The Annals of Applied Probability
1996, Vol. 6, No. 2, 650–670

A CENTRAL LIMIT THEOREM FOR SINGULARLY PERTURBED NONSTATIONARY FINITE STATE MARKOV CHAINS

BY Q. ZHANG[1] AND G. YIN[2]

University of Georgia and Wayne State University

This work is concerned with the asymptotic properties of a singular perturbed nonstationary finite state Markov chain. In a recent paper of the authors, it was shown that as the fluctuation rate of the Markov chain goes to ∞, the probability distribution of the Markov chain converges to its time-dependent quasi-equilibrium distribution. In addition, asymptotic expansion of the probability distribution was obtained. This paper is a continuation of our effort in this direction. Upon using the asymptotic expansion, a suitably scaled sequence is examined in detail. Asymptotic normality is obtained. It is shown that the accumulated difference between the indicator process and the quasi-equilibrium distribution converges to a Gaussian process with zero mean. An explicit formula for the covariance function of the Gaussian process is obtained, which depends crucially on the asymptotic expansion.

1. Introduction. The main goal of this work is to develop asymptotic properties, more specifically, asymptotic normality of a singularly perturbed nonstationary Markov chain with finite state space. Consider a finite state Markov chain $\alpha^\varepsilon(t) \in \mathscr{M} := \{0, 1, \dots, m\}$, $t \geq 0$, for a positive integer m. This process has an infinitesimal generator

$$\frac{1}{\varepsilon}Q(t) = \frac{1}{\varepsilon}\begin{pmatrix} -q_0(t) & q_{01}(t) & \cdots & q_{0m}(t) \\ q_{10}(t) & -q_1(t) & \cdots & q_{1m}(t) \\ \vdots & \vdots & \dots & \vdots \\ q_{m0}(t) & q_{m1}(t) & \cdots & -q_m(t) \end{pmatrix}, \tag{1}$$

where $q_{ij}(t) \geq 0$ if $i \neq j$, $q_i(t) = \sum_{j\neq i} q_{ij}(t)$ and $\varepsilon > 0$ is a small parameter. In this paper we show that, as $\varepsilon \to 0$, a normalized sequence of occupation times of $\alpha^\varepsilon(t)$, $t \geq 0$, converges weakly to a Gaussian process; we characterize the limit process by giving explicit formulas for the mean and covariance functions.

The process $\alpha^\varepsilon(t)$ arises from a wide range of applications which involve a rapidly fluctuating finite state Markov chain. In many problems the underlying Markov chain depends on a small parameter $\varepsilon > 0$. Roughly, as ε gets smaller and smaller, the Markov chain $\alpha^\varepsilon(\cdot)$ fluctuates more and more rapidly.

Received November 1994; revised January 1996.

AMS 1991 *subject classifications*. 60F05, 60J27, 34E05, 93E20.

Key words and phrases. Finite state Markov chain, singular perturbation, Gaussian process.

[1]Research supported in part by the Natural Sciences and Engineering Research Council of Canada Grant A4169, and in part by Office of Naval Research Grant N00014-96-1-0263.

[2]Research supported in part by NSF Grants DMS-92-24372 and DMS-95-29378, and in part by Wayne State University.

Such Markov chains can be used to represent fast changing processes in many real-world problems (see Sethi and Zhang [12]). We are interested in the limit behavior as $\varepsilon \to 0$. To proceed, an example in hierarchical production planning of a manufacturing system is given below.

EXAMPLE 1. Let us consider the production planning of a failure-prone manufacturing system. The system consists of a single machine that has $m+1$ states denoted by $\mathscr{M} = \{0, 1, \dots, m\}$. Let $\alpha^\varepsilon(t)$ denote the machine capacity process with $\alpha^\varepsilon(t) \in \mathscr{M}$. Let $x(t)$, $u(t)$ and z (constant) denote the surplus, the rate of production and the rate of demand, respectively. The system equation is

$$\frac{d}{dt} x(t) = u(t) - z, \qquad x(0) = x.$$

Our objective is to choose the rate of production $u(\cdot) = \{u(t)\colon\ t \ge 0\}$, subject to the production constraint $0 \le u(t) \le \alpha^\varepsilon(t)$, over time so as to minimize a cost functional

$$J(x, \alpha, u(\cdot)) = E \int_0^\infty e^{-\rho t} G(x(t), u(t))\, dt,$$

where (x, α) is the initial state of $(x(t), \alpha^\varepsilon(t))$, $\rho > 0$ is a discount factor and $G(\cdot)$ is a cost-to-go function.

In general, the optimal solution of the aforementioned problem is not available. Moreover, the generator of the Markov chain may depend on the production rate in a rather complex way; for example, the Markov chain $\alpha^\varepsilon(t)$ is generated by $\varepsilon^{-1} Q(u(t))$ (see Sethi and Zhang [12]). It is shown in [12] that, as ε tends to 0, the problem given above can be approximated by a limiting control problem. To be more specific, let Γ denote a set given by

$$\Gamma = \{U = (u^0, u^1, \dots, u^m)\colon\ 0 \le u^i \le i,\ i = 0, 1, \dots, m\}.$$

Consider a control space $\mathscr{A}^0 = \{U(t) \in \Gamma\}$, where $U(t) = \{u^0(t), u^1(t), \dots, u^m(t)\}$ is a deterministic process. For any $U \in \mathscr{A}^0$, let $\bar{Q}(U) = (q_{ij}(u^i))$. The limiting problem can be described as

$$\text{minimize} \qquad J^0(x, U(\cdot)) = \int_0^\infty e^{-\rho t} \sum_{i=0}^m \nu_i(U(t)) G(x(t), u^i(t))\, dt,$$

$$\text{subject to} \qquad \dot{x}(t) = \sum_{i=0}^m \nu_i(U(t)) u^i(t) - z, \qquad x(0) = x, \qquad U(\cdot) \in \mathscr{A}^0,$$

where $\nu_i(U(t))$ is the quasi-equilibrium distribution of $\bar{Q}(U(t))$, given as in Definition 1 to follow. Let $U^*(x)$ denote a Lipschitz optimal feedback control for the limiting problem. Then, using $U^*(x) = (u^{*,0}(x), \dots, u^{*,m}(x))$, one can construct a control for the original problem as follows:

$$u^*(x, \alpha) = \sum_{i=0}^m \chi_{\{\alpha = i\}} u^{*,i}(x),$$

where χ_F is the indicator function of a set F. It can be shown that $u^*(x, \alpha)$ is feasible, that is, $u^*(x, \alpha) \in \Gamma$, and $u^\varepsilon(t) = u^*(x(t), \alpha^\varepsilon(t))$ is asymptotically optimal in the sense

$$\lim_{\varepsilon \to 0}\Big(J(x, \alpha, u^\varepsilon(\cdot)) - \min_{u(\cdot)} J(x, \alpha, u(\cdot))\Big) = 0.$$

As indicated in [12], the fast changing process $\alpha^\varepsilon(\cdot)$ in the physical system is normally hard to analyze. The desired limiting properties, however, provide us with an alternative. We can replace the physical process by its "average." This approach has significant practical value.

In studying such problems, the asymptotic behavior of the Markov chain $\alpha^\varepsilon(t)$ has a major influence. Further investigation and understanding of the asymptotic properties of $\alpha^\varepsilon(t)$ play an important role for in-depth study of the hierarchical controls of stochastic manufacturing systems and many other applications, such as queueing networks, random fatigue analysis, system reliability and related fields. This brings us to the current work. To proceed, we first recall the definition of quasi-equilibrium distribution.

DEFINITION 1 (Quasi-equilibrium distribution; see [9]). Let $\nu(t) = (\nu_0(t), \ldots, \nu_m(t))$ denote a vector with nonnegative components for all $t \geq 0$. Here $\nu(t)$ is called a quasi-equilibrium distribution of a Markov chain $\alpha(t)$ with generator $Q(t)$ if

$$\nu(t)Q(t) = 0 \quad \text{and} \quad \sum_{i=0}^{m} \nu_i(t) = 1.$$

In a recent paper of Khasminskii, Yin and Zhang [9] (see also the related work [8]), using singular perturbation techniques, the authors examined the following system:

$$\begin{aligned} &\frac{d}{dt} y^\varepsilon(t) = y^\varepsilon(t) Q^\varepsilon(t), \qquad y^\varepsilon(t) \in \mathbb{R}^{m+1}, \\ &y^\varepsilon(0) = y_0 = (y_{0,0}, \ldots, y_{0,m}) \\ &\qquad\qquad \text{with } y_{0,i} \geq 0 \text{ for } 0 \leq i \leq m \text{ and } \sum_{i=0}^{m} y_{0,i} = 1, \end{aligned} \tag{2}$$

where $Q^\varepsilon(t) = Q(t)/\varepsilon$, $y^\varepsilon(t) = (y_0^\varepsilon(t), \ldots, y_m^\varepsilon(t))$ with $y_i^\varepsilon(t) = P(\alpha^\varepsilon(t) = i)$. It was proven that under suitable conditions $y^\varepsilon(t) \to \nu(t)$ as $\varepsilon \to 0$ for each $t \in [0, T]$, and $y^\varepsilon(t)$ admits an asymptotic expansion in terms of ε which is more far reaching than the convergence to the quasi-equilibrium distribution alone. The asymptotic series involves a regular part as well as boundary layer corrections. It is conceivable that such a result will help us to develop further asymptotic properties for the control problem mentioned above, in particular, to obtain error bounds on the difference between the value function and the actual cost function with the asymptotic optimal control substituted.

To gain further insight, one may ask whether there is a central limit theorem type of result associated with the $\alpha^\varepsilon(\cdot)$-process. This paper provides an affirmative answer to this question. Our effort is devoted to proving a weak convergence theorem for a suitably scaled sequence of occupation times. Owing to the asymptotic expansion, the scaling factor is $\sqrt{\varepsilon}$. We show that the limit process is Gaussian with zero mean, and give explicit representation for the covariance of the limit process. The covariance function, which depends on the asymptotic expansion in an essential way, reflects one of the distinct features of the central limit theorem. It can be seen in the sequel that it is impossible in general to obtain the central limit results without the help of the asymptotic expansion. This reveals a very different characteristic than that of the existing results of central limit type.

The rest of the paper is arranged as follows. Section 2 concerns the precise formulation of the problem, and Section 3 presents the assumptions and main results. In addition to the central limit theorem, we obtain certain exponential bounds and moment estimates as by-products. Loosely, although the prelimit process has nonzero mean and is nonstationary, by using the result of [9], after a short time period of the order ε, the quasi-equilibrium regime is established. Thus, effectively away from the ε-layer (i.e., for $\varepsilon < t \le T$), the mean of the process of interest is essentially 0. Due to the asymptotic expansion, the process $\alpha^\varepsilon(\cdot)$ satisfies an "ε-mixing" condition as ε goes to 0. We prove that the sequence is tight and that the sample paths of the limit process are continuous in probability by estimating the fourth moment. Similar to Khasminskii [7], the finite-dimensional distributions can be calculated and shown to be Gaussian. We then prove that the mean of the limit process is 0, and calculate explicitly the covariance representation for the limit process. In addition, we give an example to illustrate the results in this paper. Finally, some additional remarks are made in Section 4. The proof of a technical lemma is included in the Appendix.

2. Problem formulation. Let $(\Omega, \mathscr{F}, P)$ denote a probability space. Let $\alpha^\varepsilon(\cdot)$ be a nonstationary Markov chain on $(\Omega, \mathscr{F}, P)$ with finite state space $\mathscr{M} = \{0, 1, \dots, m\}$ as given in the Introduction. Suppose that the infinitesimal generator of the chain is as in (1).

Let $\beta_i(s)$, $i \in \mathscr{M}$, denote a bounded Borel-measurable deterministic process. Consider the process $\Lambda^\varepsilon(t, i)$ defined as

$$\begin{aligned} \Lambda^\varepsilon(t, i) &= \frac{1}{\sqrt{\varepsilon}} \int_0^t \big(\chi_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)\big)\beta_i(s)\, ds, \\ \Lambda^\varepsilon(t) &= (\Lambda^\varepsilon(t, 0), \dots, \Lambda^\varepsilon(t, m)). \end{aligned} \tag{3}$$

It is expected that the accumulated process of the indicator function should demonstrate certain "central limit type" phenomena. The main goal of this work is to study the asymptotic properties of $\Lambda^\varepsilon(\cdot)$ as $\varepsilon \to 0$. To be more specific, we will show that $\Lambda^\varepsilon(\cdot)$ converges to a Gaussian process $\Lambda(\cdot)$ as $\varepsilon \to 0$.

Notice that a special choice of $\beta_i(\cdot)$ is $\beta_i(t) \equiv 1$. Inserting $\beta_i(\cdot)$ in the scaled sequence allows one to treat more general situations in various applications. For example, in the manufacturing problem, $\beta_i(t)$ is given by a function of a control process. Throughout the paper we will need the notation of weak irreducibility as defined in Khasminskii, Yin and Zhang [9].

DEFINITION 2. For each $t \geq 0$, the matrix $Q(t)$ is weakly irreducible if

$$\nu(t)Q(t) = 0 \quad \text{and} \quad \sum_{i=0}^{m} \nu_i(t) = 1 \tag{4}$$

has a unique nonnegative solution $\nu(t)$.

NOTATION. To proceed, a word about the notation is in order. In the sequel vectors are referred to as row vectors. The transpose of a vector or a matrix A will be denoted by A'. Subscripts will be used to denote the components of a vector, for instance, $x = (x_0, x_1, \ldots, x_m)$. The indicator function of a set F is written as χ_F. Here $O(y)$ is a function of y such that $\sup_y |O(y)|/|y| < \infty$. For notational simplicity, C is utilized to denote a generic positive constant (which does not depend on $\varepsilon > 0$). Its values, however, may change for different usages. Thus, $C + C = C$ and $CC = C$ are understood in an appropriate sense.

3. Main results. Let $p^\varepsilon(t) = (P(\alpha^\varepsilon(t) = 0), \ldots, P(\alpha^\varepsilon(t) = m))$ and let $p^\varepsilon_{ij}(t, t_0) = P(\alpha^\varepsilon(t) = j | \alpha^\varepsilon(t_0) = i)$ for all $i, j \in \mathscr{M}$. We use $P^\varepsilon(t, t_0)$ to denote the transition matrix $(p^\varepsilon_{ij}(t, t_0))$.

We make the following assumptions.

ASSUMPTION 1. For each $t \in [0, T]$, $Q(t)$ is weakly irreducible.

ASSUMPTION 2. $Q(\cdot)$ is three times continuously differentiable on $[0, T]$.

Next we give two lemmas on asymptotic expansions. The proof of the first lemma can be found in [9], whereas the second one is a consequence of the first lemma.

LEMMA 1. *Under Assumptions* 1 *and* 2, *by use of* (2), *we obtain*, *for each fixed* $0 \leq T < \infty$,

$$p^\varepsilon(t) = p^{\{0\}}(t) + \varepsilon p^{\{1\}}(t) + q^{\{0\}}\left(\frac{t}{\varepsilon}\right) + \varepsilon q^{\{1\}}\left(\frac{t}{\varepsilon}\right) + O(\varepsilon^2),$$

uniformly in $t \in [0, T]$, *where* $p(\cdot)$ *and* $q(\cdot)$ *can be constructed via*

$$\begin{aligned} p^{\{0\}}(t) &= \nu(t), \\ p^{\{1\}}(t) Q(t) &= \frac{d}{dt}\nu(t), \\ \sum_{i=0}^{m} p^{\{1\}}(t) &= 0, \end{aligned}$$

$$\begin{aligned} \frac{d}{ds} q^{\{0\}}(s) &= q^{\{0\}}(s) Q(0), \qquad s \geq 0, \\ q^{\{0\}}(0) &= y_0 - \nu(0), \end{aligned}$$

where y_0 *is the initial probability distribution of* $\alpha^\varepsilon(t)$, *and*

$$\begin{aligned} \frac{d}{ds} q^{\{1\}}(s) &= q^{\{1\}}(s) Q(0) + s q^{\{0\}}(s) Q^{(1)}(0), \qquad s \geq 0, \\ q^{\{1\}}(0) &= -p^{\{1\}}(0), \end{aligned}$$

where $s := t/\varepsilon$ *is a stretched time variable. Moreover,* $p^{\{i\}}(\cdot)$ *is* $(3-i)$ *times continuously differentiable on* $[0, T]$ *for* $i = 0, 1$, *and there exist constants* $C > 0$ *and* $\kappa > 0$ *such that*

$$\left| q^{\{0\}}\left(\frac{t}{\varepsilon}\right) \right| \leq C e^{-\kappa t/\varepsilon} \quad \text{and} \quad \left| q^{\{1\}}\left(\frac{t}{\varepsilon}\right) \right| \leq C e^{-\kappa t/\varepsilon}.$$

LEMMA 2. *Under Assumptions* 1 *and* 2, *we have, for each fixed* $0 \leq T < \infty$,

$$P^\varepsilon(t, t_0) = P^{\{0\}}(t) + \varepsilon P^{\{1\}}(t) + Q^{\{0\}}\left(\frac{t - t_0}{\varepsilon}, t_0\right) + \varepsilon Q^{\{1\}}\left(\frac{t - t_0}{\varepsilon}, t_0\right) + O(\varepsilon^2),$$

uniformly in $t_0, t \in [0, T]$, *where*

$$P^{\{0\}}(t) = \begin{pmatrix} \nu(t) \\ \vdots \\ \nu(t) \end{pmatrix}, \qquad P^{\{1\}}(t) = \begin{pmatrix} p^{\{1\}}(t) \\ \vdots \\ p^{\{1\}}(t) \end{pmatrix},$$

$$\begin{aligned} \frac{d}{ds} Q^{\{0\}}(s, t_0) &= Q^{\{0\}}(s, t_0) Q(t_0), \qquad s \geq 0, \\ Q^{\{0\}}(0, t_0) &= I - P^{\{0\}}(t_0) \end{aligned}$$

and

$$\begin{aligned} \frac{d}{ds} Q^{\{1\}}(s, t_0) &= Q^{\{1\}}(s, t_0) Q(t_0) + s Q^{\{0\}}(s, t_0) Q^{(1)}(t_0), \qquad s \geq 0, \\ Q^{\{1\}}(0, t_0) &= -P^{\{1\}}(t_0), \end{aligned}$$

where $Q^{(1)}(t_0)$ *is the derivative of* $Q(t)$ *at* $t = t_0$, *and* $s := (t - t_0)/\varepsilon$. *Furthermore,* $P^{\{i\}}(\cdot)$ *is* $(3 - i)$ *times continuously differentiable on* $[0, T]$ *for* $i = 0, 1$, *and there exist constants* $C > 0$ *and* $\kappa > 0$ *such that*

$$\left|Q^{\{0\}}\left(\frac{t-t_0}{\varepsilon}, t_0\right)\right| \le C\exp\left(-\frac{\kappa(t-t_0)}{\varepsilon}\right) \quad \textit{and} \quad \left|Q^{\{1\}}\left(\frac{t-t_0}{\varepsilon}, t_0\right)\right| \le C\exp\left(-\frac{\kappa(t-t_0)}{\varepsilon}\right).$$

REMARK 1. Notice that $P^{\{0\}}(\cdot)$ and $P^{\{1\}}(\cdot)$ have identical rows. This is a consequence of the convergence of $y^\varepsilon(t)$ to the quasi-equilibrium distribution.

These two lemmas show that the probability distribution of $\alpha^\varepsilon(t)$ can be approximated by its average distribution. Note that the rate of fluctuation of $\alpha^\varepsilon(t)$ goes to ∞ as $\varepsilon \to 0$.

THEOREM 1. *Suppose that Assumptions* 1 *and* 2 *are satisfied. For* $t \in [0, T]$, *the process* $\Lambda^\varepsilon(\cdot)$ *converges weakly to a Gaussian process* $\Lambda(\cdot)$ *with independent increments such that*

$$E\Lambda(t) = 0 \quad \textit{and} \quad E[\Lambda'(t)\Lambda(t)] = \int_0^t A(s)\,ds, \tag{5}$$

where $A(t) = (A_{ij}(t))$ *with*

$$A_{ij}(t) = \beta_i(t)\beta_j(t)\left[\nu_i(t)\int_0^\infty q_{ij}^{\{0\}}(r, t)\,dr + \nu_j(t)\int_0^\infty q_{ji}^{\{0\}}(r, t)\,dr\right], \tag{6}$$

where $q_{ij}^{\{0\}}(r, t)$ *denotes the ijth entry of* $Q^{\{0\}}(r, t)$.

REMARK 2. In the stationary case, that is, when $Q(t) = Q$ is a constant matrix, the central limit theorem can be obtained as in [6]. See also [10] for related discussions in the context of random evaluations.

REMARK 3. Note that in view of (5) and the independent increment property of $\Lambda(t)$, it follows that

$$E[\Lambda'(t_1)\Lambda(t_2)] = \int_0^{\min\{t_1, t_2\}} A(s)\,ds. \tag{7}$$

This is the general form of a correlation matrix.

PROOF OF THEOREM 1. We divide the proof into several steps.

Step 1. We show that the limit of the mean for $\Lambda^\varepsilon(\cdot)$ is 0.

LEMMA 3. *For each* $t \in [0, T]$, $\lim_{\varepsilon\to 0} E\Lambda^\varepsilon(t) = 0$.

PROOF. Using Lemma 1, for $i \in \mathscr{M}$,

$$\begin{aligned}
E\Lambda^\varepsilon(t,i) &= \frac{1}{\sqrt{\varepsilon}}\int_0^t (E\chi_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s))\beta_i(s)\,ds \\
&= \frac{1}{\sqrt{\varepsilon}}\int_0^t (P(\alpha^\varepsilon(s)=i) - \nu_i(s))\beta_i(s)\,ds \\
&= \frac{1}{\sqrt{\varepsilon}}\int_0^t [O(\varepsilon) + O(e^{-\kappa s/\varepsilon})]\beta_i(s)\,ds \\
&= O(\sqrt{\varepsilon}) + \frac{1}{\sqrt{\varepsilon}}\int_0^t O(e^{-\kappa s/\varepsilon})\,ds = O(\sqrt{\varepsilon}) \to 0.
\end{aligned}$$

The proof of the lemma is complete. □

Step 2. The second step is devoted to the calculation of the limiting covariance function.

LEMMA 4. *For each* $t \in [0,T]$,

$$\lim_{\varepsilon\to 0} E(\Lambda^{\varepsilon,\prime}(t)\Lambda^\varepsilon(t)) = \int_0^t A(s)\,ds,$$

where $A(t)$ *is given by* (6).

PROOF. For each $i, j \in \mathscr{M}$,

$$\begin{aligned}
&E[\Lambda^\varepsilon(t,i)\Lambda^\varepsilon(t,j)] \\
&\quad= \frac{1}{\varepsilon}E\bigg[\bigg(\int_0^t (\chi_{\{\alpha^\varepsilon(\tau)=i\}} - \nu_i(\tau))\beta_i(\tau)\,d\tau\bigg)\bigg(\int_0^t (\chi_{\{\alpha^\varepsilon(r)=j\}} - \nu_j(r))\beta_j(r)\,dr\bigg)\bigg] \\
&\quad= \frac{1}{\varepsilon}E\bigg[\int_0^t\int_0^t \big(\chi_{\{\alpha^\varepsilon(\tau)=i,\,\alpha^\varepsilon(r)=j\}} - \nu_i(\tau)\chi_{\{\alpha^\varepsilon(r)=j\}} \\
&\qquad\qquad - \nu_j(r)\chi_{\{\alpha^\varepsilon(\tau)=i\}} + \nu_i(\tau)\nu_j(r)\big)\beta_i(\tau)\beta_j(r)\,d\tau\,dr\bigg].
\end{aligned}$$

Let $D_1 = \{(\tau,r)\colon\ 0 \le r \le \tau \le t\}$ and $D_2 = \{(\tau,r)\colon\ 0 \le \tau \le r \le t\}$ and let

$$\begin{aligned}
\Phi^\varepsilon(\tau,r) = {}& P(\alpha^\varepsilon(\tau)=i,\ \alpha^\varepsilon(r)=j) - \nu_i(\tau)P(\alpha^\varepsilon(r)=j) \\
&- \nu_j(r)P(\alpha^\varepsilon(\tau)=i) + \nu_i(\tau)\nu_j(r).
\end{aligned}$$

Then

$$\begin{aligned}
E[\Lambda^\varepsilon(t,i)\Lambda^\varepsilon(t,j)] &= \frac{1}{\varepsilon}\bigg[\int_0^t\int_0^t \Phi^\varepsilon(\tau,r)\beta_i(\tau)\beta_j(r)\,d\tau\,dr\bigg] \\
&= \frac{1}{\varepsilon}\bigg(\int_{D_1} + \int_{D_2}\bigg)\Phi^\varepsilon(\tau,r)\beta_i(\tau)\beta_j(r)\,d\tau\,dr.
\end{aligned}$$

Note that if $(\tau,r) \in D_1$, then $\tau \ge r$ and

$$P(\alpha^\varepsilon(\tau)=i,\ \alpha^\varepsilon(r)=j) = P(\alpha^\varepsilon(\tau)=i \mid \alpha^\varepsilon(r)=j)P(\alpha^\varepsilon(r)=j).$$

Hence, for $(\tau, r) \in D_1$,

$$\Phi^\varepsilon(\tau, r) = [P(\alpha^\varepsilon(\tau) = i | \alpha^\varepsilon(r) = j) - \nu_i(\tau)]P(\alpha^\varepsilon(r) = j) \\ + \nu_j(r)[\nu_i(\tau) - P(\alpha^\varepsilon(\tau) = i)].$$

Using Lemma 1, we have, for $(\tau, r) \in D_1$,

$$\begin{aligned} \Phi^\varepsilon(\tau, r) = {} & \left(\varepsilon p_i^{\{1\}}(\tau) + q_{ji}^{\{0\}}\left(\frac{\tau - r}{\varepsilon}, r\right) + \varepsilon q_{ji}^{\{1\}}\left(\frac{\tau - r}{\varepsilon}, r\right) + O(\varepsilon^2)\right) \\ & \times \left(\nu_j(r) + \varepsilon p_j^{\{1\}}(r) + q_j^{\{0\}}\left(\frac{r}{\varepsilon}\right) + \varepsilon q_j^{\{1\}}\left(\frac{r}{\varepsilon}\right) + O(\varepsilon^2)\right) \\ & - \nu_j(r)\left(\varepsilon p_i^{\{1\}}(\tau) + q_i^{\{0\}}\left(\frac{\tau}{\varepsilon}\right) + \varepsilon q_i^{\{1\}}\left(\frac{\tau}{\varepsilon}\right) + O(\varepsilon^2)\right) \\ = {} & \nu_j(r) q_{ji}^{\{0\}}\left(\frac{\tau - r}{\varepsilon}, r\right) \\ & + [O(\varepsilon e^{-\kappa r/\varepsilon}) + O(\varepsilon e^{-\kappa(\tau - r)/\varepsilon}) + O(e^{-\kappa\tau/\varepsilon}) + O(\varepsilon^2)]. \end{aligned}$$

Note that

$$\int_0^t \left(\int_0^\tau e^{-\kappa\tau/\varepsilon}\, dr\right) d\tau = \int_0^t \tau e^{-\kappa\tau/\varepsilon}\, d\tau = O(\varepsilon^2),$$

$$\varepsilon \int_0^t \left(\int_0^\tau e^{-\kappa r/\varepsilon}\, dr\right) d\tau = \frac{\varepsilon^2}{\kappa} \int_0^t (1 - e^{-\kappa\tau/\varepsilon})\, d\tau = O(\varepsilon^2)$$

and

$$\varepsilon \int_0^t \left(\int_0^\tau e^{-\kappa(\tau - r)/\varepsilon}\, dr\right) d\tau = \varepsilon \int_0^t \left(\int_0^\tau e^{-\kappa r/\varepsilon}\, dr\right) d\tau = O(\varepsilon^2).$$

Thus,

$$\begin{aligned} & \int_{D_1} \Phi^\varepsilon(\tau, r)\beta_i(\tau)\beta_j(r)\, d\tau\, dr \\ & \quad = \int_0^t \left(\int_0^\tau q_{ji}^{\{0\}}\left(\frac{\tau - r}{\varepsilon}, r\right)\nu_j(r)\beta_i(\tau)\beta_j(r)\, dr\right) d\tau + O(\varepsilon^2). \end{aligned}$$

Exchanging the order of integration leads to

$$\begin{aligned} & \int_0^t \left(\int_0^\tau q_{ji}^{\{0\}}\left(\frac{\tau - r}{\varepsilon}, r\right)\nu_j(r)\beta_i(\tau)\beta_j(r)\, dr\right) d\tau \\ & \quad = \int_0^t \left(\int_r^t q_{ji}^{\{0\}}\left(\frac{\tau - r}{\varepsilon}, r\right)\nu_j(r)\beta_i(\tau)\beta_j(r)\, d\tau\right) dr \\ & \quad = \int_0^t \beta_j(r)\nu_j(r)\left(\int_r^t q_{ji}^{\{0\}}\left(\frac{\tau - r}{\varepsilon}, r\right)\beta_i(\tau)\, d\tau\right) dr. \end{aligned}$$

Making the change of variable $\tau - r = \varepsilon s$ yields

$$\int_r^t q_{ji}^{\{0\}}\left(\frac{\tau - r}{\varepsilon}, r\right)\beta_i(\tau)\, d\tau = \varepsilon \int_0^{(t-r)/\varepsilon} q_{ji}^{\{0\}}(s, r)\beta_i(r + \varepsilon s)\, ds.$$

Notice that $\beta_i(\cdot)$ is bounded and $\beta_i(r+\varepsilon s) \to \beta_i(r)$ in $L^1[0, N]$ for any finite $N > 0$ as $\varepsilon \to 0$. Since $q_{ji}^{\{0\}}(\cdot)$ is "stable" as in Lemma 2,

$$\int_0^{(t-r)/\varepsilon} q_{ji}^{\{0\}}(s, r)\beta_i(r+\varepsilon s)\, ds \to \beta_i(r) \int_0^\infty q_{ji}^{\{0\}}(s, r)\, ds.$$

Therefore,

$$\begin{aligned} (8) \qquad & \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{D_1} \Phi^\varepsilon(\tau, r)\beta_i(\tau)\beta_j(r)\, d\tau\, dr \\ & \qquad = \int_0^t \beta_i(r)\beta_j(r)\nu_j(r)\left(\int_0^\infty q_{ji}^{\{0\}}(s, r)\, ds\right) dr. \end{aligned}$$

Similarly, it can be shown

$$\begin{aligned} (9) \qquad & \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{D_2} \Phi^\varepsilon(\tau, r)\beta_i(\tau)\beta_j(r)\, d\tau\, dr \\ & \qquad = \int_0^t \beta_i(r)\beta_j(r)\nu_i(r)\left(\int_0^\infty q_{ij}^{\{0\}}(s, r)\, ds\right) dr. \end{aligned}$$

Combining (8) and (9), we obtain

$$\lim_{\varepsilon\to 0} E[\Lambda^\varepsilon(t, i)\Lambda^\varepsilon(t, j)] = \int_0^t A(s)\, ds,$$

with $A(t)$ given by (6). □

Step 3. We prove the tightness of the sequence $\Lambda^\varepsilon(\cdot)$ in $C^{m+1}[0, T]$, the space of $\mathbb{R}^{m+1}$-valued continuous functions with the sup-norm topology.

First we state a preparatory lemma. Since its proof is long, in order not to disrupt the flow of the presentation, the proof is placed at the end of the paper in the Appendix.

LEMMA 5. *There exist positive constants ε_0 and C such that, for $0 < \varepsilon \le \varepsilon_0$, $i \in \mathscr{M}$, for any deterministic process $\beta(\cdot)$ satisfying $|\beta(t)| \le 1$ for all $t \ge 0$ and for fixed $T \ge 0$,*

$$(10) \qquad E \exp\left\{ \frac{1}{\sqrt{\varepsilon}(1+T)^{3/2}} \sup_{0\le t\le T} \left| \int_0^t (\chi_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s))\beta(s)\, ds \right| \right\} \le C.$$

In fact, Lemma 5 is interesting in its own right. It gives a "uniformly exponential" estimate on the scaled sequence $\{\Lambda^\varepsilon(\cdot)\}$ (see also Remark 4). Using Lemma 5, we proceed to prove the tightness of the underlying sequence. In what follows, $D^{m+1}[0, T]$ denotes the space of $\mathbb{R}^{m+1}$-valued functions that are right continuous and have left limits endowed with the Skorohod topology.

LEMMA 6. *The collection of processes $\{\Lambda^\varepsilon(\cdot)\}$ is tight in $D^{m+1}[0, T]$.*

PROOF. By virtue of Lemma 5, we can show that, for some constants $\varepsilon_0 > 0$ and $C > 0$,

$$E\exp\biggl(\frac{1}{(1+T)^{3/2}}\sup_{0\le t\le T}|\Lambda^\varepsilon(t,i)|\biggr)\le C \tag{11}$$

for $0 < \varepsilon \le \varepsilon_0$ and $i \in \mathscr{M}$. This inequality implies

$$P\Bigl(\sup_{0\le t\le T}|\Lambda^\varepsilon(t,i)|\ge k\Bigr)\le C\exp\biggl(-\frac{k}{(1+T)^{3/2}}\biggr).$$

Hence, for each $T < \infty$,

$$\lim_{k\to\infty}\limsup_{\varepsilon\to 0}P\Bigl(\sup_{0\le t\le T}|\Lambda^\varepsilon(t,i)|\ge k\Bigr)=0. \tag{12}$$

Moreover, we have

$$\begin{aligned}
&E[|\Lambda^\varepsilon(t+s,j)-\Lambda^\varepsilon(s,j)|^2|\alpha^\varepsilon(s)=i]\\
&\quad=\frac{1}{\varepsilon}E\biggl[\biggl|\int_s^{t+s}(\chi_{\{\alpha^\varepsilon(\tau)=j\}}-\nu_j(\tau))\beta_j(\tau)\,d\tau\biggr|^2\biggl|\alpha^\varepsilon(s)=i\biggr]\\
&\quad=\frac{1}{\varepsilon}\int_s^{t+s}\int_s^{t+s}\tilde{\Phi}^\varepsilon(\tau,r)\beta_j(\tau)\beta_j(r)\,d\tau\,dr,
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\Phi}^\varepsilon(\tau,r)&=P(\alpha^\varepsilon(\tau)=j,\ \alpha^\varepsilon(r)=j|\alpha^\varepsilon(s)=i)-\nu_j(\tau)P(\alpha^\varepsilon(r)=j|\alpha^\varepsilon(s)=i)\\
&\qquad-\nu_j(r)P(\alpha^\varepsilon(\tau)=j|\alpha^\varepsilon(s)=i)+\nu_j(r)\nu_i(\tau).
\end{aligned}$$

Similarly, as in the derivation of Lemma 4, let $\tilde{D}_1=\{(\tau,r)\colon\ s\le r\le\tau\le t+s\}$ and $\tilde{D}_2=\{(\tau,r)\colon\ s\le\tau\le r\le t+s\}$. Then

$$\begin{aligned}
&E[|\Lambda^\varepsilon(t+s,j)-\Lambda^\varepsilon(s,j)|^2|\alpha^\varepsilon(s)=i]\\
&\quad=\frac{1}{\varepsilon}\biggl(\int_{\tilde{D}_1}+\int_{\tilde{D}_2}\biggr)\tilde{\Phi}^\varepsilon(\tau,r)\beta_j(\tau)\beta_j(r)\,d\tau\,dr.
\end{aligned}$$

Notice that, for $0\le t\le\delta$,

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_{\tilde{D}_1}\tilde{\Phi}^\varepsilon(\tau,r)\beta_j(\tau)\beta_j(r)\,d\tau\,dr\\
&\quad=\int_s^{t+s}\beta_j(r)\nu_j(r)\biggl(\int_0^{(t+s-r)/\varepsilon}q_{jj}^{\{0\}}(\tau,r)\beta_j(r+\varepsilon\tau)\,d\tau\biggr)dr+O(\varepsilon)\\
&\quad\le\int_s^{t+s}|\beta_j(r)\nu_j(r)|\biggl(\int_0^{(t+s-r)/\varepsilon}|q_{jj}^{\{0\}}(\tau,r)\beta_j(r+\varepsilon\tau)|\,d\tau\biggr)dr+O(\varepsilon)\\
&\quad\le\int_s^{\delta+s}|\beta_j(r)\nu_j(r)|\biggl(\int_0^{(\delta+s-r)/\varepsilon}|q_{jj}^{\{0\}}(\tau,r)\beta_j(r+\varepsilon\tau)|\,d\tau\biggr)dr+O(\varepsilon)\\
&\quad:=\rho_1^\varepsilon(\delta)+O(\varepsilon).
\end{aligned}$$

Similarly, we can show

$$\frac{1}{\varepsilon}\int_{\tilde{D}_2}\tilde{\Phi}^\varepsilon(\tau, r)\beta_j(\tau)\beta_j(r)\,d\tau\,dr \le \rho_2^\varepsilon(\delta) + O(\varepsilon).$$

In addition, $\rho_1^\varepsilon(\delta)$ and $\rho_2^\varepsilon(\delta)$ above satisfy

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0} E\rho_1^\varepsilon(\delta) = 0 \quad\text{and}\quad \lim_{\delta\to 0}\limsup_{\varepsilon\to 0} E\rho_2^\varepsilon(\delta) = 0.$$

All the conditions of Kurtz's tightness criterion ([11], Theorem 3, page 47), are satisfied. By virtue of that theorem, it follows that $\Lambda^\varepsilon(\cdot)$ is tight in $D^{m+1}[0, T]$. □

Consequently, by Prohorov's theorem we can extract a convergent subsequence. Select such a convergent subsequence and, for notational simplicity, still denote the sequence by $\{\Lambda^\varepsilon(\cdot)\}$ and denote the limit by $\Lambda(\cdot)$.

Now, we state a lemma which will be proved in the Appendix.

LEMMA 7. *For any t_0, $\delta > 0$, and $t_0 \le t \le t_0 + \delta$,*

$$E|\Lambda(t, i) - \Lambda(t_0, i)|^4 \le C(1+T)^4\delta^2.$$

In view of this lemma and Kolmogorov's criterion (see [5], page 149), it follows that the limit process $\Lambda(\cdot)$ has continuous paths with probability 1. Hence, $\Lambda^\varepsilon(t)$ is relatively compact in $C^{m+1}[0, T]$ and therefore is tight in $C^{m+1}[0, T]$.

Step 4. In order to obtain asymptotic normality, one often needs to show the sequence under consideration satisfies a mixing condition. In what follows, we show that the sequence $\Lambda^\varepsilon(\cdot)$ verifies an "ε-mixing" condition for $\varepsilon > 0$ small, which is adequate for obtaining the desired result in the sequel.

LEMMA 8. *For any $\sigma\{\alpha^\varepsilon(s)\colon 0 \le s \le t\}$-measurable ξ and $\sigma\{\alpha^\varepsilon(s)\colon s \ge t+\tau\}$-measurable η such that $|\xi| \le 1$ and $|\eta| \le 1$,*

$$|E(\xi\eta) - E\xi E\eta| \le 2\sqrt{C(\varepsilon + e^{-\kappa\tau/\varepsilon})}. \tag{13}$$

PROOF. For any $0 \le s_1 \le s_2 \le \cdots \le s_n = t \le t+\tau = t_0 \le t_1 \le \cdots \le t_l < \infty$, let

$$E_1 = \{\alpha^\varepsilon(t) = i,\ \alpha^\varepsilon(s_{n-1}) = i_{n-1},\ \ldots, \alpha^\varepsilon(s_1) = i_1\}$$

and

$$E_2 = \{\alpha^\varepsilon(t+\tau) = j,\ \alpha^\varepsilon(t_1) = j_1,\ \ldots, \alpha^\varepsilon(t_l) = j_l\}.$$

Then, owing to the Markovian property of $\alpha^\varepsilon(\cdot)$,

$$\begin{aligned} P(E_2|E_1) &= P(E_2|\alpha^\varepsilon(t) = i) \\ &= P(\alpha^\varepsilon(t+\tau) = j|\alpha^\varepsilon(t) = i)[p^\varepsilon_{j, j_1}(t_1, t+\tau)\cdots p^\varepsilon_{j_{l-1}, j_l}(t_l, t_{l-1})]. \end{aligned}$$

Moreover,

$$P(E_2) = P(\alpha^\varepsilon(t+\tau) = j)[p^\varepsilon_{j,\,j_1}(t_1, t+\tau)\cdots p^\varepsilon_{j_{l-1},\,j_l}(t_l, t_{l-1})].$$

Thus,

$$\begin{aligned}|P(E_2|E_1) - P(E_2)| &\le |P(\alpha^\varepsilon(t+\tau) = j|\alpha^\varepsilon(t) = i) - P(\alpha^\varepsilon(t+\tau) = j)|\\ &\le \left|q_{ij}\left(\frac{\tau}{\varepsilon}, t\right)\right| + \left|q_j\left(\frac{t+\tau}{\varepsilon}\right)\right| + O(\varepsilon)\\ &\le C(\varepsilon + e^{-\kappa\tau/\varepsilon})\end{aligned}$$

for some positive constants C and κ that are independent of $i, j \in \mathscr{M}$ and $t \in [0, T]$.

We can show similarly, as in the proof of Lemma 1, page 170, of [1], that

$$|E(\xi\eta) - E\xi E\eta| \le 2\sqrt{C(\varepsilon + e^{-\kappa\tau/\varepsilon})}$$

for any $\sigma\{\alpha^\varepsilon(s)\colon\ 0 \le s \le t\}$-measurable ξ and $\sigma\{\alpha^\varepsilon(s)\colon\ s \ge t+s\}$-measurable η such that $|\xi| \le 1$ and $|\eta| \le 1$. □

Step 5. We show that the limit of the finite-dimensional distributions of $\Lambda^\varepsilon(\cdot)$ are Gaussian with independent increments. This part of the proof is similar to [7] (see also [6], page 224). To prove the convergence of the finite-dimensional distributions, we use the characteristic function $E\exp(i(z, \Lambda^\varepsilon(t))$, where $i = \sqrt{-1}$, $z \in \mathbb{R}^{m+1}$ and (a, b) denotes the usual inner product in $\mathbb{R}^{m+1}$.

By using the approximate mixing property and repeated applications of Lemma 8 or by induction, for arbitrary positive real numbers s_i and t_j satisfying $0 \le s_0 \le t_0 \le s_1 \le t_1 \le s_2 \le \cdots \le s_m \le t_m$,

$$\left|E\exp\left\{i\left(z, \sum_{l=0}^{m}(\Lambda^\varepsilon(t_l) - \Lambda^\varepsilon(s_l))\right)\right\} - \prod_{l=1}^{m} E\exp\{i(z, (\Lambda^\varepsilon(t_l) - \Lambda^\varepsilon(s_l)))\}\right| \xrightarrow{\varepsilon} 0.$$

This, in turn, implies that the limit process $\Lambda(\cdot)$ has independent increments. Moreover, in view of Lemma 7, the limiting process has continuous paths with probability 1. In accordance with Skorohod [13], page 7, if a process with independent increments has continuous paths w.p.1, then it must necessarily be a Gaussian process. This implies that the limits of the finite-dimensional distribution of $\Lambda(\cdot)$ are Gaussian.

Consequently, $\Lambda(\cdot)$ is a process having Gaussian finite-dimensional distributions, with mean 0 and covariance $A(t)$ given by Lemma 4. Moreover, the limit does not depend on the chosen subsequence. Thus, we conclude that $\Lambda^\varepsilon(\cdot)$ converges weakly to the Gaussian process $\Lambda(\cdot)$. This completes the proof of the theorem. □

REMARK 4. Note that, for each $x \ge 0$ and $n = 1, 2, \dots,$ $x^{2n}e^{-x} = O(1)$. It follows from (11) that there exists a constant $C_n > 0$ such that

$$\frac{1}{(1+T)^{3n/2}} \sup_{0\le t\le T} |\Lambda^\varepsilon(t,i)|^{2n} \le C_n \exp\left[\frac{1}{(1+T)^{3/2}} \sup_{0\le t\le T} |\Lambda^\varepsilon(t,i)|\right].$$

Hence,

$$E \sup_{0\le t\le T} \left|\int_0^t (\chi_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s))\beta(s)\,ds\right|^{2n} \le CC_n\varepsilon^n(1+T)^{3n}.$$

Such an estimate is very useful for establishing the error bounds of asymptotic optimal hierarchical controls in manufacturing models (see Sethi and Zhang [12]).

To conclude this section, we give an example to illustrate the results.

EXAMPLE 2. Let $\alpha^\varepsilon(t) \in \mathscr{M} = \{0, 1\}$ be a two-state Markov chain with a generator

$$Q(t) = \begin{pmatrix} -\hat{\mu}(t) & \hat{\mu}(t) \\ \hat{\lambda}(t) & -\hat{\lambda}(t) \end{pmatrix},$$

where $\hat{\lambda}(t) \ge 0$, $\hat{\mu}(t) \ge 0$ and $\hat{\lambda}(t) + \hat{\mu}(t) > 0$ for each $t \in [0, T]$. Moreover, $\hat{\lambda}(\cdot)$ and $\hat{\mu}(\cdot)$ are three times continuously differentiable.

It is easy to see that Assumptions 1 and 2 are satisfied.

In this example,

$$\nu(t) = (\nu_0(t), \nu_1(t)) = \left(\frac{\hat{\lambda}(t)}{\hat{\lambda}(t)+\hat{\mu}(t)}, \frac{\hat{\mu}(t)}{\hat{\lambda}(t)+\hat{\mu}(t)}\right).$$

Moreover,

$$Q^{\{0\}}(s, t_0) = -\frac{\exp(-\{\hat{\lambda}(t_0)+\hat{\mu}(t_0)\}s)}{\hat{\lambda}(t_0)+\hat{\mu}(t_0)} Q(t_0).$$

Thus, the covariance matrix is

$$A(t) = \frac{2\hat{\lambda}(t)\hat{\mu}(t)}{(\hat{\lambda}(t)+\hat{\mu}(t))^3} \begin{pmatrix} (\beta_0(t))^2 & -\beta_0(t)\beta_1(t) \\ -\beta_0(t)\beta_1(t) & (\beta_1(t))^2 \end{pmatrix}.$$

It is interesting to note that either $\hat{\lambda}(t)$ or $\hat{\mu}(t)$ can be equal to 0 for a while as long as $\hat{\lambda}(t) + \hat{\mu}(t) \ge c > 0$. For example, in a manufacturing model, $\hat{\mu}(\cdot)$ may denote the repair rate of a machine. Then $\hat{\mu}(t) = 0$ corresponds to, for example, rest periods for repair workers or a waiting period needed to get the required parts.

4. Further remarks. A class of singular perturbation problems is dealt with in this work. The underlying systems arise from a wide range of applications where a finite state Markov chain is involved and where the Markov chain varies rapidly in a rather fast time scale t/ε. Under a weak irreducibility condition (see [9] and [8]) and a smoothness condition on the generator $Q(t)$, asymptotic normality is established. One of the distinct characteristics of the result is that the covariance function of the limit process is expressed in terms of the coefficients of the asymptotic expansion. It depends not only on the regular part of the asymptotic series, but also on the boundary layer terms. This is very different from the existing results of the central limit type. Finally, in view of the problem presented and the "uniform exponential" estimate, a question naturally arises: is there a large deviation principle associated with the process (suitably normalized) $\alpha^\varepsilon(\cdot)$? We believe the answer is positive. Our current effort is devoted to this and the related asymptotic distribution (not necessarily normal) of a more complex problem studied in [8].

APPENDIX

The proof of Lemma 5 is a combination of the exponential estimates derived in [14] and the specific asymptotic expansion as stated in Lemmas 1 and 2.

PROOF OF LEMMA 5. Let

$$\lambda(t) = (\chi_{\{\alpha^\varepsilon(t)=0\}}, \dots, \chi_{\{\alpha^\varepsilon(t)=m\}})'$$

and

$$w(t) = \lambda(t) - \lambda(0) - \frac{1}{\varepsilon}\int_0^t Q'(s)\lambda(s)\,ds.$$

Then it is well known (see [4]) that $w(t) = (w_0(t), \dots, w_m(t))'$, $t \geq 0$, is a $\sigma\{\alpha(\varepsilon, s)\colon\ s \leq t\}$-martingale and

$$\lambda(t) = (P^\varepsilon(t,0))'\lambda(0) + \int_0^t (P^\varepsilon(t,s))'\,dw(s).$$

By Lemma 2, for $0 \leq s \leq t$,

$$P^\varepsilon(t,s) - P^{\{0\}}(t) = O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-s)}{\varepsilon}\right)\right)$$

and

$$(P^{\{0\}}(t))'\lambda(s) = (\nu'(t)\colon \cdots \colon \nu'(t))\lambda(s) = \nu'(t)\left(\sum_{i=0}^m \chi_{\{\alpha^\varepsilon(s)=i\}}\right) = \nu'(t).$$

Hence,

$$
\begin{aligned}
\lambda(t)-\nu'(t) &= [(P^{\varepsilon}(t,0))' - (P^{\{0\}}(t))']\lambda(0) \\
&\quad + \int_0^t [(P^{\varepsilon}(t,s))' - (P^{\{0\}}(t))' + (P^{\{0\}}(s))']\,dw(s) \\
&= O\Big(\varepsilon + \exp\Big(-\frac{k_0 t}{\varepsilon}\Big)\Big) \\
&\quad + \int_0^t \Big[O\Big(\varepsilon + \exp\Big(-\frac{k_0(t-s)}{\varepsilon}\Big)\Big) + P^{\{0\}}(s))'\Big]\,dw(s) \qquad (14)\\
&= O\Big(\varepsilon + \exp\Big(-\frac{k_0 t}{\varepsilon}\Big)\Big)' \\
&\quad + \int_0^t \Big[O\Big(\varepsilon + \exp\Big(-\frac{k_0(t-s)}{\varepsilon}\Big)\Big)\Big]\,dw(s),
\end{aligned}
$$

where the last equality follows from the observation that

$$
\begin{aligned}
(P^{\{0\}}(t))'w(t) &= (P^{\{0\}}(t))'\Big[\lambda(t)-\lambda(0)-\int_0^t Q'(s)\lambda(s)\,ds\Big] \\
&= \nu'(t)-\nu'(t)-\int_0^t (P^{\{0\}}(t))'Q'(s)\lambda(s)\,ds = 0.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\int_0^t (\lambda(s)-\nu'(s))\beta(s)\,ds \\
&\quad = O(\varepsilon(t+1)) + \int_0^t\int_0^s O\Big(\varepsilon + \exp\Big(-\frac{k_0(s-r)}{\varepsilon}\Big)\Big)\,dw(r)\beta(s)\,ds \\
&\quad = O(\varepsilon(t+1)) + \int_0^t \Big(\int_r^t O\Big(\varepsilon + \exp\Big(-\frac{k_0(t-s)}{\varepsilon}\Big)\Big)\beta(s)\,ds\Big)\,dw(r) \qquad (15)\\
&\quad = O(\varepsilon(t+1)) + O(\varepsilon)\int_0^t \Big[(t-s) + \kappa_0^{-1}\Big(1-\exp\Big(-\frac{k_0(t-s)}{\varepsilon}\Big)\Big)\Big]\,dw(s).
\end{aligned}
$$

Dividing both sides by $(T+1)$, we have

$$
\frac{1}{T+1}\sup_{0\le t\le T}\left|\int_0^t (\lambda(s)-\nu'(s))\beta(s)\,ds\right| = \varepsilon O(1) + \varepsilon \sup_{0\le t\le T}\left|\int_0^t O(1)\,dw(s)\right|. \qquad (16)
$$

Therefore,

$$
\begin{aligned}
&E\exp\left\{\frac{1}{\sqrt{\varepsilon}(1+T)^{3/2}}\sup_{0\le t\le T}\left|\int_0^t (\lambda(s)-\nu'(s))\beta(s)\,ds\right|\right\} \\
&\qquad \le E\exp\left\{\frac{1}{\sqrt{\varepsilon}\sqrt{T+1}}\left[\varepsilon O(1) + \varepsilon\sup_{0\le t\le T}\left|\int_0^t O(1)\,dw(s)\right|\right]\right\}.
\end{aligned}
$$

Without loss of generality, we may assume that $|O(1)| \le 1$, because we can always replace ε by $\varepsilon\theta$ for θ small enough:

$$\begin{aligned}
&E\exp\left\{\frac{1}{\sqrt{\varepsilon}(1+T)^{3/2}}\sup_{0\le t\le T}\left|\int_0^t(\lambda(s)-\nu'(s))\beta(s)\,ds\right|\right\}\\
&\quad\le \exp\frac{\sqrt{\varepsilon}}{\sqrt{T+1}}E\exp\left\{\frac{\sqrt{\varepsilon}}{\sqrt{T+1}}\sup_{0\le t\le T}\left|\int_0^t O(1)\,dw(s)\right|\right\}\\
&\quad\le eE\exp\left\{\frac{\sqrt{\varepsilon}}{\sqrt{T+1}}\sup_{0\le t\le T}\left|\int_0^t O(1)\,dw(s)\right|\right\}.
\end{aligned}\tag{17}$$

Recall that $w(t)=(w_0(t),\ldots,w_m(t))'$. It suffices to show that, for

$$E\exp\left\{\frac{\sqrt{\varepsilon}}{\sqrt{T+1}}\sup_{0\le t\le T}\left|\int_0^t b(s,t)\,dw_i(s)\right|\right\}\le C\tag{18}$$

for all measurable functions b such that $|b(s,t)|\le 1$ and $0\le t\le T$. In fact, for each $t_0\ge 0$, let $b_0(s)=b(s,t_0)$. Using the fact that

$$Ee^{\xi}\le e+(e-1)\sum_{j=1}^{\infty}e^jP(\xi\ge j)$$

for any nonnegative random variable ξ, we have

$$\begin{aligned}
&E\exp\left\{\frac{\sqrt{\varepsilon}}{\sqrt{T+1}}\sup_{0\le t\le T}\left|\int_0^t b_0(s)\,dw_i(s)\right|\right\}\\
&\quad\le e+(e-1)\sum_{j=1}^{\infty}e^jP\left(\frac{\sqrt{\varepsilon}}{\sqrt{T+1}}\sup_{0\le t\le T}\left|\int_0^t b_0(s)\,dw_i(s)\right|\ge j\right).
\end{aligned}\tag{19}$$

Now we estimate

$$P\left(\frac{\sqrt{\varepsilon}}{\sqrt{T+1}}\sup_{0\le t\le T}\left|\int_0^t b_0(s)\,dw_i(s)\right|\ge j\right).$$

Let $\tilde{p}_i(t)=\int_0^t b_0(s)\,dw_i(s)$. In what follows, suppress the i-dependence and write it as $\tilde{p}(\cdot)$ whenever there is no confusion. Similarly, we write $\tilde{q}(\cdot)$ in lieu of $\tilde{q}_i(\cdot)$ to be defined below. Then $\tilde{p}(t)$, $t\ge 0$, is a local martingale. Let $\tilde{q}(\cdot)$ denote the only solution to the equation (see [3])

$$\tilde{q}(t)=1+\zeta\int_0^t\tilde{q}(s^-)\,d\tilde{p}(s),$$

where $\tilde{q}(s^-)$ is the left-hand limit of q at s and ζ is a positive constant to be determined later. Since $\zeta\int_0^t\tilde{q}(s^-)\,d\tilde{p}(s)$, $t\ge 0$, is a local martingale, we have $E\tilde{q}(t)\le 1$ for all $t\ge 0$. Moreover, $\tilde{q}(t)$ can be written as follows (see [3]):

$$\tilde{q}(t)=\exp(\zeta\tilde{p}(t))\prod_{s\le t}(1+\zeta\,\Delta\tilde{p}(s))\exp(-\zeta\,\Delta\tilde{p}(s)),\tag{20}$$

where $\Delta\tilde{p}(s):=\tilde{p}(s)-\tilde{p}(s^-)$, $|\Delta\tilde{p}(s)|\le 1$.

Let us now observe that there exist positive constants ζ_0 and κ_1 such that, for $0 < \zeta \le \zeta_0$ and for all $s > 0$,

$$(1 + \zeta\,\Delta\tilde{p}(s))\exp(-\zeta\,\Delta\tilde{p}(s)) \ge \exp(-\kappa_1\zeta^2). \tag{21}$$

Combining (20) and (21) yields

$$\tilde{q}(t) \ge \exp\{\zeta\,\tilde{p}(t) - \kappa_1\zeta^2 N(t)\} \quad \text{for } 0 < \zeta \le \zeta_0,\ t > 0,$$

where $N(t)$ is the number of jumps of $\tilde{p}(s)$ in $s \in [0, t]$. Since $N(t)$ is a monotone increasing process, we have

$$\sup_{0\le t\le T} \tilde{q}(t) \ge \exp\Big\{\zeta \sup_{0\le t\le T} \tilde{p}(t) - \kappa_1\zeta^2 N(T)\Big\} \quad \text{for } 0 < \zeta \le \zeta_0.$$

Note also that, for each $i = 0, \dots, m$,

$$\begin{aligned}
&P\bigg(\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \sup_{0\le t\le T}\bigg|\int_0^t b_0(s)\,dw_i(s)\bigg| \ge j\bigg)\\
&\quad = P\bigg(\sup_{0\le t\le T}|\tilde{p}(t)| \ge \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\bigg)\\
&\quad \le P\bigg(\sup_{0\le t\le T}\tilde{p}(t) \ge \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\bigg) + P\bigg(\sup_{0\le t\le T}(-\tilde{p}(t)) \ge \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\bigg).
\end{aligned}$$

Consider the first term

$$P\bigg(\sup_{0\le t\le T}\tilde{p}(t) \ge \frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}}\bigg).$$

Let $a_j = j(T+1)/(8\kappa_1\varepsilon)$. Then

$$\begin{aligned}
&P\bigg(\sup_{0\le t\le T}\tilde{p}(t) \ge \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\bigg)\\
&\quad \le P\bigg(\sup_{0\le t\le T}\tilde{q}(t) \ge \exp\bigg\{\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} - \kappa_1\zeta^2 N(T)\bigg\}\bigg)\\
&\quad \le P\bigg(\sup_{0\le t\le T}\tilde{q}(t) \ge \exp\bigg\{\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} - \kappa_1\zeta^2 N(T)\bigg\},\ N(T) \le a_j\bigg)\\
&\qquad + P(N(T) \ge a_j)\\
&\quad \le P\bigg(\sup_{0\le t\le T}\tilde{q}(t) \ge \exp\bigg(\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} - \kappa_1\zeta^2 a_j\bigg)\bigg) + P(N(T) \ge a_j)\\
&\quad \le 2\exp\bigg(-\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} + \kappa_1\zeta^2 a_j\bigg) + P(N(T) \ge a_j).
\end{aligned}$$

The last inequality follows because of the local martingale property (see [3], Theorem 4.2).

Now if we take $\zeta = 4\sqrt{\varepsilon}/\sqrt{T+1}$, then

$$\exp\left(-\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} + \kappa_1\zeta^2 a_j\right) = \exp(-2j).$$

In view of the construction of Markov chains in [2], there exists a Poisson process $N_0(\cdot)$ with parameter a/ε for some $a > 0$, such that $N(t) \le N_0(t)$. We may assume $a = 1$ because otherwise we may take ε as εa^{-1}. By using the Poisson distribution of $N_0(t)$ and Stirling's formula, we can show that, for ε small enough,

$$P(N(T) \ge a_j) \le 2\gamma^{a_j - 1},$$

where $\gamma = 8e\kappa_1 \in (0,1)/j_0$ for $j_0 > \max\{1, 8e\kappa_1\}$.

Thus, for $j \ge j_0$,

$$P\left(\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \sup_{0\le t\le T} \int_0^t b_0(s)\,dw_i(s) \ge j\right) \le 2e^{-2j} + 2\gamma^{a_j-1}.$$

Repeating the same argument for the martingale $(-\tilde{p}(\cdot))$, we get, for $j \ge j_0$,

$$P\left(\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \sup_{0\le t\le T} \left(-\int_0^t b_0(s)\,dw_i(s)\right) \ge j\right) \le 2e^{-2j} + 2\gamma^{a_j-1}.$$

Combining the above two inequalities, we obtain, for $j \ge j_0$,

$$P\left(\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \sup_{0\le t\le T} \left|\int_0^t b_0(s)\,dw_i(s)\right| \ge j\right) \le 4(e^{-2j} + \gamma^{a_j-1}).$$

Then, by (19),

$$E\exp\left\{\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \sup_{0\le t\le T} \left|\int_0^t b(s,t)\,dw_i(s)\right|\right\} \le C_0 + 4(e-1)\sum_{j=1}^{\infty} e^j(e^{-2j} + \gamma^{a_j-1}),$$

where C_0 is the sum corresponding to those terms with $j \le j_0$. Now we choose ε small enough so that $e\gamma^{1/(8\kappa_1\varepsilon)} \le 1/2$. Then

$$E\exp\left\{\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \sup_{0\le t\le T} \left|\int_0^t b(s,t_0)\,dw_i(s)\right|\right\} \le C_0 + 4e\gamma^{-1}.$$

Since t_0 is arbitrary, we may take $t_0 = t$ in the above inequality. Then

$$E\exp\left\{\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \sup_{0\le t\le T} \left|\int_0^t b(s,t)\,dw_i(s)\right|\right\} \le C_0 + 4e\gamma^{-1}.$$

Combining this inequality with (17), we obtain

$$E\exp\left\{\frac{1}{\sqrt{\varepsilon}(1+T)^{3/2}} \sup_{0\le t\le T} \left|\int_0^t (\lambda(s) - \nu'(s))\beta(s)\,ds\right|\right\} \le e(C_0 + 4e\gamma^{-1}) := C.$$

Hence, the lemma is proved. □

PROOF OF LEMMA 7. First, there exist positive constants ε_0 and C such that, for $0 < \varepsilon \le \varepsilon_0$, $i \in \mathscr{M}$, $t_0 \ge 0$, $\delta > \varepsilon_0$, $t_0 + \delta \le T$, and for any deterministic process $\beta(\cdot)$, $|\beta(t)| \le 1$, $t \ge 0$,

$$E \exp\left\{ \frac{1}{\sqrt{\varepsilon}(1+T)\sqrt{\delta}} \sup_{t_0 \le t \le t_0+\delta} \left| \int_{t_0}^{t} (\chi_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s))\beta(s)\, ds \right| \right\} \le C. \tag{22}$$

In fact, from (15), we have

$$\frac{1}{T+1} \int_0^t (\lambda(s) - \nu'(s))\beta(s)\, ds = \varepsilon O(1) + \varepsilon \int_0^t O(1)\, dw(s).$$

This implies, for $t_0 \le t \le t_0 + \delta$,

$$\frac{1}{T+1} \int_{t_0}^t (\lambda(s) - \nu'(s))\beta(s)\, ds = \varepsilon O(1) + \varepsilon \int_{t_0}^t O(1)\, dw(s).$$

Dividing both sides by $\sqrt{\delta}$ and following the proof of Lemma 5, (22) can be established.

Note that there exists a constant C such that $x^4 \le Ce^x$ for all $x \ge 0$. Using this fact together with (22), we obtain, for all $t_0 \le t \le t_0 + \delta$,

$$\begin{aligned} &\frac{1}{(T+1)^4\delta^2} E|\Lambda^\varepsilon(t,i) - \Lambda^\varepsilon(t_0,i)|^4 \\ &\qquad \le CE \exp\left\{ \frac{1}{(1+T)\sqrt{\delta}} |\Lambda^\varepsilon(t,i) - \Lambda^\varepsilon(t_0,i)| \right\} \le C. \end{aligned}$$

This implies that, for any t_0, $\delta \ge \varepsilon_0 \ge \varepsilon > 0$, $t_0 \le t \le t_0 + \delta$,

$$E|\Lambda^\varepsilon(t,i) - \Lambda^\varepsilon(t_0,i)|^4 \le C(1+T)^4\delta^2.$$

Sending $\varepsilon \to 0$, in view of Fatou's lemma,

$$E|\Lambda(t,i) - \Lambda(t_0,i)|^4 \le C(1+T)^4\delta^2. \qquad \square$$

Acknowledgments. Professor Rafail Khasminskii read an early version of the manuscript. His detailed comments and suggestions are greatly appreciated. We also thank the two anonymous referees for their insightful input.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
[2] DAVIS, M. H. A. (1993). *Markov Models and Optimization*. Chapman and Hall, New York.
[3] ELLIOTT, R. J. (1982). *Stochastic Calculus and Applications*. Springer, New York.
[4] ELLIOTT, R. J. (1985). Smoothing for a finite state Markov process. *Lecture Notes in Control and Inform. Sci.* **69** 199–206. Springer, New York.
[5] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes, Characterization and Convergence*. Wiley, New York.
[6] FRIEDLIN, M. I. and WENTZEL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.
[7] KHASMINSKII, R. Z. (1966). On stochastic processes defined by differential equations with a small parameter. *Theory Probab. Appl.* **11** 211–228.

[8] KHASMINSKII, R. Z., YIN, G. and ZHANG, Q. (1995). Constructing asymptotic series for probability of Markov chains with weak and strong interactions. *Quart. Appl. Math*. To appear.

[9] KHASMINSKII, R. Z., YIN, G. and ZHANG, Q. (1996). Asymptotic expansions of singularly perturbed systems involving rapidly fluctuating Markov chains. *SIAM J. Appl. Math.* **56** 277–293.

[10] KURTZ, T. G. (1973). A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Funct. Anal.* **12** 55–67.

[11] KUSHNER, H. J. (1984). *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems*. MIT Press.

[12] SETHI, S. P. and ZHANG, Q. (1994). *Hierarchical Decision Making in Stochastic Manufacturing Systems*. Birkhäuser, Boston.

[13] SKOROHOD, A. V. (1982). *Studies in the Theory of Random Processes*. Dover, New York.

[14] ZHANG, Q. (1995). Risk sensitive production planning of stochastic manufacturing systems: a singular perturbation approach. *SIAM J. Control Optim.* **33** 498–527.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602
E-MAIL: qingz@math.uga.edu

DEPARTMENT OF MATHEMATICS
WAYNE STATE UNIVERSITY
DETROIT, MICHGAN 48202
E-MAIL: gyin@math.wayne.edu