernel when time-interval omission is incorporated. Hawkes *et al*. (1990) derived a recursive expression for the time-interval omitted kernel, but it is computationally highly intensive and numerically unstable for even moderate sojourn times, t. Jalali & Hawkes (1992a, b) developed an asymptotic expression for the kernel as t tends to infinity, which turns out, for reasons given in Ball (1997), to be extremely good even for quite small values of t. Thus the exact and approximate expressions can be used in combination to closely approximate the kernel for all $t \geqslant 0$, and hence permit maximum likelihood parameter estimation (Hawkes *et al*. 1992). Other approximations to the kernel have also been considered (see, for example, Roux & Sauvé 1985; Crouzy & Sigworth 1990). Because of the difficulties associated with likelihood-based inference, alternative methods of inference have been considered, such as Poisson sampling (Ball *et al*. 1992).

It is clearly attractive to base inference directly on the current record, since the reconstructed sequence of open and closed sojourns is derived from it. Fredkin & Rice (1992a) considered parameter estimation directly from single-channel recordings using hidden Markov methodology (Baum *et al*. 1970; Rabiner & Juang 1986). Magleby & Weiss (1990a, b) used a simulation-based estimation procedure that explicitly modelled the effects of filtering and noise. Hidden Markov methods have also been used by Fredkin & Rice (1992b) to reconstruct the sequence of sojourns

of the channel and by Chung *et al.* (1990) for both identification of conductance levels of the channel and reconstruction. Fredkin & Rice (1992*a*) also provided a non-model-based method of determining the conductance levels and mean sojourn times within those levels.

All of the above methods of parameter estimation belong to what is usually called classical statistics. Another approach to inference, which is becoming increasingly pervasive in contemporary statistics, is the Bayesian method (see, for example, O'Hagan 1994; Lee 1989). In the Bayesian approach, it is assumed that prior to conducting an experiment the investigator has beliefs about the values of the unknown parameters, θ say, which can be expressed by a *prior* probability distribution having density, $\pi(\theta)$ say, with respect to an appropriate measure. If data x are then observed, the investigator's beliefs about θ are updated by Bayes's theorem and are now expressed by his/her *posterior* probability distribution having density $\pi(\theta \mid x)$ given by

$$\pi(\theta \mid x) = \frac{\pi(\theta) f(x \mid \theta)}{\int \pi(\theta') f(x \mid \theta') \, \mathrm{d}\theta'}, \tag{1.1}$$

where $f(x \mid \theta)$ is the likelihood of observing x when θ is the value of the unknown parameters and the integral is over the full range of possible values for θ.

A principal difficulty in using the Bayesian method in practice is the evaluation of the integral in (1.1). In many applications, the unknown parameter θ is a high-dimensional vector, increasing the difficulty of finding the integral. Often the quantity of interest is a subvector θ_1 of θ, the computation of whose posterior distribution, called the marginal posterior distribution, requires yet another difficult multidimensional integral. However, in the past decade, a group of techniques known collectively as Markov chain Monte Carlo (MCMC) methods (see, for example, Gilks *et al.* 1996) have revolutionized the implementation of the Bayesian paradigm, making it available to high-dimensional complex models that hitherto had been computationally prohibitive. The key idea underlying MCMC methods is to determine a Markov chain whose state space is that of θ and whose equilibrium distribution is $\pi(\theta \mid x)$. Simulation of such a Markov chain provides a method of sampling from $\pi(\theta \mid x)$ and thus enables the investigator to develop an MCMC estimate of $\pi(\theta \mid x)$ that is amenable to summarization and analysis. MCMC methods have been found to be very useful in inferential problems involving unobserved features and thus seem particularly appropriate for single-channel analysis. The aim of this paper is to develop and implement MCMC methods for making Bayesian inference directly from single-channel records, using the model of Fredkin & Rice (1992*a*), which incorporates a general Markov model for single-channel gating. Simultaneous with this work, Hodgson (1999) has considered an MCMC sampler for making inference directly from the current record for a two-state alternating renewal model for channel gating, with open and closed sojourns following gamma distributions, and an autoregressive process for the noise. Ball *et al.* (1996) explored the use of MCMC methods for parameter estimation from discrete sojourn-time ion-channel data with time-interval omission.

The remainder of the paper is structured as follows. A brief introduction to Bayesian inference and MCMC methods is given in § 2. The basic model underlying our inference directly from single-channel recordings is described in § 3, together with discussion of the prior and posterior distributions for its unknown parameters. The

basic MCMC sampler is described in § 4. Using the output from the MCMC sampler is discussed in § 5, where inferences about the unknown parameters and functions of these parameters, such as the mean lengths of open and closed sojourns are considered, as well as restoration of the unobserved single-channel record. Computational implementation of the MCMC sampler is considered in § 6. Extensions of the basic MCMC sampler, such as to channel-gating models that are necessarily time reversible and to models whose transition rates are specified functions of other parameters, are considered in § 7. Some examples of application of the MCMC sampler, using simulated data from known single-channel gating models, are given in § 8. The paper ends with some concluding comments in § 9.

2. Bayesian inference and Markov chain Monte Carlo methods

The aim of this section is to provide a brief introduction to Bayesian inference and Markov chain Monte Carlo (MCMC) methods, which is accessible to more numerate biologists. Readers familiar with these topics may wish to skip this section. An elementary introduction to Bayesian statistics is given in Lee (1989). More advanced treatises are given by O'Hagan (1994) and Bernardo & Smith (1994). A practical account of MCMC methods is given by Gilks *et al.* (1996).

(*a*) *Bayesian inference*

Consider the following single-ion channel-gating mechanism:

$$\mathrm{C}_2 \underset{q_{12}}{\overset{q_{21}}{\rightleftarrows}} \mathrm{O}_1 \underset{q_{31}}{\overset{q_{13}}{\rightleftarrows}} \mathrm{C}_3,$$

where O and C indicate open and closed states, respectively. Suppose that the gating behaviour of the channel is modelled by a continuous-time Markov chain, with transition rates q_{ij} as shown. Then successive open and closed sojourns of the channel are independent. Open sojourns have the probability density function

$$f_{\mathrm{O}}(t) = \lambda \exp(-\lambda t) \quad (t > 0), \tag{2.1}$$

where $\lambda = q_{12} + q_{13}$, and λ^{-1} is the mean length of an open sojourn. Closed sojourns have the more complex probability density function

$$f_{\mathrm{C}}(t) = \pi\lambda_1 \exp(-\lambda_1 t) + (1-\pi)\lambda_2 \exp(-\lambda_2 t) \quad (t > 0), \tag{2.2}$$

where $\lambda_1 = q_{21}$, $\lambda_2 = q_{31}$, so that λ_1^{-1} and λ_2^{-1} are the mean lengths of sojourns in states C_2 and C_3, respectively, and $\pi = q_{12}/(q_{12} + q_{13})$ is the proportion of closed sojourns that are in state C_2. Suppose that we have observations $t_1, s_1, t_2, s_2, \ldots, t_n, s_n$ of n successive pairs of open and closed sojourn times, and that we wish to use these as a basis for making inferences about the parameters λ, λ_1, λ_2 and π governing $f_{\mathrm{O}}(t)$ and $f_{\mathrm{C}}(t)$.

Consider first the problem of making inferences about λ, the parameter governing the open sojourn time distribution, from the observed open times $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$; the lengths of the closed sojourns clearly tell us nothing about λ. In the Bayesian method of inference it is assumed that, before observing the open sojourn times $\boldsymbol{t}$, the investigator has beliefs about the value of λ, which can be expressed by a *prior* probability distribution. Suppose that this prior probability distribution is

continuous with probability density function $p(\lambda)$ ($\lambda > 0$). The likelihood of observing the open sojourn times $\boldsymbol{t}$ is defined to be their joint probability density

$$f(\boldsymbol{t} \mid \lambda) = \prod_{i=1}^{n} f_{\mathrm{O}}(t_i) = \lambda^n \mathrm{e}^{-n\lambda\bar{t}}, \tag{2.3}$$

where

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

is the sample mean of observed open sojourn times. After observing the sojourn times, the investigator updates his/her beliefs about λ to the *posterior* probability distribution of λ given the observed data $\boldsymbol{t}$, which, by Bayes's theorem, has probability density function $p(\lambda \mid \boldsymbol{t})$ given by

$$p(\lambda \mid \boldsymbol{t}) = \frac{p(\lambda) f(\boldsymbol{t} \mid \lambda)}{f(\boldsymbol{t})} \quad (\lambda > 0), \tag{2.4}$$

where

$$f(\boldsymbol{t}) = \int_0^\infty p(\lambda) f(\boldsymbol{t} \mid \lambda)\, \mathrm{d}\lambda. \tag{2.5}$$

The posterior probability density function $p(\lambda \mid \boldsymbol{t})$ encapsulates the investigator's beliefs about the unknown parameter λ after observing the data $\boldsymbol{t}$. Any required inference about λ is then derived from this posterior distribution. For instance, either the posterior mean or mode (i.e. the mean or mode of the posterior distribution) will constitute a point estimate of λ. The posterior standard deviation is a measure of uncertainty about λ, since it measures the likely distance of λ from the posterior mean.

Note that $f(\boldsymbol{t})$ is independent of λ so (2.4) may be expressed as

$$p(\lambda \mid \boldsymbol{t}) \propto p(\lambda) f(\boldsymbol{t} \mid \lambda). \tag{2.6}$$

Thus it is straightforward to determine $p(\lambda \mid \boldsymbol{t})$ up to a constant of proportionality. However, in order to determine that constant of proportionality one needs to evaluate the integral (2.5), which often cannot be done analytically. An example of when this integral can be evaluated is if the investigator's prior beliefs about λ follow a gamma distribution, specifically

$$\begin{aligned} p(\lambda) &= \mathrm{Ga}(\lambda \mid \alpha, \beta) \\ &= \beta^\alpha \lambda^{\alpha-1} \exp(-\beta\lambda)/\Gamma(\alpha) \quad (\lambda > 0), \end{aligned}$$

where $\alpha > 0$ and $\beta > 0$ are constants and Γ denotes the gamma function, defined by

$$\Gamma(u) = \int_0^\infty v^{u-1} \mathrm{e}^{-v}\, \mathrm{d}v.$$

It then follows from (2.6) and (2.3) that

$$p(\lambda \mid \boldsymbol{t}) \propto \lambda^{\alpha+n-1} \exp(-(\beta + n\bar{t})\lambda),$$

and hence $p(\lambda \mid \boldsymbol{t}) = \mathrm{Ga}(\lambda \mid \alpha + n, \beta + n\bar{t})$. Thus the investigator's posterior beliefs about λ also follow a gamma distribution. A class of prior distributions is said to

be *conjugate* for a given likelihood if the posterior distribution belongs to that class whenever the prior distribution does. Clearly, whenever the investigator's prior beliefs can be represented by a prior (where 'prior' means 'prior distribution') from the conjugate family, this greatly simplifies a Bayesian analysis.

Now suppose that *a priori* the investigator does not feel able to say that any one value of λ is more probable than any other, which would therefore be expressed through the prior having density $p(\lambda) = c$ $(\lambda > 0)$, for some constant c, or $p(\lambda) \propto 1$. This is not mathematically possible because there is no value of c for which $p(\lambda)$ is a proper probability distribution, since $\int_0^\infty p(\lambda)\,\mathrm{d}\lambda$ is infinite if $c \neq 0$ and zero if $c = 0$. There is nevertheless a sense in which such an *improper* prior distribution can be used. Suppose again that $p(\lambda)$ is the conjugate $\mathrm{Ga}(\lambda \mid \alpha, \beta)$ density. Notice that we could have simplified $p(\lambda)$ to $p(\lambda) \propto \lambda^{\alpha-1}\exp(-\beta\lambda)$, or the even more loosely written $p(\lambda) = \lambda^{\alpha-1}\exp(-\beta\lambda)$, since the omitted constants will cancel out in (2.6). If we set $\alpha = 1$ and $\beta = 0$, we obtain $p(\lambda) \propto 1$, or $p(\lambda) = 1$, and inserting these values into the posterior distribution shows that $p(\lambda \mid \boldsymbol{t}) = \mathrm{Ga}(\lambda \mid n+1, n\bar{t}\,)$. The $\mathrm{Ga}(1,0)$ prior distribution is improper because the β parameter of a gamma distribution needs to be strictly positive, but the posterior is nevertheless a genuine proper distribution. The investigator's prior beliefs cannot logically be described by this improper distribution, but distributions such as this have a role as *approximations* to genuine prior beliefs when prior information is weak. Even though $p(\lambda) = 1$ is not a genuine prior density, any genuine prior which expressed weak prior information would be very flat over a wide range of λ values, so that $p(\lambda \mid \boldsymbol{t}) = \mathrm{Ga}(\lambda \mid n+1, n\bar{t}\,)$ will be an accurate approximation of the resulting posterior density.

(*b*) *MCMC samplers*

In some circumstances the investigator may not feel that his prior beliefs about λ can be adequately approximated by a prior chosen from the conjugate gamma family of distributions. For such cases the integral (2.5) cannot normally be evaluated analytically. In the present setting the integral could be evaluated numerically but this may not be convenient, or even possible, for situations where there are several unknown parameters so that (2.5) becomes a high-dimensional integral. MCMC methods are an alternative to numerical integration for such problems.

The basic idea behind MCMC is to draw a large sample of values of λ randomly from the posterior distribution $p(\lambda \mid \boldsymbol{t})$. If we denote this *Monte Carlo* sample by $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(N)}$, then we can obtain accurate approximations to any desired summary of the posterior distribution. For instance, the Monte Carlo sample mean

$$\bar{\lambda} = \frac{1}{N}\sum_{k=1}^{N} \lambda^{(k)}$$

will be an approximation to the posterior mean, and for large enough N can be made as accurate an approximation as the investigator desires. Similarly, the Monte Carlo sample variance will approximate the posterior variance, and so on. The question then arises of how to carry out this scheme, how to draw a random sample from $p(\lambda \mid \boldsymbol{t})$. In the case of the foregoing example, where $p(\lambda \mid \boldsymbol{t})$ is a gamma distribution, then there are well-known procedures for drawing pseudo-random samples from gamma distributions (see Devroye 1986), but it is precisely when the posterior distribution is not some standard well-known distribution that we wish to apply the Monte Carlo

Table 1. *Illustration of MCMC*

iteration k	0	1	2	3	4
current $\lambda^{(k)}$	1	6.323	6.323	7.671	6.651
proposal λ'	6.323	7.665	7.671	6.651	7.811
acceptance prob. B	0.0488	0.6858	0.8773	1	0.7295
uniform draw U	0.0159	0.6885	0.8682	—	0.6295
new $\lambda^{(k+1)}$	6.323	6.323	7.671	6.651	7.811

approach. MCMC is a tool for sampling from complex high-dimensional distributions. Instead of a sequence of *independent* draws from $p(\lambda \mid \boldsymbol{t})$, we construct a Markov chain $\{\lambda^{(k)}; k = 0, 1, \dots\}$ whose *unique* equilibrium distribution is $p(\lambda \mid \boldsymbol{t})$. Then the theory of Markov chains tells us that, conditional on any arbitrary starting value $\lambda^{(0)}$, the chain eventually *converges* to this equilibrium, in the sense that the distribution of $\lambda^{(k)}$ will be arbitrarily close to $p(\lambda \mid \boldsymbol{t})$ for all k greater than some sufficiently large M. Therefore if we simulate a realization of this Markov chain, starting from an arbitrary $\lambda^{(0)}$, we can use the sequence $\lambda^{(M)}, \lambda^{(M+1)}, \dots$ as our Monte Carlo sample. Because the successive values in a Markov chain will in general be correlated, we will require a longer sequence to achieve adequate accuracy in approximations such as $\bar{\lambda}$, but the ease with which suitable Markov chains can be constructed for complex problems makes this a powerful computational device. Note that in practice the 'burn-in time' M will not be known, and it is necessary to use empirical diagnostic methods for determining when the chain has been run sufficiently long for convergence to have taken place (see § 5 *a* for more details).

To illustrate MCMC methods, suppose that we wish to draw a sample from the posterior distribution (2.4) based on the likelihood (2.3) and an arbitrary prior $p(\lambda)$. Consider a Markov chain $\{\lambda^{(k)}; k = 0, 1, \dots\}$ constructed as follows. Given $\lambda^{(k)}$, $\lambda^{(k+1)}$ is obtained by a two-stage process in which a random candidate value is *proposed* and then a random decision is taken whether to *accept* that proposal. Before considering more fully the reasoning behind this idea, let us see how in this example it can lead to a Markov chain being produced with the required equilibrium distribution. For this example, first suppose that we generate a proposal λ' by sampling from the gamma distribution with density $\mathrm{Ga}(\lambda' \mid n+1, n\bar{t})$ and then that we accept this proposal with probability given by $B = \min(p(\lambda')/p(\lambda^{(k)}), 1)$. This acceptance condition is effected by drawing a random number U uniformly over $(0, 1)$ and accepting the proposal if $U \leqslant B$. If the proposal is accepted then $\lambda^{(k+1)} = \lambda'$, otherwise $\lambda^{(k+1)} = \lambda^{(k)}$. Table 1 gives a numerical example to illustrate a sequence of steps of this procedure. The particular numerical values are $n = 100$, $\bar{t} = 0.15$, and $p(\lambda) = 2/\{\pi(1+\lambda^2)\}$, so that the proposals are sampled from the $\mathrm{Ga}(101, 15)$ distribution and accepted with probability B, where $B = 1$ if $\lambda' \leqslant \lambda^{(k)}$ and $B = (1+\{\lambda^{(k)}\}^2)/(1+\{\lambda'\}^2)$ if $\lambda' > \lambda^{(k)}$. The chain is started from $\lambda^{(0)} = 1$. The particular random numbers would, of course, change if we were to run the example again, resulting in different realizations of $\lambda^{(1)}, \lambda^{(2)}, \dots$.

Notice that the proposal in iteration 1 is rejected because $U > B$, and hence the current value of $\lambda^{(1)}$ is retained for $\lambda^{(2)}$. At iteration 3, $B = 1$, so the proposal is automatically accepted and it is not necessary to make the random uniform draw U.

We now show that this procedure will actually have the required equilibrium distribution. For $\lambda \neq \lambda'$, let $h(\lambda, \lambda')$ denote the probability density function of $\lambda^{(k+1)} = \lambda'$

given that $\lambda^{(k)} = \lambda$. Then

$$h(\lambda, \lambda') = \mathrm{Ga}(\lambda' \mid n+1, n\bar{t}\,) \min(p(\lambda')/p(\lambda), 1).$$

It is easily checked, by considering the cases $p(\lambda) \geqslant p(\lambda')$ and $p(\lambda) < p(\lambda')$ separately, that

$$\frac{\min(p(\lambda')/p(\lambda), 1)}{\min(p(\lambda)/p(\lambda'), 1)} = p(\lambda')/p(\lambda).$$

Hence

$$\begin{aligned}\frac{h(\lambda, \lambda')}{h(\lambda', \lambda)} &= \frac{\mathrm{Ga}(\lambda' \mid n+1, n\bar{t}\,)p(\lambda')}{\mathrm{Ga}(\lambda \mid n+1, n\bar{t}\,)p(\lambda)} \\ &= \frac{(\lambda')^n \exp(-\lambda n\bar{t}\,)p(\lambda')}{\lambda^n \exp(-\lambda nt)p(\lambda)} \\ &= \frac{p(\lambda' \mid \boldsymbol{t})}{p(\lambda \mid \boldsymbol{t})},\end{aligned}$$

using (2.3) and (2.4). Hence

$$p(\lambda \mid \boldsymbol{t})h(\lambda, \lambda') = p(\lambda' \mid \boldsymbol{t})h(\lambda', \lambda) \quad (\lambda, \lambda' > 0),$$

so $p(\lambda \mid \boldsymbol{t})$ satisfies the detailed balance conditions (cf. Kelly 1979, p. 5) for $\{\lambda^{(k)}; k = 0, 1, \dots\}$. It follows that $\{\lambda^{(k)}; k = 0, 1, \dots\}$ is time reversible with equilibrium distribution having density $p(\lambda \mid \boldsymbol{t})$. Thus the investigator can build up a picture of their posterior distribution for λ, and can approximate accurately any desired summary of that distribution, by simulating a long run of the above Markov chain. (This computational procedure of creating a Markov chain to sample from the posterior distribution is, of course, quite distinct from the Markov chain that generates the ion-channel sojourn data.)

The device of a random proposal followed by a random acceptance decision may seem convoluted, but is the basis of a general algorithm called the Metropolis–Hastings algorithm (Metropolis *et al*. 1953; Hastings 1970), which allows us to design a Markov chain specifically to achieve a target equilibrium distribution. An idea of how it works can be obtained by considering how such a Markov chain should behave. The new value $\lambda^{(k)}$ should, once the chain is in equilibrium, be a random draw from $p(\lambda \mid \boldsymbol{t})$. This means that it should be most likely to fall in regions where both the likelihood and the prior density are relatively high, but should also have appropriately smaller probabilities of falling in other regions. Now consider the two stages separately. In this example, the proposal density is proportional to the likelihood, and so is designed to generate proposals λ' that fall predominantly in areas where the likelihood is high. The acceptance probability $A = \min(p(\lambda')/p(\lambda^{(k)}), 1)$ emphasizes areas where the prior density is high by giving higher probabilities of acceptance when the proposal value has higher prior density. Indeed, if the proposal has higher prior density than the current value λ, it will always be accepted.

The general Metropolis–Hastings algorithm can be described as follows. Let η and y denote the unknown parameter(s) and data, respectively, in an arbitrary statistical problem, where η and y may both be vectors, and suppose that we wish to sample from the posterior distribution $p(\eta \mid y)$. Then we construct a Markov chain $\{\eta^{(k)}; k = 0, 1, \dots\}$ as follows. Given $\eta^{(k)}$, a proposal η' is generated from a distribution having

density $g(\eta' \mid \eta^{(k)})$. That proposal is accepted with probability $A = \min(1, B)$, where

$$B = \frac{p(\eta' \mid y)g(\eta^{(k)} \mid \eta')}{p(\eta^{(k)} \mid y)g(\eta' \mid \eta^{(k)})}. \tag{2.7}$$

If the proposal is accepted, $\eta^{(k+1)} = \eta'$. Otherwise $\eta^{(k+1)} = \eta^{(k)}$. Under very general conditions (Gilks *et al.* 1996), a chain generated this way converges to the posterior $p(\eta \mid y)$ sought. (The formula (2.7) was used to derive the acceptance probability in our earlier example.)

When η is a vector, Metropolis–Hastings methods can be applied to components of η rather than to all of η at once. Suppose that $\eta = (\eta_1, \eta_2, \eta_3)$, where η_1, η_2 and η_3 may also be vectors. Consider for instance a proposal distribution for $\eta' = (\eta_1', \eta_2', \eta_3')$ such that, with probability 1, $\eta_1' = \eta_1$ and $\eta_2' = \eta_2$, so that only η_3 is proposed to change. Clearly, a chain in which this proposal distribution is used at every iteration cannot converge uniquely to the required posterior distribution. We can, however, use different proposal distributions at different iterations. We would refer to these as different kinds of 'steps' in the chain, so that a proposal in which only η_3 changes could be referred to as an η_3 step.

Different kinds of steps may be used in a deterministic sequence, for instance by using an η_1 step, an η_2 step and an η_3 step cyclically. If we think of the combination of the three steps as a single iteration, the Markov chain is homogeneous and the usual theory can be applied to confirm convergence to the required posterior. Similarly we can apply different kinds of steps by making a random choice of step type at each iteration. There is thus great flexibility available in constructing MCMC samples, as explored in Besag *et al.* (1995).

An important special case of the Metropolis–Hastings algorithm is the Gibbs sampler. Suppose that η is a vector and write $\eta = (\eta_1, \eta_2)$, where η_1 and η_2 may also be vectors. Given $\eta^{(k)}$ $(= (\eta_1^{(k)}, \eta_2^{(k)}))$, generate a proposal $\eta' = (\eta_1^{(k)}, \eta_2')$, where η_2' is generated from the distribution $p(\eta_2' \mid \eta_1^{(k)}, y)$. Then, noting that

$$p(\eta \mid y) = p(\eta_1 \mid y)p(\eta_2 \mid \eta_1, y),$$

it follows from (2.7) that $B = 1$, so the proposal η' is necessarily accepted. Proposals of this type are called Gibbs steps. A Markov chain sampler comprising only Gibbs steps is called a Gibbs sampler (see Geman & Geman 1984).

(*c*) *A more complex example*

Consider now the problem of making inferences about $\eta = (\lambda_1, \lambda_2, \pi)$, where λ_1, λ_2 and π are the parameters governing the distribution of closed sojourn times, given by (2.2). Suppose, for the sake of illustration, that the investigator's prior beliefs about λ_1, λ_2 and π are independent, so $p(\eta) = p(\lambda_1)p(\lambda_2)p(\pi)$, and that weak prior information is expressed through the uniform priors $p(\lambda_1) = 1$ $(\lambda_1 > 0)$, $p(\lambda_2) = 1$ $(\lambda_2 > 0)$ and $p(\pi) = 1$ $(0 < \pi < 1)$. The first two of these distributions are improper, and a complication involved in using improper prior distributions is that possibly the posterior distribution is improper, in which case the MCMC chain will not converge. Fortunately, for this likelihood and improper prior (and indeed in all the situations where improper priors are introduced in this paper, unless stated explicitly to the contrary), the posterior is proper, so the difficulty does not arise here.

Let $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$ denote the vector of closed sojourn times. Then

$$p(\eta \mid \boldsymbol{s}) \propto \prod_{i=1}^{n} f_{\mathrm{C}}(s_i \mid \eta),$$

where $f_{\mathrm{C}}(s_i \mid \eta)$ is given by the right-hand side of (2.2) with t replaced by s_i. Note that this expression of $p(\eta \mid \boldsymbol{s})$ as a product of sums cannot be simplified, and the task of computing it when n is realistically large renders direct MCMC methods cumbersome. The difficulty is owing to not knowing whether any given closed sojourn is in state C_2 or C_3 and it can be overcome by introducing this information in the form of a further vector parameter $\delta = (\delta_1, \delta_2, \ldots, \delta_n)$, where, for $i = 1, 2, \ldots, n$,

$$\delta_i = \begin{cases} 1, & \text{if the sojourn } s_i \text{ is in state } \mathrm{C}_2, \\ 0, & \text{if the sojourn } s_i \text{ is in state } \mathrm{C}_3. \end{cases}$$

(This technique of including unobserved information as extra parameters is known as *data augmentation* (see, for example, Tanner & Wong 1987).) It then follows that

$$\begin{aligned} p(\eta, \delta \mid \boldsymbol{s}) &\propto p(\eta, \delta, \boldsymbol{s}) \\ &= p(\eta) p(\delta \mid \eta) p(\boldsymbol{s} \mid \eta, \delta) \\ &= \prod_{i=1}^{n} \{\pi^{\delta_i} (1-\pi)^{1-\delta_i}\} \prod_{i=1}^{n} \{(\lambda_1 \exp(-\lambda_1 s_i))^{\delta_i} (\lambda_2 \exp(-\lambda_2 s_i))^{1-\delta_i}\}, \end{aligned} \quad (2.8)$$

since $p(\eta) = 1$.

Let $n_1 = \sum_{i=1}^{n} \delta_i$ and $n_2 = n - n_1$ be the numbers of sojourns in C_2 and C_3, respectively, and let

$$\bar{s}_1 = n_1^{-1} \sum_{i=1}^{n} \delta_i s_i \quad \text{and} \quad \bar{s}_2 = n_2^{-1} \sum_{i=1}^{n} (1 - \delta_i) s_i$$

be the corresponding mean sojourn times. Then the product form of (2.8) implies that

$$p(\lambda_1, \lambda_2, \pi \mid \delta, \boldsymbol{s}) = p(\lambda_1 \mid \delta, \boldsymbol{s}) p(\lambda_2 \mid \delta, \boldsymbol{s}) p(\pi \mid \delta),$$

where $p(\lambda_i \mid \delta, \boldsymbol{s}) = \mathrm{Ga}(\lambda_i \mid n_i + 1, n_i \bar{s}_i)$ $(i = 1, 2)$ and

$$\begin{aligned} p(\pi \mid \delta) &= \mathrm{Be}(\pi \mid n_1 + 1, n_2 + 1) \\ &= \frac{\pi^{n_1} (1-\pi)^{n_2}}{\beta(n_1 + 1, n_2 + 1)} \quad (0 < \pi < 1), \end{aligned}$$

where β denotes the beta function defined by

$$\beta(x, y) = \int_0^1 u^{x-1} (1-u)^{y-1} \, \mathrm{d}u.$$

Also

$$p(\delta \mid \lambda_1, \lambda_2, \pi, \boldsymbol{s}) = \prod_{i=1}^{n} p(\delta_i \mid \lambda_1, \lambda_2, \pi, s_i),$$

where, for $i = 1, 2, \ldots, n$,

$$p(\delta_i \mid \lambda_1, \lambda_2, \pi, s_i) = \frac{\{\pi \lambda_1 \exp(-\lambda_1 s_i)\}^{\delta_i} \{(1-\pi) \lambda_2 \exp(-\lambda_2 s_i)\}^{1-\delta_i}}{\pi \lambda_1 \exp(-\lambda_1 s_i) + (1-\pi) \lambda_2 \exp(-\lambda_2 s_i)} \quad (\delta_i = 0, 1).$$

We can now construct a process $\{(\eta^{(k)}, \delta^{(t)}); k = 0, 1, \dots\}$ as follows. Suppose that the current state is

$$(\eta^{(k)}, \delta^{(k)}) = (\lambda_1^{(k)}, \lambda_2^{(k)}, \pi^{(k)}, \delta_1^{(k)}, \delta_2^{(k)}, \dots, \delta_n^{(k)}).$$

We first determine $\eta^{(k+1)}$ by simulating $\lambda_1^{(k+1)}$, $\lambda_2^{(k+2)}$ and $\pi^{(k+1)}$ independently from the distributions

$$\mathrm{Ga}(\lambda_1^{(k+1)} \mid n_1^{(k)} + 1, n_1^{(k)}\bar{s}_1^{(k)}), \quad \mathrm{Ga}(\lambda_2^{(k+1)} \mid \eta_2^{(k)} + 1, n_2^{(k)}\bar{s}_2^{(k)}),$$
$$\mathrm{Be}(\pi^{(k+1)} \mid n_1^{(k)} + 1, n_2^{(k)} + 1),$$

respectively. We then determine $\delta_1^{(k+1)}, \delta_2^{(k+1)}, \dots, \delta_n^{(k+1)}$ by simulating independently from the distributions

$$p(\delta_i^{(k+1)} \mid \lambda_1^{(k+1)}, \lambda_2^{(k+1)}, \pi^{(k+1)}, s_i) \quad (i = 1, 2, \dots, n).$$

Finally, we use $\delta_1^{(k+1)}, \delta_2^{(k+1)}, \dots, \delta_n^{(k+1)}$ to determine $n_i^{(k+1)}$, $\bar{s}_i^{(k+1)}$ $(i = 1, 2)$. Thus at each iteration η and δ are both being updated by Gibbs steps, and this Markov chain is therefore a Gibbs sampler. A process $\{(\eta^{(k)}, \delta^{(k)}); k = 1, 2, \dots\}$ constructed in this fashion is a realization of a Markov chain whose equilibrium distribution is $p(\eta, \delta \mid \boldsymbol{s})$. Thus, once this realization has settled into equilibrium, the process $\{\eta^{(k)}; k = 0, 1, \dots\}$ can be used to sample from $p(\eta \mid \boldsymbol{s})$.

(*d*) *Dimension changes*

For a time-reversible Markov model of single-channel gating that is in equilibrium, the length of a typical sojourn in the closed class of states follows a mixture of negative exponential random variables (Kijima & Kijima 1987) and has a probability density function of the form

$$f_{\mathrm{C}}(t) = \sum_{i=1}^{k_{\mathrm{C}}} \pi_i \lambda_i \exp(-\lambda_i t) \quad (t > 0), \tag{2.9}$$

where $\pi_i > 0$ $(i = 1, 2, \dots, k_{\mathrm{C}})$ and $\sum_{i=1}^{k_{\mathrm{C}}} \pi_i = 1$. (A similar result holds for open sojourns.) The number of components k_{C} comprising the mixture gives a lower bound on the number of closed states in the gating mechanism and normally this bound is attained. Thus experimenters attempt to estimate the numbers of open and closed states by fitting distributions of the form (2.9) to data on open and closed sojourns.

The Gibbs sampler described above for the case $k_{\mathrm{C}} = 2$ can easily be extended to the case $k_{\mathrm{C}} > 2$, provided that k_{C} is known. However, k_{C} is unknown in the above setting and thus needs to be treated as another parameter. It is still possible to develop a Metropolis–Hastings MCMC sampler when k_{C} is a parameter, but note now that as k_{C} changes so does the number of other parameters $\{\pi_i, \lambda_i : i = 1, 2, \dots, k_{\mathrm{C}}\}$ required to describe the mixture density. A consequence of this is that for proposals that change the value of k_{C}, the expression for B in (2.7) needs to be multiplied by an appropriate Jacobian, to create a reversible jump step (Green 1995). See Richardson & Green (1997) for a state-of-the-art account of MCMC methods for mixture distributions with an unknown number of components.

In this paper we develop an MCMC sampler for making Bayesian inference for ion-channel gating mechanisms directly from a single-channel current record. The data are the channel record, denoted by y. The unknown parameters are the transition

rates of the underlying continuous-time Markov chain used to model the channel, denoted by q, and parameters associated with the recording process, denoted by θ. The aim is to develop a method for sampling from the posterior distribution $p(q, \theta \mid y)$. However, similar to the mixture problem described above, the analysis would be greatly simplified if we knew the realization, x say, of the channel gating process. Thus we develop an MCMC sampler for $p(x, q, \theta \mid y)$. Our sampler involves a deterministic cycle, comprising a q step, then a θ step, then an x step, but each of these steps is further subdivided, and there is random choice in the x step. Note that the dimension of x is determined by the number of gating events that occur during the single-channel recording, which is unknown. Thus the reversible jump methodology of Green (1995) is required when updating x.

3. Model, prior and posterior

(a) *Model*

The gating mechanism of a single-ion channel is modelled by a continuous-time Markov chain $\{X(t); t \geqslant 0\}$, with finite state space $S = \{1, 2, \ldots, m\}$. Thus $X(t)$ denotes the state of the channel at time t and S denotes the possible states of the channel. The state space S is partitioned into $\mathrm{O} = \{1, 2, \ldots, m_{\mathrm{O}}\}$ and $\mathrm{C} = \{m_{\mathrm{O}} + 1, m_{\mathrm{O}} + 2, \ldots, m\}$, corresponding to the receptor channel being open and closed, respectively. Let Q be the transition rate matrix of $\{X(t); t \geqslant 0\}$. Thus Q is an $m \times m$ matrix with elements q_{ij}, where, for $i \neq j$, q_{ij} is the transition rate of the channel from state i to state j and the diagonal elements of Q are defined so that the row sums are all zero, i.e. $q_{ii} = -\sum_{j \neq i} q_{ij}$. We shall assume that $\{X(t); t \geqslant 0\}$ possesses an equilibrium distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m)$, where π_i is the equilibrium probability that the channel is in state i. The equilibrium distribution $\boldsymbol{\pi}$ can be obtained from Q by solving the equations

$$\boldsymbol{\pi} Q = \mathbf{0}, \quad \boldsymbol{\pi} \mathbf{1} = 1.$$

Throughout the paper $\mathbf{0}$ and $\mathbf{1}$ denote column vectors of zeros and ones, respectively, whose dimension, in the present case m, is apparent from the context. In many applications, $\{X(t); t \geqslant 0\}$ is necessarily time reversible, in which case $\boldsymbol{\pi}$ also satisfies the detailed balance equations

$$\pi_i q_{ij} = \pi_j q_{ij} \quad (i, j \in S).$$

See Läuger (1995) for a discussion of time reversibility in the ion-channel context.

In practice, the process $\{X(t); t \geqslant 0\}$ is not observed. Instead, the current flowing across the channel is recorded at times $0, \Delta, 2\Delta, \ldots, N\Delta$, where Δ denotes the sampling interval. Let $T = N\Delta$, $X = \{X(t); 0 \leqslant t \leqslant T\}$ and $x = \{x(t); 0 \leqslant t \leqslant T\}$ be the unobserved realization of X. For $l = 0, 1, \ldots, N$, let Y_l be a random variable describing the current flowing across the channel at time $l\Delta$. Following Fredkin & Rice (1992*a*), the observed process is modelled by assuming that

$$Y_l = c(x(l\Delta)) + \varepsilon_l \quad (l = 0, 1, \ldots, N), \tag{3.1}$$

where

$$c(x(l\Delta)) = \begin{cases} \mu_{\mathrm{O}}, & \text{if } x(l\Delta) \in \mathrm{O}, \\ \mu_{\mathrm{C}}, & \text{if } x(l\Delta) \in \mathrm{C} \end{cases}$$

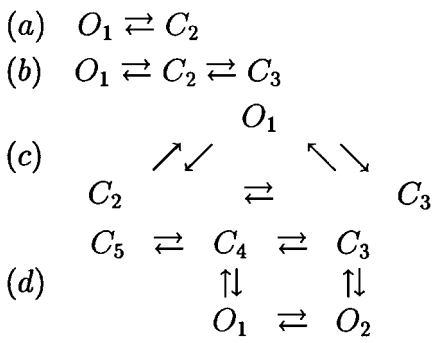

Figure 1. Examples of graphs associated with single-channel gating mechanisms.

and $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_N$ are independent normal random variables, each having mean zero, with variances given by

$$\mathrm{var}(\varepsilon_l) = \begin{cases} \sigma_{\mathrm{O}}^2, & \text{if } x(l\Delta) \in \mathrm{O}, \\ \sigma_{\mathrm{C}}^2, & \text{if } x(l\Delta) \in \mathrm{C}. \end{cases}$$

Thus the model assumes that there are no subconductance levels. If the channel does have subconductance levels, then our methodology can easily be extended to incorporate that phenomenon. As pointed out by Fredkin & Rice (1992*a*), the model is not quite faithful to the data, since it ignores low-pass filtering. Again, our methodology can be extended to incorporate this phenomenon. We briefly outline these extensions in § 9.

The transition rate matrix Q induces a directed graph, G say, on S, in which, for any ordered pair of states (i, j), there is a directed arc joining i to j if and only if $q_{ij} > 0$. Let $E = \{(i, j) : q_{ij} > 0\}$ be the set of (directed) edges in G and let $q = \{q_{ij} : (i, j) \in E\}$. Clearly q determines Q, and hence also $\boldsymbol{\pi}$. Thus we write $\boldsymbol{\pi} = \boldsymbol{\pi}(q) = (\pi_1(q), \pi_2(q), \ldots, \pi_m(q))$.

Examples of graphs G for some gating mechanisms are shown in figure 1. In these mechanisms open states are indicated by O and closed states by C, with subscripts indicating the state label in S. The vectors q for these examples are: (*a*) $q = (q_{12}, q_{21})$; (*b*) $q = (q_{12}, q_{21}, q_{23}, q_{32})$; (*c*) $q = (q_{12}, q_{21}, q_{23}, q_{32}, q_{31}, q_{13})$ and (*d*) $q = (q_{12}, q_{21}, q_{14}, q_{41}, q_{23}, q_{32}, q_{34}, q_{43}, q_{45}, q_{54})$. Let $\theta = (\mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2, \mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2)$. Our aim is to make inferences concerning q and θ from an observation, $y = (y_0, y_1, \ldots, y_N)$ say, of $Y = (Y_0, Y_1, \ldots, Y_N)$.

(*b*) *Prior*

Let $p(q, \theta)$ be a density describing our prior beliefs about q and θ. We shall assume that our prior beliefs about q and θ are independent, i.e. that

$$p(q, \theta) = p(q)p(\theta), \tag{3.2}$$

and that

$$p(q) = \prod_{(i,j)\in E} p(q_{ij}), \tag{3.3}$$

with

$$p(q_{ij}) = \mathrm{Ga}(q_{ij} \mid \alpha_{ij}, \beta_{ij}) = \beta_{ij}^{\alpha_{ij}} q_{ij}^{\alpha_{ij}-1} \exp(-\beta_{ij} q_{ij})/\Gamma(\alpha_{ij}) \quad (0 < q_{ij} < \infty), \tag{3.4}$$

and

$$p(\theta) = p(\mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2, \mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2) = (\sigma_{\mathrm{O}}^2 \sigma_{\mathrm{C}}^2)^{-1} \quad (-\infty < \mu_{\mathrm{O}},\ \mu_{\mathrm{C}} < \infty,\ 0 < \sigma_{\mathrm{O}}^2,\ \sigma_{\mathrm{C}}^2 < \infty). \tag{3.5}$$

The constants α_{ij}, β_{ij} in (3.4) satisfy $\alpha_{ij} > 0$, $\beta_{ij} > 0$.

Thus the prior beliefs about the elements of the q are that they follow independent gamma distributions, with q_{ij} having mean α_{ij}/β_{ij} and variance α_{ij}/β_{ij}^2. This form of prior is chosen for q as it is conjugate for the problem under consideration. The constants α_{ij}, β_{ij} need to be specified by the researcher, reflecting his/her prior beliefs. We also allow the possibility that $\alpha_{ij} = 1$, $\beta_{ij} = 0$, in which case q_{ij} has an improper vague prior. The prior beliefs for θ given by (3.5) are that $\mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2, \mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2$ follow independent improper vague priors. It should be noted that in the present ion-channel setting the dimension $N + 1$ of the data y is typically very large. In these circumstances the posterior for (q, θ) is heavily dominated by the data and any sensible prior is likely to lead to similar conclusions.

(*c*) *Posterior*

Let $f(y \mid q, \theta)$ denote the joint probability density function of Y given the parameters q and θ. Then, by Bayes's theorem, the posterior density of (q, θ) given the observed data y is given by

$$p(q, \theta \mid y) = \frac{f(y \mid q, \theta) p(q, \theta)}{p(y)}, \tag{3.6}$$

where

$$p(y) = \int f(y \mid q, \theta) p(q, \theta)\, \mathrm{d}q\, \mathrm{d}\theta, \tag{3.7}$$

with the integral being over the full range of possible values for q and θ.

Fredkin & Rice (1992*a*) give an algorithm for evaluating $f(y \mid q, \theta)$. However, there are two problems in using it to determine the posterior density $p(q, \theta \mid y)$. Firstly, the algorithm of Fredkin & Rice (1992*a*) is computationally highly intensive. Secondly, even if $f(y \mid q, \theta)$ were cheaply available one still needs to evaluate the integral in (3.7) to determine $p(q, \theta \mid y)$.

The difficulty in evaluating $f(y \mid q, \theta)$ is due to the realization x of the channel-gating process being unobservable. If x were observable then it would be straightforward to determine $p(q, \theta \mid y)$. The above problems can, to a certain extent, be overcome by treating x as an unknown parameter. Note that x is a random function of q, involving no other parameters, so we do not have to specify a prior for x; indeed, it would be incorrect to do so. Put another way, our prior beliefs about q and the model automatically determine our prior beliefs concerning x.

The form of the model (3.1) and prior (3.2) implies that

$$p(x, q, \theta, y) = p(q) p(\theta) f(x \mid q) f(y \mid x, \theta), \tag{3.8}$$

where $f(x \mid q)$ denotes the probability density of the realization x given the parameter q. Further, by Bayes's theorem,

$$p(x, q, \theta \mid y) = p(x, q, \theta, y)/p(y), \tag{3.9}$$

where now

$$p(y) = \int p(x, q, \theta, y)\, \mathrm{d}x\, \mathrm{d}q\, \mathrm{d}\theta. \tag{3.10}$$

The integral in (3.10) is now over the full range of possible values of x, q and θ, and again it cannot be easily evaluated. (Even if it could, we would then have to integrate $p(x, q, \theta \mid y)$ over all possible x to obtain $p(q, \theta \mid y)$.) However, the form of $p(x, q, \theta, y)$ in (3.8) implies that the conditional densities $p(x \mid q, \theta, y)$, $p(q \mid x, \theta, y)$ and $p(\theta \mid x, q, y)$ all admit relatively simple forms. It is this fact that makes MCMC methods particularly appropriate for the present ion-channel problem.

4. Markov chain Monte Carlo sampler

(*a*) *Notation and overview*

In order to describe our sampler, we first need to parametrize the unobserved realization $x = \{x(t); 0 \leqslant t \leqslant T\}$ explicitly. Let n be the number of jumps made by x, where a jump of x is a change of state (which may or may not involve a change of conductance level of the channel). Let $\chi_0 = x(0)$, $s_0 = 0$ and $s_{n+1} = T$. Provided $n > 0$, for $l = 1, 2, \ldots, n$ let $s_l = \min_{t > s_{l-1}}\{t : x(t) \neq \chi_{l-1}\}$ and $\chi_l = x(s_l)$. For $l = 0, 1, 2, \ldots, n$, let $t_l = s_{l+1} - s_l$. Then x is parametrized by $x = (n, \boldsymbol{t}, \boldsymbol{\chi})$, where $\boldsymbol{t} = (t_0, t_1, \ldots, t_n)$ is a vector containing the lengths of successive sojourns of x and $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_n)$ gives the states visited in these sojourns.

We shall construct a Markov chain $\{(x^{(k)}, q^{(k)}, \theta^{(k)}); k = 0, 1, \ldots\}$ whose equilibrium distribution is $p(x, q, \theta \mid y)$, where, in obvious notation, $x^{(k)}$ is parametrized by $x^{(k)} = (n^{(k)}, \boldsymbol{t}^{(k)}, \boldsymbol{\chi}^{(k)})$. We shall do this by developing a Metropolis–Hastings sampler that sequentially updates q, θ and x. We now describe the methods for updating q, θ and x. In describing these methods we shall assume that the current state of the sampler is $(x^{(k)}, q^{(k)}, \theta^{(k)}) = (x, q, \theta)$, i.e. for notational simplicity we drop the explicit dependence of (x, q, θ) on k.

(*b*) *Update q*

For $l = 0, 1, \ldots, n$ and $i = 1, 2, \ldots, m$ let

$$\delta_{li} = \begin{cases} 1, & \text{if } \chi_l = i, \\ 0, & \text{if } \chi_l \neq i. \end{cases}$$

For $i = 1, 2, \ldots, m$ let $\tilde{t}_i = \sum_{l=0}^{n} t_l \delta_{li}$ be the total time during $[0, T]$ spent by x in state i. For $i, j = 1, 2, \ldots, m$ $(i \neq j)$ let $n_{ij} = \sum_{l=0}^{n-1} \delta_{li}\delta_{l+1,j}$ be the total number of jumps from state i to state j made by x. Then, since the successive sojourns of X follow conditionally independent negative exponential distributions and $P(\chi_{l+1} =$

$j \mid \chi_l = i) = -q_{ii}^{-1} q_{ij}$,

$$f(x \mid q) = \pi_{\chi_0}(q) \left\{ \prod_{l=0}^{n-1} \left[-q_{\chi_l \chi_l} \exp(q_{\chi_l \chi_l} t_l) \left(\frac{q_{\chi_l \chi_{l+1}}}{-q_{\chi_l \chi_l}} \right) \right] \right\} \exp(q_{\chi_n \chi_n} t_n) \tag{4.1}$$

$$= \pi_{\chi_0}(q) \prod_{(i,j) \in E} q_{ij}^{n_{ij}} \exp(-\tilde{t}_i q_{ij}). \tag{4.2}$$

Thus, recalling the prior for q given by (3.3) and (3.4),

$$p(q \mid x) = \frac{f(x \mid q) p(q)}{p(x)} \propto \pi_{\chi_0}(q) \prod_{(i,j) \in E} \tilde{p}(q_{ij} \mid x),$$

where

$$\tilde{p}(q_{ij} \mid x) = \mathrm{Ga}(q_{ij} \mid n_{ij} + \alpha_{ij}, \tilde{t}_i + \beta_{ij}) \quad ((i,j) \in E). \tag{4.3}$$

Note that if the initial state χ_0 of x was known, rather than assumed to be sampled from the equilibrium distribution $\boldsymbol{\pi}(q)$, then $p(q \mid x)$ would be

$$\prod_{(i,j) \in E} \tilde{p}(q_{ij} \mid x).$$

Thus we generate a proposal, $q' = \{q'_{ij}; (i,j) \in E\}$, for q by sampling q'_{ij} independently from the gamma distribution $\tilde{p}(q_{ij} \mid x)$ $((i,j) \in E)$. Using (2.7), the acceptance probability for this proposal is $A = \min(1, B)$, where

$$B = \frac{p(x, q', \theta \mid y) \prod_{(i,j) \in E} \tilde{p}(q_{ij} \mid x)}{p(x, q, \theta \mid y) \prod_{(i,j) \in E} \tilde{p}(q'_{ij} \mid x)} = \pi_{\chi_0}(q') / \pi_{\chi_0}(q),$$

using (3.3), (3.4), (3.8), (3.9) and (4.2).

(c) *Update θ*

We update $\theta = (\mu_{\mathrm{O}}, \sigma^2_{\mathrm{O}}, \mu_{\mathrm{C}}, \sigma^2_{\mathrm{C}})$ using single-component Gibbs steps on μ_{O}, σ^2_{O}, μ_{C} and σ^2_{C}, i.e. by sampling successively from the distributions $p(\mu_{\mathrm{O}} \mid x, q, y, \sigma^2_{\mathrm{O}}, \mu_{\mathrm{C}}, \sigma^2_{\mathrm{C}})$, $p(\sigma^2_{\mathrm{O}} \mid x, q, y, \mu_{\mathrm{O}}, \mu_{\mathrm{C}}, \sigma^2_{\mathrm{C}})$, $p(\mu_{\mathrm{C}} \mid x, q, y, \mu_{\mathrm{O}}, \sigma^2_{\mathrm{O}}, \sigma^2_{\mathrm{C}})$ and $p(\sigma^2_{\mathrm{C}} \mid x, q, y, \mu_{\mathrm{O}}, \sigma^2_{\mathrm{O}}, \mu_{\mathrm{C}})$. Thus we now determine these conditional distributions. First, some more notation is required.

For $l = 0, 1, \ldots, N$, let

$$\delta_l = \begin{cases} 1, & \text{if } x(l\Delta) \in \mathrm{O}, \\ 0, & \text{if } x(l\Delta) \in \mathrm{C}. \end{cases} \tag{4.4}$$

Further, let

$$y_{\mathrm{O}} = \sum_{l=0}^{N} y_l \delta_l, \quad n_{\mathrm{O}} = \sum_{l=0}^{N} \delta_l, \quad y_{\mathrm{O},2} = \sum_{l=0}^{N} y_l^2 \delta_l, \quad \bar{y}_{\mathrm{O}} = y_{\mathrm{O}} / n_{\mathrm{O}}, \quad s^2_{\mathrm{O}} = y_{\mathrm{O},2} - n_{\mathrm{O}} \bar{y}^2_{\mathrm{O}}$$

and

$$y_{\mathrm{C}} = \sum_{l=0}^{N} y_l(1-\delta_l), \qquad n_{\mathrm{C}} = \sum_{l=0}^{N}(1-\delta_l), \qquad y_{\mathrm{C},2} = \sum_{l=0}^{N} y_l^2(1-\delta_l),$$
$$\bar{y}_{\mathrm{C}} = y_{\mathrm{C}}/n_{\mathrm{C}}, \qquad s_{\mathrm{C}}^2 = y_{\mathrm{C},2} - n_{\mathrm{C}}\bar{y}_{\mathrm{C}}^2.$$

Then it follows from (3.1) that

$$f(y \mid x, \theta) = \prod_{l=0}^{N} \phi_{\mathrm{O}}(y_l \mid \theta)^{\delta_l} \phi_{\mathrm{C}}(y_l \mid \theta)^{1-\delta_l}, \tag{4.5}$$

where

$$\phi_{\mathrm{O}}(y_l \mid \theta) = (2\pi\sigma_{\mathrm{O}}^2)^{-1/2} \exp[-(y_l-\mu_{\mathrm{O}})^2/(2\sigma_{\mathrm{O}}^2)],$$
$$\phi_{\mathrm{C}}(y_l \mid \theta) = (2\pi\sigma_{\mathrm{C}}^2)^{-1/2} \exp[-(y_l-\mu_{\mathrm{C}})^2/(2\sigma_{\mathrm{C}}^2)].$$

Further, using (3.8),

$$\begin{aligned} p(\theta \mid x, q, y) &\propto p(\theta) f(y \mid x, \theta) \\ &\propto (\sigma_{\mathrm{O}}^2)^{-(n_{\mathrm{O}}/2)-1} \exp[-\tfrac{1}{2}\{s_{\mathrm{O}}^2 + n_{\mathrm{O}}(\mu_{\mathrm{O}} - \bar{y}_{\mathrm{O}})^2\}/\sigma_{\mathrm{O}}^2] \\ &\quad \times (\sigma_{\mathrm{C}}^2)^{-(n_{\mathrm{C}}/2)-1} \exp[-\tfrac{1}{2}\{s_{\mathrm{C}}^2 + n_{\mathrm{C}}(\mu_{\mathrm{C}} - \bar{y}_{\mathrm{C}})^2\}/\sigma_{\mathrm{C}}^2]. \end{aligned} \tag{4.6}$$

Thus

$$\begin{aligned} p(\theta \mid x, q, y) &= p(\mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2, \mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2 \mid x, y) \\ &= p(\mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2 \mid x, y) p(\mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2 \mid x, y). \end{aligned}$$

The posterior densities $p(\mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2 \mid x, y)$ and $p(\mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2 \mid x, y)$ are those arising from two independent normal samples, whose means and variances are both unknown. Thus, from standard theory (see, for example, Lee 1989, § 2.12),

$$\begin{aligned} &p(\mu_{\mathrm{O}} \mid x, q, y, \sigma_{\mathrm{O}}^2, \mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2) \\ &\quad = p(\mu_{\mathrm{O}} \mid \sigma_{\mathrm{O}}^2, x, y) \\ &\quad = (2\pi\sigma_{\mathrm{O}}^2/n_{\mathrm{O}})^{-1/2} \exp\{-\tfrac{1}{2} n_{\mathrm{O}}(\mu_{\mathrm{O}} - \bar{y}_{\mathrm{O}})^2/\sigma_{\mathrm{O}}^2\} \quad (-\infty < \mu_{\mathrm{O}} < \infty) \end{aligned}$$

and

$$\begin{aligned} &p(\sigma_{\mathrm{O}}^2 \mid x, q, y, \mu_{\mathrm{O}}, \mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2) \\ &\quad = p(\sigma_{\mathrm{O}}^2 \mid \mu_{\mathrm{O}}, x, y) \\ &\quad \propto (\sigma_{\mathrm{O}}^2)^{-(n_{\mathrm{O}}/2)-1} \exp[-\tfrac{1}{2}\{s_{\mathrm{O}}^2 + n_{\mathrm{O}}(\mu_{\mathrm{O}} - \bar{y}_{\mathrm{O}})^2\}/\sigma_{\mathrm{O}}^2] \quad (\sigma_{\mathrm{O}}^2 > 0), \end{aligned}$$

with similar expressions holding for

$$p(\mu_{\mathrm{C}} \mid x, q, y, \mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2, \sigma_{\mathrm{C}}^2) \quad \text{and} \quad p(\sigma_{\mathrm{C}}^2 \mid x, q, y, \mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2, \mu_{\mathrm{C}}).$$

Hence θ may be updated as follows.

Let

$$\theta^{(k)} = (\mu_{\mathrm{O}}^{(k)}, \sigma_{\mathrm{O}}^{2(k)}, \mu_{\mathrm{C}}^{(k)}, \sigma_{\mathrm{C}}^{(k)})$$

denote the current value of θ. Then $(\mu_{\mathrm{O}}^{(k+1)}, \sigma_{\mathrm{O}}^{2(k+1)})$ is obtained by first sampling

$\mu_{\mathrm{O}}^{(k+1)}$ from the normal distribution with mean $\bar{y}_{\mathrm{O}}$ and variance $\sigma_{\mathrm{O}}^{2(k)}/n_{\mathrm{O}}$ and then sampling z from the χ-squared distribution with n_{O} degrees of freedom and then setting

$$\sigma_{\mathrm{O}}^{2(k+1)} = (s_{\mathrm{O}}^2 + n_{\mathrm{O}}(\mu_{\mathrm{O}}^{(k+1)} - \bar{y}_{\mathrm{O}})^2)/z.$$

The updates $(\mu_{\mathrm{C}}^{(k+1)}, \sigma_{\mathrm{C}}^{2(k+1)})$ are obtained similarly.

Note that it is possible to update $(\mu_{\mathrm{O}}, \sigma_{\mathrm{O}}^2)$ (and also $(\mu_{\mathrm{C}}, \sigma_{\mathrm{C}}^2)$) jointly, rather than in two separate Gibbs steps, by first sampling

$$\sigma_{\mathrm{O}}^{2(k+1)} \quad \text{from } p(\sigma_{\mathrm{O}}^2 \mid x, y)$$

and then sampling

$$\mu_{\mathrm{O}}^{(k+1)} \quad \text{from } p(\mu_{\mathrm{O}} \mid \sigma_{\mathrm{O}}^{2(k+1)}, x, y).$$

Explicit forms for these posterior distributions are given, for example, in Lee (1989, § 2.12). However, since in practice we usually want to constrain $\mu_{\mathrm{O}} > \mu_{\mathrm{C}}$ (see § 7 *a*), and in that case the above option is not available, we stick to separate Gibbs steps.

(*d*) *Update x*

The unobserved realization x of the channel-gating process is updated using a Metropolis–Hastings algorithm, with proposals x' that change x on a sojourn basis. The possible move types for obtaining x' are described in § 4 *d* (i). When describing the move types, the notation is local to a given move type. The same notation is used when giving the acceptance probabilities for the various move types, in § 4 *d* (ii).

(i) *Move types for updating x*

(1) *Move a boundary between two successive sojourns of x.* We first decide which boundary to move by sampling j from the random variable J having a uniform distribution on $\{0, 1, \ldots, n-1\}$ (i.e. $P(J = j) = n^{-1}$, $j = 0, 1, \ldots, n-1$). We then sample u from $U(0, t_j + t_{j+1})$, i.e. from the uniform distribution on $(0, t_j + t_{j+1})$. The proposal x' is then obtained by moving the boundary between the jth and $(j+1)$th sojourns from s_{j+1} to $s_j + u$. Thus $x' = (n', \boldsymbol{t}', \boldsymbol{\chi}')$ is given by $n' = n$, $t'_j = u$, $t'_{j+1} = t_j + t_{j+1} - u$, $t'_l = t_l$ $(l \neq j, j+1)$ and $\chi'_l = \chi_l$ $(l = 1, 2, \ldots, n')$.

(2) *Insert a sojourn.* Sample l^* from the uniform distribution on $\{0, 1, \ldots, n\}$ and let $i = \chi_{l^*}$. Determine the state for the inserted sojourn by sampling j from the distribution $P(J = j) = \tilde{p}_2(j \mid i)$ $(j \in S,\ j \neq i)$. (The distribution $\tilde{p}_2(j \mid i)$ needs to be specified by the user and should satisfy $\tilde{p}_2(j \mid i) > 0$ if and only if $q_{ij} > 0$ and $q_{ji} > 0$. It is possible to make $\tilde{p}$ depend on q, and more generally on x, i.e. have $\tilde{p}_2(j \mid i, q, x)$, though we shall treat the simple case in this paper.) Sample u from the $U(0, t_{l^*})$ distribution and then sample v from the negative exponential distribution with rate $-q_{jj}$ truncated at $t_{l^*} - u$, i.e. from the distribution having probability density function

$$\frac{-q_{jj}\exp(q_{jj}v)}{[1 - \exp\{q_{jj}(t_{l^*} - u)\}]} \quad (0 < v < t_{l^*} - u).$$

This can be achieved by sampling w from the $U(0, 1)$ distribution and setting

$$v = q_{jj}^{-1}\log(1 - w[1 - \exp\{q_{jj}(t_{l^*} - u)\}]).$$

Then a sojourn in state j of length v is inserted in the l^*th sojourn of x. Thus the proposal x' is given by $n' = n+2$, $(t'_l, \chi'_l) = (t_l, \chi_l)$ $(l = 0, 1, \ldots, l^*-1)$, $t'_{l^*} = u$, $t'_{l^*+1} = v$, $t'_{l^*+2} = t_{l^*} - u - v$, $\chi'_{l^*} = \chi'_{l^*+2} = \chi_{l^*}$ $(= i)$, $\chi'_{l^*+1} = j$, $(t'_l, \chi'_l) = (t_{l-2}, \chi_{l-2})$ $(l = l^*+3, l^*+4, \ldots, n+2)$, with an obvious modification if $l^* = 0$ or n.

(3) *Delete an intermediate sojourn whose two adjacent sojourns are in the same state.* Sample l^* from the uniform distribution on $\{1, 2, \ldots, n-1\}$. If $\chi_{l^*-1} = \chi_{l^*+1}$ then the l^*th sojourn of x is deleted, i.e. x' is obtained by setting $x'(t) = \chi_{l^*-1}$ $(s_{l^*} \leqslant t < s_{l^*+1})$. Thus x' is given by $n' = n-2$, $(t'_l, \chi'_l) = (t_l, \chi_l)$ $(l = 0, 1, \ldots, l^*-2)$, $t'_{l^*-1} = t_{l^*-1} + t_{l^*} + t_{l^*+1}$, $\chi'_{l^*-1} = \chi_{l^*-1}$, $(t'_l, \chi'_l) = (t_{l+2}, \chi_{l+2})$ $(l = l^*, l^*+1, \ldots, n-2)$, with an obvious modification if $j = 1$ or $n-1$. If $\chi_{l^*-1} \neq \chi_{l^*+1}$ then x is left unchanged, i.e. $x^{(k+1)} = x^{(k)}$.

(4) *Split an intermediate sojourn.* Sample l^* from the uniform distribution on $\{1, 2, 3, \ldots, n-1\}$, and let $i_1 = \chi_{l^*}$ and $i_3 = \chi_{l^*+1}$. If $q_{i_1 j} q_{j i_3} = 0$ for all $j \in S \backslash \{i_1, i_3\}$ then x is left unchanged. Otherwise sample u from the $U(0, t_{l^*})$ distribution and j from the user-specified distribution $\tilde{p}_4(j \mid i_1, i_3)$ $(j \in S,\ j \neq i_1$ or $i_3)$, satisfying $\tilde{p}_4(j \mid i_1, i_3) > 0$ if and only if $q_{i_1 j} q_{j i_3} > 0$. The proposal x' is then given by $n' = n+1$, $(t'_l, \chi'_l) = (t_l, \chi_l)$ $(l = 0, 1, \ldots, l^*-1)$, $t'_{l^*} = u$, $\chi'_{l^*} = \chi_l$, $t'_{l^*+1} = t_{l^*+1} - u$, $\chi'_{l^*+1} = j$, $(t'_l, \chi'_l) = (t_{l-1}, \chi_{l-1})$ $(l = l^*+2,\ l^*+3, \ldots, n+1)$.

(5) *Delete an intermediate sojourn whose two adjacent sojourns are in distinct states.* The procedure for move type (5) is the same as that for move type (3), except that, letting $i = \chi_{l^*-1}$ and $j = \chi_{l^*+1}$, if $i \neq j$ and $q_{ij} > 0$ then the l^*th sojourn of x is deleted, i.e. $n' = n-1$, $(t'_l, \chi'_l) = (t_l, \chi_l)$ $(l = 0, 1, \ldots, l^*-2)$, $t'_{l-1} = t_{l-1} + t_l$, $\chi'_{l-1} = \chi_{l-1}$, $(t'_l, \chi'_l) = (t_{l+1}, \chi_{l+1})$ $(l = l^*,\ l^*+1, \ldots, n-1)$. Alternatively, if $i = j$ or $q_{ij} = 0$ then x is left unchanged.

(6) *Split the initial sojourn of x.* Sample u from the $U(0, t_1)$ distribution and let $i = \chi_0$. Sample j from the distribution $P(J = j) = \tilde{p}_6(j \mid i)$ $(j \in S,\ j \neq i)$, where the distribution $\tilde{p}_6(j \mid i)$, which satisfies $\tilde{p}_6(j \mid i) > 0$ if and only if $q_{ji} > 0$, is specified by the user. The proposal x' is then given by $n' = n+1$, $t'_0 = u$, $\chi'_0 = j$, $t'_1 = t_0 - u$, $\chi'_1 = i$, $(t'_l, \chi'_l) = (t_{l-1}, \chi_{l-1})$ $(l = 2, 3, \ldots, n+1)$.

(7) *Delete the initial sojourn of x.* This move type is deterministic. The initial sojourn of x is just deleted. Thus the proposal x' is given by $n' = n-1$, $t'_0 = t_0 + t_1$, $\chi'_0 = \chi_1$, $(t'_l, \chi'_l) = (t_{l+1}, \chi_{l+1})$ $(l = 1, 2, \ldots, n-1)$.

(8) *Split the final sojourn of x.* Sample u from the $U(0, t_n)$ distribution and let $i = \chi_n$. Sample j from the user-specified distribution $\tilde{p}_8(j \mid i)$ $(j \in S,\ j \neq i)$, satisfying $\tilde{p}_8(j \mid i) > 0$ if and only if $q_{ij} > 0$. The proposal x' is then given by $n' = n+1$, (t_l, χ_l) $(l = 0, 1, \ldots, n-1)$, $t'_n = u$, $\chi'_n = \chi_n$, $t'_{n+1} = t_n - u$, $\chi'_{n+1} = j$.

(9) *Delete the final sojourn of x.* This move type is deterministic. The final sojourn of x is just deleted, so the proposal x' is given by $n' = n-1$, $(t'_l, \chi'_l) = (t_l, \chi_l)$ $(l = 0, 1, \ldots, n-2)$, $t'_{n-1} = t_{n-1} + t_n$, $\chi'_{n-1} = \chi_{n-1}$.

(10) *Insert a cycle of sojourns.* This move type is model dependent. For clarity, we assume that the graph G of X possesses two cycles of length $\geqslant 3$, both having length 4, $i_1 \to i_2 \to i_3 \to i_4 \to i_1$ and $i_1 \to i_4 \to i_3 \to i_2 \to i_1$ say, where i_1, i_2, i_3 and i_4 are distinct. This is the case for the mechanism of figure 1*d*. Sample l^* from the uniform distribution on $\{0, 1, \ldots, n\}$ and let $\chi_{l^*} = i$. If $i \in \{i_1, i_2, i_3, i_4\}$, then choose

one of the two cycles going through i at random, each cycle having probability $\frac{1}{2}$ of being chosen; say, choose $j_1 \to j_2 \to j_3 \to j_4 \to j_1$ with $j_1 = i$, sample u_1, u_2, u_3, u_4 independently from the $U(0, t_{l^*})$ distribution, let $u_{(1)} < u_{(2)} < u_{(3)} < u_{(4)}$ be the corresponding order statistics (i.e. the u's arranged in increasing order) and insert the corresponding cycle in the l^* sojourn of x. Thus the proposal x' is given by $n' = n + 4$, $(t'_l, \chi'_l) = (t_l, \chi_l)$ $(l = 0, 1, \ldots, l^* - 1)$, $(t'_{l^*+r-1}, \chi'_{l^*+r-1}) = (u_{(r)} - u_{(r-1)}, j_r)$ $(r = 1, 2, 3, 4)$, $(u_{(0)} = 0)$, $t'_{l^*+4} = t'_{l*} - u_{(4)}$, $\chi'_{l^*+4} = j_1$, $(t'_l, \chi'_l) = (t_{l-4}, \chi_{l-4})$ $(l = l^* + 5, l^* + 6, \ldots, n + 4)$. If $i \notin \{i_1, i_2, i_3, i_4\}$ then x is left unchanged.

(11) *Delete a cycle of sojourns*. This move type is model dependent, so we make the same assumptions as in move type (10). Sample l^* from the uniform distribution on $\{0, 1, \ldots, n - 4\}$ and let $j_r = \chi_{l^*+r-1}$ $(r = 1, 2, 3, 4, 5)$. If $\{j_r; r = 1, 2, 3, 4, 5\}$ forms a cycle in G (i.e. $j_1 = j_5$, j_1, j_2, j_3, j_4 are distinct and $q_{j_r j_{r+1}} > 0$ $(r = 1, 2, 3, 4)$) then that cycle is deleted, so the proposal x' is given by $n' = n - 4$, $(t'_l, \chi'_l) = (t_l, \chi_l)$ $(l = 0, 1, \ldots, l^* - 1)$, $t'_{l^*} = t_{l^*} + t_{l^*+1} + t_{l^*+2} + t_{l^*+3} + t_{l^*+4}$, $\chi'_{l^*} = \chi_{l^*}$, $(t'_l, \chi'_l) = (t_{l+4}, \chi_{l+4})$ $(l = l^* + 1, l^* + 2, \ldots, n - 4)$. If $\{j_r; r = 1, 2, 3, 4, 5\}$ does not form a cycle in G then x is left unchanged.

(ii) *Acceptance probabilities*

For $i = 1, 2, \ldots, 11$, let $R^{(i)}(x \to x'; q, \theta)$ be the density of the transition kernel associated with the proposal x' for move type (i) and let p_i be the probability of choosing move type (i). The choice of $p_1, p_2, \ldots, p_{11}$ is discussed in § 6 *b*, under implementation. Note that move types (3), (4), (6), (8) and (10) are discrete, so $R^{(i)}(x \to x'; q, \theta)$ is in fact a probability, rather than strictly a density, for those move types. Taking into account the probabilities of choosing the different move types, the density of the transition kernel for the proposal x' can be written as

$$R(x \to x'; q, \theta) = \sum_{i=1}^{11} p_i R^{(i)}(x \to x'; q, \theta),$$

where $R^{(i)}(x \to x'; q, \theta) = 0$ if x' cannot be obtained from x by move type (i).

Move type (1) does not alter the dimension of x so by (2.7) its acceptance probability is $A = \min(1, B)$, where

$$B = \frac{p(x', q, \theta \mid y) R(x' \to x; q, \theta)}{p(x, q, \theta \mid y) R(x \to x'; q, \theta)}. \tag{4.7}$$

Move types (2)–(11) all alter the dimension of x, so, from the penultimate paragraph of § 2, B now takes the form

$$B = \frac{p(x', q, \theta \mid y) R(x' \to x; q, \theta)}{p(x, q, \theta \mid y) R(x \to x'; q, \theta)} |J|$$

for an appropriate Jacobian J. It turns out that $|J| = 1$ for each of move types (2)–(11), so the acceptance probabilities for all of the move types can be calculated from (4.7).

From (3.8) and (3.9),

$$\frac{p(x', q', \theta \mid y)}{p(x, q, \theta \mid y)} = \frac{f(x' \mid q) f(y \mid x', \theta)}{f(x \mid q) f(y \mid x, \theta)}.$$

Thus we may write $B = B_1 B_2 B_3$, where

$$B_1 = \frac{f(x' \mid q)}{f(x \mid q)}, \quad B_2 = \frac{f(y \mid x', \theta)}{f(y \mid x, \theta)}, \quad B_3 = \frac{R(x' \to x; q, \theta)}{R(x \to x'; q, \theta)}.$$

Now B_1 and B_2 can be determined using (4.2) and (4.5), respectively, and B_3 can be calculated from the descriptions given for the different move types. The values of B_1, B_2 and B_3 for each of the move types (1)–(11) are given below. The notation corresponds to that used when describing the respective move types. The details are straightforward, so only the results are given. We first introduce some more notation.

Recall the notation $x = (n, \boldsymbol{t}, \boldsymbol{\chi})$, where n is the number of jumps made by x (so x contains $n+1$ sojourns) and $\boldsymbol{t} = (t_0, t_1, \ldots, t_n)$ and $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_n)$ are vectors giving the lengths and states of successive sojourns of x. Recall also that $s_0 = 0$ and $s_l = \sum_{r=0}^{l-1} t_r$ $(l = 1, 2, \ldots, n+1)$, so $s_{n+1} = T$. For $l = 0, 1, \ldots, n$, let $E_l = \{r \in \boldsymbol{Z} : s_l \leqslant r\Delta < s_{l+1}\}$ and $n_l = |E_l|$. Here, $\boldsymbol{Z}$ denotes the integers and $|E_l|$ denotes the number of elements of E_l. Thus n_l is the number of sampling points in the lth sojourn of x. For $l = 0, 1, \ldots, n$, let $\tilde{y}_l = \sum_{r \in E_l} y_r$ and $\tilde{y}_{l,2} = \sum_{r \in E_l} y_r^2$. Define n'_l, $\tilde{y}'_l$, $\tilde{y}'_{l,2}$ $(l = 0, 1, \ldots, n')$ similarly for x'.

For $i_1, i_2 \in S$, let

$$B_2 = \begin{cases} 1, & \text{if } i_1, i_2 \in \mathrm{O} \text{ or } i_1, i_2 \in \mathrm{C}, \\ D, & \text{if } i_1 \in \mathrm{O} \text{ and } i_2 \in \mathrm{C}, \\ D^{-1}, & \text{if } i_1 \in \mathrm{C} \text{ and } i_2 \in \mathrm{O}, \end{cases} \tag{4.8}$$

where i_1, i_2 and D will be determined by the move type. Also, let $\eta_1 = \sigma_{\mathrm{O}}/\sigma_{\mathrm{C}}$,

$$\eta_2 = \frac{1}{2}\left(\frac{1}{\sigma_{\mathrm{O}}^2} - \frac{1}{\sigma_{\mathrm{C}}^2}\right), \quad \eta_3 = \frac{\mu_{\mathrm{O}}}{\sigma_{\mathrm{O}}^2} - \frac{\mu_{\mathrm{C}}}{\sigma_{\mathrm{C}}^2}, \quad \eta_4 = \frac{1}{2}\left(\frac{\mu_{\mathrm{O}}^2}{\sigma_{\mathrm{O}}^2} - \frac{\mu_{\mathrm{C}}^2}{\sigma_{\mathrm{C}}^2}\right).$$

Move type (1). Let $i_1 = \chi_j$ and $i_2 = \chi_{j+1}$. Then $B_1 = \exp\{(t_j - u)(q_{i_2 i_2} - q_{i_1 i_1})\}$, $B_3 = 1$ and B_2 is given by (4.8) with

$$D = \eta_1^{n_l - n_{l'}} \exp\{-\eta_2(\tilde{y}'_{l,2} - \tilde{y}_{l,2}) + \eta_3(\tilde{y}'_l - \tilde{y}_l) - \eta_4(n'_l - n_l)\}.$$

Move type (2). Let $i_1 = i$ and $i_2 = j$. Then $B_1 = q_{ij} q_{ji} \exp\{(q_{jj} - q_{ii})v\}$, B_2 is given by (4.8) with

$$D = \eta_1^{n'_{l^*+1}} \exp\{\eta_2 \tilde{y}'_{l^*+1,2} - \eta_3 \tilde{y}'_{l^*+1} + \eta_4 n'_{l^*+1}\},$$

and

$$B_3 = \frac{-p_3 t_{l^*}[1 - \exp\{q_{jj}(t_{l^*} - u)\}]}{\{p_2 \tilde{p}_2(j \mid i) q_{jj} \exp(q_{jj} v)\}}.$$

Note that move type (2) changes the dimension of x, so we need to calculate the associated Jacobian J. Now, given l^*, u and v, $\boldsymbol{t}'$ is obtained from $\boldsymbol{t}$ using the invertible deterministic function

$$\boldsymbol{t}'(\boldsymbol{t}, u, v) = (t_0, t_1, \ldots, t_{l^*-1}, u, v, t_{l^*} - u - v, t_{l^*+1}, t_{l^*+2}, \ldots, t_n).$$

Thus

$$J = \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{vmatrix} = 1.$$

Similar calculations hold for move types (3)–(11), but are omitted.

Move type (3). Let $i_1 = \chi_{l^*-1}$ and $i_2 = \chi_{l^*}$. Then

$$B_1 = (q_{i_1 i_2} q_{i_2 i_1})^{-1} \exp\{t_{l^*}(q_{i_1 i_1} - q_{i_2 i_2})\},$$

B_2 is given by (4.8) with

$$D = \eta_1^{-n_{l^*}} \exp\{-\eta_2 \tilde{y}_{l^*,2} + \eta_3 \tilde{y}_{l^*} - \eta_4 n_{l^*}\},$$

and

$$B_3 = \frac{-p_2 \tilde{p}_2(i_2 \mid i_1) q_{i_2 i_2} \exp(q_{i_2 i_2} t_{l^*})}{p_3(t_{l^*-1} + t_{l^*} + t_{l^*+1})\{1 - \exp(q_{i_2 i_2}(t_{l^*} + t_{l^*+1})\}}.$$

Move type (4). Let $i_2 = j$. Then

$$B_1 = q_{i_1 i_2} q_{i_2 i_3} \exp\{(q_{i_2 i_2} - q_{i_1 i_1})(t_{l^*} - u)\}/q_{i_1 i_3},$$

B_2 is given by (4.8) with $D = \eta_1^{n'_{l^*+1}} \exp\{\eta_2 \tilde{y}'_{l^*+1,2} - \eta_3 \tilde{y}'_{l^*+1} + \eta_4 n'_{l^*+1}\}$, and

$$B_3 = \frac{(n-1) p_5 t_{l^*}}{\{p_4 \tilde{p}_4(j \mid i_1, i_3) n\}}.$$

Move type (5). Let $i_1 = i$ and $i_2 = \chi_{l^*}$. Then

$$B_1 = \frac{q_{ij} \exp\{t_{l^*}(q_{ii} - q_{i_2 i_2})\}}{(q_{i i_2} q_{i_2 j})},$$

B_2 is as for move type (3) and

$$B_3 = \frac{(n-1) p_4 \tilde{p}_4(i_2 \mid i, j)}{\{(n-2) p_5 (t_{l^*-1} + t_{l^*})\}}.$$

Move type (6). Let $i_1 = j$ and $i_2 = i$. Then

$$B_1 = \pi_j q_{ji} \exp\{(q_{jj} - q_{ii}) u\}/\pi_i,$$

B_2 is given by (4.8) with $D = \eta_1^{-n'_{\mathrm{O}}} \exp\{-\eta_2 \tilde{y}'_{0,2} + \eta_3 \tilde{y}'_0 - \eta_4 n'_0\}$, and

$$B_3 = \frac{p_4 t_0}{(p_6 \tilde{p}_6(j \mid i))}.$$

Move type (7). Let $i_1 = \chi_0$ and $i_2 = \chi_1$. Then

$$B_1 = \frac{\pi_{i_2} \exp\{(q_{i_2 i_2} - q_{i_1 i_1}) t_0\}}{(\pi_{i_1} q_{i_1 i_2})},$$

B_2 is given by (4.8) with $D = \eta_1^{n_0} \exp\{\eta_2 \tilde{y}_{2,0} - \eta_3 \tilde{y}_0 + \eta_4 n_0\}$, and

$$B_3 = \frac{p_6 \tilde{p}_6(i_1 \mid i_2)}{\{p_7(t_0 + t_1)\}}.$$

Move type (8). Let $i_1 = i$, $i_2 = j$ and $r = n'$. Then

$$B_1 = q_{ij} \exp\{(q_{jj} - q_{ii})(t - u)\},$$

B_2 is given by (4.8) with $D = \eta_1^{n'_r} \exp\{\eta_2 \tilde{y}'_{n',2} - \eta_3 \tilde{y}'_{n'} + \eta_4 n'_r\}$, and

$$B_3 = \frac{p_9 t_n}{(p_8 \tilde{p}_8(j \mid i))}.$$

Move type (9). Let $i_1 = \chi_{n-1}$ and $i_2 = \chi_n$. Then

$$B_1 = \exp\{(q_{i_1 i_1} - q_{i_2 i_2})t_n\}/q_{i_1 i_2},$$

B_2 is given by (4.8) with $D = \eta_1^{-n_n} \exp\{-\eta_2 \tilde{y}_{n,2} + \eta_3 \tilde{y}_n - \eta_4 n_n\}$, and

$$B_3 = \frac{p_8 \tilde{p}_8(i_2 \mid i_1)}{\{p_9(t_{n-1} + t_n)\}}.$$

Move type (10). To illustrate the calculation of the acceptance probabilities for move types (10) and (11), we consider the specific examples given in the descriptions of those move types. Let $v_1 = u_{(1)} - u_{(4)}$, $v_2 = u_{(2)} - u_{(1)}$, $v_3 = u_{(3)} - u_{(2)}$ and $v_4 = u_{(4)} - u_{(3)}$. Then

$$B_1 = q_{j_1 j_2} q_{j_2 j_3} q_{j_3 j_4} q_{j_4 j_1} \exp\{q_{j_1 j_1} v_1 + q_{j_2 j_2} v_2 + q_{j_3 j_3} v_3 + q_{j_4 j_4} v_4\}$$

and

$$B_3 = \frac{p_{11} t_{l^*}^4}{(\frac{1}{2} p_{10} 4!)}.$$

For $l = 0, 1, \ldots, n'$, let

$$\tilde{\delta}'_l = \begin{cases} 1, & \text{if } \chi'_l \in \mathrm{O}, \\ 0, & \text{if } \chi_l \in \mathrm{C}. \end{cases} \tag{4.9}$$

Further, if $j_1 \in \mathrm{O}$, let

$$n^* = \sum_{l=l^*+1}^{l^*+3} (1 - \tilde{\delta}'_l) n'_l, \qquad y_2^* = \sum_{l=l^*+1}^{l^*+3} (1 - \tilde{\delta}'_l) \tilde{y}'_{l,2}, \qquad y^* = \sum_{l=l^*+1}^{l^*+3} (1 - \tilde{\delta}'_l) \tilde{y}'_l,$$

while if $j_1 \in C$, let

$$n^* = \sum_{l=l^*+1}^{l^*+3} \tilde{\delta}'_l n'_l, \qquad y_2^* = \sum_{l=l^*+1}^{l^*+3} \tilde{\delta}'_l \tilde{y}'_{l,2}, \qquad y^* = \sum_{l=l^*+1}^{l^*+3} \tilde{\delta}'_l \tilde{y}'_l.$$

Then

$$B_2 = \begin{cases} D, & \text{if } j_1 \in \mathrm{O}, \\ D^{-1}, & \text{if } j_1 \in \mathrm{C}, \end{cases} \tag{4.10}$$

where

$$D = \eta_1^{n^*} \exp(\eta_2 y_2^* - \eta_3 y^* + \eta_4 n^*).$$

Note that $B_2 = 1$ if $\delta'_l = \delta'_{l^*}$ $(l = l^* + 1,\ l^* + 2,\ l^* + 3)$.

Move type (11). Let n^*, y_2^* and y_1^* be as for move type (10) but with $\tilde{\delta}'_l, \tilde{y}'_{l,2}, \tilde{y}'_l$ replaced by $\tilde{\delta}_l, \tilde{y}_{l,2}, \tilde{y}_l$ $(l = l^* + 1,\ l^* + 2,\ l^* + 3)$, and $\tilde{\delta}_l$ is given by (4.9) with χ'_l replaced by χ_l. Then

$$B_1 = \frac{\exp\{q_{j_1 j_1}(t_{l^*+1} + t_{l^*+2} + t_{l^*+3}) - q_{j_2 j_2} t_{l^*+1} - q_{j_3 j_3} t_{l^*+2} - q_{j_4 j_4} t_{l^*+3}\}}{(q_{j_1 j_2} q_{j_2 j_3} q_{j_3 j_4} q_{j_4 j_1})},$$

B_2 is given by (4.10) with $D = \eta_1^{-n^*} \exp(-\eta_2 y_2^* + \eta_3 y^* - \eta_4 n^*)$, and

$$B_3 = \frac{\frac{1}{2} 4! p_{10}}{(p_{11}(t'_{l^*})^4)}.$$

5. Using the output from an MCMC sampler

(a) Convergence and related issues

The preceding section provides a method of simulating a realization

$$\{(x^{(k)}, q^{(k)}, \theta^{(k)}); k = 0, 1, \ldots, M\}$$

of a Markov chain whose equilibrium distribution is $p(x, q, \theta \mid y)$. In §§ 5 *b* and 5 *c* we describe how the above realization can be used to draw inferences concerning q, θ and single-channel properties, and reconstruct the unobserved sequence of open and closed sojourns of the channel, respectively. Before doing so, we briefly discuss two issues that need to be addressed prior to using the output of an MCMC sampler, namely convergence of the sampler and correlation between successive (x, q, θ) realizations from the sampler.

In order to implement the MCMC sampler an initial value $(x^{(0)}, q^{(0)}, \theta^{(0)})$ must first be obtained. A method for doing this is described in § 6 *a*. However, the resulting $(x^{(0)}, q^{(0)}, \theta^{(0)})$ will not be distributed according to $p(x, q, \theta \mid y)$. (If it were straightforward to simulate from the distribution $p(x, q, \theta \mid y)$ then there would be no need to use MCMC.) Nevertheless, the distribution of $(x^{(k)}, q^{(k)}, \theta^{(k)})$ converges to $p(x, q, \theta \mid y)$ as $k \to \infty$, so $(x^{(k)}, q^{(k)}, \theta^{(k)})$ will be approximately distributed according to $p(x, q, \theta \mid y)$ provided that k is sufficiently large. There is no simple way of determining how large k should be before the above approximation becomes adequate. In practice, it is usual to examine time-series plots of the components of $q^{(k)}$ and $\theta^{(k)}$, and perhaps also of some statistics based on $x^{(k)}$. These will usually indicate a 'burn-in' period, $k = 0, 1, \ldots, k^*$ say, during which $(x^{(k)}, q^{(k)}, \theta^{(k)})$ is approaching equilibrium, after which it can reasonably be assumed that $(x^{(k)}, q^{(k)}, \theta^{(k)})$ is distributed according to $p(x, q, \theta \mid y)$. However, care must be taken in interpreting such time-series plots since $\{(x^{(k)}, q^{(k)}, \theta^{(k)}); k = 0, 1, \ldots\}$ can display long-range fluctuations.

There is a growing literature on the topic of diagnosing convergence of MCMC samplers. However, assessing convergence is still something of an art, with conflicting opinions on the utility of various approaches (see Cowles & Carlin (1996) for a recent review).

It should be noted that, even after the burn-in period, successive realizations of $(x^{(k)}, q^{(k)}, \theta^{(k)})$ will typically be highly correlated. This correlation can be reduced by subsampling, i.e. by using only every pth realization when making inferences, etc. However, in many situations subsampling may not be optimal in that it can increase the Monte Carlo variance of estimates (see, for example, Geyer 1992). In the following two subsections we assume that inference is based on $\{(\tilde{x}^{(k)}, \tilde{q}^{(k)}, \tilde{\theta}^{(k)}); k = 1, 2, \ldots, \tilde{M}\}$, where

$$(\tilde{x}^{(k)}, \tilde{q}^{(k)}, \tilde{\theta}^{(k)}) = (x^{(k^*+1+(k-1)p)}, q^{(k^*+1+(k-1)p)}, \theta^{(k^*+1+(k-1)p)}) \quad (k = 1, 2, \ldots, \tilde{M}),$$

although, for ease of exposition, we omit the tildes.

(b) Inferences concerning q, θ and single-channel properties

Recall that $q = \{q_{ij} : (i, j) \in E\}$ and $\theta = (\mu_{\mathrm{O}}, \sigma^2_{\mathrm{O}}, \mu_{\mathrm{C}}, \sigma^2_{\mathrm{C}})$, and write

$$q^{(k)} = \{q^{(k)}_{ij} : (i, j) \in E\} \quad \text{and} \quad \theta^{(k)} = (\mu^{(k)}_{\mathrm{O}}, \sigma^{2(k)}_{\mathrm{O}}, \mu^{(k)}_{\mathrm{C}}, \sigma^{2(k)}_{\mathrm{C}})$$

in the obvious fashion. The realizations $(q^{(k)}, \theta^{(k)})$ $(k = 1, 2, \ldots, M)$ can be viewed as a sample from the posterior distribution $p(q, \theta \mid y)$, and thus they can be used to estimate $p(q, \theta \mid y)$.

To be specific, consider the problem of making inferences about μ_{O}, the mean current flowing across an open channel. Then $\mu_{\mathrm{O}}^{(k)}$ $(k = 1, 2, \ldots, M)$ is a sample from $p(\mu_{\mathrm{O}} \mid y)$. Thus the posterior distribution of μ_{O} given the current record y can be estimated, either pictorially by a histogram of $\mu_{\mathrm{O}}^{(k)}$ $(k = 1, 2, \ldots, M)$ or, more formally, by a kernel-density estimate of the form

$$\hat{p}(\mu_{\mathrm{O}} \mid y) = \frac{1}{Mh} \sum_{k=1}^{M} K\left(\frac{\mu_{\mathrm{O}} - \mu_{\mathrm{O}}^{(k)}}{h}\right), \tag{5.1}$$

where K is a usually symmetric kernel and h is the window width (or smoothing parameter) (see, for example, Silverman (1986) for further details of density estimation). Joint posterior distributions, for example $p(q \mid y)$, can be estimated by similar methods, although the results are often more difficult to interpret. Posterior means, variances and covariances of q, θ given y can be estimated by corresponding sample means, variances and covariances of $(q^{(k)}, \theta^{(k)})$ $(k = 1, 2, \ldots, M)$.

It is sometimes possible to reduce the variance of MCMC estimates of posterior means, variances, etc., by using a technique known as Rao–Blackwellization (see, for example, Gelfand & Smith (1990), Liu *et al.* (1994) and, in an ion-channel context, Ball *et al.* (1996)). For example, suppose we wish to estimate the posterior mean open-current level

$$E[\mu_{\mathrm{O}} \mid y] = \int p(\mu_{\mathrm{O}} \mid y)\, \mathrm{d}\mu_{\mathrm{O}}.$$

Now

$$E[\mu_{\mathrm{O}} \mid y] = E[E[\mu_{\mathrm{O}} \mid y, x]]$$

and, from § 4 *c*, $E[\mu_{\mathrm{O}} \mid y, x] = \bar{y}_{\mathrm{O}}$. Thus, an unbiased estimator for $E[\mu_{\mathrm{O}} \mid y]$ is

$$M^{-1} \sum_{k=1}^{M} E[\mu_{\mathrm{O}} \mid y, x^{(k)}] = M^{-1} \sum_{k=1}^{M} \bar{y}_{\mathrm{O}}^{(k)},$$

where we have reintroduced the explicit dependence of x, and hence $\bar{y}_{\mathrm{O}}$, on k. This estimator typically has smaller variance than the standard MCMC estimator given by

$$M^{-1} \sum_{k=1}^{M} \mu_{\mathrm{O}}^{(k)}.$$

More generally, if we wish to estimate $E[g(q, \theta) \mid y]$, where g is any specified real-valued function such that the expectation exists, and $E[g(q, \theta) \mid y, x]$ can be evaluated, then the unbiased MCMC estimator

$$M^{-1} \sum_{k=1}^{M} E[g(q, \theta) \mid y, x^{(k)}]$$

typically has smaller variance than the usual MCMC estimator

$$M^{-1} \sum_{k=1}^{M} g(q^{(k)}, \theta^{(k)}).$$

For example, such an approach can be used to estimate $E[\sigma_{\mathrm{O}}^2 \mid y]$, and also $E[q \mid y]$ provided we assume χ_0 to be known (cf. the discussion immediately following (4.3)).

The output from our MCMC sampler can also be used to estimate the posterior distributions of various model single-channel properties. For example, let π_{O} be the equilibrium probability that the channel is in an open state and $\tilde{\mu}_{\mathrm{O}}$ be the equilibrium mean length of a sojourn of the channel in the open states. Then

$$\pi_{\mathrm{O}} = \sum_{i \in O} \pi_i$$

and

$$\tilde{\mu}_{\mathrm{O}} = \frac{\pi_{\mathrm{O}}}{\{\sum_{i \in O} \pi_i \sum_{j \in C} q_{ij}\}}$$

(see Colquhoun & Hawkes 1977, eqn (73)).

Recall that the equilibrium distribution $\boldsymbol{\pi}$ is a function of q. Thus π_{O} and $\tilde{\mu}_{\mathrm{O}}$ are both functions of q and we may write $\pi_{\mathrm{O}} = \pi_{\mathrm{O}}(q)$ and $\tilde{\mu}_{\mathrm{O}} = \tilde{\mu}_{\mathrm{O}}(q)$. For $k = 1, 2, \ldots, M$, let $\pi_{\mathrm{O}}^{(k)} = \pi_{\mathrm{O}}(q^{(k)})$ and $\tilde{\mu}_{\mathrm{O}}^{(k)} = \tilde{\mu}_{\mathrm{O}}(q^{(k)})$. Then

$$\pi_{\mathrm{O}}^{(k)} \quad (k = 1, 2, \ldots, M) \quad \text{and} \quad \tilde{\mu}_{\mathrm{O}}^{(k)} \quad (k = 1, 2, \ldots, M)$$

can be viewed as samples from the posterior distributions $p(\pi_{\mathrm{O}} \mid y)$ and $p(\tilde{\mu}_{\mathrm{O}} \mid y)$, respectively. Hence these posterior distributions can be estimated by the methods described above for $p(q, \theta \mid y)$. Clearly, estimates of posterior distributions for other model channel properties can be obtained in a similar fashion.

(c) *Reconstruction of the single-channel record*

The output from our MCMC sampler can also be used to provide a model-based Bayesian reconstruction of the unobserved single-channel record, corresponding to the marginal Bayes restoration of Fredkin & Rice (1992*b*). Recall the definition (4.4) of δ_l $(l = 0, 1, \ldots, N)$. For $k = 1, 2, \ldots, M$ and $l = 0, 1, \ldots, N$ let

$$\delta_l^{(k)} = \begin{cases} 1, & \text{if } x^{(k)}(l\Delta) \in \mathrm{O}, \\ 0, & \text{if } x^{(k)}(l\Delta) \in \mathrm{C}. \end{cases}$$

For $l = 0, 1, \ldots, N$, let

$$\bar{\delta}_l = M^{-1} \sum_{k=1}^{M} \delta_l^{(k)}.$$

Then $\bar{\delta}_l$ is an MCMC estimator of $p(\delta_l = 1 \mid y)$. An MCMC marginal Bayes estimator of δ_l, $\hat{\delta}_l$ say, is then given by

$$\hat{\delta}_l = \begin{cases} 1, & \text{if } \bar{\delta}_l \geqslant 0.5, \\ 0, & \text{if } \bar{\delta}_l < 0.5. \end{cases} \tag{5.2}$$

6. Implementation

(a) *Initial state* $(x^{(0)}, q^{(0)}, \theta^{(0)})$ *of the MCMC sampler*

In view of the convergence issues discussed in § 5 *a*, it is clearly important that the initial state $(x^{(0)}, q^{(0)}, \theta^{(0)})$ of the MCMC sampler be distributed as close to

$p(x, q, \theta \mid y)$ as possible. A poor choice for $(x^{(0)}, q^{(0)}, \theta^{(0)})$ results in a sampler that can take a very long time to converge. Note from § 4 *c* that when updating θ only the current value of x is used. Also, from § 4 *b*, the algorithm for updating q requires only the current values of x and q, but the current value of q is required only in order to calculate the acceptance probability of the proposal q'. Thus, if, just for the initial update, we were always to accept the proposal then again only the current value of x would be required. Thus once an initial value for x is available it is straightforward to generate initial values of q and θ. Since ion-channel records are typically long, the behaviour of $q^{(k)}$ and $\theta^{(k)}$ is largely determined by that of $x^{(k)}$. Thus it is essential that a sensible choice is made for $x^{(0)}$.

Our method for determining $x^{(0)}$ proceeds in three distinct steps:

Step 1. Determine the class, open or closed, of $x^{(0)}(l\Delta)$ $(l = 0, 1, \ldots, N)$;

Step 2. Given these classes, determine $x^{(0)}(l\Delta)$ $(l = 0, 1, \ldots, N)$;

Step 3. Given $x^{(0)}(l\Delta)$ $(l = 0, 1, \ldots, N)$, fill in the missing paths $\{x^{(0)}(t); l\Delta < t < (l+1)\Delta\}$ $(l = 0, 1, \ldots, N-1)$.

We now describe each of these steps in turn.

Step 1. Determine a threshold current level y^*, either from past experience or by examining the current record y, and then determine the class, A_l say, of $x^{(0)}(l\Delta)$ by setting

$$A_l = \begin{cases} \mathrm{O}, & \text{if } y_l \geqslant y^*, \\ \mathrm{C}, & \text{if } y_l < y^*. \end{cases}$$

Step 2. Make a prior guess of the transition-rate matrix Q of $\{X(t); t \geqslant 0\}$ and determine the corresponding equilibrium distribution $\boldsymbol{\pi}$. Let $P = \exp(\Delta Q)$ be the corresponding transition matrix of the discrete time-Markov chain $\{X(l\Delta); l = 0, 1, \ldots\}$. Partition

$$Q = \begin{bmatrix} Q_{\mathrm{OO}} & Q_{\mathrm{OC}} \\ Q_{\mathrm{CO}} & Q_{\mathrm{CC}} \end{bmatrix}, \quad \boldsymbol{\pi} = [\boldsymbol{\pi}_{\mathrm{O}}, \boldsymbol{\pi}_{\mathrm{C}}], \quad P = \begin{bmatrix} P_{\mathrm{OO}} & P_{\mathrm{OC}} \\ P_{\mathrm{CO}} & P_{\mathrm{CC}} \end{bmatrix},$$

where, for example, Q_{OO} corresponds to transitions of X that remain within the open states and Q_{OC} to transitions from the open states to the closed states. Also, write $P = [p_{ij}]$.

For $l = 0, 1, \ldots, N-1$ and $i \in A_l$, let

$$\begin{aligned} \alpha_{li} &= \Pr(X(k\Delta) \in A_k,\ k = l+1, l+2, \ldots, N \mid X(l\Delta) = i) \\ &= \left\{ \left[\prod_{k=l}^{N-1} P_{A_k A_{k+1}} \right] \mathbf{1} \right\}_i, \end{aligned} \tag{6.1}$$

where $\mathbf{1}$ is an $m_{A_N} \times 1$ column vector of ones. Then

$$\Pr(X(0) = i \mid X(k\Delta) \in A_k,\ k = 0, 1, \ldots, N) = \frac{\pi_i \alpha_{0i}}{\sum_{j \in A_0} \pi_j \alpha_{0j}} \quad (i \in A_0), \tag{6.2}$$

for $k = 1, 2, \ldots, N-1$,

$$\begin{aligned} \Pr(X(k\Delta) = i \mid X(l\Delta) = i_l,\ l = 0, 1, \ldots, k-1 \text{ and } X(l\Delta) \in A_l,\ l = k, k+1, \ldots, N) \\ = \frac{p_{i_{k-1} i} \alpha_{ki}}{\sum_{j \in A_k} p_{i_{k-1} j} \alpha_{kj}} \quad (i \in A_k), \end{aligned} \tag{6.3}$$

$$\Pr(X(N\Delta)=i \mid X(l\Delta)=i_l,\ l=0,1,\ldots,N-1 \text{ and } X(N\Delta)\in A_N) = \frac{p_{i_{N-1}i}}{\sum_{j\in A_N} p_{i_{N-1}j}} \quad (i\in A_N). \quad (6.4)$$

Thus $x^{(0)}(l\Delta)$ $(l=0,1,\ldots,N)$ can be determined by sampling successively from the distributions given by (6.2), (6.3) with $k=1,2,\ldots,N-1$ and (6.4).

When N is large, underflow errors are likely to occur when computing the α_{li}. However, α_{li} only enters (6.2) and (6.3) via the ratio α_{li}/α_{lj}, and consequently these underflow errors can be overcome by appropriate scaling. We have found that the following scaling works well in practice.

Define vectors $\boldsymbol{\alpha}''_{N-1}, \boldsymbol{\alpha}''_{N-2}, \ldots, \boldsymbol{\alpha}''_0$ as follows. Let

$$\boldsymbol{\alpha}''_{N-1} = P_{A_{N-1}A_N}\mathbf{1}, \quad \boldsymbol{\alpha}'_{N-1} = \boldsymbol{\alpha}''_{N-1}/\|\boldsymbol{\alpha}''_{N-1}\| \quad (6.5)$$

and, for $l=N-2, N-3, \ldots, 0$, let

$$\boldsymbol{\alpha}''_l = P_{A_lA_{l+1}}\boldsymbol{\alpha}'_{l+1}, \quad \boldsymbol{\alpha}'_l = \boldsymbol{\alpha}''_l/\|\boldsymbol{\alpha}_l\|. \quad (6.6)$$

(For a vector, $\boldsymbol{u}=(u_1,u_2,\ldots,u_p)^{\mathrm{T}}$ say, where 'T' denotes transpose, $\|\boldsymbol{u}\|$ denotes the usual norm of $\boldsymbol{u}$, given by $\|\boldsymbol{u}\|=(\sum_{i=1}^p u_i)^{1/2}$.) Further, let

$$\boldsymbol{\alpha}_l = \left[\prod_{k=l}^{N-1} P_{A_kA_{k+1}}\right]\mathbf{1} \quad (l=0,1,\ldots,N-1).$$

Then the vectors $\boldsymbol{\alpha}_l$ $(l=0,1,\ldots,N-1)$ contain the elements $\boldsymbol{\alpha}_{li}$ $(l=0,1,\ldots,N-1; i\in A_l)$ defined in (6.1). Also,

$$\boldsymbol{\alpha}_{N-1}=\boldsymbol{\alpha}''_{N-1}, \quad \boldsymbol{\alpha}_l = (\|\boldsymbol{\alpha}''_{l+1}\|\|\boldsymbol{\alpha}''_{l+2}\|\cdots\|\boldsymbol{\alpha}''_{N-1}\|)\boldsymbol{\alpha}''_l \quad (l=0,1,\ldots,N-2).$$

It follows that (6.2) and (6.3) hold with α_{kj} replaced by α''_{kj}, etc. Note that (6.5) and (6.6) provide a simple way of computing $\boldsymbol{\alpha}''_l$ $(l=0,1,\ldots,N-1)$.

Step 3. First note that, by the Markov property, for $k=0,1,\ldots,N-1$ and $i,j\in S$,

$$\{X(t); k\Delta<t<(k+1)\Delta \mid X(k\Delta)=i, \quad X((k+1)\Delta)=j,\ X(t)=x^{(0)}(t)\ (t\notin[k\Delta,(k+1)\Delta])\}$$

has the same distribution as

$$\{X(t); k\Delta<t<(k+1)\Delta \mid X(k\Delta)=i,\ X((k+1)\Delta)=j\}.$$

Thus $x^{(0)}(t)$ $(k\Delta<t<(k+1)\Delta)$ can be obtained by simulating from

$$\{X(t); k\Delta<t<(k+1)\Delta \mid X(k\Delta)=i,\ X((k+1)\Delta)=j\}.$$

We shall describe how to do this for the case $k=0$. The cases $k=1,2,\ldots,N-1$ are done similarly.

For $v>0$, let J_v be the random variable defined by

$$J_v = \begin{cases} 1, & \text{if } X \text{ jumps in } (0,v),\\ 0, & \text{if } X \text{ does not jump in } (0,v).\end{cases}$$

Then, for $i,j\in S$ and $v>0$, let

$$p_J(v \mid i,j) = \Pr(J_v=1 \mid X(0)=i,\ X(v)=j) = \begin{cases} 1, & \text{if } i\neq j,\\ 1-\exp(q_{ii}v)/p_{ii}(v), & \text{if } i=j,\end{cases} \quad (6.7)$$

where $P(v) = [p_{ij}(v)] = \exp(vQ)$. If $J_v = 1$, let $U = \min_{t>0}\{t : X(t) \neq X(0)\}$ be the time of the first jump of X. For $i, j \in S$ and $v > 0$, let $f(u \mid i, j, v)$ $(0 < u < v)$ be the probability density function of U given $J_v = 1$, $X(0) = i$ and $X(v) = j$, and let

$$p(k \mid i, j, u, v) = \Pr(X(U) = k \mid U = u,\ X(0) = i,\ X(v) = j,\ J_v = 1) \qquad (k \in S,\ k \neq i).$$

Then

$$f(u \mid i, j, v) = \sum_{k \neq i} f_{ijk}(u \mid v) \quad (0 < u < v;\ i, j \in S) \tag{6.8}$$

and

$$p(k \mid i, j, u, v) = \frac{f_{ijk}(u \mid v)}{f(u \mid i, j, v)} \quad (k \in S,\ k \neq i), \tag{6.9}$$

where

$$f_{ijk}(u \mid v) = \frac{\exp(q_{ii}u)q_{ik}p_{kj}(v-u)}{1 - [\exp(q_{ii}v)/p_{ii}(v)]\delta_{ij}} \quad (0 < u < v;\ i, j, k \in S,\ k \neq i), \tag{6.10}$$

and

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

The path $x^{(0)}(t)$ $(0 < t < \Delta)$, given $x^{(0)}(0) = i$ and $x^{(0)}(\Delta) = j$, is simulated as follows. Sample J_1 from the distribution $p_J(\Delta \mid i, j)$, given by (6.7) with $v = \Delta$. If $J_1 = 0$ then set $x^{(0)}(t) = i$ $(0 < t < \Delta)$, and the required path is now simulated. If $J = 1$ then sample u_1 from the distribution $f(u_1 \mid i, j, \Delta)$ (i.e. from the distribution having probability density function given by (6.8) with $v = \Delta$) and then sample k_1 from the distribution $p(k_1 \mid i, j, u_1, \Delta)$, given by (6.9) with $v = \Delta$. Set $x^{(0)}(t) = i$ $(0 < t < u_1)$ and $x^{(0)}(u_1) = k_1$. Now determine whether or not $x^{(0)}$ jumps in (u, Δ), by sampling J_2 from the distribution $p_J(\Delta - u_1 \mid k_1, j)$. If $J_2 = 0$, so $x^{(0)}$ does not jump in (u_1, Δ), then set $x^{(0)}(t) = k_1$ $(u_1 < t < \Delta)$ and the required path is now simulated. If $J_2 = 1$ then sample u_2 from the distribution $f(u_2 \mid k_1, j, \Delta - u_1)$ and then sample k_2 from the distribution $p(k_2 \mid k_1, j, u_2, \Delta - u_1)$, and set $x^{(0)}(t) = k_2$ $(u_1 < t < u_1 + u_2)$ and $x^{(0)}(u_1 + u_2) = k_2$. This procedure is continued until eventually $J_l = 0$, for some l, in which case $x^{(0)}$ makes no further jumps in $(0, \Delta)$. A procedure for sampling from the distribution $f(u \mid i, j, v)$ is described in the appendix.

Note that when performing step 3 the description $((t_l^{(0)}, \chi_l^{(0)}); l = 0, 1, \ldots, n^{(0)})$ of $x^{(0)}$ can be determined as the intervals $[k\Delta, (k+1)\Delta]$ are gone through in turn.

For models with more than two states, i.e. having $m > 2$, we have found it beneficial to first fit the two-state model (figure 1*a*) and use the corresponding reconstruction (5.2) to determine A_l $(l = 0, 1, \ldots, N)$. This will provide an initial realization $x^{(0)}$ whose distribution is closer to $p(x \mid y)$ than one obtained by using step 1. Consequently, the resulting MCMC sampler will have a shorter burn-in period.

When analysing data for which the sampling interval Δ is small relative to the lengths of typical open and closed sojourns of the channel, the computing time required to obtain an initial $x^{(0)}$ can sometimes be reduced by fitting a model with a larger Δ to a corresponding fraction of the data. Of course, once an initial $x^{(0)}$ has been obtained, the complete data should be used in the subsequent MCMC run.

(*b*) *Choice of move types*

The move types (1)–(11) for updating x, described in § 4 *d* (i), can be grouped into three classes, $A = \{1\}$, $B = \{2, 4, 6, 8, 10\}$ and $C = \{3, 5, 7, 9, 11\}$ corresponding to boundary move, insertion and deletion, respectively. Note that the move types in C are the respective inverses of those in B, while the inverse of a boundary move is also a boundary move. The set of move types used in a particular application is dictated by its graph G. Move types (1), (6)–(9) should always be used. Move types (2) and (3) should be used whenever G contains a cycle of length 2. (For an integer $r \geqslant 2$, $i_1 \to i_2 \to \cdots \to i_r \to i_1$ forms a cycle of length r in G if $i_1, i_2, \ldots, i_r$ are distinct and $(i_l, i_{l+1}) \in E$ $(l = 1, 2, \ldots, r)$, where $i_{r+1} = i_1$. Thus G contains a cycle of length 2 if and only if there exists $i, j \in S$ such that $q_{ij} > 0$ and $q_{ji} > 0$. Note that G necessarily contains cycles of length 2 if X is time reversible.) Move types (4) and (5) should be used if there are distinct $i, j, k \in S$ with $(i, j) \in E$, $(j, k) \in E$ and $(i, k) \in E$. Note that if X is time reversible then move types (4) and (5) should be used if G contains cycles of length 3. Move types (10) and (11) should be used if G contains cycles of length $\geqslant 3$.

The move types chosen for successive x updates are obtained by sampling independently from the distribution p_i $(i = 1, 2, \ldots, 11)$ (cf. § 4 *d* (ii)) as follows. First the class of the move type is determined by sampling from the distribution (p_A, p_B, p_C), where, for example, p_B is the probability that a class-B move is chosen. Then if a class-B or class-C move is chosen, the actual move is found by sampling $\tilde{i}$ from a distribution $(\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_5)$ and choosing move type $2\tilde{i}$ if a class-B move is required and move type $2\tilde{i} + 1$ if a class-C move is required. Thus $p_1 = p_A$ and $p_{2k} = p_B \tilde{p}_k$, $p_{2k+1} = p_C \tilde{p}_k$ $(k = 1, 2, \ldots, 5)$.

The distributions (p_A, p_B, p_C) and $(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4, \tilde{p}_5)$ need to be specified by the user. Clearly, for $k = 1, 2, 5$, $\tilde{p}_k$ should be zero if the corresponding move type is precluded by G. It seems sensible to choose $\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4$ to be in approximately the same ratio as the number of sojourns of the corresponding type in x. Recall that x has $n + 1$ sojourns. If the graph G dictates that $\tilde{p}_2 = 0$ then $(\tilde{p}_1, \tilde{p}_3, \tilde{p}_4)$ could be chosen to be directly proportional to $(n - 1, 1, 1)$. Otherwise, let n_1^* be the number of intermediate sojourns of x whose two adjacent sojourns are in the same state and n_2^* be the number of intermediate sojourns of x whose two adjacent sojourns are in distinct states, i, j say, such that $(i, j) \in E$. Then $(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4)$ could be chosen to be directly proportional to $(n_1^*, n_2^*, (n_1^* + n_2^*)/n, (n_1^* + n_2^*)/n)$. Of course, this is impossible in practice since x is unobserved. However, one can either let $\tilde{p}_i$ $(i = 1, 2, \ldots, 5)$ depend on the current x realization $x^{(k)}$ (in which case the acceptance probabilities given in § 4 *d* (ii) need to be modified accordingly) or run the sampler with a sensible initial choice of $\tilde{p}_i$ $(i = 1, 2, \ldots, 5)$ until it has converged and then use the 'converged' realization $x^{(k)}$ to reset $\tilde{p}_i$ $(i = 1, 2, \ldots, 5)$. However, $\tilde{p}_i$ should never be set to zero if the corresponding move type is permissible in G.

In examples described in § 8, which are based on simulated data, $(p_A, p_B, p_C) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, if G contains cycles of length greater than or equal to 3 then $\tilde{p}_5 = 0.4$ and $\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4$ are chosen using the simulated realization x, as described above.

In ion-channel records, the observation period $[0, T]$ is typically long relative to the individual state sojourns of x. In such circumstances the variances of the posterior distributions $p(q \mid x, \theta, y)$ and $p(\theta \mid x, q, y)$ are small, so consequently the behaviour of $\{x^{(k)}; k = 0, 1, \ldots, M\}$ dictates to a large extent the behaviour of the complete sampler $\{(x^{(k)}, q^{(k)}, \theta^{(k)}); k = 0, 1, \ldots, M\}$. Thus it seems sensible to update x con-

siderably more frequently than q and θ. Hence, we recommend that x be updated n^* times for each update of (q, θ). Clearly, there are no hard and fast rules for the choice of n^*, (p_A, p_B, p_C) and $(\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_5)$, and experimentation may be required to find values that work well for a given application.

(*c*) *Computation aspects*

The MCMC sampler is straightforward to implement numerically on a computer. Recall that the sampler delivers a realization $\{(x^{(k)}, q^{(k)}, \theta^{(k)}); k = 0, 1, \ldots, M\}$ of a Markov chain having equilibrium distribution $p(x, q, \theta \mid y)$. In order to implement the sampler we store the current values of $x^{(k)}$, $q^{(k)}$ and $\theta^{(k)}$, together with various sufficient statistics based on $x^{(k)}$, and update (overwrite) accordingly each time x, q or θ is updated.

The information concerning x can be stored efficiently as follows. Suppose that the $n+1$ sojourns of x are labelled $0, 1, \ldots, n$. Then for $l = 0, 1, \ldots, n$ the following are stored: the length and state of the lth sojourn, t_l and χ_l, the number of sampling points in the lth sojourn n_l and the corresponding statistics $\tilde{y}_l$ and $\tilde{y}_{l,2}$, used for calculating the acceptance probabilities of moves that change x (see § 4 *d* (ii)). All the above variables can be stored in arrays indexed by l. However, there is no need to store the successive sojourns of x sequentially. Indeed, it is computationally cheaper not to do so. Instead, two pointers are attached to each sojourn indicating the indices of its preceding and succeeding sojourns in x. Initially, the sojourns of $x^{(0)}$ are stored in sequential order. Then, inserted sojourns are added at the end of the list of sojourns, while deleted sojourns are removed and replaced by sojourns taken from the end of the list. In both cases, appropriate changes need to be made to the pointers of the deleted/inserted sojourns and their preceding and succeeding sojourns in x. As well as the above sojourn-based information, one should also store the current values of the sufficient statistics n_{ij} $(i, j = 1, 2, \ldots m,\ i \neq j)$ and $\tilde{t}_i$ $(i = 1, 2, \ldots, m)$ for updating q (see § 4 *b*) and $n_{\mathrm{O}}, y_{\mathrm{O}}, y_{\mathrm{O},2}, n_{\mathrm{C}}, y_{\mathrm{C}}, y_{\mathrm{C},2}$ for updating θ (see § 4 *c*). All these sufficient statistics are very easy to update each time x changes.

The above indicates how to simulate $\{(x^{(k)}, q^{(k)}, \theta^{(k)}); k = 0, 1, \ldots, M\}$ efficiently. Of course, one also needs to store the output of the MCMC sampler for analysis as appropriate. Thus, if our aim is to make inferences concerning q and θ then we need to save successive values of $q^{(k)}$ and $\theta^{(k)}$, possibly after subsampling. Alternatively, if our aim is to reconstruct the unobserved single-channel record, as in § 5 *c*, then we need to determine and save $\delta_l^{(k)}$ $(l = 0, 1, \ldots, N)$ for successive values of k.

7. Constraints on (q, θ)

In many applications constraints may be imposed on the parameters q and θ, for example, owing to $\{X(t); t \geqslant 0\}$ being time reversible. In this section, we describe how such constraints can be incorporated into our MCMC sampler.

(*a*) *Constraint on mean current levels* μ_{O} *and* μ_{C}

Clearly, one would expect *a priori* that the mean open current level is larger than the mean closed current level, and consequently there is a strong case for incorporating this information into the prior distribution for θ. Suppose that the prior (3.5)

for θ is replaced by

$$p(\mu_{\mathrm{O}}, \sigma^2_{\mathrm{O}}, \mu_{\mathrm{C}}, \sigma^2_{\mathrm{C}}) = (\sigma^2_{\mathrm{O}}\sigma^2_{\mathrm{C}})^{-1} \quad (-\infty < \mu_{\mathrm{C}} < \mu_{\mathrm{O}} < \infty,\ 0 < \sigma^2_{\mathrm{O}},\ \sigma^2_{\mathrm{C}} < \infty),$$

which is still improper. Then the functional form of the posterior $p(\theta \mid x, q, y)$ is still given by (4.6), except now it is only valid over the region $\{-\infty < \mu_{\mathrm{C}} < \mu_{\mathrm{O}} < \infty,\ 0 < \sigma^2_{\mathrm{O}},\ \sigma^2_{\mathrm{C}} < \infty\}$. It still follows that

$$p(\sigma^2_{\mathrm{O}}, \sigma^2_{\mathrm{C}} \mid x, q, y, \mu_{\mathrm{O}}, \mu_{\mathrm{C}}) = p(\sigma^2_{\mathrm{O}} \mid x, y, \mu_{\mathrm{O}})p(\sigma^2_{\mathrm{C}} \mid x, y, \mu_{\mathrm{C}}),$$

so σ^2_{O} and σ^2_{C} can be updated as before. However, μ_{O} and μ_{C} are no longer *a posteriori* independent given $x, q, y, \sigma^2_{\mathrm{O}}, \sigma^2_{\mathrm{C}}$. Indeed, it follows from (4.6), with the constraints $\mu_{\mathrm{C}} < \mu_{\mathrm{O}}$, that

$$p(\mu_{\mathrm{O}}, \mu_{\mathrm{C}} \mid x, q, y, \sigma^2_{\mathrm{O}}, \sigma^2_{\mathrm{C}}) \propto \exp\left\{-\frac{n_{\mathrm{O}}(\mu_{\mathrm{O}} - \bar{y}_{\mathrm{O}})^2}{2\sigma^2_{\mathrm{O}}} - \frac{n_{\mathrm{C}}(\mu_{\mathrm{C}} - \bar{y}_{\mathrm{C}})^2}{2\sigma^2_{\mathrm{C}}}\right\}$$

for $-\infty < \mu_{\mathrm{C}} < \mu_{\mathrm{O}} < \infty$. Thus, *a posteriori* (given $x, q, y, \sigma^2_{\mathrm{O}}, \sigma^2_{\mathrm{C}}$) μ_{O} and μ_{C} follow 'independent' normal distributions conditioned so that $\mu_{\mathrm{O}} > \mu_{\mathrm{C}}$. Hence, $(\mu_{\mathrm{O}}, \mu_{\mathrm{C}})$ can be updated by repeatedly sampling $(\mu^*_{\mathrm{O}}, \mu^*_{\mathrm{C}})$ from the pair of independent normal distributions, $N(\bar{y}_{\mathrm{O}}, \sigma^2_{\mathrm{O}}/n_{\mathrm{O}})$ and $N(\bar{y}_{\mathrm{C}}, \sigma^2_{\mathrm{C}}/n_{\mathrm{C}})$, until a pair satisfying $\mu^*_{\mathrm{O}} > \mu^*_{\mathrm{C}}$ is obtained, and then updating $(\mu_{\mathrm{O}}, \mu_{\mathrm{C}})$ to $(\mu^*_{\mathrm{O}}, \mu^*_{\mathrm{C}})$. In most applications, the values of $\bar{y}_{\mathrm{O}}$, $\sigma^2_{\mathrm{O}}/n_{\mathrm{O}}$, $\bar{y}_{\mathrm{C}}$ and $\sigma^2_{\mathrm{C}}/n_{\mathrm{C}}$ will be such that the first pair of $(\mu^*_{\mathrm{O}}, \mu^*_{\mathrm{C}})$ will be accepted almost all of the time.

(*b*) *Reversibility*

In many applications the underlying single-channel gating process $\{X(t); t \geqslant 0\}$ is necessarily time reversible. This places constraints on the set of permissible q, which need to be reflected in the prior $p(q)$.

If $\{X(t); t \geqslant 0\}$ is time reversible then, for $i, j \in S$, $q_{ij} > 0$ if and only if $q_{ji} > 0$. It follows that the directed graph G may be replaced by an undirected graph, G^* say, in which, for any $i, j \in S$, there is an (undirected) arc between i and j if and only if $q_{ij} > 0$. Kolmogorov's criterion for reversibility states that $\{X(t); t \geqslant 0\}$ is time reversible if and only if for any cycle of states in G^*, the product of the transition rates q_{ij} one way round the cycle is the same as the product the other way round (see Kelly 1979, theorem 1.8). Thus if G^* is a tree, as in the mechanisms of figure 1*a*, *b*, $\{X(t); t \geqslant 0\}$ is necessarily time reversible. Consequently, no constraints need to be imposed on q and we can proceed as before. However, if G^* contains cycles, Kolmogorov's criterion imposes constraints on q that must be incorporated into $p(q)$.

Suppose, for simplicity, that G^* contains just one cycle, as in the mechanisms of figure 1*c*, *d*. Then one of the q_{ij}, say $q_{i_0 j_0}$, can be expressed as a function of the others. For example, for the mechanism of figure 1*c*, we can set $i_0 = 1$, $j_0 = 2$ and Kolmogorov's criterion implies that $q_{12}q_{23}q_{31} = q_{13}q_{32}q_{21}$, so that $q_{12} = (q_{13}q_{32}q_{21})/(q_{23}q_{31})$. Let $E' = E\backslash\{(i_0, j_0)\}$ be the set of edges in G excluding (i_0, j_0) and $q_{E'} = \{q_{ij} : (i, j) \in E'\}$. Then, as a consequence of Kolmogorov's criterion, we can write $q_{i_0 j_0} = q_{i_0 j_0}(q_{E'})$. Thus if we specify a prior for $q_{E'}$, $p(q_{E'})$ say, then that will induce a prior for q. However, our prior beliefs about $\{q_{ij} : (i, j) \in E\}$ can no longer be independent, and consequently we may wish our prior beliefs about

$q_{i_0j_0}$ to influence $p(q_{E'})$. One approach would be to take

$$p(q_{E'}) \propto \left[\prod_{(i,j)\in E'} p(q_{ij})\right] p(q_{i_0j_0}(q_{E'})), \tag{7.1}$$

where $p(q_{ij})$ $((i,j)\in E)$ is given by (3.4). However, there is arbitrariness in the choice of (i_0,j_0) and it is easily verified that a different choice for (i_0,j_0) in (7.1) would yield a different distribution for q, owing to the nonlinear nature of the constraints imposed on q by Kolmogorov's criterion.

The above undesirable property of $p(q_{E'})$ can be avoided as follows. For $(i,j)\in E$ let $r_{ij} = \ln q_{ij}$ and write $r = \{r_{ij} : (i,j)\in E\}$ and $r_{E'} = \{r_{ij} : (i,j)\in E'\}$. Equations (3.3) and (3.4) induce a prior, $p(r)$ say, for r, from which a prior $p(r_{E'})$ can be derived analogously to the method of (7.1). Note that the constraint imposed by Kolmogorov's criterion is linear in r, and hence the induced prior on (time reversible) r is invariant to the particular choice of (i_0,j_0). Finally, we obtain a prior $p(q_{E'})$ by the retransformation $q_{ij} = \exp(r_{ij})$ $((i,j)\in E')$. Following this recipe yields

$$p(q_{E'}) \propto \left[\prod_{(i,j)\in E'} p(q_{ij})\right] q_{i_0j_0}(q_{E'}) p(q_{j_0j_0}(q_{E'})). \tag{7.2}$$

It is readily verified that the distribution on time reversible q induced by (7.2) is invariant to the choice of (i_0,j_0).

More generally, if the graph G^* contains two or more cycles, the edge set E can be partitioned into $E = E'\cup E''$, with the transition rates $q_{E'} = \{q_{ij} : (i,j)\in E'\}$ being free and the transition rates $q_{E''} = \{q_{ij} : (i,j)\in E''\}$ being a function of $q_{E'}$, determined by Kolmogorov's criterion. The above method of constructing a prior for $q_{E'}$ then yields

$$p(q_{E'}) \propto \prod_{(i,j)\in E'} \{q_{ij}^{\alpha_{ij}-1}\exp(-\beta_{ij}q_{ij})\} \prod_{(i,j)\in E''} \{[q_{ij}(q_{E'})]^{\alpha_{ij}}\exp(-\beta_{ij}q_{ij}(q_{E'}))\}, \tag{7.3}$$

where (α_{ij},β_{ij}) $((i,j)\in E)$ satisfy the conditions following (3.5).

Turning now to our MCMC sampler, to update $q_{E'}$ we generate a proposal $q'_{E'} = \{q'_{ij} : (i,j)\in E'\}$ by sampling, for $(i,j)\in E'$, q'_{ij} independently from the gamma distribution $\tilde{p}(q_{ij}\mid x)$ given by (4.3). It follows from (2.7) that the acceptance probability for this proposal is $A' = \min(1,B')$, where

$$\begin{aligned} B' &= \frac{p(x,q'_{E'},\theta\mid y)\prod_{(i,j)\in E'}\tilde{p}(q_{ij}\mid x)}{p(x,q_{E'},\theta\mid y)\prod_{(i,j)\in E'}\tilde{p}(q'_{ij}\mid x)} \\ &= \frac{\pi_{\chi_0}(q'_{E'})\prod_{(i,j)\in E''}\{q_{ij}(q'_{E'})\}^{\alpha_{ij}+n_{ij}}\exp\{-(\tilde{t}_i+\beta_{ij})q_{ij}(q'_{E'})\}}{\pi_{\chi_0}(q_{E'})\prod_{(i,j)\in E''}\{q_{ij}(q_{E'})\}^{\alpha_{ij}+n_{ij}}\exp\{-(\tilde{t}_i+\beta_{ij})q_{ij}(q_{E'})\}}, \end{aligned}$$

using (3.8), (3.9), (4.2) (with q replaced by q_E) and (7.3). Note that there is arbitrariness in the choice of partition $E = E'\cup E''$, although all choices of partition yield the same prior for q. Thus different choices of partition can be made in successive updates of q.

A disadvantage of the above method of obtaining a prior $p(q)$ when $\{X(t); t\geqslant 0\}$ is time reversible is that the resulting prior $p(q_{E'})$ does not follow a convenient

distributional form, and consequently it is not possible to determine the prior mean, variances and covariances of the individual q_{ij}. This can be overcome by replacing the initial prior $p(q)$ (given by equation (3.3)) by a product of lognormal densities, namely

$$p(q) = \prod_{(i,j)\in E} \{(q_{ij}\sqrt{2\pi}\sigma_{ij})^{-1} \exp[-(\log q_{ij} - \mu_{ij})^2/2\sigma_{ij}^2]\}$$

and then deriving a prior $p(q_{E'})$ as before. Thus we now assume that, without the reversibility constraints, *a priori* $\log q_{ij} \sim N(\mu_{ij}, \sigma_{ij})$ $((i,j) \in E)$, independently. The multiplicative reversibility constraints on the q_{ij} become additive constraints on the $\log q_{ij}$, and it follows that $p(q_{E'})$ is a multivariate lognormal density. Indeed, taking account of the reversibility constraints, q *a priori* follows a singular multivariate lognormal distribution, and explicit expressions can be obtained for the prior means, variances and covariances of the q_{ij}. A drawback of the lognormal prior is that $p(q)$ is no longer nearly conjugate for $f(x \mid q)$. However, this is not a great problem. The parameter $q_{E'}$ can be updated by generating a proposal $q'_{E'}$ by sampling, for $(i,j) \in E'$, q'_{ij} from the gamma distribution $\tilde{p}(q_{ij} \mid x)$ given by (4.3), but with $\alpha_{ij} = 1$ and $\beta_{ij} = 0$, and modifying the acceptance probability accordingly. In practice, the number of sojourns in the data will be sufficiently large to ensure that the gamma distributions are much narrower than the prior distribution, and hence nearly all sampled q will have acceptance probabilities equal to or close to 1. Thus very little efficiency is lost by this approach.

(*c*) *Models specified in terms of parameters governing q*

In many applications, the transition rates of the continuous-time Markov chain used to model single-channel gating are themselves functions of a set of channel kinetic parameters, ψ say. For example, the mechanism of figure 1*d*, which has been used by Colquhoun & Sakmann (1985) for the frog muscle nAChR, is often presented in the form shown in figure 2.

In figure 2, C represents the closed channel, O the open channel and A the agonist. The six independent parameters are K_{B}, the agonist binding constant; K_{01}, the $\mathrm{CA} \rightleftharpoons \mathrm{OA}$ equilibrium constant; K_{02} the $\mathrm{CA}_2 \rightleftharpoons \mathrm{OA}_2$ equilibrium constant; k_{on}, the association rate of agonist and receptor; and h_1 and h_2, the channel opening rates when one and two molecules of the agonist are bound to the channel, respectively. The parameter a denotes the concentration of the agonist, which is controlled by the experiment and thus is known. Hence, for this mechanism, we may write $q = q(\psi)$, where $\psi = (K_{\mathrm{B}}, K_{01}, K_{02}, k_{\mathrm{on}}, h_1, h_2)$.

For models like the one above, it is natural to express our prior beliefs in terms of a prior, $p(\psi)$ say, for ψ, rather than by a prior $p(q)$. (Of course, $p(\psi)$ will induce a prior on q.) We shall then require an MCMC sampler $\{(x^{(k)}, \psi^{(k)}, \theta^{(k)}); k = 0, 1, \dots\}$ whose equilibrium distribution is the posterior $p(x, \psi, \theta \mid y)$. Such a sampler can be obtained by using the same method of updating θ and x as before, except q is replaced by $q(\psi)$ when updating x, though clearly the method for updating q needs modifying. We now describe a method for updating ψ for the model of figure 2, but using notation that reflects the general case.

For the model of figure 2, the transition rates q determine the parameter ψ in that we have, for example,

$$\left.\begin{array}{c} K_{\mathrm{B}} = 2q_{43}/(aq_{34}), \quad K_{01} = q_{41}/q_{14}, \quad K_{02} = q_{32}/q_{23}, \\ k_{\mathrm{on}} = q_{43}/a, \quad h_1 = q_{41}, \quad h_2 = q_{32}, \end{array}\right\} \tag{7.4}$$

where the states are labelled as in figure 1*d*.

Let

$$E' = \{(4,3),(3,4),(4,1),(1,4),(3,2),(2,3)\},$$
$$E'' = E \backslash E' = \{(1,2),(2,1),(4,5),(5,4)\},$$
$$q_{E'} = \{q_{ij} : (i,j) \in E'\}, \quad q_{E''} = \{q_{ij} : (i,j) \in E''\}.$$

Then we may write $\psi = \psi(q_{E'})$, where the function $\psi(q_{E'})$ is given by (7.4).

Let $(x^{(k)}, \psi^{(k)}, \theta^{(k)}) = (x, \psi, \theta)$ denote the current state of our MCMC sampler. Then we generate a proposal ψ' for ψ by first making a proposal $q'_{E'}$ for $q_{E'}$ by sampling, for $(i,j) \in E'$, q'_{ij} independently from the gamma distribution $p'(q_{ij} \mid x) = \mathrm{Ga}(q_{ij} \mid n_{ij}+1, \tilde{t}_i)$, and then setting $\psi' = \psi(q'_{E'})$. Thus the density of the transition kernel for the proposal ψ' is

$$R(\psi \to \psi'; x, \theta) = \left\{ \prod_{(i,j)\in E'} p'(q'_{ij}(\psi') \mid x) \right\} |J(\psi')|, \tag{7.5}$$

where the Jacobian $J(\psi)$ is given by

$$J(\psi) = \begin{vmatrix} \frac{\partial q_{43}}{\partial K_{\mathrm{B}}} & \frac{\partial q_{43}}{\partial K_{01}} & \frac{\partial q_{43}}{\partial K_{02}} & \frac{\partial q_{43}}{\partial k_{\mathrm{on}}} & \frac{\partial q_{43}}{\partial h_1} & \frac{\partial q_{43}}{\partial h_2} \\ \frac{\partial q_{34}}{\partial K_{\mathrm{B}}} & \frac{\partial q_{34}}{\partial K_{01}} & \frac{\partial q_{34}}{\partial K_{02}} & \frac{\partial q_{34}}{\partial k_{\mathrm{on}}} & \frac{\partial q_{34}}{\partial h_1} & \frac{\partial q_{34}}{\partial h_2} \\ \frac{\partial q_{41}}{\partial K_{\mathrm{B}}} & \frac{\partial q_{41}}{\partial K_{01}} & \frac{\partial q_{41}}{\partial K_{02}} & \frac{\partial q_{41}}{\partial k_{\mathrm{on}}} & \frac{\partial q_{41}}{\partial h_1} & \frac{\partial q_{41}}{\partial h_2} \\ \frac{\partial q_{14}}{\partial K_{\mathrm{B}}} & \frac{\partial q_{14}}{\partial K_{01}} & \frac{\partial q_{14}}{\partial K_{02}} & \frac{\partial q_{14}}{\partial k_{\mathrm{on}}} & \frac{\partial q_{14}}{\partial h_1} & \frac{\partial q_{14}}{\partial h_2} \\ \frac{\partial q_{32}}{\partial K_{\mathrm{B}}} & \frac{\partial q_{32}}{\partial K_{01}} & \frac{\partial q_{32}}{\partial K_{02}} & \frac{\partial q_{32}}{\partial k_{\mathrm{on}}} & \frac{\partial q_{32}}{\partial h_1} & \frac{\partial q_{32}}{\partial h_2} \\ \frac{\partial q_{23}}{\partial K_{\mathrm{B}}} & \frac{\partial q_{23}}{\partial K_{01}} & \frac{\partial q_{23}}{\partial K_{02}} & \frac{\partial q_{23}}{\partial k_{\mathrm{on}}} & \frac{\partial q_{23}}{\partial h_1} & \frac{\partial q_{23}}{\partial h_2} \end{vmatrix}$$
$$= \frac{2ak_{\mathrm{on}}h_1h_2}{(K_{\mathrm{B}}K_{01}K_{02})^2}.$$

Using (2.7), the acceptance probability for this proposal is $A = \min(1, B)$, where

$$B = \frac{p(x,\psi',\theta \mid y) R(\psi' \to \psi; x, \theta)}{p(x,\psi,\theta \mid y) R(\psi \to \psi'; x, \theta)}$$
$$= \frac{p(\psi')\pi_{\chi_0}(q(\psi'))|J(\psi)| \prod_{(i,j)\in E''} [\{q_{ij}(\psi')\}^{n_{ij}} \exp\{-q_{ij}(\psi')\tilde{t}_i\}]}{p(\psi)\pi_{\chi_0}(q(\psi))|J(\psi')| \prod_{(i,j)\in E''} [\{q_{ij}(\psi)\}^{n_{ij}} \exp\{-q_{ij}(\psi)\tilde{t}_i\}]}, \tag{7.6}$$

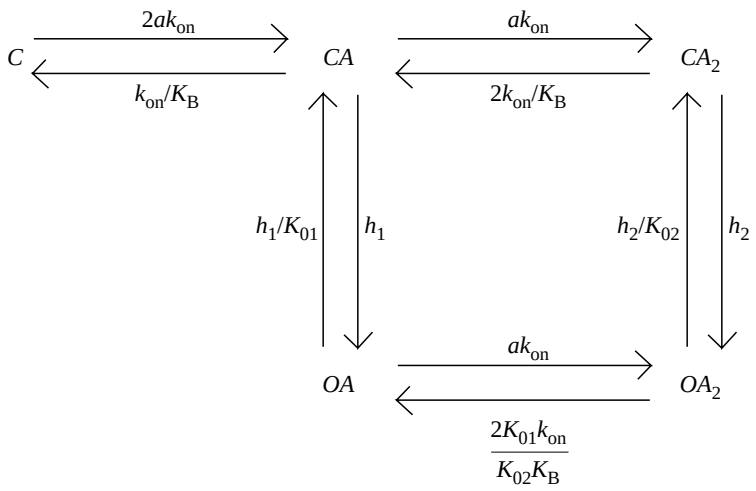

Figure 2. A five-state model with transition rates.

using (3.8), (3.9) and (4.2), with obvious modification, and (7.5).

In order to implement the above we need to specify the prior $p(\psi)$. One possibility is to use the improper vague prior $p(\psi) = 1$, in which case the terms $p(\psi)$ and $p(\psi')$ cancel in (7.6).

Note that again there is arbitrariness in the choice of the partition $E = E' \cup E''$, and different partitions could be used for successive updates of ψ. For the example of figure 2, it seems sensible that $(4, 1)$, $(1, 4)$, $(3, 2)$ and $(2, 3)$ always belong to E' and that E' is completed by choosing one edge from each of $(5, 4)$, $(4, 3)$, $(1, 2)$ and $(4, 5)$, $(3, 4)$, $(2, 1)$, perhaps by maximizing n_{ij}, the number of transitions from i to j in the current $x^{(k)}$. This approach aims at putting the maximum posterior information into the gamma proposal distributions, hence maximizing acceptance probabilities.

In general there may not be a one-to-one correspondence between q and ψ, in which case the function $\psi(q_{E'})$ may not be well defined, i.e. for given $q_{E'}$ there may be several ψ such that $q_{E'}(\psi) = q_{E'}$. Let $\mathcal{Q}_{E'}$ and $\boldsymbol{\psi}$ denote the spaces of possible values for $q_{E'}$ and ψ, respectively. If q and ψ are not in one-to-one correspondence, the above algorithm for sampling ψ can be modified as follows. First note that the prior distribution on ψ induces a prior $p(q_{E'})$ on $q_{E'}$. In the case of a one-to-one correspondence between ψ and $q_{E'}$, this is $p(q_{E'}) = p(\psi)|J(\psi)|^{-1}$, but otherwise $p(q_{E'})$ will typically be more complex. Given $q_{E'}$, there is also an induced conditional distribution $p(\psi \mid q_{E'})$ with support on the set $\Psi(q_{E'})$ of ψ values for which $q_{E'}(\psi) = q_{E'}$. Since the likelihood depends on ψ only through $q_{E'}$, this conditional distribution is not modified by the data, so the posterior conditional distribution is the same as the prior $p(\psi \mid q_{E'})$. (One consequence is that it is necessary that these prior conditional distributions be proper. Otherwise the posterior distribution would also be improper.)

In implementing the MCMC sampler we generate only $q_{E'}$ values, and do not at this stage convert a sampled $q_{E'}$ into a value of ψ. We sample $q_{E'}$ as above, by sampling q_{ij} for each $(i, j) \in E'$ independently from the gamma distribution $p'(q_{ij} \mid x)$. The new value $q'_{E'}$ is accepted with probability $A = \min(1, B)$, where B is given by (7.6) except that $p(\psi)|J(\psi)|^{-1}$ is replaced by $p(q_{E'})$, and $p(\psi')|J(\psi')|^{-1}$

by $p(q'_{E'})$. We do not need to go further and generate a value for ψ at this stage, because $q_{E'}$ implies q, which is all that is needed for updating x and θ.

We only generate ψ values to correspond to the $q_{E'}$ values that are retained in the final MCMC sample. This saves computation because ψ need not be generated during burn-in, nor for any iterations after burn-in that are not retained if we draw the final sample by thinning. For each retained $q_{E'}$, we sample ψ within $\Psi(q_{E'})$ from the conditional distribution $p(\psi \mid q_{E'})$. (Note that this step would be impossible if $p(\psi \mid q_{E'})$ were improper.) In fact, the Rao–Blackwellization device mentioned in § 5 *b* means that we could make inferences about ψ by averaging these conditional distributions, rather than actually generating ψ values from them.

8. Examples

We tested the performance of our MCMC sampler, using simulated data from the four gating mechanisms depicted in figure 1. For each gating model and each set of parameters, we simulated a realization x of the gating process and then obtained a current record by adding Gaussian noise as described by (3.1). We then ran our sampler on the resulting data.

The method of choosing the probabilities of the different move types for updating x has been discussed in § 6 *b*. Move types 2, 4, 6 and 8 require further probability distributions to be specified, namely, in the notation of § 4 *d* (i), $\tilde{p}_2(j \mid i)$, $\tilde{p}_4(j \mid i_1, i_3)$, $\tilde{p}_6(j \mid i)$ and $\tilde{p}_8(j \mid i)$. These were all taken to be uniform on the corresponding set of permissible states. In describing the results, an iteration of the MCMC sampler consists of single updates of q and θ, followed by n^* updates of x (see § 6 *b*), where $n^* = 200$ for the two-state model and $n^* = 400$ for the other models.

(*a*) *Two-state model*

We first considered the two-state model of figure 1*a*, for which parameter estimation for time-interval omitted sojourn time data has been studied in depth by Yeo *et al.* (1988). We took $q_{12} = 500$ and $q_{21} = 100$ (where the units are transitions per second) and simulated a current record with $N = 10^4$, $\Delta = 10^{-4}$ s, $\mu_{\mathrm{O}} = 1$, $\mu_{\mathrm{C}} = 0$ and $\sigma_{\mathrm{O}} = \sigma_{\mathrm{C}} = 0.4$. This gave rise to a channel record consisting of 88 open sojourns and 89 closed sojourns. The above parameter values correspond broadly to those used by Fredkin & Rice (1992*b*, example 4.1), although the variance of the noise is slightly smaller.

For this example we did not use the method of § 6 *a* for choosing the initial value $x^{(0)}$. Instead we set

$$x^{(0)}(l\Delta) = \begin{cases} 1, & \text{if } y_l \geqslant 0.5, \\ 2, & \text{if } y_l < 0.5, \end{cases}$$

to yield data in the form of discrete time sojourns. These were then converted into continuous-time sojourns by assuming that if $x^{(0)}(l\Delta) = x^{(0)}((l+1)\Delta)$ then the same sojourn spans $l\Delta$ and $(l+1)\Delta$, while if $x^{(0)}(l\Delta) \neq x^{(0)}((l+1)\Delta)$ then $x^{(0)}$ has a single jump in $(l\Delta, (l+1)\Delta)$, at a point sampled uniformly in that interval. Although this initial value is unlikely to be as good as one obtained by the method of § 6 *a*, it should provide an indication of the mixing properties of our sampler.

We assumed an improper vague prior for q and ran the sampler for 70 000 iterations. The results are shown in figure 3, where a burn-in of 5000 iterations was used

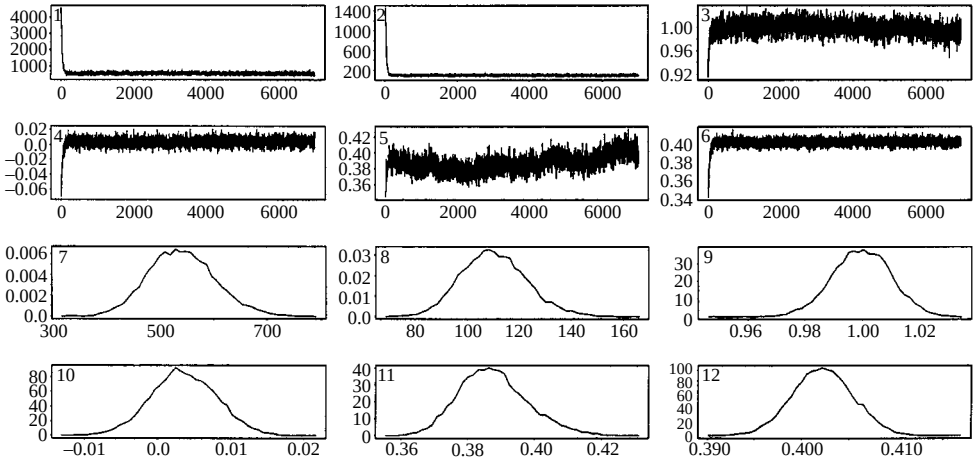

Figure 3. (*a*) Convergence plots and marginal plots for parameters of a two-state model. Convergence plot of q_{12} (1), q_{21} (2), μ_{O} (3), μ_{C} (4), σ_{O} (5), σ_{C} (6). Marginal plot of q_{12} (7), q_{21} (8), μ_{O} (9), μ_{C} (10), σ_{O} (11), σ_{C} (12).

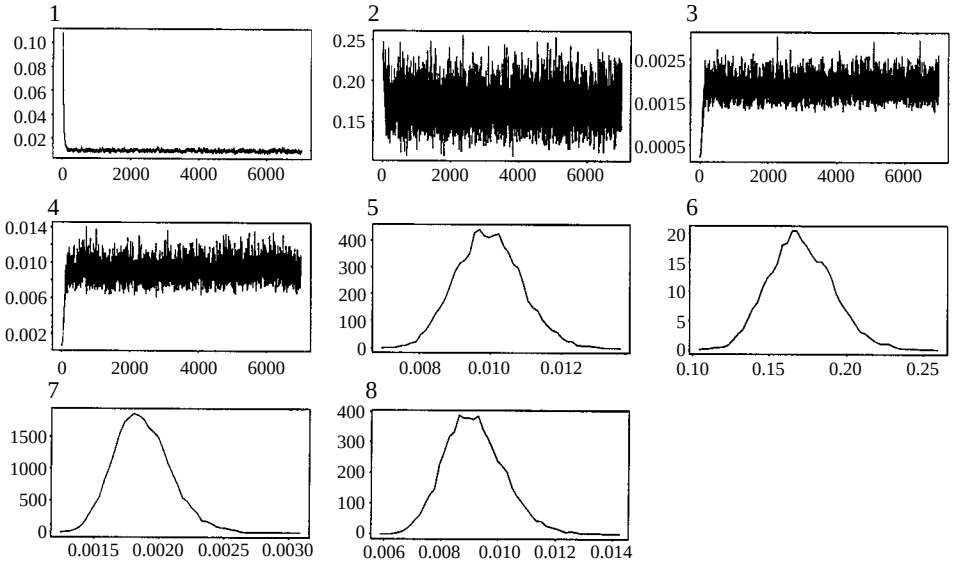

Figure 3. (*b*) Convergence plots and marginal plots for features of a two-state model. Convergence plot of p_{MIS} (1), π_0 (2), $\tilde{\mu}_0$ (3), $\tilde{\mu}_{\mathrm{C}}$ (4). Marginal plot of p_{MIS} (5), π_0 (6), $\tilde{\mu}_0$ (7), $\tilde{\mu}_{\mathrm{C}}$ (8).

for the marginal plots and the Bayesian reconstruction, after which the samples were collected every tenth iteration. (All the MCMC output figures in this paper were produced using the statistical package S-PLUS. The marginal plots are kernel density estimates (see § 5 *b*), using a Gaussian kernel K and the S-PLUS default window width h. Thus, for example, the marginal plot of μ_{O} is a graph of $\tilde{p}(\mu_{\mathrm{O}} \mid y)$ as defined by (5.1).) It is clear from the convergence and marginal plots in figure 3*a* that the sampler is working satisfactorily for this model. Note the rapid movement in the convergence plots from the poor initial values of the parameters to their correspond-

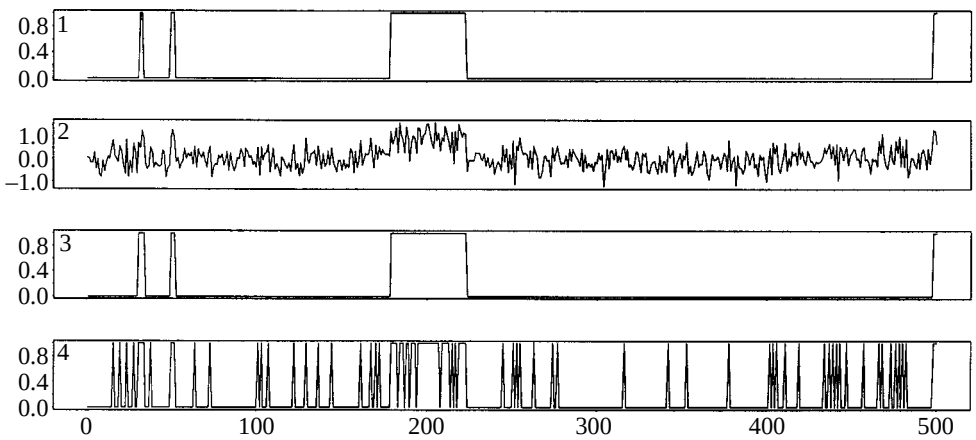

Figure 3. (*c*) Restoration of single-channel record for a two-state model. Original sequence (1), noisy sequence (2), Bayesian reconstruction of r (3), crude threshold reconstruction of r (4).

ing equilibrium regions. The values of $q_{12}^{(k)}$ and $q_{21}^{(k)}$, for small k, were far too large because the crude threshold algorithm used for determining $x^{(0)}$ yielded a realization having far too many jumps, owing to the noise in the current record y. Note also that the marginal posterior distributions of μ_{C} and σ_{C} are more concentrated than those for μ_{O} and σ_{O}. This is because, for the parameter values chosen, the channel spends more time being closed than being open, so consequently the current record y contains more information about $(\mu_{\mathrm{C}}, \sigma_{\mathrm{C}})$ than about $(\mu_{\mathrm{O}}, \sigma_{\mathrm{O}})$.

Figure 3*b* shows the convergence and marginal plots for the proportion of misclassified sampling points in $x^{(k)}$ (denoted by p_{MIS}), the equilibrium probability π_{O} that the channel is open, and the equilibrium mean length of a sojourn of the channel in open and closed states, $\tilde{\mu}_{\mathrm{O}}$ and $\tilde{\mu}_{\mathrm{C}}$. Again these are very satisfactory. The estimated means and variances of the marginal distributions were 0.1704 and 4.282×10^{-4} for π_{O}, 1.8709×10^{-3} s and 4.695×10^{-8} s^2 for $\tilde{\mu}_{\mathrm{O}}$ and 9.169×10^{-3} s and 1.095×10^{-6} s^2 for $\tilde{\mu}_{\mathrm{C}}$. The true values of these channel properties were $\pi_{\mathrm{O}} = 0.1667$, $\tilde{\mu}_{\mathrm{O}} = 2 \times 10^{-3}$ s and $\tilde{\mu}_{\mathrm{C}} = 10^{-2}$ s.

Figure 3*c* shows, for the first 501 sampling points, the true sequence of open and closed sojourns of the channel, the corresponding noisy current record, the MCMC marginal Bayesian reconstruction given by (5.2) and the crude threshold reconstruction, which decides whether the channel is open or closed at time $l\Delta$ according to whether $y_l \geqslant \frac{1}{2}$ or $y_l < \frac{1}{2}$. The MCMC marginal Bayes reconstruction performs very well. The proportion of misclassified sampling points in this reconstruction was 6.099×10^{-3}, compared with 0.1095 in the crude threshold reconstruction and an average of 9.962×10^{-3} in $x^{(k)}$, as k ranges over the samples collected after burn-in. When comparing the performances of the MCMC marginal Bayes and crude threshold reconstructions it should be borne in mind that the latter would never be used in practice, since, if presented with data resembling the simulated records of the examples, an experimentalist would either filter the data before thresholding and/or use some form of time-course fitting, both of which would significantly reduce the number of misclassified sampling points.

This example took 84 min to run on a Sun Sparcstation 10.

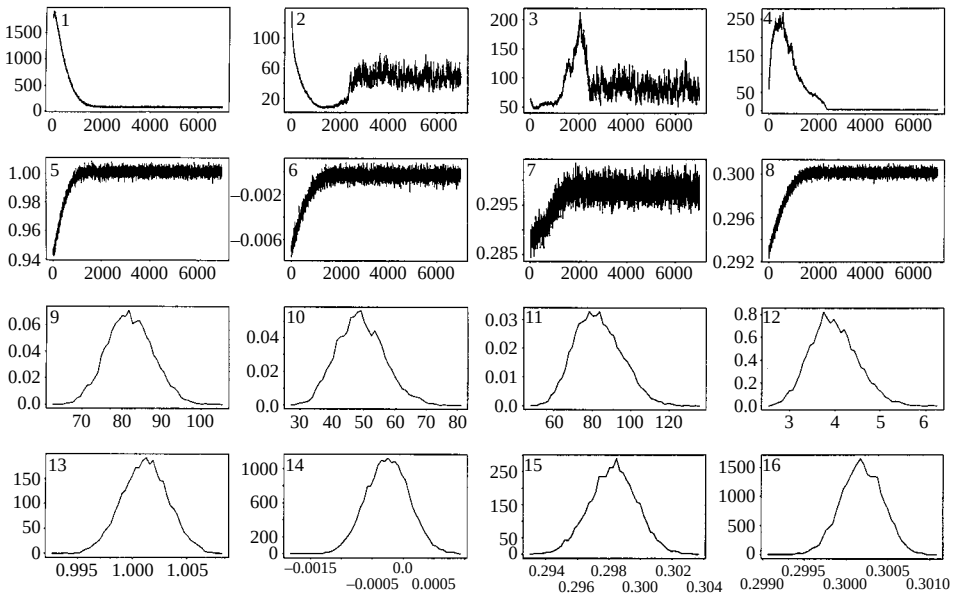

Figure 4. (*a*) Convergence plots and marginal plots for parameters of a three-state linear model with parameter set 1. Convergence plot of q_{12} (1), q_{21} (2), q_{23} (3), q_{32} (4), μ_{O} (5), μ_{C} (6), σ_{O} (7), σ_{C} (8). Marginal plot of q_{12} (9), q_{21} (10), q_{21} (11), q_{32} (12), μ_{O} (13), μ_{C} (14), σ_{O} (15), σ_{C} (16).

(*b*) *Three-state linear model*

The three-state model of figure 1*b* was used by Fredkin & Rice (1992*a*) to examine the performance of their hidden Markov maximum-likelihood algorithm. They considered three sets of values for the transition rates q, only one of which could be estimated satisfactorily. The failure to estimate the other two sets satisfactorily is not an inherent problem with their procedure but rather a consequence of the near unidentifiability of the model with those parameter values. We tested our MCMC sampler on simulated data from two of the three sets of parameter values considered by Fredkin & Rice. The unit of time in those examples was a second, and is not explicitly shown in the following discussion.

We first considered parameter set 1, in which $q_{12} = 94$, $q_{21} = 50$, $q_{23} = 91$ and $q_{32} = 4$, which Fredkin & Rice (1992*a*) could resolve. We simulated a current record with $N = 7.5 \times 10^5$, $\Delta = 1.28 \times 10^{-4}$, $\mu_{\mathrm{O}} = 1$, $\mu_{\mathrm{C}} = 0$ and $\sigma_{\mathrm{O}} = \sigma_{\mathrm{C}} = 0.3$, yielding a channel record consisting of 210 open sojourns and 211 closed sojourns. (The values for Δ, μ_{O}, μ_{C}, σ_{O} and σ_{C} were taken from Fredkin & Rice (1992*a*).) We determined the initial $x^{(0)}$ by first fitting a two-state model and then using the method described in § 6 *a*, with the prior guess for q being twice its true value (so leading to an initial reconstruction with too many sojourns). We took a fairly uninformative prior by setting $\alpha_{ij} = 0.01 q_{ij}$, $\beta_{ij} = 0.01$ ($(i, j) \in E$). We ran the sampler for 70 000 iterations, with a burn-in of 25 000 iterations, after which the samples were collected every tenth iteration. The results are shown in figure 4. Their discussion is delayed to later.

We also considered parameter set 2, in which $q_{12} = 94$, $q_{21} = 50$, $q_{23} = 4$ and $q_{32} = 91$, which Fredkin & Rice (1992*a*) could not resolve satisfactorily. We simulated

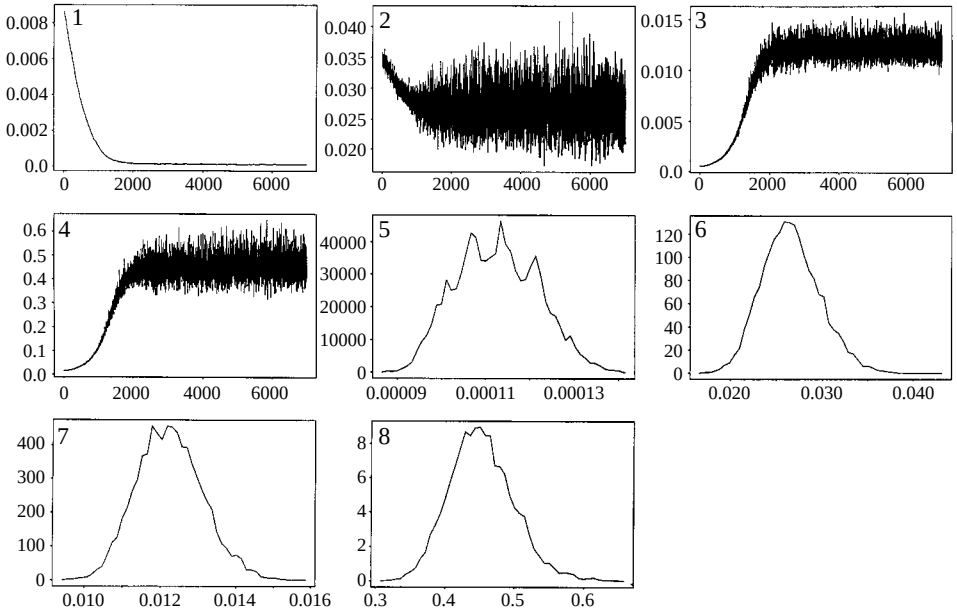

Figure 4. (*b*) Convergence plots and marginal plots for features of a three-state linear model with parameter set 1. Convergence plot of p_{MIS} (1), π_0 (2), $\tilde{\mu}_0$ (3), $\tilde{\mu}_{\mathrm{C}}$ (4). Marginal plot of p_{MIS} (5), π_0 (6), $\tilde{\mu}_0$ (7), $\tilde{\mu}_{\mathrm{C}}$ (8).

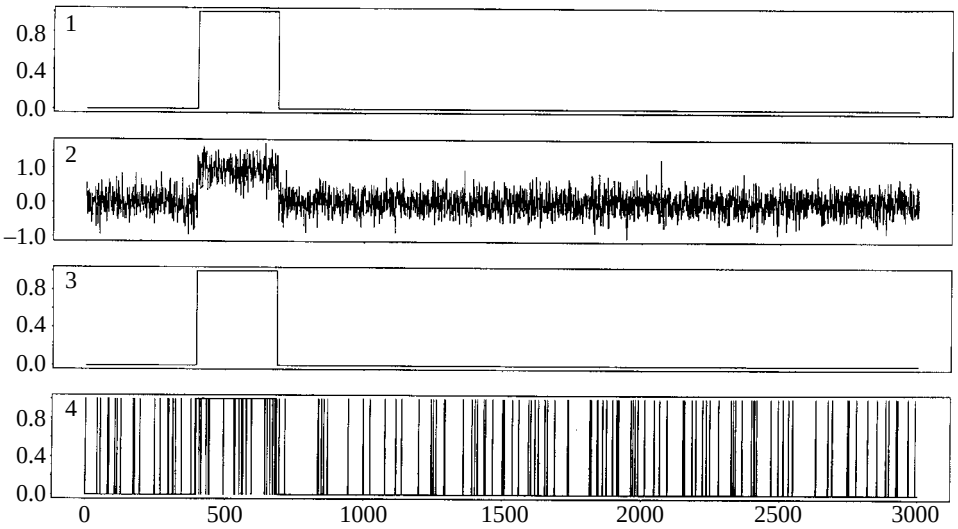

Figure 4. (*c*) Restoration of single-channel record for a three-state linear model with parameter set 1. Original sequence (1), noisy sequence (2), Bayesian reconstruction (3), crude threshold reconstruction (4).

a current record with $N = 3.2 \times 10^5$ and Δ, μ_{O}, σ_{O}, μ_{C}, σ_{C} as before. This yielded a channel record consisting of 1290 open sojourns and 1291 closed sojourns. The initial $x^{(0)}$ and the prior $p(q)$ were determined using the same method as for parameter

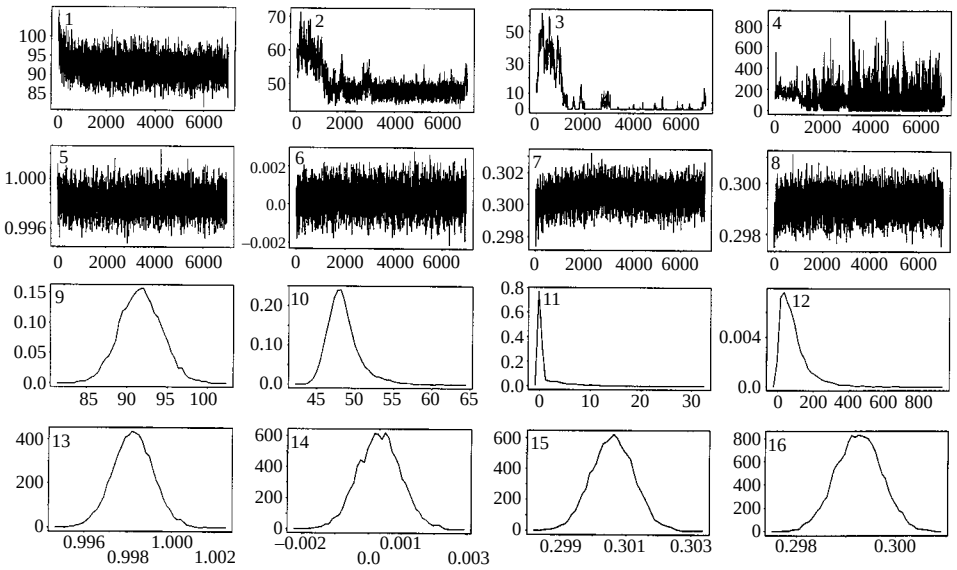

Figure 5. (*a*) Convergence plots and marginal plots for parameters of a three-state linear model with parameter set 2. Convergence plot of q_{12} (1), q_{21} (2), q_{23} (3), q_{32} (4), μ_{O} (5), μ_{C} (6), σ_{O} (7), σ_{C} (8). Marginal plot of q_{12} (9), q_{21} (10), q_{21} (11), q_{32} (12), μ_{O} (13), μ_{C} (14), σ_{O} (15), σ_{C} (16).

set 1. We ran the sampler for 560 000 iterations, with a burn-in of 120 000 iterations, after which the samples were collected every 80th iteration. The results are shown in figure 5.

It is clear from the convergence and marginal plots of figures 4*a* and 5*a* that the MCMC sampler is performing reasonably satisfactorily for both sets of parameter values. However, closer examination of the plots for parameter set 2 shows that, unlike for parameter set 1, the variances of the posterior distributions $p(q_{23} \mid y)$ and $p(q_{32} \mid y)$ are both relatively large. This is related to the aforementioned identifiability problems experienced by Fredkin & Rice (1992*a*). Note that the plots for the MCMC runs give an indication of which parameters are being estimated well and which are not. This feature of the MCMC plots was more marked in other runs that we performed, although the results are not presented here as the sampling interval Δ was unrealistically large.

The plots for π_{O}, $\tilde{\mu}_{\mathrm{O}}$ and $\tilde{\mu}_{\mathrm{C}}$, shown in figure 4*b*, *c*, are all highly satisfactory for both sets of parameter values. For parameter set 1, the estimated means and variances of the marginal posterior distributions were 0.026 60 and 1.0145×10^{-5} for π_{O}, 0.012 25 s and 7.4916×10^{-7} s^2 for $\tilde{\mu}_{\mathrm{O}}$ and 0.4528 s and 2.1612×10^{-3} s^2 for $\tilde{\mu}_{\mathrm{C}}$. The true values of these channel properties were $\pi_{\mathrm{O}} = 0.021\,91$, $\tilde{\mu}_{\mathrm{O}} = 0.010\,64$ s and $\tilde{\mu}_{\mathrm{C}} = 0.4750$ s. For parameter set 2, the estimated means and variances of the marginal posterior distributions were 0.3422 and 8.0023×10^{-5} for π_{O}, 0.01092 s and 9.4449×10^{-8} s^2 for $\tilde{\mu}_{\mathrm{O}}$ and 0.020 99 s and 3.6765×10^{-7} s^2 for $\tilde{\mu}_{\mathrm{C}}$. The true values were $\pi_{\mathrm{O}} = 0.3375$, $\tilde{\mu}_{\mathrm{O}} = 0.010\,64$ s and $\tilde{\mu}_{\mathrm{C}} = 0.020\,88$ s. Note that for both parameter sets, all of the above channel properties are well estimated by our MCMC procedure, in spite of the difficulties associated with estimating some of the q_{ij} for parameter set 2. The estimates of channel properties for parameter set 2 are more

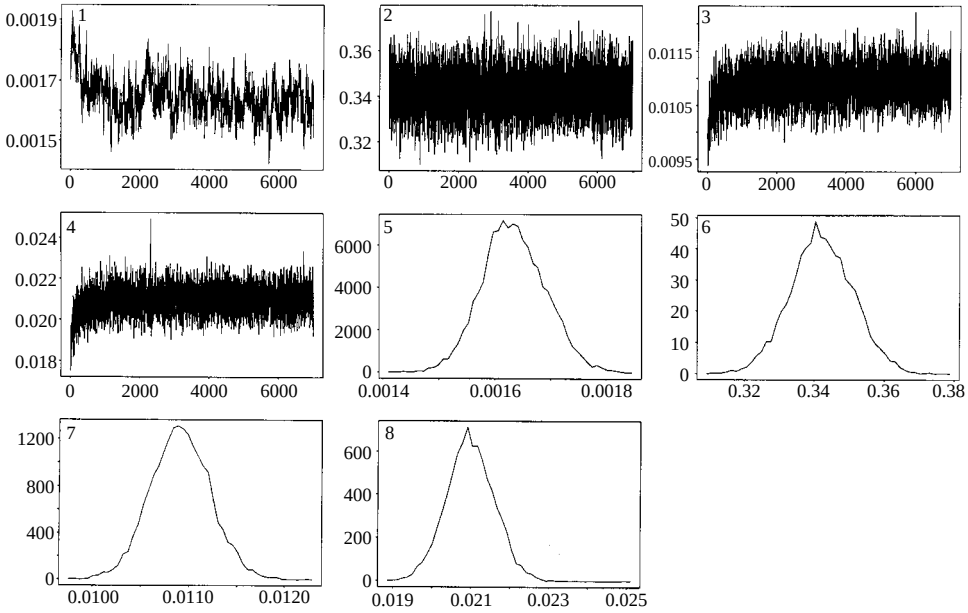

Figure 5. (*b*) Convergence plots and marginal plots for features of a three-state linear model with parameter set 2. Convergence plot of p_{MIS} (1), π_0 (2), $\tilde{\mu}_0$ (3), $\tilde{\mu}_{\mathrm{C}}$ (4). Marginal plot of p_{MIS} (5), π_0 (6), $\tilde{\mu}_0$ (7), $\tilde{\mu}_{\mathrm{C}}$ (8).

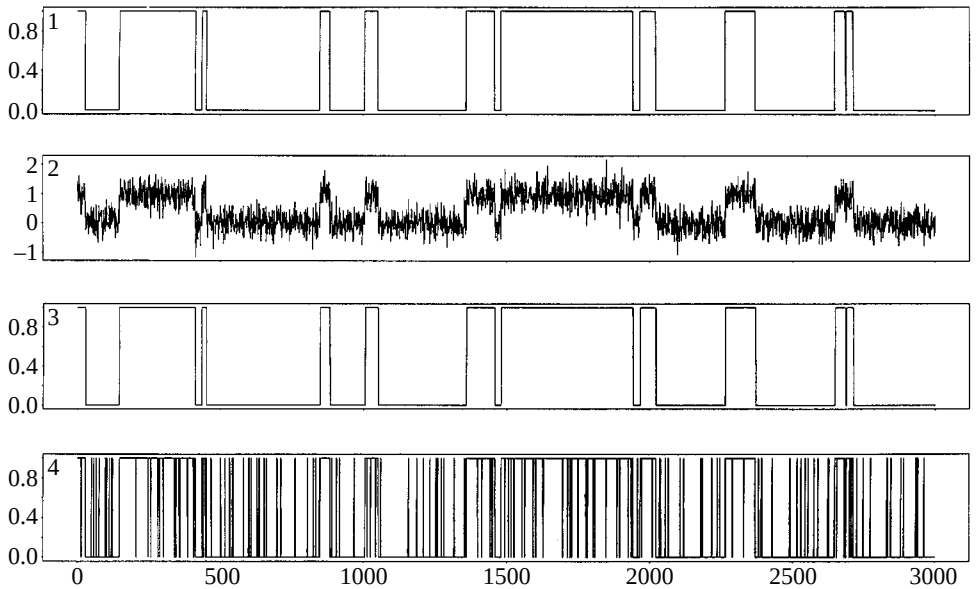

Figure 5. (*c*) Restoration of single-channel record for a three-state linear model with parameter set 1. Original sequence (1), noisy sequence (2), Bayesian reconstruction (3), crude threshold reconstruction (4).

accurate than those for parameter set 1 because the simulated channel record for parameter set 2 contains considerably more open and closed sojourns than that for parameter set 1.

Figures 4*c* and 5*c* show, for the first 3001 sampling points, various reconstructions of the single-channel records for the two sets of parameter values. Again the MCMC marginal Bayes reconstruction performs extremely well. For parameter set 1 (parameter set 2) the proportion of misclassified sampling points in the marginal Bayes reconstruction was 8.5333×10^{-6} (1.0750×10^{-3}), compared with 4.7712×10^{-2} (4.7625×10^{-2}) in the crude threshold reconstruction and an average of 1.1254×10^{-4} (1.6356×10^{-3}) in $x^{(k)}$, as k ranges over the samples collected after burn-in.

For parameter set 1, it took 21 h on a Sun Sparcstation 10 to determine the initial $x^{(0)}$ by fitting a two-state model and a further 26.75 h to then run the MCMC sampler for the three-state model. The corresponding times for parameter set 2 were 9.5 and 42 h, respectively. Although the time to obtain $x^{(0)}$ may seem relatively large, experience shows that a poorly chosen $x^{(0)}$ typically leads to burn-in times far in excess of these figures. Therefore time taken to ensure a more realistic $x^{(0)}$ is a good investment.

(*c*) *Three-state cyclic model*

The three-state model of figure 1*c* is the simplest model containing a cycle of length greater than or equal to 3, so we used it to test the performance of our sample on a model that contains cycles. We took $q_{12}=50$, $q_{13}=90$, $q_{21}=30$, $q_{23}=72$, $q_{31}=60$ and $q_{32}=80$ (again the units are in seconds and are omitted) and simulated a current record with $N=10^5$, $\Delta=10^{-4}$, $\mu_{\mathrm{O}}=1$, $\mu_{\mathrm{C}}=0$ and $\sigma_{\mathrm{O}}=\sigma_{\mathrm{C}}=0.5$. This yielded a channel record consisting of 334 open sojourns and 334 closed sojourns. The initial $x^{(0)}$ was determined as for the three-state linear model, using a prior guess for q obtained by adding six to each of the q_{ij}. The prior $p(q)$ was chosen by setting $\alpha_{ij}=0.01q_{ij}$, $\beta_{ij}=0.01$ ($(i,j)\in E$) and using the method described in § 7 *b* to take account of reversibility. We ran the sampler for 2.1×10^5 iterations, with a burn-in of 3×10^4 iterations, after which the samples were collected every 30th iteration. The results are shown in figure 6.

The convergence and marginal plots for q and θ are shown in figure 6*a*. The plots for q_{23} are not given since reversibility implies that q_{23} is determined by the other five elements of q. The plots for $\theta=(\mu_{\mathrm{O}},\sigma^2_{\mathrm{O}},\mu_{\mathrm{C}},\sigma^2_{\mathrm{C}})$ are generally very satisfactory, though there may be some doubt as to whether the convergence plots for μ_{O} and σ^2_{O} have reached equilibrium. The plots for q are less good, although the marginal posterior plots are broadly centred about the corresponding known true values. However, the convergence plots for q indicate that the time-series $\{q^{(k)}\}$ clearly has long-range dependence and that equilibrium may not yet have been attained. In contrast, the convergence plots for π_{O}, $\tilde{\mu}_{\mathrm{O}}$ and $\tilde{\mu}_{\mathrm{C}}$ shown in figure 6*b* are extremely well behaved and the corresponding marginal plots are nicely centred around their respective 'true values'. The estimated means of the marginal posterior distributions $p(\pi_{\mathrm{O}}\mid y)$, $p(\tilde{\mu}_{\mathrm{O}}\mid y)$ and $p(\tilde{\mu}_{\mathrm{C}}\mid y)$ were 0.2408, 7.2162×10^{-3} s and 2.2796×10^{-2} s, respectively. The true values of these channel properties were $\pi_{\mathrm{O}}=0.2400$, $\tilde{\mu}_{\mathrm{O}}=7.1429\times10^{-3}$ s and $\tilde{\mu}_{\mathrm{C}}=2.2619\times10^{-2}$ s. The MCMC marginal Bayes reconstruction shown in figure 6*c* is again excellent. The proportion of misclassified sampling points in this reconstruction was 4.5230×10^{-3}, compared with 1.5792×10^{-1} in the crude threshold

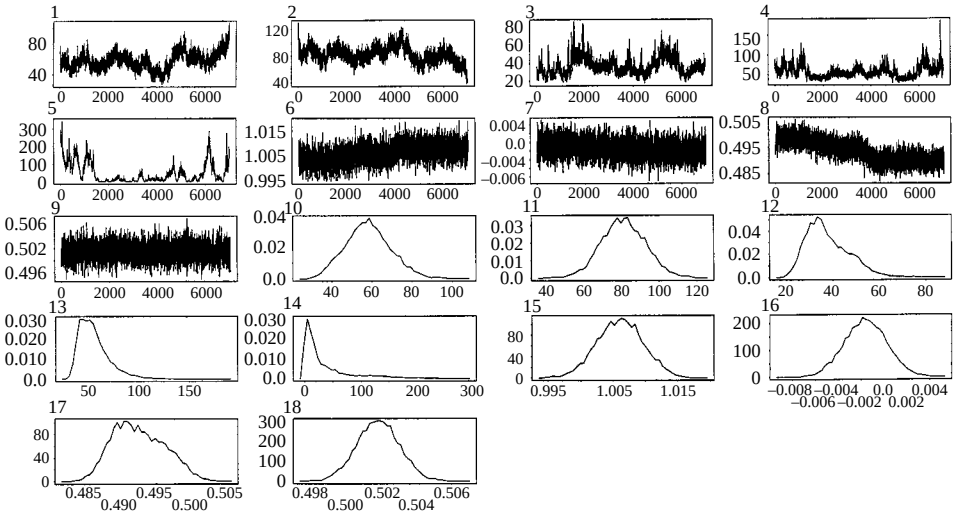

Figure 6. (*a*) Convergence plots and marginal plots for a reversible three-state cyclic model. Convergence plot of q_{12} (1), q_{13} (2), q_{21} (3), q_{31} (4), q_{32} (5), μ_{O} (6), μ_{C} (7), σ_{O} (8), σ_{C} (9). Marginal plot of q_{12} (10), q_{13} (11), q_{21} (12), q_{31} (13), q_{32} (14), μ_{O} (15), μ_{C} (16), σ_{O} (17), σ_{C} (18).

reconstruction and an average of 6.6266×10^{-3} in $x^{(k)}$, as k ranges over the samples collected after burn-in. For this example, it took 4 h on a Sun Sparcstation 10 to determine the initial $x^{(0)}$ by fitting a two-state model, and a further 13 h to then run the MCMC sampler on the full model.

(*d*) *Five-state model*

Our final example is the five-state model of figure 1*d*, which has been discussed previously in § 7 *c*. Using the parametrization described in that section, we took $K_{\mathrm{B}} = 1.3 \times 10^4\ \mathrm{M}^{-1}$, $K_{01} = 0.32$, $K_{02} = 32.0$, $k_{\mathrm{on}} = 10^5\ \mathrm{ms}^{-1}\mathrm{M}^{-1}$, $h_1 = 3.1\ \mathrm{ms}^{-1}$, $h_2 = 31\ \mathrm{ms}^{-1}$ and agonist concentration $a = 10^{-5}\ \mathrm{M}$ (these parameter values are taken from Ball & Sansom (1989)), and simulated a current record with $N = 10^5$, $\Delta = 0.001\ \mathrm{ms}$, $\mu_{\mathrm{O}} = 1$, $\mu_{\mathrm{C}} = 0$ and $\sigma_{\mathrm{O}} = \sigma_{\mathrm{C}} = 0.4$, to yield a channel record consisting of 74 open sojourns and 74 closed sojourns. The initial $x^{(0)}$ was determined as for the three-state models, with a prior guess for ψ given by $K_{\mathrm{B}} = 1.31 \times 10^4$, $K_{01} = 0.5$, $K_{02} = 50.0$, $k_{\mathrm{on}} = 1.0011 \times 10^5$, $h_1 = 4.6$ and $h_2 = 46.0$. An improper uniform prior was chosen for ψ. We ran the sampler for 70 000 iterations, with a burn-in of 10 000 iterations, using the method for updating ψ described in § 7 *c*. The results are shown in figure 7.

The convergence and marginal plots shown in figure 7*a*, *b* are all satisfactory. The convergence plots all appear to have reached their equilibrium and the marginal plots are all roughly centred around their corresponding true values. The estimated means of the marginal posterior distributions $p(\pi_{\mathrm{O}} \mid y)$, $p(\tilde{\mu}_{\mathrm{O}} \mid y)$ and $p(\tilde{\mu}_{\mathrm{C}} \mid y)$ were 0.3582, 0.4465 and 0.8687, respectively. The true values of these channel properties were $\pi_{\mathrm{O}} = 0.3282$, 0.4692 and 0.9601. The MCMC marginal Bayes reconstruction shown in figure 7*c* is again very good. The proportion of misclassified sampling points in this reconstruction was 6.6700×10^{-4}, compared with 1.0639×10^{-1} in the crude

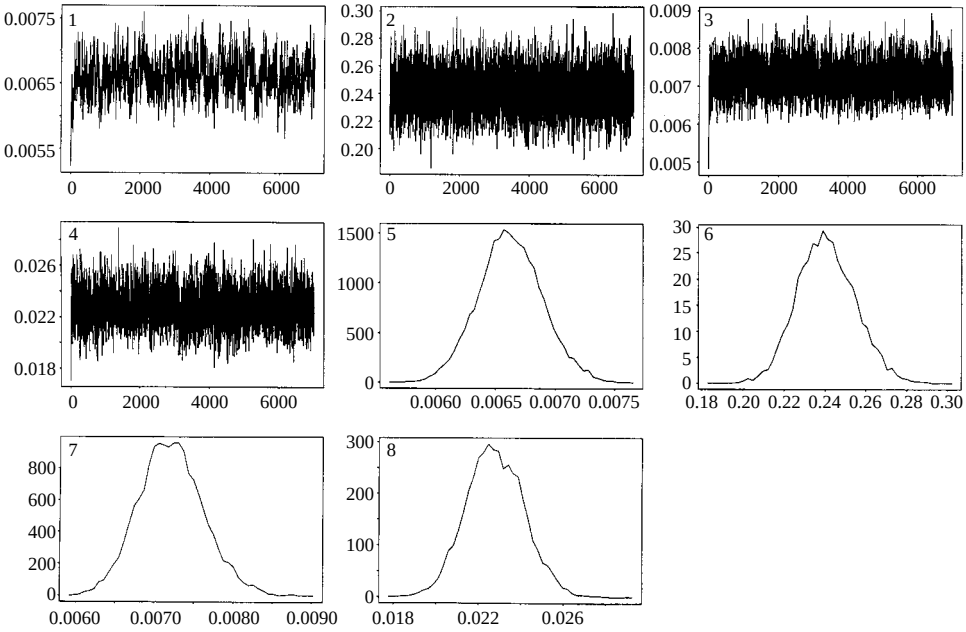

Figure 6. (*b*) Convergence plots and marginal plots for a reversible three-state cyclic model. Convergence plot of p_{MIS} (1), π_0 (2), $\tilde{\mu}_0$ (3), $\tilde{\mu}_{\mathrm{C}}$ (4). Marginal plot of p_{MIS} (5), π_0 (6), $\tilde{\mu}_0$ (7), $\tilde{\mu}_{\mathrm{C}}$ (8).

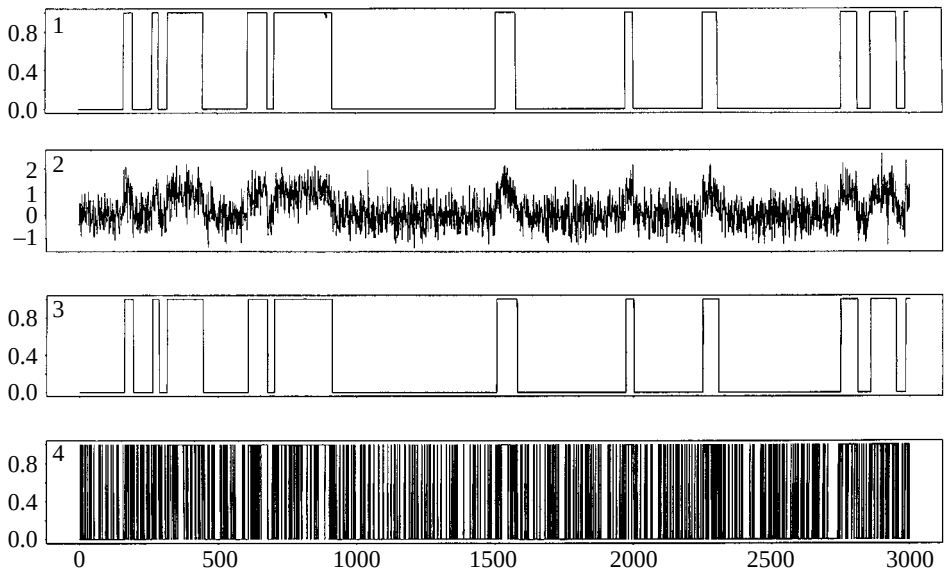

Figure 6. (*c*) Restoration of single-channel record for a reversible three-state cyclic model. Original sequence (1), noisy sequence (2), Bayesian reconstruction (3), crude threshold reconstruction (4).

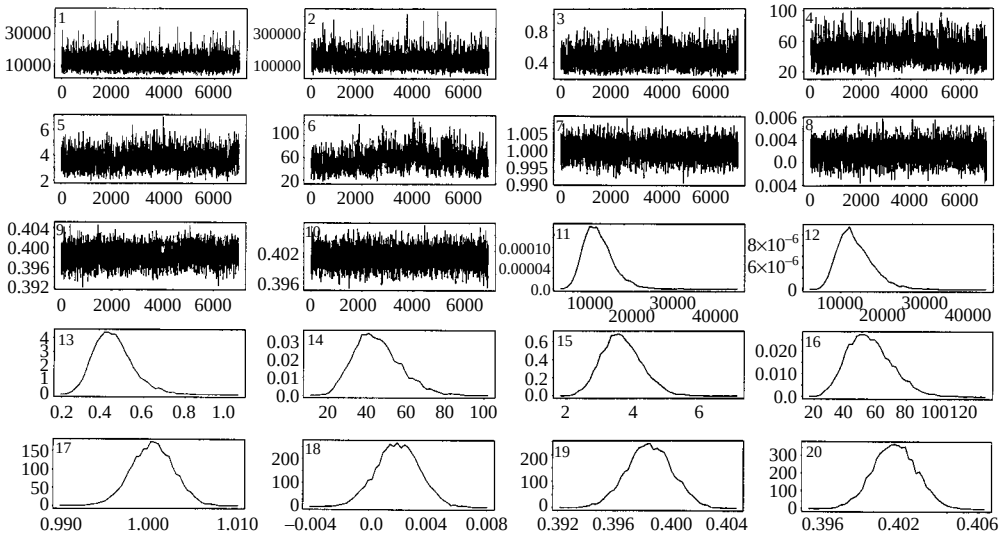

Figure 7. (*a*) Convergence plots and marginal plots for a five-state model. Convergence plot of K_{B} (1), k_{on} (2), K_{01} (3), K_{02} (4), h_1 (5), h_2 (6), μ_{O} (7), μ_{C} (8), σ_{O} (9), σ_{C} (10). Marginal plot of K_{B} (11), k_{on} (12), K_{01} (13), K_{02} (14), h_1 (15), h_2 (16), μ_{O} (17), μ_{C} (18), σ_{O} (19), σ_{C} (20).

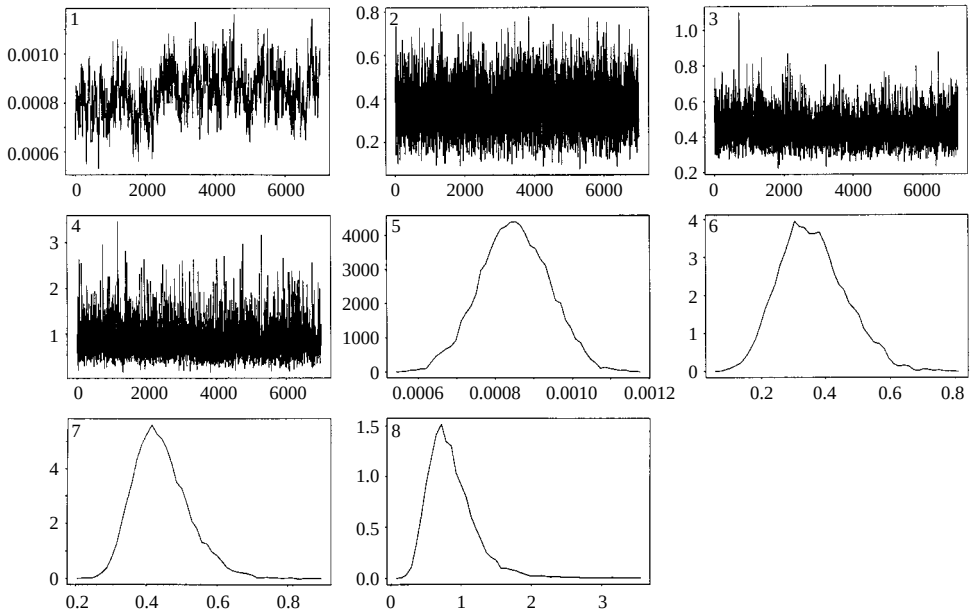

Figure 7. (*b*) Convergence plots and marginal plots for a five-state model. Convergence plot of p_{MIS} (1), π_0 (2), $\tilde{\mu}_0$ (3), $\tilde{\mu}_{\mathrm{C}}$ (4). Marginal plot of p_{MIS} (5), π_0 (6), $\tilde{\mu}_0$ (7), $\tilde{\mu}_{\mathrm{C}}$ (8).

threshold reconstruction and an average of 8.4017×10^{-4} in $x^{(k)}$, as k ranges over the samples collected after burn-in. The running time for this example on a Sun Sparcstation 10 was 8.25 h to determine the initial $x^{(0)}$ by fitting a two-state model, and a further 7 h to run the MCMC sampler on the full model. Admittedly, the value

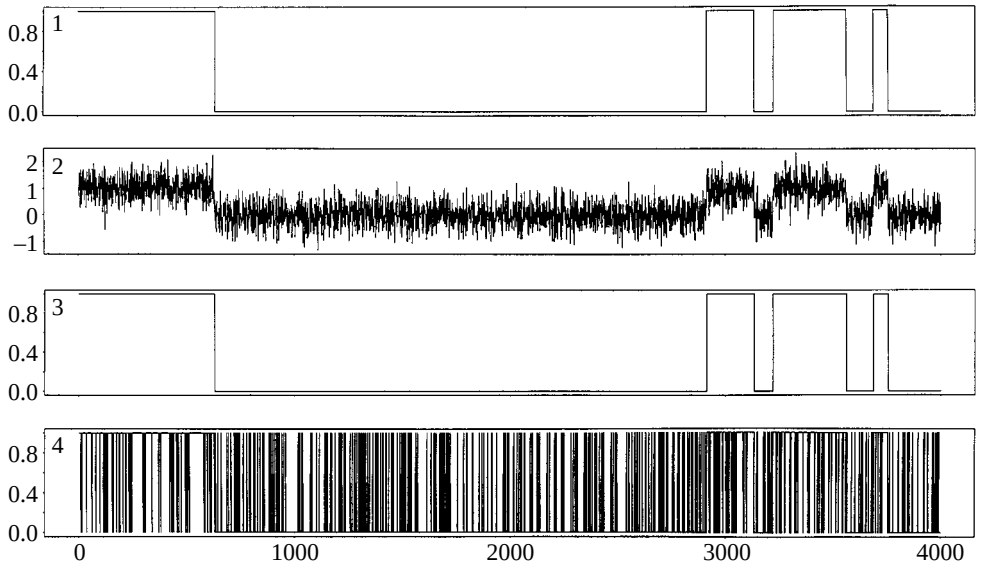

Figure 7. (*c*) Restoration of single-channel record. Original sequence (1), noisy sequence (2), Bayesian reconstruction (3), crude threshold reconstruction (4).

of Δ used in this example is unrealistically small. We plan to explore examples with more realistic values of Δ. However, we are experiencing difficulties with convergence in this example.

9. Concluding comments

We have demonstrated that our MCMC sampler can be successfully used for making inferences directly from single-channel recordings. The sampler generally performed very satisfactorily for estimating the mean and variance of conductance levels corresponding to open and closed sojourns, and also for reconstructing the unobserved single-channel record. Its performance for estimating the transition rate q depends on how identifiable q is from the actual sequence of open and closed sojourns of the channel. However, examination of the convergence and marginal plots for the components of q will often indicate whether identifiability problems are present and, if so, which components of q are most affected.

Identifiability problems are an important aspect of single-channel analysis, although they have not been directly addressed in this paper. Identifiability problems will arise if distinct underlying single-channel gating processes give rise to aggregated processes (i.e. sequences of successive open and closed sojourns) that are probabilistically indistinguishable, or very similar (see Kienker (1989) and Larget (1994) for detailed discussions, and also Fredkin *et al.* (1985), Fredkin & Rice (1992*a*) and Edeson *et al.* (1994)). If an improper uniform prior is assumed for $p(q)$, then unidentifiability will be reflected in the posterior $p(q \mid y)$, in that $p(q \mid y)$ will have two or more peaks of equal, or nearly equal height, or $p(q \mid y)$ will not have a sharp maximum but will have a 'ridge' of q values for which $p(q \mid y)$ is almost constant. As noted above, the MCMC convergence and marginal plots for q may detect such behaviour, though, if $p(q \mid y)$ has several modes, there is a danger that the MCMC

sampler will not visit all of them if the run is not sufficiently long (cf. Ball *et al.* 1996). Whether or not these problems persist for other choices of $p(q)$ depends on the strength of the information in the prior. Time-interval omission can also induce identifiability problems even when there is no such difficulty in the underlying gating process (see, for example, Yeo *et al.* (1988) and Ball *et al.* (1994) for detailed discussion in the context of the two-state model of figure 1*a*). It is not known whether the channel current model, described by (3.1), can give rise to unidentifiability, although an extensive investigation of the corresponding likelihood surface for the two-state model, using the algorithm of Fredkin & Rice (1992*a*), failed to find more than one peak.

The MCMC marginal plots for q and θ give the experimenter an estimate of the posterior uncertainty attached to these parameters. It is only an estimate in the sense that the MCMC plot for q_{ij}, say, approximates the posterior beliefs $p(q_{ij} \mid y)$. If the MCMC sampler is run for sufficiently long then this approximation will become extremely good. This is in sharp contrast to the classical approximate standard errors obtained from the curvature of the likelihood function at its maximum, which are likely to be unreliable in the present ion-channel setting (Fredkin & Rice 1992*a*).

An important issue in single-channel analysis is the identification of the 'best' model from a range of competing models (see, for example, Horn 1987; Ball & Sansom 1989). Given the Bayesian formulation of our method, it is natural to employ Bayesian model choice theory. In principle, we can regard the model identifier simply as another parameter ω, taking possible values $1, 2, \ldots, M$, where M is the number of competing models. However, changing ω entails changes to q and x, since these take model-specific values, and furthermore the dimension of q depends on ω. This is another situation where the reversible jump MCMC method of Green (1995) is appropriate. One simply adds the possibility of ω changes to the Markov chain, and a single MCMC run covers all of the competing models. We intend to explore this approach in future work. In practice, the implementation of such a sampler requires care in formulating the model jumps. The values of q and x in the new model should be such that the jump acceptance probability is not too low. Other difficulties may arise when prior information on the parameters within models is weak: there is work in progress to adapt the fractional Bayes factor of O'Hagan (1995) to MCMC computation, so that such problems may also be overcome.

The basic MCMC sampler can be extended to capture other possible features in ion-channel data. Following Fredkin & Rice (1992*b*), the effects of low-pass filtering can be incorporated by replacing the model (3.1) by

$$Y_l = \sum_{k=0}^{m-1} a_k c(x((l-k)\Delta)) + \varepsilon_l \quad (l = m-1, m, \ldots, N),$$

where the sequence $(a_0, a_1, \ldots, a_{m-1})$ is a discretization of the impulse response of the filter, and thus are known, and $\varepsilon_{m-1}, \varepsilon_m, \ldots, \varepsilon_N$ are independent and identically distributed normal random variables, with mean 0 and variance σ^2. The MCMC sampler is easily adapted to this model, where now $\theta = (\mu_{\mathrm{O}}, \mu_{\mathrm{C}}, \sigma^2)$. The method for updating q is unchanged. The parameter θ can still be updated by single-component Gibbs steps, though of course the details are now different. The proposals for updating x are the same as before, though their acceptance probabilities are different owing to the new model for the recorded current y.

Hodgson (1999) considered an alternative model for the filter, in which the noise is filtered with the signal. He considered a single-channel gating model in which successive open and closed sojourns follow independent gamma distributions. He developed an MCMC sampler for his model, which has many features in common with our sampler. He used a form of simulated tempering to improve the mixing properties of his MCMC sampler. This involved splitting the observation period $[0, T]$ up into a number of shorter intervals and considering a model in which, in our notation, q can vary from one interval to another in a carefully specified fashion. He then only used those realizations of his MCMC in which $q^{(k)}$ is the same for all of the intervals. Provided things are set up correctly this will yield realizations from $p(x, q, \theta \mid y)$. It would be interesting to try a similar approach with our sampler.

The sampler is easily extended to the situation when the channel has more than two conductance levels. Indeed, in some circumstances, such multiconductance-level ion-channel data may prove easier to analyse since the state space of the underlying continuous-time Markov chain is partitioned into more aggregates and consequently more information may be available about the underlying process. The sampler can also be extended in principle to data from multiple ion channels, where the total current flowing across $p > 1$ channels is observed (cf., for example, Yeo *et al*. 1989; Colquhoun & Hawkes 1990; Fredkin & Rice 1991). We simply treat the unobserved realizations of the p-channel gating mechanisms as unknown parameters in the MCMC. However, there are likely to be mixing problems with the ensuing sampler. The sampler can also be generalized to incorporate baseline drift, by allowing the mean open and closed conductance levels, μ_{O} and μ_{C}, to vary slowly through the channel record according to a specified random process. Other artefacts of the recording process, such as breakdown of the electrical seal, present a greater modelling challenge. Finally, although we have described our sampler in the context of equilibrium channel data, its extension to perturbation experiments and models with desensitization is immediate.

A disadvantage of our present MCMC sampler is the large amount of computing time required to fit models, even for records with relatively few openings. The key to the performance of our sampler is the mixing properties of the simulated process $\{x^{(k)}; k = 0, 1, \dots\}$. The burn-in time depends very much on the choice of $x^{(0)}$ and there may be scope for obtaining a good $x^{(0)}$ more cheaply than described, for example by filtering the current record before determining the class A_l of $x^{(0)}(l\Delta)$ $(l = 0, 1, \dots, N)$. There is considerable scope for varying the parameters and move types associated with the updating of x, in order to increase acceptance probabilities and improve the mixing of our sampler, with a consequence that the MCMC sampler would not have to be run for so long in order to obtain adequate results.

In some of the examples, useful information about the unknown parameters has been obtained from very short stretches of data. An important practical question is how long a record does the experimenter need to collect in order to estimate adequately the parameters of the underlying single-channel gating mechanism. This question does not admit a simple answer. In broad terms, in the absence of strong prior information, the standard deviation of the posterior distribution for a given parameter is roughly inversely proportional to the square root of the length of the current record. However, the constant of proportionality varies appreciably from one model to another. For example, in the three-state linear model, the parameters governing the lengths of closed sojourns are fitted better for parameter set 1 than

for parameter set 2, even though the latter was based on about six times as many sojourns. As noted previously, the MCMC convergence and marginal plots provide the experimenter with information about identifiability issues and the precision of parameter estimates, and hence they can be used to give an indication of whether or not more data need to be collected.

In summary, the MCMC methods presented in this paper provide a promising and flexible approach to the analysis of ion-channel data in a wide variety of settings.

This research was supported by the Engineering and Physical Sciences Research Council, under research grant numbers GR/J99780 and GR/K76696.

Appendix A.

In the appendix we describe a method for simulating a realization of a random variable, U say, having probability density function $f(u \mid i, j, v)$ given by (6.8) and (6.10). We suppose that the transition-rate matrix Q is diagonalizable with eigenvalues $\lambda_1 = 0, \lambda_2, \ldots, \lambda_m$ and corresponding linearly independent right eigenvectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m$. This will necessarily be the case if Q is the transition-rate matrix of a time-reversible continuous-time Markov chain, in which case the eigenvalues and right eigenvectors are all real and $\lambda_k < 0$ $(k = 2, 3, \ldots, m)$. Let B be the $m \times m$ matrix whose ith column is $\boldsymbol{b}_i$ $(i = 1, 2, \ldots, m)$ and let $C = B^{-1}$. Further, for $i = 1, 2, \ldots, m$, let $\boldsymbol{c}_i$ denote the ith row of C. Then Q admits the spectral representation

$$Q = \sum_{k=1}^{m} \lambda_k E_k,$$

where $E_k = \boldsymbol{b}_k \boldsymbol{c}_k$ $(k = 1, 2, \ldots, m)$. Moreover, the matrices $E_1, E_2, \ldots, E_m$ satisfy

$$E_i E_j = \begin{cases} E_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases}$$

and

$$E_1 + E_2 + \cdots + E_m = I,$$

where I is the $m \times m$ identity matrix.

It then follows that

$$\begin{aligned} P(v - u) &= \exp((v - u)Q) \\ &= E_1 + \sum_{l=2}^{m} \exp(\lambda_l (v - u)) E_l, \end{aligned} \tag{A 1}$$

thus providing a method of computing $P(v - u)$.

For $i, j \in S$ and $0 < u < v$, let

$$F(u \mid i, j, v) = \int_0^u f(w \mid i, j, v)\, \mathrm{d}w$$

be the distribution function of U. Then from (6.8) and (6.10),

$$F(u \mid i, j, v) = \sum_{k \neq i} F_{ijk}(u \mid v),$$

where

$$F_{ijk}(u \mid v) = \frac{\displaystyle\int_0^u \exp(q_{ii}w) q_{ik} p_{kj}(v-w)\,\mathrm{d}w}{1 - [\exp(q_{ii}-v)/p_{ii}(v)]\delta_{ij}}.$$

Using (A 1),

$$\int_0^u \exp(q_{ii}-w) q_{ik} p_{kj}(v-w)\,\mathrm{d}w = \frac{q_{ik}}{q_{ii}}(E_1)_{kj}(\exp(q_{ii}u)-1) + \sum_{l=2}^{m} q_{ik}(E_l)_{kj}\exp(\lambda_l v)\frac{(\exp\{(q_{ii}-\lambda_l)u\}-1)}{(q_{ii}-\lambda_l)},$$

provided that $\lambda_l \neq q_{ii}$ $(l = 2, 3, \ldots, m)$. (If $\lambda_l = q_{ii}$ then the expression

$$\frac{(\exp\{(q_{ii}-\lambda_l)u\}-1)}{(q_{ii}-\lambda_l)}$$

in the above is replaced by u.) Thus $F_{ijk}(u \mid v)$, and hence $F(u \mid i, j, v)$ can be computed.

We can simulate a realization of the random variable U by simulating a realization, u^* say, from the $U(0,1)$ distribution and then solving $F(u \mid i, j, v) = u^*$ for u. This latter step is easily accomplished numerically, since $F(u \mid i, j, v)$ is strictly increasing with u, $F(0 \mid i, j, v) = 0$ and $F(v \mid i, j, v) = 1$.

References

Ball, F. G. 1997 Empirical clustering of bursts of openings in Markov and semi-Markov models of single channel gating incorporating time interval omission. *Adv. Appl. Prob.* **29**, 909–946.

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

Ball, F. G., Milne, R. K. & Yeo, G. F. 1991 Aggregated semi-Markov processes incorporating time interval omission. *Adv. Appl. Prob.* **23**, 727–797.

Ball, F. G., Chen, A. & Sansom, M. S. P. 1992 Poisson sampling based inference for single ion channel data with time interval omission. *Proc. R. Soc. Lond.* B **250**, 263–269.

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

Ball, F. G., Davies, S. S. & Sansom, M. S. P. 1994 Ion channel gating and time interval omission: statistical inference for a two-state Markov model. *Proc. R. Soc. Lond.* B **255**, 267–272.

Ball, F. G., Cai, Y., Kadane, J. B. & O'Hagan, A. 1996 MCMC methods for discrete sojourn time ion channel data. Nottingham Statistics Group Research Report 96-12.

Baum, L. E., Petrie, T., Soules, G. & Weiss, N. 1970 A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41**, 164–171.

Bernardo, J. M. & Smith, A. F. M. 1994 *Bayesian theory.* Chichester: Wiley.

Besag, J., Green, P., Higdon, D. & Mengersen, K. 1995 Bayesian computation and stochastic systems (with discussion). *Stat. Sci.* **10**, 3–66.

Chay, T. T. 1988 Kinetic modelling for the channel gating process from single channel patch clamp data. *J. Theor. Biol.* **132**, 449–468.

Chung, S. H., Moore, J. B., Xia, L., Premkumar, L. S. & Gage, P. W. 1990 Characterisation of single channel currents using digital signal processing techniques based on hidden Markov models. *Phil. Trans. R. Soc. Lond.* B **329**, 265–285.

Colquhoun, D. & Hawkes, A. G. 1977 Relaxation and fluctuations of membrane currents that flow through drug-operated channels. *Proc. R. Soc. Lond.* B **199**, 231–262.

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

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

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

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

Cowles, M. K. & Carlin, B. P. 1996 Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Statist. Ass.* **91**, 883–904.

Crouzy, S. C. & Sigworth, F. J. 1990 Yet another approach to the dwell-time omission problem of single-channel analysis. *Biophys. Jl* **58**, 731–743.

Devroye, L. 1986 *Non-uniform random variate generation*. New York: Springer.

Edeson, R. O., Ball, F. G., Yeo, G. F., Milne, R. K. & Davies, S. S. 1994 Model properties underlying non-identifiability in single channel inference. *Proc. R. Soc. Lond.* B **255**, 21–29.

Fredkin, D. R., Montal, M. & Rice, J. A. 1985 Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. In *Proc. Berkeley Conf. in Honor of J. Neyman and J. Kiefer* (ed. L. M. LeCam & R. A. Olshen), vol. 1, pp. 269–289. Monterey, CA: Wadsworth.

Fredkin, D. & Rice, J. A. 1991 On the superposition of currents from ion channels. *Phil. Trans. R. Soc. Lond.* B **334**, 347–356.

Fredkin, D. R. & Rice, J. A. 1992*a* Maximum likelihood estimation and identification directly from single channel recordings. *Proc. R. Soc. Lond.* B **249**, 125–132.

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

Gelfand, A. E. & Smith, A. F. M. 1990 Sampling based approaches to calculating marginal densities. *J. Am. Statist. Ass.* **85**, 398–409.

Geman, S. & Geman, D. 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattn Analysis Mach. Intell.* **6**, 721–741.

Geyer, C. J. 1992 Practical Markov chain Monte Carlo (with discussion). *Stat. Sci.* **7**, 473–511.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (eds) 1996 *Markov chain Monte Carlo in practice*. London: Chapman & Hall.

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

Hastings, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Hawkes, A. G., Jalali, A. & Colquhoun, D. 1990 The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected. *Phil. Trans. R. Soc. Lond.* A **332**, 511–538.

Hawkes, A. G., Jalali, A. & Colquhoun, D. 1992 Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil. Trans. R. Soc. Lond.* B **337**, 383–404.

Hodgson, M. E. A. 1999 A Bayesian restoration of an ion channel signal. *J. R. Statist. Soc.* B **61**, 95–114.

Horn, R. 1987 Statistical methods for model discrimination: applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophys. Jl* **51**, 255–263.

Horn, R. & Lange, K. 1983 Estimating kinetic constants from single channel data. *Biophys. Jl* **43**, 207–233.

Jalali, A. & Hawkes, A. G. 1992*a* The distribution of apparent occupancy times in a two-state Markov process in which brief events cannot be detected. *Adv. Appl. Prob.* **24**, 288–301.

Jalali, A. & Hawkes, A. G. 1992*b* Generalised eigenproblems arising in aggregated Markov processes allowing for time interval omission. *Adv. Appl. Prob.* **24**, 302–321.

Kelly, F. P. 1979 *Reversibility and stochastic networks*. Chichester: Wiley.

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

Kijima, S. & Kijima, H. 1987 Statistical analysis of channel current from a membrane patch. I. Some stochastic properties of ion channels or molecular systems in equilibrium. *J. Theor. Biol.* **128**, 423–434.

Larget, B. R. 1994 Equivalence of aggregated Markov processes with application to ion channels. PhD thesis, University of California at Berkeley.

Läuger, P. 1995 Conformation transitions of ionic channels. In *Single-channel recording* (ed. B. Sakmann & E. Neher), 2nd edn, pp. 651–662. New York: Plenum.

Lee, P. M. 1989 *Bayesian statistics: an introduction*. London: Edward Arnold.

Liu, J. S., Wong, W. H. & Kong, A. 1994 Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81**, 27–40.

Magleby, K. L. & Weiss, D. S. 1990*a* Optimating kinetic parameters for single channels with simulation: a general opthod that resolves the missed events problem and accounts for noise. *Biophys. Jl* **58**, 1411–1426.

Magleby, K. L. & Weiss, D. S. 1990*b* Identifying kinetic gating mechanisms for ion channels by using two-dimensional distributions of simulated dwell times. *Proc. R. Soc. Lond.* B **241**, 220–228.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1091.

O'Hagan, A. 1994 *Kendall's advanced theory of statistics: Bayesian inference*, vol. 2B. London: Edward Arnold.

O'Hagan, A. 1995 Fractional Bayes factors for model comparison (with discussion). *J. R. Statist. Soc.* B **57**, 99–138.

Rabiner, L. R. & Juang, B. H. 1986 An introduction to hidden Markov models. *IEEE ASSP Magazine* **3**, 4–16.

Richardson, S. & Green, P. J. 1997 On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. R. Statist. Soc.* B **59**, 731–792.

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

Sakmann, B. & Neher, E. (eds) 1995 *Single-channel recording*, 2nd edn. New York: Plenum.

Silverman, B. W. 1986 *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Tanner, M. & Wong, W. 1987 The calculation of posterior distributions by data augmentation (with discussion). *J. Am. Statist. Ass.* **82**, 528–550.

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

Yeo, G. F., Edeson, R. O., Milne, R. K. & Madsen, B. W. 1989 Superposition of independent ion channels. *Proc. R. Soc. Lond.* B **238**, 155–170.