Exponential and Strong Ergodicity for Markov Processes with an Application to Queues**

Yuanyuan LIU*   Zhenting HOU*

Abstract For an ergodic continuous-time Markov process with a particular state in its space, the authors provide the necessary and sufficient conditions for exponential and strong ergodicity in terms of the moments of the first hitting time on the state. An application to the queue length process of M/G/1 queue with multiple vacations is given.

Keywords Markov processes, Queueing theory, Exponential ergodicity, Strong ergodicity
2000 MR Subject Classification 60J25, 60K25

1 Introduction

Necessary and sufficient criteria of exponential and strong ergodicity for continuous-time Markov chains (i.e. continuous-time Markov processes on a countable space), based on the moments of the first hitting time, have been developed in [1, 2], while for continuous-time Markov processes on a general space, the given criteria (see [3, 4]) are sufficient but not necessary. In this paper, we aim to find the necessary and sufficient conditions for both forms of ergodicity for Markov processes with a particular state in their spaces, using different methods, and apply the results to the study of the queue length of the M/G/1 queue with vacations.

Throughout the paper, we denote by R_+ the non-negative real number set, Z_+ the non-negative integer set and N_+ the positive integer set. Let $(\Phi_t)_{t \in R_+}$ be a time-homogeneous continuous-time Markov process on a locally compact separable metric space X, endowed with the Borel σ-field $\mathcal{B}(X)$. We denote by $P(t, x, A)$, $t \in R_+, A \in \mathcal{B}(X)$ the transition probability function of the Markov process:

$$P(t, x, A) = P_x[\Phi_t \in A] = E_x[I_{\{\Phi_t \in A\}}].$$

Here, P_x and E_x denote respectively the probability and expectation of the Markov process Φ_t under the initial condition $\Phi_0 = x$. We write $P(t, x, x) = P(t, x, \{x\})$.

The Markov process Φ_t is said to be ergodic if there exists (the unique) invariant probability measure π such that

$$\lim_{t \to \infty} \|P(t, x, \cdot) - \pi\| = 0 \tag{1.1}$$

for all $x \in X$, where $\| \cdot \|$ denotes the usual total variation norm; exponentially ergodic if it is ergodic, and there exists some $r > 0$ such that

$$\lim_{t \to \infty} e^{rt} \| P(t, x, \cdot) - \pi \| = 0 \qquad (1.2)$$

for π-a.e. $x \in X$; strongly ergodic if it is ergodic, and

$$\lim_{t \to \infty} \sup_{x \in X} \| P(t, x, \cdot) - \pi \| = 0. \qquad (1.3)$$

To obtain our results, we shall make the following assumption on the Markov process.

Assumption 1.1 *There exists a state x_0 such that whenever the Markov process Φ_t hits x_0, it will sojourn there for a random time that is positive and finite with probability 1.*

The assumption seems a little strong, but there are still plenty of Markov processes satisfying it, for example many queueing processes and all the totally stable continuous-time Markov chains. For a Markov process satisfying the assumption, due to the Markov property and the homogeneity, it can be easily proved that the sojourn time T_{x_0} in x_0 is exponentially distributed with some parameter λ, $0 < \lambda < \infty$. Z. T. Hou, et al [5] first investigated subgeometric convergence for such a process. Roughly speaking, subgeometric convergence is a kind of convergence quicker than ordinary ergodicity and slower than exponential ergodicity. As its continuation, we study exponential and strong ergodicity in the paper.

2 Exponential and Strong Ergodicity

In this section, we study exponential and strong ergodicity for Markov processes in terms of its discrete-time skeleton chains. In the following, we first review some definitions and results of discrete-time Markov chains.

Let Φ_n be a discrete-time Markov chain on X and define $\tau_x = \inf\{n \in N_+ : \Phi_n = x\}$ to be the first return time to the state x. The chain Φ_n is called ergodic, geometrically ergodic, and strongly ergodic if (1.1)–(1.3) hold for $t = n$, respectively. The following proposition states the known criteria of geometric and strong ergodicity, of which part (i) is from [6, Proposition 1] and part (ii) is from [7, Theorem 16.0.2].

Proposition 2.1 *Let Φ_n be an ergodic Markov chain on X with invariant probability measure π. Suppose that there exists a state $x_0 \in X$ such that $\pi_{x_0} > 0$. Then*

(i) *Φ_n is geometrically ergodic if and only if $E_{x_0}[e^{\alpha \tau_{x_0}}] < \infty$ for some $\alpha > 0$,*

(ii) *Φ_n is strongly ergodic if and only if $\sup_{x \in X} E_x[\tau_{x_0}] < \infty$.*

For the Markov process Φ_t, we define $\tau_x = \inf\{t > 0 : \Phi_t = x\}$ to be the first hitting time on x, and $\delta_x = \inf\{t > J_1 : \Phi_t = x\}$ to be the first return time to x, where J_1 is the first jump time of Φ_t. Define $\tau_{x_0}(h) = h \inf\{n \in N_+ : \Phi_{nh} = x_0\}$ to be the first hitting time on x_0 of its skeleton chain Φ_{nh}. Due to the convention, we adopt the notation τ_x for both discrete-time chains and continuous-time processes. The notation's meaning in the paper is clear and should not cause any confusion for understanding.

The following lemma reveals the relationship between the moments of the first hitting time of the Markov process Φ_t and those of its skeleton chain Φ_{nh}, which plays a crucial role in proving Theorem 2.1.

Lemma 2.1 *Suppose that the Markov process Φ_t satisfies Assumption* 1.1. *Then*

(i) $E_{x_0}[e^{r\delta_{x_0}}] < \infty$ *for some $r > 0$ if and only if $E_{x_0}[e^{\alpha \delta_{x_0}(h)}] < \infty$ for some $\alpha > 0$ and $h > 0$;*

(ii) $\sup\limits_{x \in X} E_x[\tau_{x_0}] < \infty$ *if and only if $\sup\limits_{x \in X} E_x[\tau_{x_0}(h)] < \infty$ for any $h > 0$.*

Proof Due to the assumption on the process, we know that $P(t, x_0, x_0) \geq e^{-\lambda t} > 0$ for all $t > 0$. We now prove that $P(t, x_0, x_0) < 1$ for all $t > 0$. We conversely assume that $P(\widehat{t}, x_0, x_0) = 1$ for some $\widehat{t} > 0$. Then for $s < \widehat{t}$, by Chapman-Kolmogorov equation we have

$$
\begin{aligned}
0 &= P(\widehat{t}, x_0, X - \{x_0\}) \\
&= \int_X P(\widehat{t} - s, x_0, \mathrm{d}y) P(s, y, X - \{x_0\}) \\
&\geq P(\widehat{t} - s, x_0, x_0) P(s, x_0, X - \{x_0\}).
\end{aligned}
$$

Since $P(\widehat{t} - s, x_0, x_0) > 0$, it implies that $P(s, x_0, X - \{x_0\}) = 0$, so $P(s, x_0, x_0) = 1$. And for $s > \widehat{t}$, choose some n such that $\frac{s}{n} < \widehat{t}$, we have $P(s, x_0, x_0) \geq [P(\frac{s}{n}, x_0, x_0)]^n = 1$. Hence, we get that $P(t, x_0, x_0) = 1$ for all $t > 0$, thus it conflicts Assumption 1.1.

(1) With the proved fact that $0 < P(t, x_0, x_0) < 1$ for all $t > 0$, we can prove part (i) by copying the proof of Lemma 6.2 in [1, Chapter 6], so we omit the proof.

(2) To prove (ii), we first prove the sufficiency. Suppose that $\sup\limits_{x \in X} E_x[\tau_{x_0}(h)] < \infty$ for any $h > 0$. It is possible that the skeleton chain Φ_{nh} can miss visits of the continuous-time process to x_0, and so result in $\tau_{x_0} \leq \tau_{x_0}(h)$. Hence we have $\sup\limits_{x \in X} E_x[\tau_{x_0}] < \infty$.

Next prove the necessity. Suppose that $\sup\limits_{x \in X} E_x[\tau_{x_0}] < \infty$. Then we have

$$
E_{x_0}[\delta_{x_0}] = \int_{X \backslash x_0} P(J_1, x_0, \mathrm{d}y) E_y[\tau_{x_0}] + E_{x_0}[J_1] \leq \sup_{x \in X} E_x[\tau_{x_0}] + \frac{1}{\lambda} < \infty. \tag{2.1}
$$

Assume that $\Phi_0 = x$. Once the process Φ_t arrives at x_0, it must stay at x_0 for a positive length, and then repeat leaving and returning infinitely. Let D_k be the kth sojourn time in x_0 and W_k be the length of the interval between the kth exit from x_0 and the next visit to x_0.

Note that W_k are independent and that D_k are independent of each other and the W_k. Moreover, D_k are identically exponentially distributed with parameter λ. Define $N = \min\{n \geq 1 \mid$ the h-skeleton is in state x_0 during the interval $D_n\}$. Then we have

$$
\begin{aligned}
E_x[\tau_{x_0}(h)] &\leq E_x[\tau_{x_0}] + E_{x_0}\Big[\sum_{i=1}^{N-1} (D_i + W_i) + h \Big] \\
&\leq (h + E_x[\tau_{x_0}]) + \sum_{n=1}^{\infty} E_{x_0}\Big[\sum_{i=1}^{n-1} (D_i + W_i) I_{\{N=n\}} \Big] \\
&\leq (h + E_x[\tau_{x_0}]) + \sum_{n=1}^{\infty} E_{x_0}\Big[\sum_{i=1}^{n-1} (D_i + W_i) I_{\bigcap\limits_{k=1}^{n-1} \{D_k \leq h\}} \Big]
\end{aligned}
$$

$$\leq (h + E_x[\tau_{x_0}]) + \sum_{n=1}^{\infty} E_{x_0}\left[(n-1)h + \sum_{i=1}^{n-1} W_i\right](1 - e^{-\lambda h})^{n-1}$$

$$\leq (h + E_x[\tau_{x_0}]) + \sum_{n=1}^{\infty}(n-1)(h + E_{x_0}[\delta_{x_0}])(1 - e^{-\lambda h})^{n-1}. \tag{2.2}$$

From (2.1) and (2.2), we have that $\sup_{x \in X} E_x[\tau_{x_0}(h)] < \infty$.

It was shown by [5] that the Markov process Φ_t is ergodic, subgeometrically convergent if and only if so is any skeleton chain Φ_{nh} of Φ_t. Combing this fact with the following lemma, we can say that the Markov process Φ_t has almost the same convergence as any skeleton chain of Φ_t.

Lemma 2.2 *The Markov process Φ_t is exponentially* (*resp. strongly*) *ergodic if and only if any skeleton chain Φ_{nh} of Φ_t is geometrically* (*resp. strongly*) *ergodic.*

Proof The necessity is obvious. In fact, if Φ_t is exponentially (resp. strongly) ergodic, by putting $t = nh$ in (1.2) (resp. (1.3)), then we get that the skeleton chain Φ_{nh} is geometrically (resp. strongly) ergodic.

For the sufficiency, [6, Theorem 1] has shown that if any skeleton chain Φ_{nh} of Φ_t is geometrically ergodic, then Φ_t is exponentially ergodic. Similarly, we can prove that if the skeleton chain Φ_{nh} of Φ_t is strongly ergodic, then so is Φ_t.

We are now in a position to establish our main result.

Theorem 2.1 *Suppose that the Markov process Φ_t is ergodic and satisfies Assumption* 1.1. *Then*

(ⅰ) *Φ_t is exponentially ergodic if and only if $E_{x_0}[e^{r\delta_{x_0}}] < \infty$ for some $r > 0$;*

(ⅱ) *Φ_t is strongly ergodic if and only if $\sup_{x \in X} E_x[\tau_{x_0}] < \infty$.*

Proof (ⅰ) In the proof Theorem 2.1 of [5], we have shown that $\pi_{x_0} > 0$. Suppose that Φ_t is exponentially ergodic. By Lemma 2.2, we know that any skeleton chain Φ_{nh} of Φ_t is geometrically ergodic, and by Proposition 2.1 we know that there exists some $\alpha > 0$ such that $E_{x_0}[e^{\alpha \delta_{x_0}(h)}] < \infty$. Hence, it follows from Lemma 2.1 that there exists some $r > 0$ such that $E_{x_0}[e^{r\delta_{x_0}}] < \infty$.

Conversely, suppose that $E_{x_0}[e^{r\delta_{x_0}}] < \infty$ for some $r > 0$. By Lemma 2.1 we know that there exists some $\alpha > 0$, $h > 0$ such that $E_{x_0}[e^{\alpha \tau_{x_0}(h)}] < \infty$, and by Proposition 2.1 and Lemma 2.2 we know that Φ_t is exponentially ergodic.

(ⅱ) Following the same lines as the proof of part (ⅰ), we easily have that part (ⅱ) holds from Proposition 2.1, Lemmas 2.1 and 2.2.

Remark 2.1 Part (ⅰ) of Theorem 2.1 is not an entirely new result. The sufficiency of part (ⅰ) can also be proved with different methods. (For more details, see [4, Theorem 5.2 and Theorem 6.2].)

3 Length of the M/G/1 Queue with Multiple Vacations

There is much literature (e.g. [8, 9]) on M/G/1 queues with vacations. These queues include, for example the M/G/1 queue with step-up time, with N-policy, with single vacation, or with multiple vacations. In this section, we only study the most complicated case: the M/G/1 queue with multiple vacations, which is denoted simply by M/G/1(E, MV) (see e.g. [10–12]), and the corresponding results for the other queues can be easily obtained by the same method.

M/G/1(E, MV) is gotten by introducing the strategy of exhaustive service and multiple vacations to the classical M/G/1 queue: once the system has no customers, the server begins a vacation of random length V immediately. If, when the vacation ends, the system still has no customers, then the server continues with further independent, identically distributed vacations that do not end until the system has customers queueing when a vacation ends. Here V is always assumed to be a non-negative random variable, with distribution function $V(x)$, that has finite first moment, i.e., $E[V] < \infty$. For M/G/1(E, MV), the customers arrive according to a Poisson process with the parameter λ, $0 < \lambda < \infty$ and the service time B has a general distribution $B(x)$.

Let Q_b be the number of customers in the system when one busy period begins. Then

$$P[Q_b = j] = \frac{v_j}{1 - v_0}, \quad j \in N_+,$$

where

$$v_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dV(t), \quad j \in Z_+$$

is the probability that j customers join the queue during a vacation. Denote by D_v the busy period of M/G/1(E, MV) and by D the busy period of the classical M/G/1 queue. It is easy to see that

$$\{D_v \mid Q_b = k\} = \{D_1 + D_2 + \cdots + D_k\}, \tag{3.1}$$

where D_k is the busy period of the classical M/G/1 queue caused by the kth customer and D_i's are independent and identically distributed. Note that D_k has the same distribution as D. Let J be the number of vacations during a series of consecutive vacations. Then

$$P[J = j] = v_0^{j-1}(1 - v_0), \quad j \in N_+.$$

Let V_v be the vacation period of M/G/1(E, MV). Then

$$\{V_v \mid J = j\} = \{V_1 + V_2 + \cdots + V_j\}, \tag{3.2}$$

where V_i denotes the ith vacation and the V_i's are independent and identically distributed. Define

$$\frac{1}{\mu} = \int_0^\infty x \, dB(x) \quad \text{and} \quad \rho = \frac{\lambda}{\mu}.$$

Let L_t be the queue length process of M/G/1(E, MV). It is known that L_t is not a Markov process unless $B(x)$ is exponentially distributed. We introduce a supplementary variable as

follows:

$$\theta_t = \text{the elapsed service time of the customer being served at time } t$$
$$(= 0 \text{ if the server is idle at time } t).$$

Then (L_t, θ_t) becomes a continuous-time Markov process on the two-dimensional state space $X = Z_+ \times R_+$. It is easy to see that when (L_t, θ_t) hits the state $(0,0)$, it will stay there for a random length which is exponentially distributed with the parameter λ, so (L_t, θ_t) satisfies Assumption 1.1 with $x_0 = (0,0)$. Several types of ergodicity for the discrete-time embedded chain of $L(t)$ were studied in [13], and polynomial convergence for (L_t, θ_t) was investigated in [5]. By [5, Theorem 3.1], we know that (L_t, θ_t) is ergodic if and only if $\rho < 1$.

For a given constant $r > 0$, denote by $\mathcal{G}^+(r)$ the set of all distributions such that

$$\int_0^\infty e^{rx} dF(x) < \infty,$$

and by $\mathcal{G}^+$ the set of all nonnegative distributions with finite exponential moments, i.e.

$$\mathcal{G}^+ = \bigcup_{r>0} \mathcal{G}^+(r).$$

Lemma 3.1 *Suppose that $\rho < 1$ for the classical* M/G/1 *queue. Then its busy period distribution $D(x) \in \mathcal{G}^+$ if and only if its service time distribution $B(x) \in \mathcal{G}^+$.*

Proof Let $W(t)$ be the virtual waiting time of the classical M/G/1 queue. Then $W(t)$ is a Markov process satisfying Assumption 1.1 and the state 0 is the particular state x_0. From [6, Theorem 2], we know that $W(t)$ is ergodic if and only if $\rho < 1$. Moreover, $W(t)$ is exponentially ergodic if and only if $B(x)$ is in $\mathcal{G}^+$. Now suppose that $\rho < 1$. Then it follows from Theorem 2.1 that $W(t)$ is exponentially ergodic if and only if $E_0[e^{r\delta_0}] = E[e^{rD}] < \infty$, or equivalently, $D(x)$ is in $\mathcal{G}^+$. Hence, $D(x)$ is in $\mathcal{G}^+$ if and only if $B(x)$ is in $\mathcal{G}^+$.

Remark 3.1 Lemma 3.1 has been obtained with a different method (see, e.g. [14]), which is important in the queue literature. Here, we display a new and short proof of it.

Theorem 3.1 *Suppose that (L_t, θ_t) is ergodic. Then*
(i) *(L_t, θ_t) is exponentially ergodic if and only if both $V(x)$ and $B(x)$ are in $\mathcal{G}^+$,*
(ii) *(L_t, θ_t) is not strongly ergodic.*

Proof (i) If both $V(x)$ and $B(x)$ belong to $\mathcal{G}^+$, then there exists some $r > 0$, such that

$$E[e^{rB}] < +\infty \quad \text{and} \quad E[e^{rV}] < +\infty.$$

Since $E[e^{rB}] < +\infty$, it implies from Lemma 3.1 that there exists some $r_1 > 0$, such that $E[e^{r_1 D}] < \infty$. Thus the functions $E[e^{sD}]$ and $E[e^{sV}]$ are continuous in s when $0 \le s \le \min\{r, r_1\}$, so we can choose an appropriate r_2 that is greater than, but sufficiently close to, 0 such that

$$\lambda(E[e^{r_2 D}] - 1) < r, \quad v_0 E[e^{r_2 V}] < 1.$$

Thus from (3.1), we have

$$
\begin{aligned}
E[\mathrm{e}^{r_2 D_v}] &= \sum_{k=1}^{\infty} P\{Q_b = k\} E[\mathrm{e}^{r_2(D_1 + D_2 + \cdots + D_{Q_b})} \mid Q_b = k] \\
&= \sum_{k=1}^{\infty} \frac{v_k}{1 - v_0} (E[\mathrm{e}^{r_2 D}])^k \\
&= \sum_{k=1}^{\infty} \frac{(E[\mathrm{e}^{r_2 D}])^k}{1 - v_0} \int_0^{\infty} \frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t} \mathrm{d}V(t) \\
&\le \frac{1}{1 - v_0} \int_0^{\infty} \mathrm{e}^{\lambda(E[\mathrm{e}^{r_2 D}] - 1)t} \mathrm{d}V(t) \\
&< \infty,
\end{aligned}
\tag{3.3}
$$

and from (3.2), we get

$$
\begin{aligned}
E[\mathrm{e}^{r_2 V_v}] &= \sum_{j=1}^{\infty} P\{J = j\} E[\mathrm{e}^{r_2(V_1 + V_2 + \cdots + V_J)} \mid J = j] \\
&= \sum_{j=1}^{\infty} (1 - v_0) v_0^{j-1} (E[\mathrm{e}^{r_2 V}])^j \\
&< \infty.
\end{aligned}
\tag{3.4}
$$

Hence,

$$
E_{(0,0)}[\mathrm{e}^{r_2 \delta_{(0,0)}}] = E_{(0,0)}[\mathrm{e}^{r_2(D_v + V_v)}] = E[\mathrm{e}^{r_2 V_v}] E[\mathrm{e}^{r_2 D_v}] < \infty,
\tag{3.5}
$$

and by Theorem 2.1 we see that (L_t, θ_t) is exponentially ergodic.

On the other hand, if (L_t, θ_t) is exponentially ergodic, then by Theorem 2.1 we know that for some $r > 0$,

$$
E_{(0,0)}[\mathrm{e}^{r \delta_{(0,0)}}] < \infty.
$$

We get from (3.3)–(3.5) that both $V(x)$ and $B(x)$ are in $\mathcal{G}^+$.

(ii) Since

$$
\sup_{x \in Z_+ \times R_+} E_x[\delta_{(0,0)}] \ge \sup_{i \in Z_+} E_{(i,0)}[\delta_{(0,0)}] \ge \sup_{i \in Z_+} i E[D] = \infty,
$$

it follows from Theorem 2.1 that (L_t, θ_t) is not strongly ergodic.

Remark 3.2 Combing Theorem 3.1 with [5, Theorem 3.3], we know that exponential (resp. polynomial) moments of V and B determine the corresponding convergence of (L_t, θ_t). It should be noted that usually queue length processes are not strongly ergodic.

References

[1] Anderson, W. J., Continuous-Time Markov Chains, An Applications-Oriented Approach, Springer-Verlag, New York, 1991.

[2] Chen, M. F., From Markov Chains to Non-equilibrium Particle Systems, World Scientific, Singapore, 1992.

[3] Meyn, S. P. and Tweedie, R. L., Stability of Markovian processes III: Foster-Lyaunov criteria for Continuous-time rocesses, *Adv. Appl. Prob.*, **25**, 1993, 518–548.

[4] Down, D., Meyn, S. P. and Tweedie, R. L., Exponential and uniform egoddicity of Markov processes, *Ann. Prob.*, **23**, 1995, 1671–1691.

[5] Hou, Z. T., Liu, Y. Y. and Zhang, H. J., Subgeometric rates of convergence for a class of continuous-time Markov processes, *J. Appl. Prob.*, **42**, 2005, 698–712.

[6] Tuominen, P. and Tweedie, R. L., Exponential ergodicity in Markovian queueing and dam models, *J. Appl. Prob.*, **16**, 1979, 867–880.

[7] Meyn, S. P. and Tweedie, R. L., Markov Chains and Stochastic Stability, Springer, Berlin, 1993.

[8] Doshi, B., An M/G/1 queue with variable vacation, Modeling Techniques and Tools for Performance Analysis'85, Abu el Ata N. (ed.), North-Holland, Amsterdam, 1986, 67–81.

[9] Levy, H. and Kleinrock, L., A queue with starter and a queue with vacations: Delay annlysis by decompostition, *Opns. Res.*, **34**, 1986, 426–436.

[10] Levy, Y. and Yechiali, U., Utilization of idle time in an M/G/1 queueing system, *Manag. Sci.*, **22**, 1975, 202–211.

[11] Fuhrman, S., A note on the M/G/1 queue with server vacation, *Opns. Res.*, **32**, 1984, 1368–1373.

[12] Harris, C. and Marchal, W., State dependence in M/G/1 server vacation models, *Opns. Res.,* **36**, 1988, 560–565.

[13] Hou, Z. T. and Liu, Y. Y., Explicit criteria for several types of ergodicity of the embedded M/G/1 and GI/M/n queues, *J. Appl. Prob.*, **41**, 2004, 778–790.

[14] Foss, S. and Sapozhnikov, A., On the existence of moments for the busy period in a single-server queue, *Math. Oper. Res.*, **29**, 2004, 592–601.