

Aggregated Markov Processes and Channel Gating Kinetics

Donald R. Fredkin and John A. Rice
University of California, San Diego; La Jolla, CA 92093

Accepted: July 1, 1985

A finite state Markov process is aggregated into several groups. Rather than observing the underlying Markov process, one is only able to observe the aggregated process. What can be learned about the underlying process from the aggregated one? Such questions arise in the study of gating mechanisms in ion channels in muscle and nerve cell membranes. We discuss some recent results and their implications.

Key words: aggregated Markov process; channel gating kinetics.

1. Introduction

We are concerned with the following mathematical problem: a finite state Markov process in continuous time is aggregated, by which we mean that its states are grouped into a smaller number of aggregates. The Markov process is assumed to be in equilibrium. One is not able to observe the process itself, but only what aggregate the process is in as time goes along. From the aggregated process we wish to draw inferences about the underlying Markov process. How much can we learn? This is a rather general question and more sharply defined questions can be asked. For example, to what extent is the graph structure of the Markov process identifiable? Are some aspects of it identifiable and some not? (By the graph we mean a diagram showing the states and the interconnections between them, but not the numerical values of the transition rates.) Can some graph structures be ruled out as incompatible with the aggregated process? If

the graph structure is known or hypothesized *a priori*, what functionals of the various rate constants can be identified?

These are questions of identifiability. There are also problems related to the efficient statistical use of observation of the aggregated process over a finite time interval.

Problems such as these arise in modeling and data analysis for biophysical studies of gating mechanisms in ion channels in muscle and nerve cell membranes. The next section of this paper briefly describes the biophysical context. In the third section we summarize the mathematical results that we have been able to obtain and discuss their implications and possible applications. In the fourth section we make some concluding remarks and mention several open questions.

2. Biophysical Background

Ion channels are transmembrane proteins with the ability to open a pore through which ions can flow with high conductance; in the absence of such pores, the lipid bilayer membranes of cells are virtually impermeable to most charged particles. Most channels are either voltage "gated" (controlled), such as the sodium channel in nerve axons,

About the Authors: Donald R. Fredkin is with the Department of Physics and John A. Rice with the Department of Mathematics at the University of California, San Diego : La Jolla.

which is the fundamental non-linear circuit element involved in the propagation of nerve impulses, or chemically gated, such as the post-synaptic acetylcholine receptor. We have already reviewed the biophysical context [1]¹ and a broad review of ion channels is now available [2]. We shall therefore confine ourselves here to an illustrative example.

The present work started with the desire to extract information from experiments on the chemically gated acetylcholine receptor. Chemically extracted protein is incorporated in an artificial lipid bilayer membrane separating two compartments containing electrolyte solution. The voltage difference between the two sides of the membrane is fixed, and the current through the membrane is measured. In the presence of an agonist (chemical stimulant, such as acetylcholine), the current is found to fluctuate randomly between two levels, reflecting the open or closed state of the channel. (Great pain is taken to arrange to have only one active channel, as shown by the absence of time intervals with current a multiple of the minimum quantum.) It is known that the agonist must bind to the channel to permit opening, and the time scale for channel opening and closing is much shorter than the time scale for agonist binding, as shown by chemical kinetics studies, so the simplest kinetics would be a Markov process with three states: C_1 has no bound agonist and the channel is closed, C_2 has bound agonist and the channel is still closed, and O has bound agonist and the channel is open, and transitions $C_1 \leftrightarrow C_2 \leftrightarrow O$. The transition $C_2 \leftrightarrow O$ is visible in the experiment as a jump in electric current through the membrane. The transition $C_1 \leftrightarrow C_2$, which involves the binding or dissociation of agonist, is invisible because it does not involve a change in channel conductance. Under these circumstances, we have the simplest example of an aggregated Markov process, with aggregates $\{C_1, C_2\}$ and $\{O\}$. We would like to see if the scheme $C_1 \leftrightarrow C_2 \leftrightarrow O$ is consistent with the data, and, if it is, we would like to estimate the four transition rates in the scheme from the data. In this particular case, it is easy to see, using the results described in the next section, that the transition rates can be estimated from the data, and in fact it is sufficient to use the one dimensional densities for this purpose.

The model we have just described is radically oversimplified and inconsistent with experiment. One feature which is accessible via these experiments but not via agonist binding studies is that the one dimensional density for the channel open "state" (actually, aggregate) is the sum of at least two exponentials [3]. According to the next section, this demonstrates the existence of at least two open states O_1 and O_2 . The aggregates are now $\{C_1, C_2\}$ and $\{O_1, O_2\}$. We would like to accept or reject schemes like $C_1 \leftrightarrow C_2 \leftrightarrow O_1 \leftrightarrow O_2$, $\{C_1 \leftrightarrow C_2 \leftrightarrow O_1, C_2 \leftrightarrow O_2\}$, and $C_1 \leftrightarrow C_2 \leftrightarrow O_1 \leftrightarrow O_2 \leftrightarrow C_2$. The theorems of the next section imply immediately that the one dimensional densities contain all the information available if

any of these schemes is correct, that the third scheme, which contains a cycle, is not identifiable, and that any of these schemes can be excluded if correlation is observed between two consecutive durations of channel opening, which it is [4,5].

3. Results

We first introduce some notation. $P(t)$ will denote the transition matrix of the Markov process. We will assume throughout that the process is in equilibrium. As is standard, we let

$$Q = \lim_{t \rightarrow 0} \frac{P(t) - I}{t}, \text{ where } I \text{ is the identity matrix.}$$

The aggregates will be indexed by lower case Greek letters; n_α is the number of states in aggregate α . We order the states so that states in the same aggregate are contiguous, and we partition the matrix Q into sub-matrices $Q_{\alpha\beta}$. We will assume throughout that the submatrices $Q_{\alpha\alpha}$ are diagonalizable, which holds if the law of detailed balance is valid for the system.

Supposing that the process enters aggregate α at time $t=0$, we denote the probability density of the length of time T spent in that aggregate before leaving by $f_\alpha(t)$. It is shown in [6] and [1] that

$$f_\alpha(t) = \pi_\alpha e^{Q_{\alpha\alpha} t} q_\alpha$$

where $q_\alpha = \sum_{\beta \neq \alpha} Q_{\alpha\beta} u_\beta$, u_β is a column vector of n_β ones, and π_α is a row vector giving the probabilities that aggregate α is entered via each of its states.

Under the assumption that $Q_{\alpha\alpha}$ is diagonalizable this one dimensional density can be re-expressed as:

$$f_\alpha(t) = \sum_{i=1}^{n_\alpha} a_\alpha^i e^{-\lambda_\alpha^i t}.$$

An implication of this result is that a lower bound on n_α can be obtained by counting the number of exponential components. This result has been widely used in channel gating studies by fitting experimental results to sums of exponentials and judging how many exponentials to include by a chi-squared statistic.

Two dimensional densities can also be used to obtain information about the Markov process. It can be shown that the density of spending a length of time s in aggregate α and then a length of time t in aggregate β is

$$f_{\alpha\beta}(s, t) = \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} a_{\alpha\beta}^{ij} e^{-\lambda_\alpha^i s - \lambda_\beta^j t}.$$

(Here we use i and j as indices, not powers.) It is noteworthy

¹Numbers in brackets indicate literature references.

that the same exponential parameters occur in both the one and two dimensional densities. This might be used to judge the plausibility of an underlying Markov process model.

The matrix of linear coefficients with ij entry $a_{\alpha\beta}^{ij}$ yields information about how the aggregates are interconnected:

Theorem A. [7] *Let $A_{\alpha\beta} = [a_{\alpha\beta}^{ij}]$ be the matrix of coefficients of a two dimensional density above. Then the rank of $A_{\alpha\beta}$ is less than or equal to the rank of $Q_{\alpha\beta} = P_{\alpha\beta}$, say, and $A_{\alpha\beta}$ depends on at most $p_{\alpha\beta}(n_{\alpha} + n_{\beta} - p_{\alpha\beta})$ parameters. The rank of $Q_{\alpha\beta}$ is less than or equal to the smaller of: the number of states in α which are linked directly to states in β , and the number of states in β which are linked directly to states in α . It may be possible to empirically obtain a lower bound on the complexity of interconnection by fitting two dimensional densities and using chi-square tests. We hope to implement and test such a procedure in the future.*

Higher dimensional densities can be considered also. The following theorem shows that under certain conditions no more information about the process can be obtained from these densities, however.

Theorem B. [7] *If, for each aggregate α , the eigenvalues λ_{α}^i are distinct and $\alpha_i \neq 0$ for all i , the higher dimensional densities $f_{\alpha_0, \dots, \alpha_r}(t_0, \dots, t_r)$, $r > 1$ are completely determined by the two-dimensional densities.*

Thus, in principle, all the available information about the underlying process can be extracted from the two dimensional densities if the hypotheses of the theorem are satisfied. By counting the number of independent parameters involved in the two dimensional densities, we have the following theorem:

Theorem C. [7] *Under the assumptions of Theorem B, the finite dimensional distributions depend on at most $\sum_{\alpha \neq \beta} p_{\alpha\beta}(n_{\alpha} + n_{\beta} - p_{\alpha\beta})$ parameters.*

Thus for example, if there are two aggregates, there is information on at most $2p(n_{\alpha} + n_{\beta} - p)$ parameters. If a model depends on more than this many parameters, its parameters are not uniquely identifiable.

Theorem A above suggests one way to study the complexity of interconnection between two aggregates. Labarca et al. [5] have also used certain correlation functions for this purpose. For a particular aggregate α , say, the sequence of dwell times in that aggregate, T_1, T_2, \dots is a stationary process, and it can be shown that the covariance function of that process is of the form given in the following theorem:

Theorem D. [1] *The covariance function is of the form*

$$\Gamma_{\alpha}(k) = \sum_{i=1}^{M-1} u_i \kappa_i^{|k|}$$

where $0 \leq \kappa_i < 1$ and $k \neq 0$ and where M is the rank of the matrix Q_{α} which is composed of the off-diagonal blocks corresponding to aggregate α in the matrix Q . If $M=1$, $\Gamma(k)=0$ for $k \neq 0$.

The rank of Q_{α} is less than or equal to the smaller of: the number of states in α which are linked directly to states in any other aggregate, and the number of states in other aggregates which are linked directly to states in α .

Correlation functions can thus be used to obtain information similar to that in the two dimensional densities as in Theorem A. In the case that there are more than two aggregates, the two dimensional densities contain finer information, however, since a lower bound on the rank of each of the matrices $Q_{\alpha\beta}$ which constitute Q_{α} can be obtained.

It is interesting that the covariance function of Theorem D is of the form of the covariance function of a moving average-autoregressive process, although the stationary process is distinctly non-Gaussian. It may well be that techniques developed for order estimation in the time series literature can be used to estimate M . Labarca et al. [5] were primarily interested in testing whether M was greater than 1, which is relatively simple since the large sample distribution of the empirical correlation coefficients in the case $M=1$ can be used.

4. Further Considerations

We believe that although the results summarized in the previous section are useful, there are still many open and interesting questions.

We have developed some necessary conditions for identifiability which can sometimes be used to conclude that hypothetical models are unidentifiable. It would be useful to have checkable sufficient conditions as well.

Our analysis applies to stationary Markov processes. The nonstationary case is of both theoretical and practical interest, and we hope to consider that situation in the future. Records from the sodium channel are typically nonstationary because of the presence of an absorbing inactivated state.

We have only begun to explore practical consequences of our results for the analysis of experimental data [3]. It is tempting to consider estimating the two dimensional distributions and basing further analysis on them. Horn and Lange [8] have proposed likelihood analysis of sample paths and Horn and Vandenburg [9] have applied these techniques to data from the sodium channel. Advantages of their approach are that it is applicable to nonstationary data and multi-channel data. The likelihood method is computationally intensive, however; Horn uses an array processor on a VAX 11/730 and reports that days of computer time are necessary. An analysis based on the two dimensional distributions would be much faster; it is not clear what loss of statistical efficiency would be incurred.

Finally, it would be desirable to develop more data-analytic and model-free methods for analyzing experimental records to give qualitative insights that might suggest phys-

ical mechanisms. Labarca et al. [3] have used box-plots to advantage in analyzing data from the chloride channel.

5. References

- [1] Fredkin, D.R.; M. Montal and J. Rice, Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Ed. L. Le Cam and R. Olshen, Vol. I, pp 269-290, Wadsworth Publishing Co., Belmont, CA (1985).
- [2] Hille, B., *Ionic Channels of Excitable Membranes*, Sinauer Associates, Inc. (Sunderland, MA) (1984).
- [3] Labarca, P.; J. Rice, D.R. Fredkin and M. Montal, Kinetic analysis of channel gating: application to the cholinergic receptor channel and the chloride channel from *Torpedo californica*. *Biophys. J.* **47** 469-478. (1985).
- [4] Jackson, M.B.; B.S. Wong, C.E. Morris, H. Lecar and C.N. Christian. Successive openings of the same acetylcholine receptor channel are correlated in their open times. *Biophys. J.* **42** 109-114. (1983).
- [5] Labarca, P.; J. Lindstrom and M. Montal, The acetylcholine receptor channel from *Torpedo californica* has two open states. *J. Neurosci.* **4** 473-496. (1984).
- [6] Colquhoun, D., and A.G. Hawkes, On the stochastic properties of single ion channels. *Proc. R. Soc. London B* **211** 205-235 (1981).
- [7] Fredkin, D.R., and J. Rice, On aggregated Markov processes. To appear in *J. Appl. Prob.* (1984).
- [8] Horn, R. and K. Lang, Estimating kinetic constants from single channel data. *Biophys. J.* **43** 207-223 (1983).
- [9] Horn, R. and Vandenberg, C., Statistical properties of single sodium channels. *J. Gen., Physiol.* **84** 505-533 (1984).