13

Systems with Markovian Parameters

This chapter concerns the identification of systems with time-varying parameters. The parameters are vector-valued and take values in a finite set. As in the previous chapters, only binary-valued observations are available.

Our study is motivated by applications in the areas of smart sensors, sensor networks, networked mobile agents, distributed power generation networks, etc. For instance, consider an array of mobile sensors being dispatched to survey an area for potential land contamination. Each sensor travels along a trajectory, measures a surface, and communicates the measured values via a wireless network to the command center. Some of the features include

(1) The parameter of interest takes only a few possible values representing regions such as "no contamination," "low contamination," and "high contamination."

(2) When the sensor travels, the parameter values switch randomly depending on the actual contamination.

(3) Due to communication limitations, only quantized measurements are available. Here when the sensor moves slowly, the parameter values switch infrequently. This problem may be described as a system with an unknown parameter that switches over finite possible values randomly. This application represents problems in ocean survey, detection of water pipe safety, mobile robots for bomb, chemical, biological threats, etc.

L.Y. Wang et al., *System Identification with Quantized Observations*, Systems & Control: Foundations & Applications, DOI 10.1007/978-0-8176-4956-2_13,

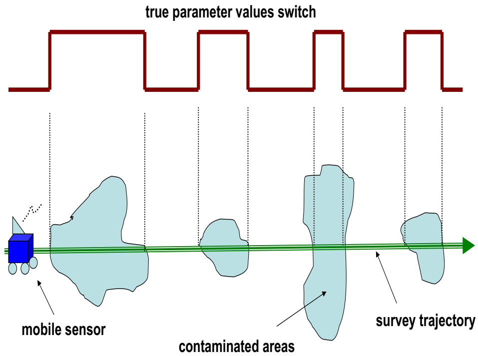

FIGURE 13.1. Mobile sensor systems for area survey

To capture the essence of such problems as those above, we formulate a class of identification problems with randomly switching parameters and binary-valued observations. We shall focus on time-varying parameters modeled by a discrete-time Markov chain with a finite state space. The limited information due to binary-valued sensors makes identification a difficult task. Our approach for identifying regime-switching systems with binary observations relies on the basic idea of Wonham-type filters. Based on the key ideas of such filters, we derive mean-square estimators and analyze their error bounds. To obtain the error bounds for mean-square estimators, we utilize asymptotic distributional results. We first establish weak convergence of functional central limit results, followed by strong approximation of the scaled sequences. Then these distributional results are used to obtain error bounds.

In applications, the frequency of the switching processes plays a crucial role. Consider two typical cases for tracking and identification. The first case is concerned with Markov chains whose switching movements occur infrequently. Here, the time-varying parameter takes a constant value for a relatively long time and switches to another value at a random time. The jumps happen relatively infrequently. We develop maximum posterior (MAP) estimators and obtain bounds on estimation or tracking errors based on Wonham filters. We also point out that a simplified estimator can be developed using empirical measures. The second class of systems aims at treating fast-switching systems. One motivation of such systems is the discretization of a fast-varying Markovian system in continuous time. Suppose the precise transition probabilities are unknown. When parameters frequently change their values, the system becomes intractable if one insists on tracking instantaneous changes. In fact, if the jump parameter switches too frequently, it would be impossible to identify the instantaneous jumps even with regular linear sensors, let alone binary observations. As a result, an alternative approach is suggested. Instead of tracking the moment-by-moment changes, we examine the averaged behavior of the system. The rationale is as follows: Because the Markov chain varies at a fast pace, within a short period of time, it should settle down at a stationary or steady state. In the steady state, the underlying system is a weighted average with the weighting factors the components of the stationary distribution of the Markov chains.

Section 13.1 begins with the setup of the tracking and identification problem with a Markov parameter process. Section 13.2 presents Wonham-type filters for the identification problem. Section 13.3 concerns mean-square criteria. Section 13.4 proceeds with the study of infrequently switching systems. Section 13.5 takes up the issue of fast-switching systems.

13.1 Markov Switching Systems with Binary Observations

Consider a single-input–single-output (SISO), discrete-time system represented by

$$y_k = \phi_k' \theta_k + d_k, \tag{13.1}$$

where $\phi_k = (u_k, \dots, u_{k-n_0+1})'$ and $\{d_k\}$ is a sequence of random disturbances. The θ_k is a Markov chain that takes m_0 possible vector values $\theta^{(j)} \in \mathbb{R}^{n_0}$, $j = 1 \dots, m_0$. y_k is measured by a binary-valued sensor with the known threshold C. After applying an input u, the output s_k is measured for $k = 0, \dots, N-1$ with observation length $N \geq n_0$. We will use the following assumptions throughout this chapter.

(A13.1) The time-varying process $\{\theta_k\}$ is a discrete-time Markov chain with a transition probability matrix P and a finite state space $\mathcal{M} = \{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(m_0)}\}$.

(A13.2) The $\{d_k\}$ is a sequence of i.i.d. random variables with a continuously differentiable distribution function $F(\cdot)$ whose density function is denoted by $f(\cdot)$. The inverse $F^{-1}(\cdot)$ exists and is continuous, and the moment generating function of d_k exists.

13.2 Wonham-Type Filters

Tracking θ_k or identifying the system under binary-valued observations is a nonlinear filtering problem. A crucial step toward this goal is to build a good estimator of the probability distribution given the observations. The identification problem may be stated as follows.

Denote the observation data up to k by $S_k = \{s_l, l \leq k\}$ and the sequence of increasing σ-algebras of the observations up to time k by $\mathcal{F}_{S_k} = \sigma\{s_l : l \leq k\}$. Note that $\mathcal{F}_{S_0} \subset \mathcal{F}_{S_1} \cdots \subset \mathcal{F}_{S_k}$. Similarly, denote the sequence of

σ-algebras generated by θ_k as $\mathcal{F}_{\Theta_k} = \sigma\{\theta_l : l \leq k\}$, and the σ-algebras generated by the observation noise as $\mathcal{F}_{D_k} = \sigma\{d_l : l \leq k\}$. We wish to find the probabilities

$$w_N^{\{j\}} = P(\theta_N = \theta^{(j)} | \mathcal{F}_{S_N}), \quad N \geq 0, \; j = 1, \ldots, m_0. \tag{13.2}$$

Denote the initial probability distribution by $p_0^{\{j\}} = P(\theta_0 = \theta^{(j)})$. Recall that $P = (p^{\{ij\}}) \in \mathbb{R}^{m_0 \times m_0}$, with

$$p^{\{ij\}} = P(\theta_N = \theta^{(j)} | \theta_{N-1} = \theta^{(i)}), \; i, j = 1, \ldots, m_0,$$

are the entries in the transition matrix P. The development uses the Wonham filter techniques in [59], which is a discrete version of the original Wonham filter in [114]. Nevertheless, in our case, we only have binary-valued observations. The noise does not appear additive either. For each $j = 1, \ldots, m_0$, we denote

$$\begin{aligned} G^{\{j\}}(s_N) \; &:= P(s_N | \theta_N = \theta^{(j)}) = I_{\{s_N = 1\}} F(C - \phi_N' \theta^{(j)}) \\ &\quad + [1 - I_{\{s_N = 1\}}](1 - F(C - \phi_N' \theta^{(j)}), \end{aligned} \tag{13.3}$$

which is a function of the random variable s_N.

Theorem 13.1. *Assume* (A13.1) *and* (A13.2). *The Wonham-type filter for the binary-valued observations can be constructed as*

$$w_0^{\{j\}} = \frac{p_0^{\{j\}} G^{\{j\}}(s_0)}{\sum_{j_1=1}^{m_0} p_0^{\{j_1\}} G^{\{j_1\}}(s_0)}, \; j = 1, \ldots, m_0 \tag{13.4}$$

and

$$w_N^{\{j\}} = \frac{G^{\{j\}}(s_N) \sum_{i=1}^{m_0} p^{\{ij\}} w_{N-1}^{\{i\}}}{\sum_{i=1}^{m_0} \sum_{j_1=1}^{m_0} G^{\{j_1\}}(s_N) p^{\{ij_1\}} w_{N-1}^{\{j_1\}}}, \; j = 1, \ldots, m_0. \tag{13.5}$$

Proof. To verify (13.4), applying Bayes' theorem leads to

$$w_0^{\{j\}} = \frac{P(s_0 | \theta_0 = \theta^{(j)}) P(\theta_0 = \theta^{(j)})}{\sum_{j_1=1}^{m_0} P(s_0, \theta_N = \theta^{(j_1)})} = \frac{p_0^{\{j\}} G^{\{j\}}(s_0)}{\sum_{j_1=1}^{m_0} p_0^{\{j_1\}} G^{\{j_1\}}(s_0)}.$$

To prove (13.5), we first introduce the one-step prediction

$$w_{N|N-1}^{\{j\}} = P(\theta_N = \theta^{(j)} | \mathcal{F}_{S_{N-1}}). \tag{13.6}$$

Since $\{d_k\}$ is a sequence of i.i.d. random variables and $\{\theta_N\}$ is Markovian, we have

$$P(\theta_N = \theta^{(j)}|\theta_{N-1} = \theta^{(j_1)}, \mathcal{F}_{S_{N-1}}) = \theta^{(j_1)}) = p^{\{j_1 j\}}.$$

By the law of total probability,

$$E(w_N^{\{j\}}|\mathcal{F}_{S_{N-1}}) = P(\theta_N = \theta^{(j)}|\mathcal{F}_{S_{N-1}}) = \sum_{j_1=1}^{m_0} p^{\{j_1 j\}} w_{N-1}^{\{j_1\}}. \tag{13.7}$$

Now, by Bayes' theorem and (13.7),

$$\begin{aligned} w_N^{\{j\}} &= P(\theta_N = \theta^{(j)}|s_N, \mathcal{F}_{S_{N-1}}) \\ &= \frac{P(s_N|\theta_N = \theta^{(j)}, \mathcal{F}_{S_{N-1}})P(\theta_N = \theta^{(j)}|\mathcal{F}_{S_{N-1}})}{\sum_{j_1=1}^{m_0} P(s_N|\theta_N = \theta^{(j_1)}, \mathcal{F}_{S_{N-1}})P(\theta_N = \theta^{(j_1)}|\mathcal{F}_{S_{N-1}})} \\ &= \frac{G^{\{j\}}(s_N)\sum_{i=1}^{m_0} p^{\{ij\}} w_{N-1}^{\{i\}}}{\sum_{j_1=1}^{m_0} G^{j_1}(s_N)\sum_{i=1}^{m_0} p^{\{ik\}} w_{N-1}^{\{i\}}}. \end{aligned}$$

The last line above follows from (13.6). □

13.3 Tracking: Mean-Square Criteria

Based on Wonham-type filters, under different criteria, we may develop several different estimators. First, consider the following optimization problem: Choose θ to minimize the mean-square errors conditioned on the information up to time N. That is, find θ to minimize $\min_\theta E(|\theta_N - \theta|^2|S_N)$. Just as in the usual argument for Kalman filters, bearing in mind the use of conditional expectation, we obtain the minimizer of the cost, which leads to the following mean-square estimator:

$$\widehat{\theta}_N = E(\theta_N|\mathcal{F}_{S_n}) = \sum_{j=1}^{m_0} \theta^{(j)} w_N^{\{j\}}.$$

To derive the error estimates of $\widehat{\theta}_N - \theta_N$, we need the associated asymptotic distribution for

$$e_N = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} (\theta_k - \widehat{\theta}_k).$$

Note that $\{e_N\}$ is a sequence of centered and scaled deviations of the Markov chain from its mean-square tracker with a scaling factor $\sqrt{N}$. For future use, we note that

$$\begin{aligned} e_N &= \frac{1}{\sqrt{N}} \sum_{j\in\mathcal{M}} \sum_{k=0}^{N-1} \theta^{(j)} \{[I_{\{\theta_k=\theta^{(j)}\}} - P(\theta_k = \theta^{(j)})] \\ &\quad + [P(\theta_k = \theta^{(j)}) - w_k^{\{j\}}]\}. \end{aligned} \tag{13.8}$$

To examine the deviation, in lieu of working with a discrete-time formula directly, we focus on a continuous-time interpolation of the form

$$v_N(t) = \frac{1}{\sqrt{N}} \sum_{j\in\mathcal{M}} \sum_{k=0}^{\lfloor Nt\rfloor-1} \theta^{(j)} [I_{\{\theta_k=\theta^{(j)}\}} - w_k^{\{j\}}], \ t \in [0,1], \tag{13.9}$$

where $\lfloor z\rfloor$ denotes the integer part of $z \in \mathbb{R}$. We shall show that the limit of $v_N(\cdot)$ is a Brownian motion, whose properties help us to derive the desired error bounds. Obtaining the weak convergence to the Brownian motion requires verifying that the sequence under consideration is tight (or compact). Then we characterize the limit by means of martingale problem formulation.

To use weak convergence theory, it is common and more convenient to use the so-called D space, which is a space of functions that are right continuous and have left limits, with a topology weaker than uniform convergence, known as the Skorohod topology. The main advantage of using such a setup is that it enables one to verify the tightness or compactness relatively easily. The exact definitions of these are somewhat technical; we refer the reader to [55, Chapter 7] for further reference.

Lemma 13.2. *Assume the conditions of Theorem* 13.1*, and suppose the Markov chain is irreducible.*

(a) *Then for each $\delta > 0$, each $t \geq 0$, and each $s > 0$ with $0 \leq s \leq \delta$,*

$$\sup_N E|v_N(t+s) - v_N(t)|^2 \leq Ks, \tag{13.10}$$

for some $K > 0$.

(b) *The sequence $v_n(\cdot)$ is tight in $D([0,1];\mathbb{R}^{m_0})$, the space of $\mathbb{R}^{m_0}$-valued functions that are right continuous, have left limits, and are endowed with the Skorohod topology.*

Proof. We first prove (a). In view of the second line of (13.8), for each $\delta > 0$, for $t > 0$, $s > 0$ satisfying $0 \leq s \leq \delta$, we have

$$E|v_N(t+s) - v_N(t)|^2 \leq L_{N,1} + L_{N,2},$$

where

$$
\begin{aligned}
L_{N,1} &= \frac{2}{N} \sum_{j_1,j_2 \in \mathcal{M}} \sum_{k=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \sum_{i=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \theta^{(j_1),\prime} \theta^{(j_2)} E I_k^{\{j_1\}} I_i^{\{j_2\}}, \\
L_{N,2} &= \frac{2}{N} \sum_{j_1,j_2 \in \mathcal{M}} \sum_{k=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \sum_{i=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \theta^{(j_1),\prime} \theta^{(j_2)} E \breve{w}_k^{\{j_1\}} \breve{w}_i^{\{j_2\}}, \\
I_k^{\{j\}} &= I_{\{\theta_k = \theta^{(j)}\}} - P(\theta_k = \theta^{(j)}), \\
\breve{w}_k^{\{j\}} &= P(\theta_k = \theta^{(j)}) - w_k^{\{j\}}, \\
\widetilde{w}_k^{\{j\}} &= I_{\{\theta_k = \theta^{(j)}\}} - w_k^{\{j\}}.
\end{aligned}
\tag{13.11}
$$

To proceed, we estimate $L_{N,1}$ and $L_{N,2}$. We need only look at a fixed pair j_1 and j_2. First, consider $L_{N,1}$ without the first sum. Without loss of generality, assume $k \geq i$. Then we obtain that for fixed j_1 and $j_2 \in \mathcal{M}$,

$$
\begin{aligned}
&\left| \sum_{i=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \sum_{k=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \theta^{(j_1),\prime} \theta^{(j_2)} E I_k^{\{j_1\}} I_i^{\{j_2\}} \right| \\
&\quad \leq 2 \sum_{i=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \sum_{i \leq k}^{\lfloor N(t+s) \rfloor -1} \theta^{(j_1),\prime} \theta^{(j_2)} \left| E[I_i^{\{j_2\}} E_i I_k^{\{j_1\}}] \right|,
\end{aligned}
\tag{13.12}
$$

where E_l denotes the expectation conditioned on $\mathcal{F}_l = \sigma\{d_k, \theta_k : k \leq l\}$, the past information up to l. Since the Markov chain is irreducible, it is ergodic. That is, there is a row vector ν, the stationary distribution of the Markov chain such that

$$|P^N - \mathbb{1}\nu| \leq K\lambda^N \quad \text{for some} \quad 1 > \lambda > 0,$$

where $\mathbb{1}$ is a column vector with all its component being 1. Using this spectrum gap estimate,

$$|E_i I_k^{\{j_1\}}| \leq \lambda^{k-i}.$$

It then follows that for the term in (13.12), we have

$$
\begin{aligned}
&\left| \sum_{i=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \sum_{k=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \theta^{(j_1),\prime} \theta^{(j_2)} E I_k^{\{j_1\}} I_i^{\{j_2\}} \right| \\
&\quad \leq K \left(\lfloor N(t+s) \rfloor - \lfloor Nt \rfloor \right).
\end{aligned}
$$

Dividing the above by N leads to $\sup_N L_{N,1} \leq Ks$. As for the terms involved in $L_{N,2}$,

$$\left| \sum_{i=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \sum_{i \le k}^{\lfloor N(t+s) \rfloor -1} \theta^{(j_1),\prime} \theta^{(j_2)} E \breve{w}_k^{\{j_1\}} \breve{w}_i^{\{j_2\}} \right| \le K \sum_{i=\lfloor Nt \rfloor}^{\lfloor N(t+s) \rfloor -1} \sum_{k \ge i}^{\lfloor N(t+s) \rfloor -1} \lambda^{k-i} \le KNs.$$

Dividing the above by N and taking $\sup_N$ yields $\sup_N L_{N,2} \le Ks$. Thus, (a) is true.

By using (a), with arbitrary $\delta > 0$ and the chosen t and s, we have

$$\lim_{\delta \to 0} \limsup_{N \to \infty} E|v_N(t+s) - v_N(t)|^2 \le \lim_{\delta \to 0} K\delta = 0.$$

Thus, the tightness follows from the criterion [53, Theorem 3, p. 47]. The lemma is proved. □

To proceed, let us point out:

(i) Since $v_N(\cdot)$ is tight, we can extract weakly convergent subsequences by means of Prohorov's theorem (see [55, Chapter 7]). Loosely, sequential compactness enables us to extract convergent subsequences. Without loss of generality, still index the selected subsequence by N, and assume $v_N(\cdot)$ itself is the weakly convergent subsequence. Denote the limit by $v(\cdot)$. We shall characterize the limit process.

(ii) From the defining relationship of $v_N(t)$, it is readily seen that

$$\begin{aligned} Ev_N(t) &= \frac{1}{\sqrt{N}} \sum_{j \in \mathcal{M}} \sum_{k=0}^{\lfloor Nt \rfloor -1} \theta^{(j)} \Big\{ E\Big[I_{\{\theta_k = \theta^{(j)}\}} - P\Big(\theta_k = \theta^{(j)}\Big)\Big] \\ &\quad + E\Big[P\Big(\theta_k = \theta^{(j)}\Big) - w_k^{\{j\}} \Big] \Big\} = 0 \quad \text{for each} \quad t \ge 0. \end{aligned} \tag{13.13}$$

To determine the limit process, we consider a vector-valued process $\widetilde{v}_N(t) = (\widetilde{v}_N^1(t), \ldots, \widetilde{v}_N^{m_0}(t)) \in \mathbb{R}^{m_0}$, where

$$\widetilde{v}_N^{\{i\}}(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{\lfloor Nt \rfloor -1} \theta^{(i)} \Big[I_{\{\theta_k = \theta^{(i)}\}} - w_k^{\{i\}} \Big].$$

Define $\Sigma_N(t) = (\Sigma^{\{ij\}}(t)) = E\widetilde{v}_N(t)\widetilde{v}_N'(t)$, where $\Sigma_N^{\{ij\}}(t)$ denotes the ijth entry of the partitioned matrix $\Sigma_N(t)$, namely,

$$\Sigma_N^{\{ij\}}(t) = \frac{1}{N} \sum_{k=0}^{\lfloor Nt \rfloor -1} \sum_{l=0}^{\lfloor Nt \rfloor -1} E\zeta_k^{\{i\}} \zeta_l^{\{j\},\prime}, \tag{13.14}$$

where

$$\zeta_k^{\{i\}} = \theta^{(i)} \left[I_{\{\theta_k = \theta^{(i)}\}} - w_k^{\{i\}} \right] \in \mathbb{R}^{m_0}.$$

Using the notation of $\widetilde{v}_N(t)$, we can rewrite $v_N(t)$ as $v_N(t) = \mathbb{1}'_{m_0} \widetilde{v}_N(t)$, where $\mathbb{1}'_{m_0} = (1, \ldots, 1) \in \mathbb{R}^{1 \times m_0}$. To proceed, we first determine the limit covariance function of $E v_N(t) v'_N(t) = \mathbb{1}'_{m_0} [E \widetilde{v}_N(t) \widetilde{v}'_N(t)] \mathbb{1}_{m_0}$. From the above expression, it is seen that to accomplish this goal, we need only consider the limit covariance of $E \widetilde{v}_N(t) \widetilde{v}'_N(t)$. The following lemma details the calculation of the asymptotic covariance.

Lemma 13.3. *Assume the conditions of Lemma* 13.2. *Then*

(a) *the limit covariance of* $\widetilde{v}_N(t)$ *is given by*

$$\begin{aligned} &\lim_{N \to \infty} \Sigma_N(t) = t \Sigma_0, \quad \Sigma_0 = \operatorname{diag}\left(\Sigma_0^{\{11\}}, \ldots, \Sigma_0^{\{m_0 m_0\}} \right), \\ &\Sigma_0^{\{i\}} \stackrel{\text{def}}{=} \Sigma_0^{\{ii\}} = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N} E \zeta_k^{\{i\}} \zeta_k^{\{i\},\prime}, \ i \in \mathcal{M}; \end{aligned} \tag{13.15}$$

(b) *as* $N \to \infty$,

$$E v_N(t) v'_N(t) \to t \overline{\Sigma} = t \sum_{i=1}^{m_0} \Sigma_0^{\{i\}}.$$

Proof. To prove (a), it suffices to work with the partitioned matrix $\Sigma_N^{\{ij\}}(t)$. Note that

$$\sum_{k<l}^{\lfloor Nt \rfloor - 1} E \zeta_k^{\{i\}} \zeta_l^{\{j\},\prime} = \sum_{k<l}^{\lfloor Nt \rfloor - 1} E \left[\zeta_k^{\{i\}} E_k \zeta_l^{\{j\},\prime} \right] = 0 \quad \text{for} \quad i \neq j.$$

Likewise,

$$\sum_{l<k}^{\lfloor Nt \rfloor - 1} E \zeta_k^{\{i\}} \zeta_l^{\{j\},\prime} = 0 \quad \text{for} \quad i \neq j.$$

This leads to

$$\begin{aligned} \Sigma_N^{\{ij\}}(t) &= \delta_{ij} \frac{1}{N} \sum_{k=0}^{\lfloor Nt \rfloor - 1} E \zeta_k^{\{i\}} \zeta_k^{\{j\},\prime} + \frac{1}{N} \sum_{l=0}^{\lfloor Nt \rfloor - 1} \sum_{k<l}^{\lfloor Nt \rfloor - 1} E \zeta_k^{\{i\}} \zeta_l^{\{j\},\prime} \\ &\quad + \frac{1}{N} \sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{l<k}^{\lfloor Nt \rfloor - 1} E \zeta_k^{\{i\}} \zeta_l^{\{j\},\prime} \\ &= \delta_{ij} \frac{\lfloor Nt \rfloor}{N} \frac{1}{\lfloor Nt \rfloor} \sum_{k=0}^{\lfloor Nt \rfloor - 1} E \zeta_k^{\{i\}} \zeta_k^{\{j\},\prime} \\ &\to \begin{cases} t \Sigma_0^{\{i\}}, & \text{if } i = j, \\ 0, & \text{otherwise}, \end{cases} \end{aligned} \tag{13.16}$$

as $N \to \infty$, where $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise, and $\Sigma_0^{\{ij\}}$ denotes the ijth partitioned matrix in Σ_0.

Finally, (b) is a direct consequence of Lemma 13.3. The lemma is thus proved. □

Note that by (13.15), $\Sigma_0^{\{i\}} = \Sigma_0^{\{ii\}}$, the partitioned matrix of Σ_0. To proceed, we prove that $v_N(\cdot)$ converges weakly to a Brownian motion. We characterize the limit process by means of identifying the limit covariance function. The analysis is carried out by using the martingale problem formulation. For a twice continuously differentiable function $h : \mathbb{R}^{m_0} \mapsto \mathbb{R}$, define an operator as

$$\mathcal{L}h(v) = \frac{1}{2}\mathrm{tr}\left[\overline{\Sigma}h_{vv}(v)\right], \tag{13.17}$$

where h_{vv} denotes the Hessian matrix (the second partial derivatives with respect to v). We have the following result.

Theorem 13.4. *Assume the conditions of Lemma* 13.2. *Then*

(a) $v_N(\cdot)$ *converges weakly to* $v(\cdot)$*, which is a Brownian motion with covariance* $\overline{\Sigma}t$*;*

(b) $v_N(1)$ *converges in distribution to a normal random variable with mean 0 and covariance* $\overline{\Sigma}$*.*

Proof. Part (b) is a direct consequence of (a). Thus, we need only prove (a). Since $v_N(\cdot)$ converges weakly, there is a convenient device known as the Skorohod representation (see [55, Chapter 7]) that enables us to work with w.p.1 convergence on an enlarged space. Without loss of generality and with a slight abuse of notation, we may assume $v_N(\cdot) \to v(\cdot)$ in the sense of w.p.1. We want to show that $v(\cdot)$ is a solution to the martingale problem with operator $\mathcal{L}$ defined in (13.17). To this end, it suffices to show that

$$h(v(t)) - h(v(0)) - \int_0^t \mathcal{L}h(v(\rho))d\rho \ \text{ is a martingale.}$$

To verify the above, it only needs to be shown (see [55]) that for any bounded and continuous function $H(\cdot)$, any $t, s > 0$, any integers ℓ, and any $t_\iota \le t$,

$$\begin{aligned} EH(v(t_\iota) : \iota \le \ell)[\, h(v(t+s)) - h(v(t)) \\ - \int_t^{t+s} \mathcal{L}h(v(\rho))d\rho] = 0. \end{aligned} \tag{13.18}$$

To verify (13.18), use $v_n(\cdot)$. By the weak convergence and the Skorohod representation, as $N \to \infty$,

$$\begin{aligned} &EH(v_N(t_\iota) : \iota \le \ell)[h(v_N(t+s)) - h(v_N(t))] \\ &\quad \to EH(v(t_\iota) : \iota \le \ell)[h(v(t+s)) - h(v(t))]. \end{aligned} \tag{13.19}$$

On the other hand, direct computation reveals that

$$\begin{aligned}
&\lim_{N\to\infty} EH(v_N(t_\iota):\iota\le\ell)[h(v_N(t+s))-v_N(t))]\\
&=\lim_{N\to\infty} EH(v_N(t_\iota):\iota\le\ell)\Bigg[\frac{1}{\sqrt{N}}\sum_{i\in\mathcal{M}} h'_v(v_N(t))\theta^{(i)}\sum_{k=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\widetilde{w}_k^{\{i\}}\\
&\quad+\frac{1}{2N}\sum_{i\in\mathcal{M}}\sum_{i_1\in\mathcal{M}}\mathrm{tr}[h_{vv}(v_N(t))\theta^{(i)}\theta^{(i_1)}]\sum_{k=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\sum_{l=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\widetilde{w}_k^{\{i\}}\widetilde{w}_l^{i_1}\Bigg],
\end{aligned}\tag{13.20}$$

where $\widetilde{w}_k^{\{i\}}$ is given by (13.11). Using nested expectation and inserting $E_{\lfloor Nt\rfloor}$, since $v_N(t_\iota):\iota\le\ell$ and $v_N(t)$ are all $\mathcal{F}_{\lfloor Nt\rfloor}$-measurable, by inserting $E_{\lfloor Nt\rfloor}$ we have

$$\begin{aligned}
&EH(v_N(t_\iota):\iota\le\ell)\Bigg[\frac{1}{\sqrt{N}}\sum_{i\in\mathcal{M}} h'_v(v_N(t))\theta^{(i)}E_{\lfloor Nt\rfloor}\sum_{k=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\widetilde{w}_k^{\{i\}}\Bigg]\\
&\quad\to 0 \quad\text{as}\quad N\to\infty.
\end{aligned}$$

Likewise,

$$\begin{aligned}
&EH(v_N(t_\iota):\iota\le\ell)\Bigg[\frac{1}{2N}\sum_{i\in\mathcal{M}}\sum_{i_1\in\mathcal{M}}\mathrm{tr}[h_{vv}(v_N(t))\theta^{(i)}\theta^{(i_1)}]\\
&\qquad\times\sum_{k=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\sum_{l=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\widetilde{w}_k^{\{i\}}\widetilde{w}_l^{i_1}\Bigg]\\
&\quad=EH(v_N(t_\iota):\iota\le\ell)\Bigg[\frac{1}{2N}\sum_{i\in\mathcal{M}}\sum_{i_1\in\mathcal{M}}\mathrm{tr}[h_{vv}(v_N(t))\theta^{(i)}\theta^{(i_1)}]\\
&\qquad\times E_{\lfloor Nt\rfloor}\sum_{k=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\sum_{l=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\widetilde{w}_k^{\{i\}}\widetilde{w}_l^{i_1}\Bigg].
\end{aligned}$$

Dividing the cases into $l\le k$ and $k<l$, we can handle the last equation above as in the proof of Lemma 13.3 by inserting E_l and E_k, respectively. It follows that the double summations above reduce to a single one. The last two equations together with (13.20) then imply that

$$\begin{aligned}
&EH(v_N(t_\iota):\iota\le\ell)\Big[\frac{1}{2N}\sum_{i\in\mathcal{M}}\sum_{i_1\in\mathcal{M}}\mathrm{tr}[h_{vv}(v_N(t))\theta^{(i)}\theta^{(i_1)}]\\
&\qquad\times E_{\lfloor Nt\rfloor}\sum_{k=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\sum_{l=\lfloor Nt\rfloor}^{\lfloor N(t+s)\rfloor-1}\widetilde{w}_k^{\{i\}}\widetilde{w}_l^{i_1}\Big]\\
&\quad\to EH(v(t_\iota):\iota\le\ell)\Big[\int_t^{t+s}\mathcal{L}h(\rho)d\rho\Big]\quad\text{as}\quad N\to\infty.
\end{aligned}$$

This establishes the desired theorem. □

By virtue of Theorem 13.4, we further obtain a strong approximation result. This strong approximation will aid us in obtaining error bounds in what follows.

Lemma 13.5. *Under the conditions of Theorem* 13.4, *there is a constant* $\gamma > 0$ *such that*

$$\sup_{0\le t\le 1} |v_N(t) - v(t)| = o(N^{-\gamma}) \quad \textit{w.p.1.}$$

Proof. Note that

$$E_{k-1}\zeta_k^{\{i\}} = E_{k-1}[I_{\{\theta_k=\theta^{(i)}\}} - w_k^{\{i\}}] = 0.$$

Thus, it is a martingale difference sequence. Using the martingale version of the Skorohod representation (see [41, p. 269]), we can establish the result. The details are omitted. □

We next show that the tracking error in the average sense is exponentially small. The result is based on part (b) in Theorem 13.4 and large deviations for normal random variables. There are different ways to obtain the error bounds. We do one as follows, whose proof is also in the appendix.

Theorem 13.6. *Under the conditions of Lemma* 13.2,

$$\begin{aligned} P\left(\frac{1}{\sqrt{N}}\left|\sum_{i\in\mathcal{M}}\sum_{k=0}^{N}\theta^{(i)}[I_{\{\theta_k=\theta^{(i)}\}} - w_k^{\{i\}}]\right|_1 \ge \varepsilon\right) \\ \le 2m_0 \exp\left(-\frac{N\varepsilon^2}{2{m_0}^2\sigma^2_{v^{\{i\}}(1)}}\right), \end{aligned} \tag{13.21}$$

where $|\cdot|_1$ *denotes the* l_1 *norm.*

Proof. Note that $v^{\{i\}}(1) = e_i'v(1)$, where e_i is the ith standard unit vector. Note also that $v^{\{i\}}(1)$ is normally distributed with mean 0 and variance $\sigma^2_{v^{\{i\}}(1)} = e_i'\overline{\Sigma}e_i$. We then have that for any $\alpha > 0$,

$$\begin{aligned} P\left(\frac{1}{\sqrt{N}}|v^{\{i\}}(1)| \ge \frac{\varepsilon}{m_0}\right) &\le \exp\left(-\frac{\alpha\varepsilon}{m_0}\right) E\exp\left(\frac{\alpha|v^{\{i\}}(1)|}{\sqrt{N}}\right) \\ &\le 2\exp\left(-\frac{\alpha\varepsilon}{m_0} + \frac{\sigma^2_{v^{\{i\}}(1)}\alpha^2}{2N}\right). \end{aligned} \tag{13.22}$$

Choosing the α to minimize the index in the exponent leads to

$$\alpha = (N\varepsilon/(m_0\sigma^2_{v^{\{i\}}(1)})).$$

Using this in (13.22) yields the upper bound

$$P\left(\frac{1}{\sqrt{N}}|v^{\{i\}}(1)| \geq \frac{\varepsilon}{m_0}\right) \leq 2\exp\left(-\frac{N\varepsilon^2}{2{m_0}^2\sigma^2_{v^{\{i\}}(1)}}\right).$$

Thus,

$$P\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{m_0}|v^{\{i\}}(1)| \geq \varepsilon\right) \leq 2m_0\exp\left(-\frac{N\varepsilon^2}{2{m_0}^2\sigma^2_{v^{\{i\}}(1)}}\right). \qquad (13.23)$$

Note that

$$\begin{aligned}P&\left(\frac{1}{\sqrt{N}}\left|\sum_{i\in\mathcal{M}}\sum_{k=0}^{N}\theta^{(i)}\widetilde{w}_k^{\{i\}}\right|_1 \geq \varepsilon\right)\\ &= P\left(\exp\left(\frac{\alpha}{\sqrt{N}}\left|\sum_{i\in\mathcal{M}}\sum_{k=0}^{N}\theta^{(i)}\widetilde{w}_k^{\{i\}}\right|_1\right) \geq \exp(\alpha\varepsilon)\right).\end{aligned}$$

We can approximate the

$$(\alpha/\sqrt{N})\sum_{k=0}^{N}\theta^{(i)}\widetilde{w}_k^{\{i\}}$$

by $v^{\{i\}}(t)$ by using Lemma 13.5. Adding and subtracting $v^{\{i\}}(t)$ in the above and using the triangle inequality yield that

$$\begin{aligned}P&\left(\frac{1}{\sqrt{N}}\left|\sum_{i\in\mathcal{M}}\sum_{k=0}^{N}\theta^{(i)}\widetilde{w}_k^{\{i\}}\right|_1 \geq \varepsilon\right)\\ &\leq \exp\left(-\frac{\alpha\varepsilon}{m_0}\right)E\exp\left(\frac{\alpha}{\sqrt{N}}\left|\sum_{i\in\mathcal{M}}\sum_{k=0}^{N}\{\theta^{(i)}\widetilde{w}_k^{\{i\}} - v^{\{i\}}(t)\}\right|_1\right.\\ &\qquad\qquad\left. + \frac{\alpha}{\sqrt{N}}\sum_{i\in\mathcal{M}}|v^{\{i\}}(t)|_1\right)\\ &\leq \exp\left(-\frac{\alpha\varepsilon}{m_0}\right)E\exp(o(N^{-\gamma}))\exp\left(\frac{\alpha}{\sqrt{N}}\sum_{i\in\mathcal{M}}|v^{\{i\}}(t)|_1\right).\end{aligned}$$

Using (13.23) in the above estimate, the desired result then follows. □

13.4 Tracking Infrequently Switching Systems: MAP Methods

Here, we construct a sequence of estimates of the Markov chain by maximizing the *a posterior* probabilities. The estimator is given by

$$\widehat{\theta}_N = \theta^{(j_N)}, \quad j_N = \text{argmax}_{j\in\mathcal{M}} w_N^{\{j\}}. \qquad (13.24)$$

Our goal is to derive an error bound on $P(\widehat{\theta}_N \neq \theta_N)$. We are interested in the case that for each $i \in \mathcal{M}$, $\sum_{j\neq i} p_{ij} = \varepsilon$ and $p_{ii} = 1-\varepsilon$, for ε sufficiently small. One such model assumes the transition probability of the Markov chain to be $P^\varepsilon = I + \varepsilon Q$, where Q is a generator of a continuous-time Markov chain. It indicates that most of the time, the system will remain at a constant value, but it has infrequent jumps from one parameter value to another at random times. This is a class of "infrequent switching" systems. It is intuitively understood that the data size n should be neither too small for lack of information from data nor too large since old data will contain diminishing information about the current θ_N. It is also conceivable that for smaller ε, a larger N may be used. It is our desire to establish a concrete relationship between N and ε to guarantee a desired accuracy of identification. For a selected N, the implementation is the standard moving-window method: To identify θ_k, the data in the time window $l = k - N$, $k - N + 1$, …, k, will be used. It is noted for a large window size N, the initial distribution of θ_{k-N} will have diminishing effects on the MAP estimates $\widehat{\theta}_N$, at the end of the window $k - 1 \leq l \leq N$. Consequently, one may choose any initial distribution, such as the uniform distribution, to start the MAP algorithm. The following discussion is generic for a given moving window with a chosen initial distribution. For simplicity, we make the following assumption.

(A13.3) $y_k = \theta_k + d_k$ and the initial probability distribution of the Markov chain satisfies $p^{\{j\}}(0) > 0$ for $j = 1, \ldots, m_0$.

In what follows, we choose ε to be sufficiently small and N sufficiently large. Let $p_j = F(C - \theta^{(j)})$ and $\delta = \min_{i\neq j} |p_i - p_j|$. Define

$$\xi_{N+1} = \frac{1}{N+1} \sum_{k=0}^{N} s_k, \tag{13.25}$$

and denote the data set by $S_N = \{s_k, k = 0, \ldots, N\}$. For a given $\beta < \delta/2$, define

$$\begin{aligned} M_N^{\{j\}} &= \{s_k : 0 \leq k \leq N,\ |\xi_{N+1} - p_j| < \beta\}, \\ M_N &= \bigcup_{j=1}^{m_0} M_N^{\{j\}}. \end{aligned}$$

Lemma 13.7. *For sufficiently small ε and sufficiently large N and some constant $c > 0$,*

$$P(M_N) \geq (1 - e^{-N\beta^2 c})(1 - \varepsilon)^N, \tag{13.26}$$

which implies

$$\lim_{N\to\infty} \lim_{\varepsilon\to 0} P(M_N) = 1.$$

Proof. Note that

$$\begin{aligned} P(M_N^{\{j\}}) &= \sum_{i=1}^{m_0} P(M_N^{\{j\}} | \theta_0 = \theta^{(i)}) p_0^{\{i\}} \\ &= \left[p_0^{\{j\}} P(M_N^{\{j\}} | \theta_0 = \theta^{(j)}) + \sum_{i \neq j} P(M_N^{\{j\}} | \theta_0 = \theta^{(i)}) p_0^{\{i\}} \right] \\ &\geq p_0^{\{j\}} P(M_N^{\{j\}} | \theta_0 = \theta^{(j)}). \end{aligned}$$

Then,

$$\begin{aligned} &P(M_N^{\{j\}}) \geq p_0^{\{j\}} P(M_N^{\{j\}} | \theta_k = \theta^{(j)}, k = 0, \ldots, N) \\ &\quad \times P(\theta_k = \theta^{\{j\}}, k = 1, \ldots, N | \theta_0 = \theta^{(j)}) \\ &= p_0^{\{j\}} P(M_N^{\{j\}} | \theta_k = \theta^{(j)}, k = 0, \ldots, N)(1 - \varepsilon)^N \\ &= p_0^{\{j\}} P(|\xi_{N+1} - p_j| \leq \beta | \theta_k = \theta^{(j)}, k = 0, \ldots, N)(1 - \varepsilon)^N. \end{aligned}$$

By the large deviations principle,

$$P(|\xi_{N+1} - p_j| \leq \beta | \theta_k = \theta^{(j)}, k = 0, \ldots, N) \geq 1 - e^{-N\beta^2 c}$$

for some $c > 0$. This implies

$$P(M_N^{\{j\}}) \geq p_0^{\{j\}} (1 - e^{-N\beta^2 c})(1 - \varepsilon)^N.$$

Since $\beta < \delta/2$, $M_N^{\{j\}}$, $j = 1, \ldots, m_0$, are disjoint. Hence,

$$P(M_N) = \sum_{j=1}^{m_0} P(M_N^{\{j\}})$$

and (13.26) follows. This completes the proof. □

A sequence in $\{M_N\}$ is called a *typical sequence*. Lemma 13.7 indicates that for small ε and large N, the probability for a sequence to be typical is nearly 1. For this reason, to derive error bounds in probability, we may consider only the data set in M_N.

Lemma 13.8. *For sufficiently small ε and β, and sufficiently large N, if $S_N \in M_N^{\{j\}}$, then $\widehat{\theta}_N = \theta^{(j)}$.*

Proof. Suppose $S_N \in M_N^{\{j\}}$. Using the MAP estimator, $\widehat{\theta}_N = \theta^{(j)}$ if and only if

$$P(\theta_N = \theta^{(j)} | S_N) > P(\theta_N = \theta^{(i)} | S_N),\ i \neq j.$$

Since

$$P(\theta_N = \theta^{(i)} | S_N) = \frac{P(\theta_N = \theta^{(i)}, S_N)}{P(S_N)}$$

the conclusion is true if

$$P(\theta_N = \theta^{(j)}, S_N) > P(\theta_N = \theta^{(i)}, S_N),\ i \neq j.$$

In the following derivation, K_1, K_2, and K are some positive constants. Now

$$\begin{aligned}
&P(\theta_N = \theta^{(i)}, S_N) \\
&\quad = \sum_{l=1}^{m_0} P(\theta_N = \theta^{(i)}, S_N | \theta_0 = \theta^{(l)}) p_0^l \\
&\quad = p_0^{\{i\}} P(\theta_N = \theta^{(i)}, S_N | \theta_0 = \theta^{(i)}) + \varepsilon K_1 \\
&\quad = p_0^{\{i\}} P(S_N | \theta_k = \theta^{(i)}, k = 0, \ldots, N) \\
&\qquad \times P(\theta_k = \theta^{(i)}, k = 1, \ldots, N-1 | \theta_0 = \theta^{(i)}) + \varepsilon K_2 + \varepsilon K_1 \\
&\quad = p_0^{\{i\}} P(S_N | \theta_k = \theta^{(i)}, k = 0, \ldots, N)(1-\varepsilon)^N + \varepsilon K.
\end{aligned}$$

By the definition of ξ_{N+1}, S_N contains $(N+1)\xi_{N+1}$ of 1's and $(N+1)(1-\xi_{N+1})$ of 0's:

$$P(S_N | \theta_k = \theta^{(i)}, k = 0, \ldots, N) = p_i^{(N+1)\xi_{N+1}} (1-p_i)^{(N+1)(1-\xi_{N+1})}.$$

Consequently,

$$\begin{aligned}
&P(\theta_N = \theta^{(i)}, S_N) \\
&\quad = \frac{1}{m_0} p_i^{(N+1)\xi_{N+1}} (1-p_i)^{(N+1)(1-\xi_{N+1})} (1-\varepsilon)^N + \varepsilon K.
\end{aligned} \tag{13.27}$$

For sufficiently small ε, the first term is dominant. As a result, to prove

$$P(\theta_N = \theta^{(j)}, S_N) > P(\theta_N = \theta^{(i)}, S_N),\ i \neq j,$$

we need only show

$$\begin{aligned}
&p_j^{(N+1)\xi_{N+1}} (1-p_j)^{(N+1)(1-\xi_{N+1})} \\
&\quad > p_i^{(N+1)\xi_{N+1}} (1-p_i)^{(N+1)(1-\xi_{N+1})},
\end{aligned}$$

or equivalently, if

$$\begin{aligned}
&\xi_{N+1} \log p_j + (1-\xi_{N+1}) \log(1-p_j) \\
&\quad > \xi_{N+1} \log p_i + (1-\xi_{N+1}) \log(1-p_i).
\end{aligned} \tag{13.28}$$

Since $S_N \in M_N^{\{j\}}$, $p_j - \beta \leq \xi_{N+1} \leq p_j + \beta$. Now, the convex inequality [74, p. 643], which is, in fact, the relative entropy or Kullback–Leibler distance [22, p. 18],

$$p_j \log p_j + (1-p_j) \log(1-p_j) > p_j \log p_i + (1-p_j) \log(1-p_i),\ p_i \neq p_j,$$

and the continuity imply that for sufficiently small β, (13.28) holds. This concludes the proof. □

We now derive error bounds on the MAP algorithm.

Theorem 13.9. *Let $0 < \beta < \delta/2$ be a sufficiently small constant. For sufficiently small ε and sufficiently large N,*

$$P(\widehat{\theta}_N = \theta_N) \geq (1-\beta)2^{-(N+1)\beta}(1-\varepsilon)^N. \tag{13.29}$$

Proof. By Lemma 13.7, we may focus on $S_N \in M_N$. The probability of correct identification of θ_N is

$$\begin{aligned} P(\widehat{\theta}_N = \theta_N) &= \sum_{S_N} P(\widehat{\theta}_N = \theta_N | S_N) P(S_N) \\ &\geq \sum_{S_N \in M_N} P(\widehat{\theta}_N = \theta_N | S_N) P(S_N) \\ &= \sum_{j=1}^{m_0} \sum_{S_N \in M_N^{\{j\}}} P(\widehat{\theta}_N = \theta_N | S_N) P(S_N). \end{aligned}$$

By Lemma 13.8, for $S_N \in M_N^{\{j\}}$, $\widehat{\theta}_N = \theta^{(j)}$, which implies

$$P(\widehat{\theta}_N = \theta_N) \geq \sum_{j=1}^{m_0} \sum_{S_N \in M_N^{\{j\}}} P(\theta_N = \theta^{(j)} | S_N) P(S_N).$$

By (13.27),

$$\begin{aligned} &P(\theta_N = \theta^{(j)} | S_N) P(S_N) = P(\theta_N = \theta^{(j)}, S_N) \\ &\quad = p_0^{\{j\}} p_j^{(N+1)\xi_{N+1}} (1-p_j)^{(N+1)(1-\xi_{N+1})} (1-\varepsilon)^N + \varepsilon K. \end{aligned}$$

Let $\lambda_N^{\{j\}}$ be the cardinality of $M_N^{\{j\}}$. Then,

$$\begin{aligned} &\sum_{S_N \in M_N^{\{j\}}} p_j^{(N+1)\xi_{N+1}} (1-p_j)^{(N+1)(1-\xi_{N+1})} \\ &\quad = \lambda_N^{\{j\}} p_j^{(N+1)\xi_{N+1}} (1-p_j)^{(N+1)(1-\xi_{N+1})}. \end{aligned}$$

By [22, Theorem 3.1.2, p. 51], for sufficiently small ε and sufficiently large N,

$$\lambda_N^{\{j\}} \geq (1-\beta)2^{(N+1)(H(p_j)-\beta)}$$

for any small β, where

$$H(p_j) = -p_j \log p_j - (1-p_j)\log(1-p_j)$$

is the entropy of p_j. Denote

$$H(p_j, \xi_{N+1}) = -\xi_{N+1} \log p_j - (1 - \xi_{N+1}) \log(1 - p_j).$$

Since $\xi_N \to p_j$ w.p.1 as $N \to \infty$, $H(p_j, \xi_{N+1}) \to H(p_j)$ w.p.1 as $N \to \infty$. It follows that

$$\begin{aligned} &\lambda_N^{\{j\}} p_j^{(N+1)\xi_{N+1}} (1 - p_j)^{(N+1)(1-\xi_{N+1})} \\ &\quad \geq (1 - \beta) 2^{(N+1)(H(p_j) - H(p_j, \xi_{N+1}) - \beta)} \\ &\quad = (1 - \beta) 2^{-N\beta} 2^{-\beta + o(1)}, \end{aligned}$$

where $o(1) \to 0$ w.p.1 as $N \to \infty$. For sufficiently small β and sufficiently large N, $2^{-\beta+o(1)} > 1 - \beta$ since $\beta > 1 - 2^{-\beta}$. As a result, for sufficiently small ε and sufficiently large N,

$$\sum_{S_N \in M_N^{\{j\}}} P(\theta_N = \theta^{(j)} | S_N) P(S_N) > p_0^{\{j\}} (1 - \beta)^2 2^{-N\beta} (1 - \varepsilon)^N.$$

Therefore, (13.29) is obtained. □

Theorem 13.9 provides a guideline for window size selection. To achieve a required estimation accuracy for $0 < \eta < 1$, we may select the window size to be $(1 - \beta)^2 2^{-N\beta} (1 - \varepsilon)^N = \eta$ provided that ε is sufficiently small and β is sufficiently small.

13.5 Tracking Fast-Switching Systems

We begin this section by considering the scenario that the process $\theta(t)$ is a continuous-time Markov chain whose states vary on a fast pace. It is now understood that depending on the actual scenarios, only when the speed or frequency of state variations is relatively small, one expects to track the time-varying parameters with reasonable accuracy [6]. For instance, consider a continuous-time system whose observation is given by

$$y(t) = \varphi'(t)\theta(t) + w(t),$$

where $\varphi(t)$ is the input. Suppose that the parameter process $\theta(t)$ is a continuous-time Markov chain with a finite state space $\mathcal{M}$ and generator $Q^\eta = Q/\eta$ with $\eta > 0$ a small parameter. With Q being irreducible, when $\eta \to 0$, within a very short period of time $\theta(t)$ reaches its stationary distribution. In this case, it is virtually impossible to track the instantaneous variation of the process from observations of binary-valued outputs $s(t) = I_{\{y(t) \leq C\}}$. For such systems, the main goal becomes identifying an averaged system (averaging with respect to the stationary measure of the Markov chain). The main reason for focusing on the averaged system is the following: When a system performance is measured by some averaged outputs, as in most performance indices for optimal or adaptive control, the net effect of fast-switching parameters on the system performance can be approximated by using their average values.

The development of this section is motivated by the following scenario: For the above parameter process $\theta(t)$, denote its transition matrix by $P(t) = P^\eta(t)$. Then $P^\eta(t)$ satisfies the forward equation

$$\dot{P}^\eta(t) = P^\eta(t)Q/\eta.$$

A change of variables $\tau = t/\eta$ and $P(\tau) = P^\eta(t)$ leads to

$$\frac{d}{d\tau}P(\tau) = P(\tau)Q.$$

Discretizing the equation with a step size $h > 0$, we obtain a discrete matrix recursion

$$P^{k+1} = P^k[I + hQ].$$

By choosing $h > 0$ properly, $I+hQ$ becomes a one-step transition matrix of a discrete-time Markov chain θ_N and P^k represents the kth-step transition probability. In terms of the original fast-changing $\theta(t)$, we see that θ_N is corresponding to $\theta(N\eta h)$. When η is small, for a fixed time t, we have $N = t/(\eta h)$. That is, for the discrete-time system, we need to look at its property for N being large enough. We call such a chain a fast switching discrete-time Markov chain. Consequently, estimation of $\theta(t)$ for small η is reduced to estimation of θ_N for large n. For the problem treated in this section, in addition to the conditions posed previously, we make the following additional assumptions.

(A13.4) The Markov chain $\{\theta_n\}$ is irreducible and aperiodic.

It is observed that under Assumptions (A13.1), (A13.2), and (A13.4), if both $\mathcal{M}$ and P are unknown, then the stationary distribution $\nu = (\nu_1, \nu_2, \dots, \nu_{m_0})$ can be derived from P and the average w.r.t. the stationary measure can be calculated directly from

$$\overline{\theta} = \sum_{j=1}^{m_0} \nu_j \theta^{(j)}.$$

We will develop algorithms that estimate $\overline{\theta}$ without prior knowledge on P. Hence, we assume that $\mathcal{M}$ is known, but P is unknown. In this case, the goal is to identify ν from which $\overline{\theta}$ can be calculated.

13.5.1 Long-Run Average Behavior

Since $\nu_{m_0} = 1-(\nu_1+\cdots+\nu_{m_0-1})$, we need only identify m_0-1 parameters. For simplicity, we consider the observation horizon L with $L = N(m_0 - 1)$

for some positive integer N. Denote by $\mathbb{N}_0$ the following class of input signals:

$$\mathbb{N}_0 := \{u \in l^\infty : |u|_\infty \le K_u, u \text{ is } (m_0-1) - \text{periodic and full rank}\}.$$

Define the $(m_0-1)\times(m_0-1)$ matrix $\widetilde{M} = (\widetilde{m}_{ij})$, where

$$\widetilde{m}_{ij} = F(C - \phi_i'\theta^{(j)}) - F(C - \phi_i'\theta^{(m_0)}).$$

Let $\mathbb{N} := \{u \in \mathbb{N}_0\text{: } \widetilde{M} \text{ is full rank}\}$, define

$$\xi_N^{\{i\}} = \frac{1}{N}\sum_{l=0}^{N-1} s_{l(m_0-1)+i},\ i = 1,\ldots,m_0-1, \tag{13.30}$$

and denote $\xi_N = (\xi_N^{\{1\}},\ldots,\xi_N^{\{m_0-1\}})'$. It is easy to verify that

$$p_i = E\xi_N^{\{i\}} = \sum_{j=1}^{m_0-1} \nu_j(F(C-\phi_i'\theta^{(j)}) - F(C-\phi_i'\theta^{(m_0)})) + F(C-\phi_i'\theta^{(m_0)}).$$

Hence, $\xi_N^{\{i\}}$ represents the empirical measure of p_i. By defining

$$\begin{aligned} p &= [p_1,\ldots,p_{m_0-1}]',\\ b &= [F(C-\phi_1'\theta^{(m_0)}),\ldots,F(C-\phi_{m_0-1}'\theta^{(m_0)})]', \end{aligned}$$

and

$$\widetilde{\nu} = (\nu_1,\ldots,\nu_{m_0-1}) \in \mathbb{R}^{1\times(m_0-1)},$$

we obtain $p = \widetilde{M}\widetilde{\nu} + b$. This implies a relationship between p and $\widetilde{\nu}$, $\widetilde{\nu} = \widetilde{M}^{-1}(p-b)$. Since $\widetilde{M}$ and b are known from the input, this relationship implies that an estimate of $\widetilde{\nu}$ can be derived from the empirical measures of p, $\widehat{\nu}_N = \widetilde{M}^{-1}(\xi_N - b)$. From

$$\widehat{\nu}_N - \widetilde{\nu} = \widetilde{M}^{-1}(\xi_N - p),$$

the analysis of error bounds, convergence, and convergence rates of $\widehat{\nu}_N$ can be directly derived from that of ξ_N. For this reason, the remaining part of this section is devoted to the analysis of error bounds on empirical measures.

Example 13.10 The selection of inputs that will make the matrix M full rank is not difficult. For instance, suppose that the distribution is uniform with support on $[-30,30]$. Let the threshold be $C = 10$. The system has three states: $\theta_1 = [1,3]'$, $\theta_2 = [5,-3]'$, and $\theta_3 = [10,2]'$. The input is randomly selected to generate three regressors: $\phi_1' = [0.4565, 0.0185]$, $\phi_2' = [0.8214, 0.4447]$, and $\phi_3' = [0.6154, 0.7919]$. The M matrix becomes

$$\widetilde{M} = \begin{pmatrix} 0.6581 & 0.6296 & 0.5900 \\ 0.6307 & 0.6205 & 0.5149 \\ 0.6168 & 0.6550 & 0.5377 \end{pmatrix},$$

which is full rank.

13.5.2 Empirical Measure-Based Estimators

One immediate question is, what can one say about the asymptotic properties of the empirical measures defined above? From the well-known result of the Glivenko–Cantelli theorem ([8, p. 103]), in the usual empirical measure setup, the law of large numbers yields the convergence to the distribution function of the noise process if no switching is present. However, in the current setup, the empirical measures are coupled by a Markov chain. Intuitively, one would not doubt the existence of a limit. However, the additional random elements due to the Markov chain make the identification of the limit a nontrivial task. Corresponding to the above-mentioned law of large numbers, we first obtain the following result.

Theorem 13.11. *Under* (A13.1), (A13.2), *and* (A13.4),

$$\xi_N^{\{i\}} \to \sum_{j=1}^{m_0} \nu_j F(C - \phi_i'\theta^{(j)})$$

in probability as $N \to \infty$ *uniformly in* $i = 0, 1, 2, \ldots, m_0 - 1$.

Proof. For each $i = 0, 1, 2, \ldots, m_0 - 1$, the equalities $\phi_{lm_0+i} = \phi_i$ and $\widetilde{\phi}_{lm_0+i} = \widetilde{\phi}_i$ hold for all $l = 0, 1, \ldots, N-1$ due to the periodicity of the inputs, so

$$\begin{aligned} s_{lm_0+i} &= I_{\{y_{lm_0+i} \le C\}} = I_{\{\phi'_{lm_0+i}\theta_{lm_0+i} + d_{lm_0+i} \le C\}} \\ &= \sum_{j=1}^{m_0} I_{\{d_{lm_0+i} \le C - \phi_i'\theta^{(j)}\}} I_{\{\theta_{lm_0+i} = \theta^{(j)}\}} \\ &= \sum_{j=1}^{m_0} [I_{\{d_{lm_0+i} \le C - \phi_i'\theta^{(j)}\}} - \nu_j] I_{\{\theta_{lm_0+i} = \theta^{(j)}\}} \\ &\quad + \sum_{j=1}^{m_0} \nu_j I_{\{\theta_{lm_0+i} = \theta^{(j)}\}}. \end{aligned}$$

By virtue of the same argument as that of [122, p. 74], we have

$$E\left| \frac{1}{N} \sum_{l=0}^{N-1} I_{\{d_{lm_0+i} \le C - \phi_i'\theta^{(j)}\}} (I_{\{\theta_{lm_0+i} = \theta^{(j)}\}} - \nu_j) \right|^2 \to 0 \tag{13.31}$$

as $N \to \infty$, so

$$\frac{1}{N}\sum_{l=0}^{N-1} I_{\{d_{lm_0+i} \le C - \phi_i' \theta^{(j)}\}} (I_{\{\theta_{lm_0+i} = \theta^{(j)}\}} - \nu_j) \to 0$$

in probability and in the second moment as $N \to \infty$. Thus, it follows that

$$\xi_N^{\{i\}} = \sum_{j=1}^{m_0} \frac{1}{N} \sum_{l=0}^{N-1} I_{\{d_{lm_0+i} \le C - \phi_i' \theta^{(j)}\}} \nu_j + o(1), \tag{13.32}$$

where $o(1) \to 0$ in probability as $N \to \infty$. Note that

$$\frac{1}{N}\sum_{l=0}^{N-1} I_{\{d_{lm_0+i} \le C - \phi_i' \theta^{(j)}\}}$$

is the empirical distribution of the noise $\{d_N\}$ at $x = C - \phi_i' \theta^{(j)}$. Thus, by virtue of the well-known Glivenko–Cantelli theorem, for each $j = 1, \ldots, m_0$ and $i = 0, 1, \ldots, m_0 - 1$,

$$\frac{1}{N}\sum_{l=0}^{N-1} I_{\{d_{lm_0+i} \le C - \phi_i' \theta^{(j)}\}} \to F(C - \phi'(t_0 + i)\theta^{(j)}) \quad \text{as} \quad N \to \infty.$$

Thus, the desired result follows from the familiar Slutsky's result. □

The above result may be considered as the first approximation of the empirical measures to the weighted average of the distribution functions. Naturally, one would also like to know how fast the convergence will take place. This is presented in Theorem 13.12, which entails the study of the asymptotics of a centered and scaled sequence of errors or deviations. Compared with the results with a fixed parameter, it can be viewed as a hybrid coupling of discrete events with the normalized deviations. Conceptually, one expects that a rescaled sequence of the empirical measures should converge to a Brownian bridge suitably combined or coupled owing to the Markov chain in the original observation. In view of the above law of large numbers for empirical measures, one expects that the weak limit of the rescaled sequence should also be suitably combined by the stationary distributions of the Markov chain. Thus, it is not difficult to guess the limit. However, verifying this limit is not at all trivial. To illustrate, if we have two sequences $X_N^{\{1\}}$ and $X_N^{\{2\}}$ satisfying $X_N^{\{i\}} \to X^{\{i\}}$, $i = 1, 2$, in distribution as $N \to \infty$, we cannot conclude $X_N^{\{1\}} + X_N^{\{2\}} \to X^{\{1\}} + X^{\{2\}}$ in distribution generally. In our case, the difficulties are incurred by the presence of the Markov chain. In the proof of Theorem 13.12, we overcome the difficulty by establishing several claims. We first derive its asymptotic equivalence by bringing out the important part and discarding the asymptotically negligible part. We may call this step the decorrelation step. Next, we consider a suitably scaled sequence with a fixed-θ process. That is, we replace the "random jump" process with a fixed value. This replacement enables us to utilize a known result on the empirical process with a fixed parameter. The third step is to use finite-dimensional distributions convergence due to weak convergence of the empirical measures to treat an m-tuple

$$\left(\eta_N^{\{1\}}(\theta^{(1)}), \dots, \eta_N^{\{2\}}(\theta^{(m_0)})\right)$$

[the definition of $\eta_N^{\{i\}}(\theta)$ is given in (13.34) in what follows]. Finally, we use a Wold's device [8, p. 52] to finish the proof. The proof itself is interesting in its own right.

Theorem 13.12. *Assume the conditions of Theorem* 13.11. *The sequence*

$$\sqrt{N}\ \left[\xi_N^{\{i\}} - \sum_{j=1}^{m_0} \nu_j F(C - \phi_i' \theta^{(j)})\right]$$

converges weakly to

$$\sum_{j=1}^{m_0} \nu_j B(C - \phi_i' \theta^{(j)}),$$

where $B(\cdot)$ is a Brownian bridge process such that the covariance of $B(\cdot)$ (for $x_1,\ x_2 \in \mathbb{R}$) is given by

$$\begin{aligned} &EB(x_1)B(x_2) \\ &\qquad = \min(F(x_1), F(x_2)) - F(x_1)F(x_2). \end{aligned}$$

Proof. Step 1: Asymptotic equivalence: By virtue of [122, p. 74], similarly to (13.31), we can show that for each $j = 1, \dots, m_0$ and $i = 0, 1, \dots, m_0 - 1$,

$$\begin{aligned} &\frac{1}{N} \sum_{l=0}^{N-1} I_{\{d_{lm_0+i} \le C - \phi_i' \theta^{(j)}\}} I_{\{\theta_{lm_0+i} = \theta^{(j)}\}} \\ &\qquad = \frac{1}{N} \sum_{l=0}^{N-1} I_{\{d_{lm_0+i} \le C - \phi_i' \theta^{(j)}\}} \nu_j + o(1), \end{aligned}$$

where $o(1) \to 0$ in probability (and also in the second moment) as $N \to \infty$. This together with (13.32) leads to

$$\begin{aligned} &\sqrt{N} \left[\xi_N^{\{i\}} - \sum_{j=1}^{m_0} \nu_j F(C - \phi_i' \theta^{(j)})\right] \\ &= \sum_{j=1}^{m_0} \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} [I_{\{d_{lm_0+i} \le C - \phi_i' \theta^{(j)}\}} - F(C - \phi_i' \theta^{(j)})]\nu_j + o(1), \end{aligned} \tag{13.33}$$

where $o(1) \to 0$ in probability as $N \to \infty$.

Step 2: Convergence in distribution of a fixed-θ process: Consider now a typical term in the last line of (13.33). For convenience, for a fixed θ, define

$$\eta_N^{\{i\}}(\theta) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} \Big[I_{\{d_{lm_0+i} \le C - \phi_i' \theta\}} - F(C - \phi_i'\theta) \Big]. \tag{13.34}$$

It is readily seen that $\eta_N^{\{i\}}(\theta)$ is a centered empirical measure (with a fixed θ) re-scaled by $\sqrt{N}$. The results on empirical measures (see [8, p. 141], [76], and also [84]) then imply that $\eta_N^{\{i\}}(\cdot)$ converges weakly to a Brownian bridge process $B(C - \phi_i' \cdot)$ with mean 0 and covariance

$$\begin{aligned} &EB(C - \phi_i'\theta_1)B(C - \phi_i'\theta_2) \\ &\quad = \min(F(C - \phi_i'\theta_1), F(C - \phi_i'\theta_2)) - F(C - \phi_i'\theta_1)F(C - \phi_i'\theta_2). \end{aligned}$$

Step 3: Convergence of finite-dimensional distributions: Since $\eta_N^{\{i\}}(\cdot)$ converges weakly to $B(C - \phi_i'\cdot)$, its finite-dimensional distributions converge. That is, for any integer p and any $(x_1, \ldots, x_p)$, $(\eta_N^{\{i\}}(x_1), \ldots, \eta_N^{\{i\}}(x_p))$ converges in distribution to $(B(C - \phi_i'x_1), \ldots, B(C - \phi_i'x_p))$. In particular, we have that $(\eta_N^{\{i\}}(\theta^{(1)}), \ldots, \eta_N^{\{i\}}(\theta^{(p)}))$ converges in distribution to $(B(C - \phi_i'\theta^{(1)}), \ldots, B(C - \phi_i'\theta^{(p)}))$.

Step 4: The weak convergence and the form of the finite-dimensional distributional convergence in Step 3 imply that

$$(\nu_1, \ldots, \nu_{m_0})' \Big(\eta_N^{\{i\}}(\theta^{(1)}), \ldots, \eta_N^{\{i\}}(\theta^{(m_0)}) \Big)$$

converges in distribution to $\sum_{j=1}^{m_0} \nu_j B(C - \phi_i'\theta^{(j)})$ by Wold's device [8, p. 52]. Finally, putting all the steps together, the desired result follows. □

Note that a Brownian bridge is a Brownian motion tied down at both ends. Here we emphasize that the process considered is allowed to take values not just in $[0, 1]$, but in the entire real line; thus, what we have is a "stretched" Brownian bridge as discussed in Chapter 3; see also [76]. Similarly to Lemma 13.5, the next lemma provides a strong approximation result for empirical measures. Its detailed proof is omitted for brevity.

Lemma 13.13 *Under the conditions of Theorem* 13.4*, there is a constant* $\gamma > 0$ *such that*

$$\begin{aligned} &\sup_{0 \le i \le m_0 - 1} \left| \sqrt{N} \left[\xi_N^{\{i\}} - \sum_{j=1}^{m_0} \nu_j F(C - \phi_i'\theta^{(j)}) \right] - \sum_{j=1}^{m_0} \nu_j B(C - \phi_i'\theta^{(j)}) \right| \\ &\quad = o(N^{-\gamma}) \text{ w.p.1.} \end{aligned}$$

13.5.3 Estimation Errors on Empirical Measures: Upper and Lower Bounds

In the context of system identification, estimation error bounds are of crucial importance. This section obtains such bounds for the fast-varying systems. As a preparation, we first present a lemma, which is an exponential estimate for a Gaussian process.

Lemma 13.14 *Under the assumptions of Theorem* 13.11*, for N sufficiently large and for each $j = 1, \dots, m_0$,*

$$P\left(\left| \frac{1}{\sqrt{N}} B(C - \phi_i' \theta^{(j)}) \right| \geq \frac{\varepsilon}{m_0 M} \right) \leq 2 \exp\left(-\frac{2N\varepsilon^2}{{m_0}^2 M^2} \right), \qquad (13.35)$$

where $M = \max\{\nu_1, i \leq m_0\}$ and $B(\cdot)$ is given by Theorem 13.12*.*

Proof. Let $\sigma_{ij}^2 = \mathrm{Var}(B(C - \phi_i' \theta^{(j)}))$. By direct computation, one can show that for any $\alpha > 0$,

$$E \exp(\alpha | \frac{1}{\sqrt{N}} B(C - \phi_i' \theta^{(j)})|) \leq 2 \exp(\frac{\alpha^2 \sigma_{ij}^2}{2N}).$$

Thus,

$$\begin{aligned} P &\left(\alpha \left| \frac{1}{\sqrt{N}} B(C - \phi_i' \theta^{(j)}) \right| \geq \frac{\alpha\varepsilon}{m_0 M} \right) \\ &\leq \exp\left(-\frac{\alpha\varepsilon}{m_0 M} \right) E \exp\left(\alpha \left| \frac{1}{\sqrt{N}} B(C - \phi_i' \theta^{(j)}) \right| \right) \\ &\leq 2 \exp\left(\frac{\sigma_{ij}^2}{2N} \alpha^2 - \frac{\varepsilon}{m_0 M} \alpha \right). \end{aligned}$$

Choose $\alpha = N_\varepsilon / (m_0 M \sigma_{ij}^2)$ to minimize the quadratic form in the exponent above and note that

$$\sigma_{ij}^2 = F(C - \phi_i' \theta^{(j)})(1 - F(C - \phi_i' \theta^{(j)})) \leq \frac{1}{4}.$$

Then the upper bound is obtained. □

It can be seen that Lemma 13.14 derives an exponential type of upper bound on the estimation errors. To some extent, it is a large deviations result. Note that $B(\cdot)$, the Brownian bridge process, is a Gaussian process. The deviation given above indicates that the "tail" probabilities of deviations of the order $O(\sqrt{N})$ will be exponentially small. With this lemma, we can proceed to obtain the "large deviations" of the empirical measures from those of the averaged distribution functions (average with respect to the stationary distributions of the Markov chain).

Theorem 13.15. *Under the assumptions of Theorem* 13.11*, for N large enough and for any $\varepsilon > 0$,*

$$P\left(\left|\xi_N^{\{i\}} - \sum_{j=1}^{m_0} \nu_j F(C - \phi_i'\theta^{(j)})\right| \geq \varepsilon\right) \leq 2m_0 \exp\left(-\frac{2N\varepsilon^2}{{m_0}^2 M^2}\right). \tag{13.36}$$

Proof. It is easy to see that

$$P\left(\left|\frac{1}{\sqrt{N}}\sum_{j=1}^{m_0} \nu_j B(C - \phi_i'\theta^{(j)})\right| \geq \varepsilon\right) \leq P\left(\sum_{j=1}^{m_0}\left|\frac{1}{\sqrt{N}} B(C - \phi_i'\theta^{(j)})\right| \geq \frac{\varepsilon}{M}\right).$$

Recall that $\widetilde{c}_{ij}$ is given in Theorem 13.16. Observe that

$$\begin{aligned}&\left\{(\tilde{c}_{i1}, \ldots, \tilde{c}_{im_0})' : \sum_{j=1}^{m_0}\left|\frac{1}{\sqrt{N}} B(\tilde{c}_{ij})\right| \geq \frac{\varepsilon}{M}\right\}\\ &\subseteq \left\{(\tilde{c}_{i1}, \ldots, \tilde{c}_{im_0})' : \left|\frac{1}{\sqrt{N}} B(\tilde{c}_{ij})\right| \geq \frac{\varepsilon}{m_0 M} \text{ for some } j\right\}\\ &\subseteq \bigcup_{j=1}^{m_0}\left\{(\tilde{c}_{i1}, \ldots, \tilde{c}_{im})' : \left|\frac{1}{\sqrt{k}} B(\tilde{c}_{ij})\right| \geq \frac{\varepsilon}{m_0 M}\right\}.\end{aligned}$$

Thus, by (13.36),

$$P\left(\left|\frac{1}{\sqrt{N}}\sum_{j=1}^{m_0} \nu_j B(C - \phi_i'\theta^{(j)})\right| \geq \varepsilon\right) \leq \sum_{j=1}^{m_0} P\left(\left|\frac{1}{\sqrt{N}} B(\tilde{c}_{ij})\right| \geq \frac{\varepsilon}{m_0 M}\right) \leq 2m_0 \exp\left(-\frac{2N\varepsilon^2}{{m_0}^2 M^2}\right).$$

For sufficiently large N, the desired result follows from Lemma 13.13 and the same kind of argument as in the proof of Theorem 13.6. □

Next, we proceed to obtain lower bounds on the estimation error when full-rank periodic inputs are used.

Theorem 13.16. *Denote*

$$\widetilde{c}_{ij} = C - \phi_i'\theta^{(j)}$$

for each $i = 0, \ldots, m_0 - 1$ and $j = 1, \ldots, m_0$, and denote the matrix

$$\Sigma = (\tilde{\sigma}_{i_1,i_2} : i_1, i_2 = 1, \ldots, m_0),$$

where for $i_1, i_2 = 1, \ldots, m_0$,

$$\tilde{\sigma}_{i_1,i_2} = \min(F(\tilde{c}_{ii_1}), F(\tilde{c}_{ii_2})) - F(\tilde{c}_{ii_1})F(\tilde{c}_{ii_2}).$$

Let λ *be the minimum eigenvalue of the covariance matrix* Σ*. Under the assumptions of Theorem* 13.11*, for sufficiently large* N *and for any* $\varepsilon > 0$*,*

$$\begin{aligned} P &\left(\left| \xi_N^{\{i\}} - \sum_{j=1}^{m_0} \nu_j F(C - \phi_i' \theta^{(j)}) \right| \geq \varepsilon \right) \\ &\geq \sqrt{\frac{2}{\pi}} \left(\frac{\sqrt{\lambda}|\nu|}{\sqrt{N}\varepsilon} - \left(\frac{\sqrt{\lambda}|\nu|}{\sqrt{N}\varepsilon} \right)^3 \right) \exp\left(-\frac{\varepsilon^2}{2\lambda|\nu|^2} N \right). \end{aligned} \tag{13.37}$$

Proof. Note that $(B(\tilde{c}_{i1}), \ldots, B(\tilde{c}_{im_0}))$ can be regarded as a multinormal distributed vector with mean 0 and covariance Σ. Recall that

$$\nu = (\nu_1, \ldots, \nu_{m_0}) \;\text{ and }\; \check{S} = \sum_{j=1}^{m_0} \nu_j B(\tilde{c}_{ij}).$$

Then $\check{S}$ is a one-dimensional Gaussian random variable with mean 0 and variance $\nu'\Sigma\nu$. Direct computation yields that

$$P\left(\left| \frac{1}{\sqrt{N}} \sum_{j=1}^{m_0} \nu_j B(\tilde{c}_{ij}) \right| \geq \varepsilon \right) = P\left(|Z| \geq \frac{\sqrt{N}\varepsilon}{\sqrt{\nu'\Sigma\nu}} \right),$$

where $Z = \check{S}/(\sqrt{\nu'\Sigma\nu})$. Then

$$\begin{aligned} P &\left(\left| \frac{1}{\sqrt{N}} \sum_{j=1}^{m_0} \nu_j B(\tilde{c}_{ij}) \right| \geq \varepsilon \right) \geq 2P\left(Z \geq \frac{\sqrt{N}\varepsilon}{\sqrt{\lambda}|\nu|} \right) \\ &= 2\left(1 - \Phi\left(\frac{\sqrt{N}\varepsilon}{\sqrt{\lambda}|\nu|} \right) \right) \geq 2\left(\frac{\sqrt{\lambda}|\nu|}{\sqrt{N}\varepsilon} - \left(\frac{\sqrt{\lambda}|\nu|}{\sqrt{N}\varepsilon} \right)^3 \right) \check{\varphi}\left(\frac{\sqrt{N}\varepsilon}{\sqrt{\lambda}|\nu|} \right), \end{aligned}$$

where $\Phi(\cdot)$ and $\check{\varphi}(\cdot)$ are the cumulative distribution and density function of the standard normal variable, respectively. Thus, for sufficiently large N,

$$\begin{aligned} P &\left(\left| \xi_N^{\{i\}} - \sum_{j=1}^{m_0} \nu_j F(C - \phi_i' \theta^{(j)}) \right| \geq \varepsilon \right) \\ &\geq \sqrt{\frac{2}{\pi}} \left(\frac{\sqrt{\lambda}|\nu|}{\sqrt{N}\varepsilon} - \left(\frac{\sqrt{\lambda}|\nu|}{\sqrt{N}\varepsilon} \right)^3 \right) \exp\left(-\frac{\varepsilon^2}{2\lambda|\nu|^2} N \right). \end{aligned}$$

The proof is concluded. □

13.6 Notes

Although the problems considered in this chapter are centered around switching systems with binary observations, the main ideas and results can be generalized to quantized observations. Regime-switching systems often appear as integrated parts of hybrid systems, discrete-event systems, logic-based systems, finite automata, hierarchical systems, and complex systems. Consequently, our results may have potential applications in these areas as well. More information on regime-switching systems can be found in [123] and its references.

The framework here is based on our recent work [120] for tracking Markovian parameters with binary-valued observations. Several directions may be pursued. The inclusion of unmodeled dynamics is a worthwhile research direction. Quantized sensors may be treated. Optimal sensor placement in conjunction with the filters developed in this chapter may be considered. Optimal selection of the threshold values and sensor locations is an important issue. Complexity is another direction for further investigation.