

单离子通道的隐马氏模型与状态的还原*

黎培兴¹, 李彩霞², 方积乾², 关永源³

(1. 中山大学数学系, 广东 广州 510275;

2. 中山大学公共卫生学院, 广东 广州 510080;

3. 中山大学药理教研室, 广东 广州 510080)

摘要: 对细胞膜单离子通道建立隐马氏模型, 其中通道潜在的开关状态序列 $\{X_t\}$ 为马氏链, 而膜片钳记录数据 $\{Y_t\}$ 为依赖通道状态的正态分布。用 EM 算法估计模型中的转移概率以及正态分布的均值与方差等参数, 并利用隐马氏模型的性质, 用观察到的 $\{Y_t\}$ 来还原通道潜在的开关状态 $\{X_t\}$ 。

关键词: 离子通道; 隐马氏模型; EM 算法; 还原

中图分类号: Q332 **文献标识码:** A **文章编号:** 0529-6579 (2003) 02-0009-03

一个隐马氏过程 $\{X_t, Y_t\}$ 包含两部分: 一个潜在的、不可观察的(故称为隐的)有限状态马氏链 $\{X_t\}$ 和一个外显的、可观察的随机过程 $\{Y_t\}$, $\{Y_t\}$ 的分布依赖 $\{X_t\}$, 我们常通过可观察的过程 $\{Y_t\}$ 探索 $\{X_t\}$ 的统计规律。

离子通道是神经元、肌细胞等兴奋细胞组织上的特殊大分子蛋白质, 在某些构象中这些大分子形成具有选择性的孔洞, 处于开放状态, 允许一种或数种离子通过, 形成离子电流, 当构象状态不形成孔洞时, 通道处于关闭状态。因此, 离子通道具有时而开放、时而关闭的动态特性。典型的单通道电流在两个电导水平(分别对应开与关)之间随机交替变化, 它是振幅恒定, 而持续时间不定的脉冲变化结果, 而多通道电流则在好几个电导水平间交替变化。利用膜片钳技术可记录到各类离子通道活动的电信号, 但由于噪声的影响, 通道的开关状态序列 $\{X_t\}$ 不能由膜片钳记录数据 $\{Y_t\}$ 直接得到。PCLAMP 软件是用阈值法来判断, 还原通道的开关状态, 即设定一个阈值 c , 当 $Y_t < c$ 时认为通道是开的, 否则是关的。但是, 此阈值的确定带有很大的主观性, 因此也有人^[1,2]提出用 Bayes 法来还原通道的开关状态。文[3]对多通道建立隐马氏模型, 探讨通道间的交互作用。本文利用隐马氏模型的理论, 用 EM 算法估计参数, 同时提出一种新的方法还原通道的开关状态。

1 建模

现在考虑具有两个电导水平(分别对应开和关)的单离子通道。设

$$X_t = \begin{cases} 0 & \text{通道在 } t \text{ 时刻是关的} \\ 1 & \text{通道在 } t \text{ 时刻是开的} \end{cases}$$

对 $\{X_t, Y_t\}$ 建立隐马氏模型, 设 $\{X_t\}$ 的一步转移概率矩阵为

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} p_0 & 1 - p_0 \\ 1 - p_1 & p_1 \end{pmatrix}$$

并假定当 $X_t = 0$ 时, Y_t 的条件分布为正态分布 $N(\mu_0, \sigma_0^2)$, 其概率密度记为 $f_0(Y_t)$; 当 $X_t = 1$ 时, Y_t 的条件分布为正态分布 $N(\mu_1, \sigma_1^2)$, 其概率密度记为 $f_1(Y_t)$ 。给定 $\{X_t\}$, Y_t 之间是条件独立的, Y_t 的条件分布只依赖 X_t 。令 $\theta = (p_0, p_1, \mu_0, \mu_1, \sigma_0, \sigma_1)$, 则似然函数为:

$$\begin{aligned} L(\theta) &= P(Y_0, Y_1, \dots, Y_n | \theta) = \\ &= \sum_{x_0, x_1, \dots, x_n} P(Y_0, Y_1, \dots, Y_n | x_0, x_1, \dots, x_n, \theta) \cdot \\ &= P(x_0, x_1, \dots, x_n | \theta) = \\ &= \sum_{x_0, x_1, \dots, x_n} P(x_0) \prod_{i=1}^n P(x_i | x_{i-1}, \theta) P(Y_i | x_i, \theta) \end{aligned}$$

2 参数估计

文[3]利用 Baum^[4]中的算法估计模型参数。下面把 $\{X_t\}$ 看成缺失数据, 用 EM 算法^[5]求隐马氏

* 收稿日期: 2002-07-23

基金项目: 广东省自然科学基金团队资助项目; 中山大学青年教师启动基金资助项目

作者简介: 黎培兴(1971年生), 男, 博士, 讲师; E-mail: lnspx@zsu.edu.cn

链的参数的极大似然估计。令 $\theta_x = (p_0, p_1)$, $\theta_y = (\mu_0, \sigma_0, \mu_1, \sigma_1)$, 则 $\theta = (\theta_x, \theta_y)$ 。注意到 $L(\theta)$ 中, $P(X_i | X_{i-1}, \theta)$ 的计算只与参数 θ_x 有关, $P(Y_i | X_i, \theta)$ 的计算只与参数 θ_y 有关, 因此隐马氏链的 EM 算法为:

E 步: 分别将 $\sum_{i=1}^n \ln P(Y_i | X_i, \theta_y)$ 与 $\sum_{i=1}^n \ln P(X_i | X_{i-1}, \theta_x)$ 求关于 $X | Y, \theta^{(i)}$ 的条件期望。

$$\begin{aligned} Q_1(\theta_y | \theta^{(i)}) &= \\ \sum_{i=1}^n \sum_{j=0}^1 v_i(j, \theta^{(i)}) \ln P(Y_i | X_i = j, \theta_y) &= \\ \sum_{i=1}^n \sum_{j=0}^1 v_i(j, \theta^{(i)}) \ln f(Y_i, \theta_y), \\ Q_2(\theta_x | \theta^{(i)}) &= \\ \sum_{i=1}^n \sum_{j=0}^1 \sum_{k=0}^1 w_i(j, k, \theta^{(i)}) \cdot \\ \ln P(X_i = k | X_{i-1} = j, \theta_x) &= \\ \sum_{i=1}^n \sum_{j=0}^1 \sum_{k=0}^1 w_i(j, k, \theta^{(i)}) \ln p_{jk} \end{aligned}$$

其中

$$\begin{aligned} v_i(j, \theta^{(i)}) &= P(X_i = j | Y_0, Y_1, \dots, Y_n, \theta^{(i)}), \\ w_i(j, k, \theta^{(i)}) &= \end{aligned}$$

$$P(X_{i-1} = j, X_i = k | Y_0, Y_1, \dots, Y_n, \theta^{(i)})$$

关于 $v_i(j, \theta)$ 与 $w_i(j, k, \theta)$ 可由 Baum^[4] 介绍的向前概率 $\alpha_i(j, \theta)$ 与向后概率 $\beta_i(j, \theta)$ 计算得到。

$$\alpha_i(j, \theta) \triangleq P(Y_0, \dots, Y_i, X_i = j) =$$

$$f_j(Y_i) \sum_{i=0}^1 \alpha_{i-1}(i, \theta) p_{ij},$$

$$\beta_i(j, \theta) \triangleq P(Y_{i+1}, \dots, Y_n | X_i = j) =$$

$$\sum_{i=0}^1 f_i(Y_{i+1}) \beta_{i+1}(i, \theta) p_{ji}$$

M 步: 将 $Q_1(\theta_y | \theta^{(i)})$ 与 $Q_2(\theta_x | \theta^{(i)})$ 极大化, 即找一个 $\theta^{(i+1)}$ 使

$$Q_1(\theta_y^{(i+1)} | \theta^{(i)}) = \max Q_1(\theta_y | \theta^{(i)})$$

$$Q_2(\theta_x^{(i+1)} | \theta^{(i)}) = \max Q_2(\theta_x | \theta^{(i)})$$

可算得

$$\begin{aligned} \mu_j^{(i+1)} &= \left[\sum_{i=1}^n v_i(j, \theta^{(i)}) Y_i \right] / \left[\sum_{i=1}^n v_i(j, \theta^{(i)}) \right] = \\ &= \frac{\sum_{i=1}^n \alpha_i(j, \theta^{(i)}) \beta_i(j, \theta^{(i)}) Y_i}{\sum_{i=1}^n \alpha_i(j, \theta^{(i)}) \beta_i(j, \theta^{(i)})}, \\ (\sigma_j^2)^{(i+1)} &= \frac{\sum_{i=1}^n v_i(j, \theta^{(i)}) (Y_i - \mu_j^{(i+1)})^2}{\sum_{i=1}^n v_i(j, \theta^{(i)})} = \end{aligned}$$

$$\begin{aligned} &= \frac{\sum_{i=1}^n \alpha_i(j, \theta^{(i)}) \beta_i(j, \theta^{(i)}) (Y_i - \mu_j^{(i+1)})^2}{\sum_{i=1}^n \alpha_i(j, \theta^{(i)}) \beta_i(j, \theta^{(i)})}, \\ p_j^{(i+1)} &= \frac{\sum_{i=1}^n w_i(j, \theta^{(i)}) Y_i}{\sum_{i=1}^n v_i(j, \theta^{(i)})} = \\ &= \frac{\sum_{i=1}^n \alpha_{i-1}(j, \theta^{(i)}) p_j^{(i)} f_j^{(i)}(Y_i) \beta_i(j, \theta^{(i)})}{\sum_{i=1}^n \alpha_i(j, \theta^{(i)}) \beta_i(j, \theta^{(i)})} \\ & \quad j = 0, 1 \end{aligned}$$

3 开关状态的还原

利用隐马氏链的性质, 可以通过膜片钳记录数据来推断通道的状态, 还原标准: 使 $P(X_i | Y_i, \dots, Y_n)$ 达到再大的 X_i 值即为通道在 i 时刻的状态。这等价于计算通道是开的条件优势

$$\begin{aligned} odds(t) &= \frac{P(X_t = 1 | Y_1, \dots, Y_n)}{P(X_t = 0 | Y_1, \dots, Y_n)} = \\ \frac{v_t(1, \theta)}{v_t(0, \theta)} &= \frac{\alpha_t(1, \theta^{(i)}) \beta_t(1, \theta^{(i)})}{\alpha_t(0, \theta^{(i)}) \beta_t(0, \theta^{(i)})} \end{aligned}$$

当 $odds(t) > 1$ 时, 则认为通道在 t 时刻是开的, 否则是关的。

4 实例

单离子通道的膜片钳记录数据由中山大学药理学教研室提供, 其中一段电信号图形见图 1:



图 1 单离子通道的膜片钳记录数据

Fig.1 Part of single-channel patch recordings

对参数估计与状态的还原过程, 笔者编制了简单的 Fortran 程序, 程序分析结果:

$$\theta_x = (p_0, p_1) = (0.906, 0.902)$$

$$\theta_y = (\mu_0, \sigma_0, \mu_1, \sigma_1) =$$

$$(0.412, 0.247, 5.344, 1.032)$$

图 1 的状态还原如图 2。



图 2 图 1 中信号经隐马氏法的还原结果

Fig.2 Restoration of the recordings in Fig.1 by Hidden Markov method

PCLAMP 软件的阈值法: 首先在电信号图形上凭直观划出 0 状态水平线 (基线) 与 1 状态水平线, 由此得到阈值 c (两水平线的中间位置), 然后根据 Y_i 与阈值的关系判断通道状态。还原结果见图 3。



图 3 图 1 中信号经 PCLAMP 软件阈值法的还原结果

Fig. 3 Restoration of the recordings

in Fig. 1 by PCLAMP

本文对具有 2 个电导水平的单离子通道建立了隐马氏模型, 用 EM 算法估计参数, 同时提出了还原通道状态的一个新方法, 克服阈值法的主观性的缺陷。实际上, 本文所提出的状态还原方法可以类似推广至具有多个电导水平的离子通道。

参考文献:

- [1] FREDKIN D R, RICE J A. Bayesian restoration of single-channel patch clamp recordings[J]. Biometrics, 1992, 48: 427 - 448.
- [2] HODGSON M E. A Bayesian restoration of an ion channel single[J]. J R Statist Soc(B), 1999, 61(1):95 - 114.
- [3] KLEIN S, TIMMER J, HONERKAMP J. Analysis of multichannel patch clamp recordings by hidden Markov models [J]. Biometrics, 1997, 53: 870 - 884.
- [4] BAUM L E, PETRIE T, SOULES G, et al. A maximization technique occurring in the statistic analysis of probabilistic functions of Markov chains[J]. Ann Math Statist, 1970, 41: 164 - 171.
- [5] 茆诗松, 王静龙, 濮晓龙. 高等数理统计[M]. 北京: 高等教育出版社, 1998:428 - 443.

A Hidden Markov Model for Single Ion-channel and Restoration of States

LI Pei-xing¹, LI Cai-xia², FANG Ji-qian², GUAN Yong-yuan³

(1. Department of Mathematics, Sun Yat-sen University, Guangzhou 510275, China;

2. School of Public Health, Sun Yat-sen University, Guangzhou 510080, China;

3. Department of Pharmacology, Sun Yat-sen University, Guangzhou 510080, China)

Abstract: A hidden Markov model is built for single ion-channels. In this model, the underlying state sequence of the channel is a Markov chain, and the observed patch-clamp recording shave normal distributions which depend on the underlying states. By EM algorithm, the parameters, transition probabilities, means and variations, are estimated. Based on the hidden Markov model, the information of the observed sequence $\{Y_i\}$ is used to restore the underlying states $\{X_i\}$.

Key words: ion-channel; hidden Markov model; EM algorithm; restoration