State Space Model with Markov Switching for Estimating Time-Dependent Gene Regulatory Networks from Time Series Microarray Experiments

Ryo Yoshida [1]
yoshidar@ims.u-tokyo.ac.jp

Seiya Imoto [1]
imoto@ims.u-tokyo.ac.jp

Tomoyuki Higuchi[2]
higuchi@ism.ac.jp

Satoru Miyano[1]
miyano@ims.u-tokyo.ac.jp

[1] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan
[2] Research Organization of Information and Systems, Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo, 103-8569, Japan

Keywords: Time-Dependent Gene Networks, State Space Model with Markov Switching, Dimension Reduction, Time Series Microarray data

1 Introduction

In gene network estimation from time series microarray data, dynamic models such as differential equations and dynamic Bayesian networks assume that the network structure is stable through all time points, while the real network might changes its structure depending on time, affection of some shocks and so on. If the true network structure underlying the data changes at certain points, the fitting of the usual dynamic linear models fails to estimate the structure of gene network and we cannot obtain efficient information from data. To solve this problem, we propose a dynamic linear model with Markov switching for estimating time-dependent gene network structure from time series gene expression data. Using our proposed method, the network structure between genes and its change points are automatically estimated.

2 State Space Model with Markov Switching

Let $\boldsymbol{y}_t$ be a vector of d observed random variables, i.e. expression values of d genes, at time point t. The state space model (SSM) relates a collection of $\boldsymbol{y}_t$, $t = 1, \cdots, T$, to the hidden k-dimensional state vector $\boldsymbol{x}_t$ in the following way:

$$\boldsymbol{y}_t = \boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{w}_t. \tag{1}$$

Here, the $\boldsymbol{A}_t$ is $d \times k$ measurement matrix and the $\boldsymbol{w}_t$ is the Gaussian white noise, $\boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{R}_t)$, having diagonal covariance matrix $\boldsymbol{R} = \mathrm{diag}\{r_1, \cdots, r_d\}$. Usually the dimension of state vector is taken to be much smaller than that of data, i.e. $k < d$. In SSM, the time evolution of the state variables are modeled by a first-order Markov process as

$$\boldsymbol{x}_t = \boldsymbol{B}_t \boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \tag{2}$$

where $\boldsymbol{B}_t$ is $q \times q$ state transition matrix and the additive system noise is also Gaussian, $\boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{I})$. Notice that the model parameters $\{\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{R}_t\}$ depend on the time index. This implies that the underlying dynamics changes discontinuously at certain undetermined points in time.