

# A sufficient condition for the subexponential asymptotics of GI/G/1-type Markov chains with queueing applications

Hiroyuki Masuyama<sup>†</sup>

Department of Systems Science, Graduate School of Informatics, Kyoto University  
Kyoto 606-8501, Japan

## Abstract

The main contribution of this paper is to present a new sufficient condition for the subexponential asymptotics of the stationary distribution of a GI/GI/1-type Markov chain without jumps from level “infinity” to level zero. For simplicity, we call such Markov chains *GI/GI/1-type Markov chains without disasters* because they are often used to analyze semi-Markovian queues without “disasters”, which are negative customers who remove all the customers in the system (including themselves) on their arrivals. In this paper, we demonstrate the application of our main result to the stationary queue length distribution in the standard BMAP/GI/1 queue. Thus we obtain new asymptotic formulas and prove the existing formulas under weaker conditions than those in the literature. In addition, applying our main result to a single-server queue with Markovian arrivals and the  $(a, b)$ -bulk-service rule (i.e., MAP/GI<sup>(a,b)</sup>/1 queue), we obtain a subexponential asymptotic formula for the stationary queue length distribution.

**Keywords:** Subexponential asymptotics; GI/G/1-type Markov chain; disaster; bulk service; BMAP/GI/1 queue; MAP/GI<sup>(a,b)</sup>/1 queue

**Mathematics Subject Classification:** Primary 60K25; Secondary 60J10

## 1 Introduction

This paper studies the subexponential asymptotics of the stationary distribution of a GI/GI/1-type Markov chain (see, e.g., He 2014) without jumps from level “infinity” to level zero. For simplicity, we call such Markov chains *GI/GI/1-type Markov chains without disasters* because they are often used to analyze semi-Markovian queues without “disasters”, which are negative customers who remove all the customers in the system (including themselves) on their arrivals. It should be noted that every M/G/1-type Markov chain is a GI/GI/1-type Markov chain without disasters (see, e.g., He 2014).

Several researchers have studied the subexponential asymptotics of the stationary distributions of GI/GI/1-type Markov chains (including M/G/1-type ones). Asmussen and Møller

---

<sup>†</sup>E-mail: masuyama@sys.i.kyoto-u.ac.jp

(1999) derive subexponential asymptotic formulas for the stationary distribution of a M/GI/1-type Markov chain with subexponential level increments. Li and Zhao (2005) study a GI/GI/1-type Markov chain with subexponential level increments, though some of their asymptotic formulas are incorrect (for details, see Masuyama 2011). Takine (2004) presents a subexponential asymptotic formula for M/GI/1-type Markov chains, under the assumption that the integrated tail distribution of level increments is subexponential. It should be noted that Takine (2004)'s assumption does not necessarily imply the subexponentiality of level increments themselves (see, e.g., Remark 3.5 in Sigman 1999). Focusing on the period of the  $G$ -matrix, Masuyama (2011) establishes sufficient conditions for the subexponential asymptotics for M/GI/1-type Markov chains, which are weaker than those presented in the literature (Asmussen and Møller 1999; Li and Zhao 2005; Takine 2004), except for being limited to the M/G/1-type Markov chain. Masuyama (2011) also points out that Takine (2004)'s derivation of the asymptotic formula implicitly assumes the aperiodicity of the  $G$ -matrix. Kim and Kim (2012) weaken Masuyama (2011)'s sufficient condition in the case where the  $G$ -matrix is periodic. Kimura et al. (2013) present a comprehensive study on the subexponential asymptotics of GI/GI/1-type Markov chains. They study the *locally* subexponential asymptotics (Asmussen et al. 2003) as well as the (ordinarily) subexponential asymptotics. The sufficient conditions presented in Kimura et al. (2013) are weaker than those reported in the literature mentioned above.

The main result of this paper is to present a new sufficient condition for the subexponential asymptotics of the stationary distribution of a GI/GI/1-type Markov chain without disasters. This sufficient condition is weaker than the corresponding one presented in Kimura et al. (2013).

In this paper, we demonstrate the application of the main result to the stationary queue length distribution in the (standard) BMAP/GI/1 queue (see, e.g., Lucantoni 1991). According to Takine (2000), the stationary queue length distribution in the BMAP/GI/1 queue is equivalent to the stationary distribution of a certain M/G/1-type Markov chain. Combining this fact and the main result of this paper, we derive four subexponential asymptotic formulas for the stationary queue length distribution. Two of the four formulas are proved under weaker conditions than the two corresponding ones presented in Masuyama et al. (2009); and the other two formulas are shown for a BMAP/GI/1 queue with consistently varying service times, which is not considered in Masuyama et al. (2009).

We also apply the main result of this paper to a single-server queue with Markovian arrivals and the  $(a, b)$ -bulk-service rule, denoted by MAP/GI $^{(a,b)}$ /1 queue (see, e.g., Singh et al. 2013). For the MAP/GI $^{(a,b)}$ /1 queue, we construct a GI/GI/1-type Markov chain without disasters by observing the queue length process at departure points. Thus using the main result, we obtain a subexponential asymptotic formula for the stationary queue length distribution at departure points. Combining the obtained formula with the relationship between the stationary queue length distribution at departure points and that at an arbitrary time point, we have a subexponential asymptotic formula for the stationary queue length distribution at an arbitrary time point.

The rest of this paper is divided into four sections. Section 2 provides basic definitions,

notation and preliminary results. Section 3 presents the main result of this paper. Sections 4 and 5 discuss the applications of the main result.

## 2 Preliminaries

### 2.1 Basic definitions and notation

Let  $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ ,  $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$  and  $\mathbb{N} = \{1, 2, 3, \dots\}$ , respectively. For any distribution function  $F$  on  $\mathbb{R}_+ := [0, \infty)$ , let  $\overline{F} = 1 - F$  and  $F_e$  denote the equilibrium distribution function of  $F$ , i.e.,  $F_e(x) = \int_0^x \overline{F}(y)dy / \int_0^\infty \overline{F}(y)dy$  for  $x \geq 0$ , which is well-defined if  $F$  has a positive finite mean. For any nonnegative random variable  $Y$  with positive finite mean, let  $Y_e$  denote the equilibrium random variable of  $Y$  such that

$$\mathbb{P}(Y_e \leq x) = \frac{1}{\mathbb{E}[Y]} \int_0^x \mathbb{P}(Y > y)dy, \quad x \in \mathbb{Z}_+;$$

and  $Y_{\text{de}} = \lfloor Y_e \rfloor$ , which is called the discretized equilibrium random variable of  $Y$ . If  $Y$  is nonnegative integer-valued, then

$$\mathbb{P}(Y_{\text{de}} = k) = \frac{1}{\mathbb{E}[Y]} \mathbb{P}(Y > k), \quad k \in \mathbb{Z}_+.$$

We now define  $\mathbf{e}$  and  $\mathbf{I}$  as the column vector of ones and the identity matrix, respectively, with appropriate dimensions according to the context. The superscript “t” represents the transpose operator for vectors and matrices. The notation  $[\cdot]_{i,j}$  (resp.  $[\cdot]_i$ ) denotes the  $(i, j)$ th (resp.  $i$ th) element of the matrix (resp. vector) in the square brackets.

For any matrix sequence  $\{\mathbf{M}(k); k \in \mathbb{Z}\}$ , let  $\overline{\mathbf{M}}(k) = \sum_{l=k+1}^\infty \mathbf{M}(l)$  and  $\overline{\overline{\mathbf{M}}}(k) = \sum_{l=k+1}^\infty \overline{\mathbf{M}}(l)$  for  $k \in \mathbb{Z}$ . For any two matrix sequences  $\{\mathbf{M}(k); k \in \mathbb{Z}\}$  and  $\{\mathbf{N}(k); k \in \mathbb{Z}\}$  such that their products are well-defined, let  $\{\mathbf{M} * \mathbf{N}(k); k \in \mathbb{Z}\}$  denote the convolution of  $\{\mathbf{M}(k)\}$  and  $\{\mathbf{N}(k)\}$ , i.e.,

$$\mathbf{M} * \mathbf{N}(k) = \sum_{l \in \mathbb{Z}} \mathbf{M}(k-l)\mathbf{N}(l) = \sum_{l \in \mathbb{Z}} \mathbf{M}(l)\mathbf{N}(k-l), \quad k \in \mathbb{Z}.$$

In addition, for any square matrix sequence  $\{\mathbf{M}(k); k \in \mathbb{Z}\}$ , let  $\{\mathbf{M}^{*n}(k); k \in \mathbb{Z}\}$  ( $n \in \mathbb{N}$ ) denote the  $n$ -fold convolution of  $\{\mathbf{M}(k)\}$  with itself, i.e.,

$$\mathbf{M}^{*n}(k) = \sum_{l \in \mathbb{Z}} \mathbf{M}^{*(n-1)}(k-l)\mathbf{M}(l), \quad k \in \mathbb{Z},$$

where  $\mathbf{M}^{*0}(0) = \mathbf{I}$  and  $\mathbf{M}^{*0}(k) = \mathbf{O}$  for  $k \in \mathbb{Z} \setminus \{0\}$ .

Finally, for simplicity, we may write  $\mathbf{Z}(x) = o(f(x))$  and  $\mathbf{Z}(x) \overset{x}{\sim} \tilde{\mathbf{Z}}f(x)$  to represent

$$\lim_{x \rightarrow \infty} \frac{\mathbf{Z}(x)}{f(x)} = \mathbf{O}, \quad \lim_{x \rightarrow \infty} \frac{\mathbf{Z}(x)}{f(x)} = \tilde{\mathbf{Z}},$$

respectively.

The above definitions and notation for matrices are applied to vectors and scalars in an appropriate manner.

## 2.2 Stationary distribution of GI/G/1-type Markov chain

Let  $\mathbb{M}_0 = \{1, 2, \dots, M_0\}$  and  $\mathbb{M} = \{1, 2, \dots, M\}$ , where  $M_0, M \in \mathbb{N}$ . We then define  $\{(X_n, S_n); n \in \mathbb{Z}_+\}$  as a Markov chain with state space  $\mathbb{F} := (\{0\} \times \mathbb{M}_0) \cup (\mathbb{N} \times \mathbb{M})$  and transition probability matrix  $\mathbf{T}$ , which is given by

$$\mathbf{T} = \begin{pmatrix} \mathbf{B}(0) & \mathbf{B}(1) & \mathbf{B}(2) & \mathbf{B}(3) & \cdots \\ \mathbf{B}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \mathbf{A}(2) & \cdots \\ \mathbf{B}(-2) & \mathbf{A}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \cdots \\ \mathbf{B}(-3) & \mathbf{A}(-2) & \mathbf{A}(-1) & \mathbf{A}(0) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (2.1)$$

where  $\mathbf{B}(0)$  and  $\mathbf{A}(0)$  in the diagonal blocks are  $M_0 \times M_0$  and  $M \times M$  matrices, respectively. Each element of  $\mathbf{T}$  is specified by two nonnegative integers  $(k, i) \in \mathbb{F}$ , where the first variable  $k$  is called *level* and the second one  $i$  is called *phase*.

Throughout this paper, we make the following assumption:

**Assumption 2.1** (i)  $\mathbf{T}$  is irreducible and stochastic; (ii)  $\sum_{k=1}^{\infty} k\mathbf{B}(k)\mathbf{e} < \infty$ ; (iii)  $\mathbf{A} := \sum_{k \in \mathbb{Z}} \mathbf{A}(k)$  is irreducible and stochastic; (iv)  $\sum_{k \in \mathbb{Z}} |k|\mathbf{A}(k) < \infty$ ; (v)  $\sigma := \boldsymbol{\pi} \sum_{k \in \mathbb{Z}} k\mathbf{A}(k)\mathbf{e} < 0$ , where  $\boldsymbol{\pi} := (\pi_i)_{i \in \mathbb{M}}$  is the stationary probability vector of  $\mathbf{A} := \sum_{k \in \mathbb{Z}} \mathbf{A}(k)$ .

**Remark 2.1**  $\mathbf{T}$  is positive recurrent if and only if  $\sigma < 0$  and  $\sum_{k=1}^{\infty} k\mathbf{B}(k)\mathbf{e} < \infty$ , provided that  $\mathbf{T}$  and  $\mathbf{A}$  are irreducible and stochastic (see, e.g., Asmussen 2003, Chapter XI, Proposition 3.1). Therefore Assumption 2.1 is equivalent to condition (I) of Assumption 2 in Kimura et al. (2013).

**Remark 2.2** For  $k \in \mathbb{N}$ , we have  $\mathbf{B}(-k)\mathbf{e} + \sum_{l=-k+1}^{\infty} \mathbf{A}(l)\mathbf{e} = \mathbf{e}$ . Thus condition (iii) of Assumption 2.1 implies  $\lim_{k \rightarrow \infty} \mathbf{B}(-k) = \mathbf{O}$ , which shows that the one-step transition probability from level “infinity” to level zero is equal to zero, i.e., no “disasters” happen in the context of queueing models.

Let  $\boldsymbol{x} := (\boldsymbol{x}(0), \boldsymbol{x}(1), \boldsymbol{x}(2), \dots)$  denote the unique stationary probability vector of  $\mathbf{T}$ , where  $\boldsymbol{x}(0)$  (resp.  $\boldsymbol{x}(k)$ ;  $k \in \mathbb{N}$ ) is a  $1 \times M_0$  (resp.  $1 \times M$ ) subvector of  $\boldsymbol{x}$  corresponding to level zero (resp. level  $k$ ). To characterize  $\boldsymbol{x} = (\boldsymbol{x}(0), \boldsymbol{x}(1), \boldsymbol{x}(2), \dots)$ , we introduce  $\mathbf{R}$ -matrices. Let  $\mathbf{R}_0(k)$  and  $\mathbf{R}(k)$  ( $k \in \mathbb{N}$ ) denote  $M_0 \times M$  and  $M \times M$  matrices, respectively, such that

$$[\mathbf{R}_0(k)]_{i,j} = \mathbb{E} \left[ \sum_{n=1}^{T_{<k}} \mathbb{1}(X_n = k, S_n = j) \mid X_0 = 0, S_0 = i \right],$$

and for any fixed  $\nu \in \mathbb{N}$ ,

$$[\mathbf{R}(k)]_{i,j} = \mathbb{E} \left[ \sum_{n=1}^{T_{<k+\nu}} \mathbb{1}(X_n = k + \nu, S_n = j) \mid X_0 = \nu, S_0 = i \right],$$

where  $T_{<k} = \inf\{n \in \mathbb{N}; X_n < k \leq X_m \ (m = 1, 2, \dots, n-1)\}$  and  $\mathbb{1}(\cdot)$  denotes the indicator function of the event in the parentheses. For convenience, let  $\mathbf{R}_0(0) = \mathbf{O}$  and  $\mathbf{R}(0) = \mathbf{O}$ . It then follows (see, e.g., Kimura et al. 2013; Li and Zhao 2005) that

$$\mathbf{x}(k) = \mathbf{x}(0)\mathbf{R}_0 * \mathbf{F}(k), \quad k \in \mathbb{N},$$

where

$$\mathbf{F}(k) = \sum_{n=0}^{\infty} \mathbf{R}^{*n}(k), \quad k \in \mathbb{Z}_+. \quad (2.2)$$

Thus we have

$$\bar{\mathbf{x}}(k) = \mathbf{x}(0)\overline{\mathbf{R}_0 * \mathbf{F}(k)}, \quad k \in \mathbb{Z}_+, \quad (2.3)$$

and especially,

$$\bar{\mathbf{x}}(0) = \mathbf{x}(0)\mathbf{R}_0(\mathbf{I} - \mathbf{R})^{-1}, \quad (2.4)$$

where  $\mathbf{R} = \sum_{k=1}^{\infty} \mathbf{R}(k)$  and  $\mathbf{R}_0 = \sum_{k=1}^{\infty} \mathbf{R}_0(k)$ .

For the discussion in the next section, we need some more definitions and preliminary results. Let  $\mathbf{G}(k)$  ( $k \in \mathbb{N}$ ) denote an  $M \times M$  matrix such that for any fixed  $\nu \in \mathbb{N}$ ,

$$[\mathbf{G}(k)]_{i,j} = \mathbf{P}(X_{T_{<k+\nu}} = \nu, S_{T_{<k+\nu}} = j \mid X_0 = k + \nu, S_0 = i), \quad k \in \mathbb{N}.$$

Let  $\Phi(0)$  denote an  $M \times M$  matrix such that for any fixed  $\nu \in \mathbb{N}$ ,

$$[\Phi(0)]_{i,j} = \mathbf{P}(S_{T_{\downarrow\nu}} = j \mid X_0 = \nu, S_0 = i),$$

where  $T_{\downarrow\nu} = \inf\{n \in \mathbb{N}; X_n = \nu < X_m \ (m = 1, 2, \dots, n-1)\}$ . Note here that  $\sum_{n=0}^{\infty} (\Phi(0))^n = (\mathbf{I} - \Phi(0))^{-1}$  exists because  $\mathbf{T}$  is irreducible. Since Assumption 2.1 is equivalent to condition (I) of Assumption 2 in Kimura et al. (2013) (see Remark 2.1), we have the following result:

**Proposition 2.1 (Kimura et al. 2013, Lemma 3.1.1)** *Under Assumption 2.1,*

$$\sigma = -\pi(\mathbf{I} - \mathbf{R})(\mathbf{I} - \Phi(0)) \sum_{k=1}^{\infty} k\mathbf{G}(k)\mathbf{e} \in (-\infty, 0).$$

Let  $\mathbf{L}(k)$  ( $k \in \mathbb{N}$ ) denote an  $M \times M$  matrix such that for any fixed  $\nu \in \mathbb{N}$ ,

$$[\mathbf{L}(k)]_{i,j} = \mathbf{P}(S_{T_{\downarrow\nu}} = j \mid X_0 = k + \nu, S_0 = i), \quad k \in \mathbb{N}.$$

We then have

$$\mathbf{L}(k) = \sum_{m=1}^k \mathbf{G}^{*m}(k), \quad k \in \mathbb{N}.$$

In terms of  $\mathbf{L}(k)$ , the matrices  $\mathbf{R}_0(k)$  and  $\mathbf{R}(k)$  are expressed as

$$\begin{aligned} \mathbf{R}_0(k) &= \left[ \mathbf{B}(k) + \sum_{m=1}^{\infty} \mathbf{B}(k+m)\mathbf{L}(m) \right] (\mathbf{I} - \Phi(0))^{-1}, \quad k \in \mathbb{N}, \\ \mathbf{R}(k) &= \left[ \mathbf{A}(k) + \sum_{m=1}^{\infty} \mathbf{A}(k+m)\mathbf{L}(m) \right] (\mathbf{I} - \Phi(0))^{-1}, \quad k \in \mathbb{N}. \end{aligned} \quad (2.5)$$

The following proposition is used to prove Lemma 3.1 in the next section.

**Proposition 2.2 (Kimura et al. 2013, Lemma 3.1.2)** *If Assumption 2.1 holds, then*

$$\lim_{n \rightarrow \infty} \sum_{l=0}^{\tau-1} \mathbf{L}(n\tau + l) = \tau \mathbf{e}\boldsymbol{\psi},$$

where

$$\boldsymbol{\psi} = \boldsymbol{\pi}(\mathbf{I} - \mathbf{R})(\mathbf{I} - \boldsymbol{\Phi}(0))/(-\sigma), \quad (2.6)$$

and  $\tau$  denotes the period of an Markov additive process with kernel  $\{\mathbf{A}(k); k \in \mathbb{Z}\}$  (see Appendix B in Kimura et al. 2010).

**Remark 2.3** Proposition 2.1 implies that  $\boldsymbol{\psi}$  is finite.

## 2.3 Long-tailed distributions

We begin with the definitions of the long-tailed class and higher-order long-tailed classes.

**Definition 2.1** A nonnegative random variable  $U$  and its distribution  $F_U$  are said to be long-tailed if  $\mathbb{P}(U > x) > 0$  for all  $x \geq 0$  and  $\mathbb{P}(U > x + y) \stackrel{x}{\sim} \mathbb{P}(U > x)$  for some (thus all)  $y > 0$ . The class of long-tailed distributions is denoted by  $\mathcal{L}$ .

**Definition 2.2** A nonnegative random variable  $U$  and its distribution  $F_U$  are said to be the  $\mu$ -th-order long-tailed if  $U^{1/\mu} \in \mathcal{L}$ , where  $\mu \geq 1$ . The class of the  $\mu$ -th-order long-tailed distributions is denoted by  $\mathcal{L}^\mu$ . Further if  $U \in \mathcal{L}^\mu$  (resp.  $F_U \in \mathcal{L}^\mu$ ) for all  $\mu \geq 1$ , we write  $U \in \mathcal{L}^\infty$  (resp.  $F_U \in \mathcal{L}^\infty$ ) and call  $U$  (resp.  $F_U$ ) infinite-order long-tailed.

The basic properties of the higher-order long-tailed classes (including the long-tailed class) are summarized in Proposition 2.3 below.

**Proposition 2.3 (Masuyama 2013, Lemmas A.1–A.3)**

- (i)  $\mathcal{L}^{\mu_2} \subset \mathcal{L}^{\mu_1}$  for  $1 \leq \mu_1 < \mu_2$ .
- (ii) If  $U \in \mathcal{L}^\mu$  ( $\mu \geq 1$ ), then  $\mathbb{P}(U > x) = \exp\{-o(x^{1/\mu})\}$ .
- (iii)  $U \in \mathcal{L}^\mu$  ( $\mu \geq 1$ ) if and only if  $\mathbb{P}(U > x - \xi x^{1-1/\mu}) \stackrel{x}{\sim} \mathbb{P}(U > x)$  for some (thus all)  $\xi \in \mathbb{R} \setminus \{0\}$ .

Next we introduce the subexponential class, which is the largest tractable subclass of  $\mathcal{L}$ .

**Definition 2.3 (Goldie and Klüppelberg 1998; Sigman 1999)** A nonnegative random variable  $U$  and its distribution  $F_U$  are said to be subexponential if  $\mathbb{P}(U > x) > 0$  for all  $x \geq 0$  and

$$\mathbb{P}(U_1 + U_2 > x) \stackrel{x}{\sim} 2\mathbb{P}(U > x),$$

where  $U_i$ 's ( $i = 1, 2, \dots$ ) are independent copies of  $U$ . The class of subexponential distributions is denoted by  $\mathcal{S}$ .

**Remark 2.4** The class  $\mathcal{S}$  includes Pareto, heavy-tailed Weibull, lognormal, Burr, and loggamma distributions, etc (see, e.g., Goldie and Klüppelberg 1998).

The following proposition is used several times in the subsequent sections.

**Proposition 2.4 (Masuyama 2011, Proposition A.3)** Let  $\{\mathbf{M}(k); k \in \mathbb{Z}_+\}$  and  $\{\mathbf{N}(k); k \in \mathbb{Z}_+\}$  denote finite-dimensional nonnegative matrix sequences such that their convolution  $\{\mathbf{M} * \mathbf{N}(k); k \in \mathbb{Z}_+\}$  is well-defined and  $\mathbf{M} := \sum_{k=0}^{\infty} \mathbf{M}(k)$  and  $\mathbf{N} := \sum_{k=0}^{\infty} \mathbf{N}(k)$  are finite. Suppose that for some random variable  $U \in \mathcal{S}$ ,

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{M}}(k)}{\mathbb{P}(U > k)} = \widetilde{\mathbf{M}} \geq \mathbf{O}, \quad \lim_{k \rightarrow \infty} \frac{\overline{\mathbf{N}}(k)}{\mathbb{P}(U > k)} = \widetilde{\mathbf{N}} \geq \mathbf{O},$$

where  $\widetilde{\mathbf{M}} = \widetilde{\mathbf{N}} = \mathbf{O}$  is allowed. We then have

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{M} * \mathbf{N}}(k)}{\mathbb{P}(U > k)} = \widetilde{\mathbf{M}}\mathbf{N} + \mathbf{M}\widetilde{\mathbf{N}}.$$

Finally we describe two subclasses of  $\mathcal{S}$ , which are used to apply the main result of this paper to the BMAP/GI/1 queue in Section 4.

**Definition 2.4 (Shneer 2006)** A nonnegative random variable  $U$  and its distribution function  $F_U$  and cumulative hazard function  $Q_U := -\log \overline{F}_U$  belong to the subexponential concave class  $\mathcal{SC}$  with index  $\alpha$  ( $0 < \alpha < 1$ ) if the following hold: (i)  $Q_U$  is eventually concave; (ii)  $\log x = o(Q_U(x))$ ; and (iii) there exist some  $x_0 > 0$  such that  $Q_U(x)/x^\alpha$  is nonincreasing for all  $x \geq x_0$ , i.e.,

$$\frac{Q_U(x)}{Q_U(u)} \leq \left(\frac{x}{u}\right)^\alpha, \quad x \geq u \geq x_0.$$

The subexponential concave class with index  $\alpha$  is denoted by  $\mathcal{SC}_\alpha$ .

**Remark 2.5**  $\mathcal{SC}_\alpha \subset \mathcal{L}^{1/\beta}$  for all  $0 < \alpha < \beta \leq 1$  (see Lemma A.6 in Masuyama 2013). In addition, typical examples of  $Q_U \in \mathcal{SC}$  are (i)  $Q_U(x) = (\log x)^\gamma x^\alpha$  and (ii)  $Q_U(x) = (\log x)^\beta$ , where  $0 < \alpha < 1$ ,  $\beta > 1$  and  $\gamma \in \mathbb{R}$ . See Appendix A.2 in Masuyama (2013) for further remarks.

**Definition 2.5** A nonnegative random variable  $U$  and its distribution function  $F_U$  belong to the consistent variation class  $\mathcal{C}$  if  $\overline{F}_U(x) > 0$  for all  $x \geq 0$  and

$$\lim_{v \downarrow 1} \liminf_{x \rightarrow \infty} \frac{\overline{F}_U(vx)}{\overline{F}_U(x)} = 1 \quad \text{or equivalently,} \quad \lim_{v \uparrow 1} \limsup_{x \rightarrow \infty} \frac{\overline{F}_U(vx)}{\overline{F}_U(x)} = 1.$$

**Remark 2.6** It is known that (i)  $\mathcal{C} \subset \mathcal{L}^\infty$  (see Lemma A.4 in Masuyama 2013); (ii)  $\mathcal{R} \subset \mathcal{C} \subset \mathcal{L} \cap \mathcal{D} \subset \mathcal{S}$  where  $\mathcal{D}$  and  $\mathcal{R}$  denote the dominated variation class and the regular variation class, respectively (see, e.g., the introduction of Aleškevičienė et al. 2008).

### 3 Main Result

Before presenting the main result, we first show a related result.

**Proposition 3.1 (Kimura et al. 2013, Theorem 3.1.1)** *Suppose that (i) Assumption 2.1 is satisfied; and (ii) there exists some random variable  $U$  in  $\mathbb{Z}_+$  with positive finite mean such that  $U_{\text{de}} \in \mathcal{S}$  and*

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{A}}(k)\mathbf{e}}{\mathbb{P}(U > k)} = \frac{\mathbf{c}_A}{\mathbb{E}[U]}, \quad \lim_{k \rightarrow \infty} \frac{\overline{\mathbf{B}}(k)\mathbf{e}}{\mathbb{P}(U > k)} = \frac{\mathbf{c}_B}{\mathbb{E}[U]}, \quad (3.1)$$

where  $\mathbf{c}_A$  and  $\mathbf{c}_B$  are  $M \times 1$  and  $M_0 \times 1$  nonnegative vectors, respectively, satisfying  $\mathbf{c}_A \neq \mathbf{0}$  or  $\mathbf{c}_B \neq \mathbf{0}$ . We then have

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{x}}(k)}{\mathbb{P}(U_{\text{de}} > k)} = \frac{\mathbf{x}(0)\mathbf{c}_B + \overline{\mathbf{x}}(0)\mathbf{c}_A}{-\sigma} \cdot \boldsymbol{\pi}.$$

In this section, we present a more general result than the above proposition. For this purpose, we make the following assumption:

**Assumption 3.1** There exists some random variable  $Y$  in  $\mathbb{Z}_+$  such that

$$\lim_{k \rightarrow \infty} \frac{\overline{\overline{\mathbf{A}}}(k)\mathbf{e}}{\mathbb{P}(Y > k)} = \mathbf{c}_A, \quad \lim_{k \rightarrow \infty} \frac{\overline{\overline{\mathbf{B}}}(k)\mathbf{e}}{\mathbb{P}(Y > k)} = \mathbf{c}_B, \quad (3.2)$$

where  $\mathbf{c}_A$  and  $\mathbf{c}_B$  are  $M \times 1$  and  $M_0 \times 1$  nonnegative vectors, respectively, satisfying  $\mathbf{c}_A \neq \mathbf{0}$  or  $\mathbf{c}_B \neq \mathbf{0}$ .

**Remark 3.1** We suppose that (3.1) holds for some some random variable  $U$  in  $\mathbb{Z}_+$  with positive finite mean ( $U_{\text{de}} \in \mathcal{S}$  is not necessarily assumed). It then follows from (3.1) that

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{A}}(k)\mathbf{e}}{\mathbb{P}(U_{\text{de}} = k)} = \mathbf{c}_A, \quad \lim_{k \rightarrow \infty} \frac{\overline{\mathbf{B}}(k)\mathbf{e}}{\mathbb{P}(U_{\text{de}} = k)} = \mathbf{c}_B,$$

which yield

$$\lim_{k \rightarrow \infty} \frac{\overline{\overline{\mathbf{A}}}(k)\mathbf{e}}{\mathbb{P}(U_{\text{de}} > k)} = \mathbf{c}_A, \quad \lim_{k \rightarrow \infty} \frac{\overline{\overline{\mathbf{B}}}(k)\mathbf{e}}{\mathbb{P}(U_{\text{de}} > k)} = \mathbf{c}_B.$$

Thus Assumption 3.1 holds for  $Y = U_{\text{de}}$ .

The following theorem is the main result of this paper.

**Theorem 3.1** *Suppose that (i) Assumption 2.1 is satisfied; and (ii) Assumption 3.1 holds for some  $Y \in \mathcal{S}$ . We then have*

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{x}}(k)}{\mathbb{P}(Y > k)} = \frac{\mathbf{x}(0)\mathbf{c}_B + \overline{\mathbf{x}}(0)\mathbf{c}_A}{-\sigma} \cdot \boldsymbol{\pi}. \quad (3.3)$$

Before proving Theorem 3.1, we compare the above theorem with Proposition 3.1. According to Remark 3.1, condition (ii) of Proposition 3.1 is sufficient for condition (ii) of Theorem 3.1. On the other hand, the latter do not imply the former. To confirm this, we suppose that (3.2) holds for a random  $Y$  in  $\mathbb{Z}_+$  such that

$$P(Y > k) = \begin{cases} P(U_{\text{de}} > 2n), & k = 2n, n \in \mathbb{Z}_+, \\ \frac{1}{2} \{P(U_{\text{de}} > 2n) + P(U_{\text{de}} > 2n + 1)\}, & k = 2n + 1, n \in \mathbb{Z}_+, \end{cases} \quad (3.4)$$

where  $U$  is a random variable in  $\mathbb{Z}_+$  such that  $U \in \mathcal{S}$  and  $U_{\text{de}} \in \mathcal{S}$  (see Goldie and Klüppelberg 1998 and also Definition A.3 and Proposition A.2 in Masuyama 2011). It follows from  $U_{\text{de}} \in \mathcal{S}$  and (3.4) that  $P(Y > k) \stackrel{k}{\sim} P(U_{\text{de}} > k)$  and thus  $Y \in \mathcal{S}$  (Sigman 1999, Proposition 2.8), which shows that condition (ii) of Theorem 3.1 holds for  $Y \in \mathcal{S}$  defined in (3.4).

Note here that (3.2), (3.4) and  $U \in \mathcal{S} \subset \mathcal{L}$  yield

$$\begin{aligned} \overline{\mathbf{A}}(2n)\mathbf{e} &= \overline{\overline{\mathbf{A}}}(2n-1)\mathbf{e} - \overline{\overline{\mathbf{A}}}(2n)\mathbf{e} \\ &\stackrel{n}{\sim} \frac{\mathbf{c}_A}{2} \{P(U_{\text{de}} > 2n-2) + P(U_{\text{de}} > 2n-1)\} - \mathbf{c}_A P(U_{\text{de}} > 2n) \\ &= \mathbf{c}_A \left[ \frac{1}{2} \{P(U_{\text{de}} > 2n-2) - P(U_{\text{de}} > 2n-1)\} \right. \\ &\quad \left. + P(U_{\text{de}} > 2n-1) - P(U_{\text{de}} > 2n) \right] \\ &= \mathbf{c}_A \left[ \frac{1}{2} P(U_{\text{de}} = 2n-1) + P(U_{\text{de}} = 2n) \right] \\ &= \mathbf{c}_A \left( \frac{1}{2} \frac{P(U > 2n-1)}{E[U]} + \frac{P(U > 2n)}{E[U]} \right) \stackrel{n}{\sim} \frac{3\mathbf{c}_A}{2} \frac{P(U > 2n)}{E[U]}, \end{aligned} \quad (3.5)$$

and

$$\begin{aligned} \overline{\mathbf{A}}(2n+1)\mathbf{e} &= \overline{\overline{\mathbf{A}}}(2n)\mathbf{e} - \overline{\overline{\mathbf{A}}}(2n+1)\mathbf{e} \\ &\stackrel{n}{\sim} \mathbf{c}_A P(U_{\text{de}} > 2n) - \frac{\mathbf{c}_A}{2} \{P(U_{\text{de}} > 2n) + P(U_{\text{de}} > 2n+1)\} \\ &= \frac{\mathbf{c}_A}{2} \{P(U_{\text{de}} > 2n) - P(U_{\text{de}} > 2n+1)\} \\ &= \frac{\mathbf{c}_A}{2} P(U_{\text{de}} = 2n+1) = \frac{\mathbf{c}_A}{2} \frac{P(U > 2n+1)}{E[U]}. \end{aligned} \quad (3.6)$$

The equations (3.5) and (3.6) show that  $\lim_{k \rightarrow \infty} \overline{\mathbf{A}}(k)\mathbf{e}/P(U > k)$  does not exist and thus condition (ii) of Proposition 3.1 does not hold. Consequently, Theorem 3.1 is a more general result than Proposition 3.1.

In what follows, we prove Theorem 3.1. To this end, we establish three lemmas.

**Lemma 3.1** *Suppose that Assumption 2.1 is satisfied. If Assumption 3.1 holds for some  $Y \in \mathcal{L}$ ,*

then

$$\lim_{k \rightarrow \infty} \sum_{m=1}^{\infty} \frac{\overline{\mathbf{A}}(k+m) \mathbf{L}(m)}{\mathbf{P}(Y > k)} = \frac{\mathbf{c}_A \boldsymbol{\pi} (\mathbf{I} - \mathbf{R})(\mathbf{I} - \boldsymbol{\Phi}(0))}{-\sigma}, \quad (3.7)$$

$$\lim_{k \rightarrow \infty} \sum_{m=1}^{\infty} \frac{\overline{\mathbf{B}}(k+m) \mathbf{L}(m)}{\mathbf{P}(Y > k)} = \frac{\mathbf{c}_B \boldsymbol{\pi} (\mathbf{I} - \mathbf{R})(\mathbf{I} - \boldsymbol{\Phi}(0))}{-\sigma}. \quad (3.8)$$

*Proof.* See Appendix A.1.  $\square$

**Lemma 3.2** *Suppose that Assumption 2.1 is satisfied. If Assumption 3.1 holds for some  $Y \in \mathcal{L}$ , then*

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{R}}(k)}{\mathbf{P}(Y > k)} = \frac{\mathbf{c}_A \boldsymbol{\pi} (\mathbf{I} - \mathbf{R})}{-\sigma}, \quad (3.9)$$

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{R}}_0(k)}{\mathbf{P}(Y > k)} = \frac{\mathbf{c}_B \boldsymbol{\pi} (\mathbf{I} - \mathbf{R})}{-\sigma}. \quad (3.10)$$

*Proof.* From (2.5), we have

$$\overline{\mathbf{R}}(k) = \left[ \overline{\mathbf{A}}(k) + \sum_{m=1}^{\infty} \overline{\mathbf{A}}(k+m) \mathbf{L}(m) \right] (\mathbf{I} - \boldsymbol{\Phi}(0))^{-1}. \quad (3.11)$$

Further it follows from (3.2) and  $Y \in \mathcal{L}$  that

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{A}}(k)}{\mathbf{P}(Y > k)} \leq \lim_{k \rightarrow \infty} \frac{\overline{\overline{\mathbf{A}}}(k-1) \mathbf{e} \mathbf{e}^t - \overline{\overline{\mathbf{A}}}(k) \mathbf{e} \mathbf{e}^t}{\mathbf{P}(Y > k)} = \mathbf{O}.$$

Thus (3.11) yields

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{R}}(k)}{\mathbf{P}(Y > k)} = \lim_{k \rightarrow \infty} \sum_{m=1}^{\infty} \frac{\overline{\mathbf{A}}(k+m) \mathbf{L}(m)}{\mathbf{P}(Y > k)} (\mathbf{I} - \boldsymbol{\Phi}(0))^{-1}. \quad (3.12)$$

Substituting (3.7) into (3.12), we obtain (3.9). Similarly, we can prove (3.10).  $\square$

**Lemma 3.3** *Suppose that Assumption 2.1 is satisfied. If Assumption 3.1 holds for some  $Y \in \mathcal{S}$ , then*

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{F}}(k)}{\mathbf{P}(Y > k)} = \frac{(\mathbf{I} - \mathbf{R})^{-1} \mathbf{c}_A \boldsymbol{\pi}}{-\sigma}. \quad (3.13)$$

*Proof.* It follows from (2.2) that

$$\sum_{k=0}^{\infty} \mathbf{F}(k) = (\mathbf{I} - \mathbf{R})^{-1}. \quad (3.14)$$

Further combining (2.2) with Lemma 6 in Jelenković and Lazar (1998) and (3.14) yields

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{F}}(k)}{\mathbf{P}(Y > k)} = (\mathbf{I} - \mathbf{R})^{-1} \lim_{k \rightarrow \infty} \frac{\overline{\mathbf{R}}(k)}{\mathbf{P}(Y > k)} (\mathbf{I} - \mathbf{R})^{-1}.$$

From this and (3.9), we have (3.13).  $\square$

We now provide the proof of Theorem 3.1.

*Proof of Theorem 3.1.* Applying Proposition 2.4 to (2.3) and using (3.10), (3.13) and (3.14), we obtain

$$\lim_{k \rightarrow \infty} \frac{\bar{x}(k)}{\mathbb{P}(Y > k)} = \frac{\mathbf{x}(0)}{-\sigma} [\mathbf{c}_B \boldsymbol{\pi} + \mathbf{R}_0 (\mathbf{I} - \mathbf{R})^{-1} \mathbf{c}_A \boldsymbol{\pi}].$$

Substituting (2.4) into the above equation yields (3.3).  $\square$

## 4 Application to BMAP/GI/1 Queue

This section discusses the application of the main result to the standard BMAP/G/1 queue.

### 4.1 Model description

We first introduce the batch Markovian arrival process (BMAP) (Lucantoni 1991). Let  $\{J(t); t \geq 0\}$  denote a Markov chain with state space  $\mathbb{M} = \{1, 2, \dots, M\}$ , which is called background Markov chain. Let  $\{N(t); t \geq 0\}$  denote the counting process of arrivals from the BMAP. We assume that the bivariate process  $\{(N(t), J(t)); t \geq 0\}$  is a Markov chain with state space  $\mathbb{Z}_+ \times \mathbb{M}$  and the following infinitesimal generator  $\mathbf{Q}$ :

$$\mathbf{Q} = \begin{pmatrix} \mathbf{C} & \mathbf{D}(1) & \mathbf{D}(2) & \mathbf{D}(3) & \cdots \\ \mathbf{O} & \mathbf{C} & \mathbf{D}(1) & \mathbf{D}(2) & \cdots \\ \mathbf{O} & \mathbf{O} & \mathbf{C} & \mathbf{D}(1) & \cdots \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{C} & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \quad (4.1)$$

where  $\mathbf{D}(k) \geq \mathbf{O}$  ( $k \in \mathbb{N}$ ),  $[\mathbf{C}]_{i,i} < 0$  ( $i \in \mathbb{M}$ ),  $[\mathbf{C}]_{i,j} \geq 0$  ( $i \neq j, i, j \in \mathbb{M}$ ) and  $(\mathbf{C} + \sum_{k=1}^{\infty} \mathbf{D}(k)) \mathbf{e} = \mathbf{0}$ . Thus the BMAP is characterized by the rate matrices  $\{\mathbf{C}, \mathbf{D}(1), \mathbf{D}(2), \dots\}$ .

Let  $\widehat{\mathbf{D}}(z) = \sum_{k=1}^{\infty} z^k \mathbf{D}(k)$  and  $\mathbf{D} = \widehat{\mathbf{D}}(1) = \sum_{k=1}^{\infty} \mathbf{D}(k)$ . It then follows from (4.1) that

$$\mathbb{E}[z^{N(t)} \mathbb{1}(J(t) = j) \mid J(0) = i] = \left[ \exp\{(\mathbf{C} + \widehat{\mathbf{D}}(z))t\} \right]_{i,j}, \quad i, j \in \mathbb{M}, t \geq 0,$$

and that  $\mathbf{C} + \mathbf{D}$  is the infinitesimal generator of the background Markov chain  $\{J(t); t \geq 0\}$ . For analytical convenience, we assume that  $\mathbf{C} + \mathbf{D}$  is irreducible, and then define  $\boldsymbol{\varpi} := (\varpi_i)_{i \in \mathbb{M}} > \mathbf{0}$  as the unique stationary probability vector of  $\mathbf{C} + \mathbf{D}$ . In this setting, the mean arrival rate, denoted by  $\lambda$ , is given by

$$\lambda = \boldsymbol{\varpi} \sum_{k=1}^{\infty} \mathbf{D}(k) \mathbf{e}, \quad (4.2)$$

which is assumed to be strictly positive (i.e.,  $\lambda > 0$ ) in order to exclude a trivial case.

Customers are served on the first-come-first-served basis, and their service times are independent and identically distributed (i.i.d.) according to distribution function  $H$  with mean  $h \in (0, \infty)$  and  $H(0) = 0$ . We assume that the offered load  $\rho := \lambda h > 0$  satisfies

$$\rho < 1,$$

which ensures that the BMAP/GI/1 queue is stable (Loynes 1962).

Let  $\mathbf{y}(k)$  denote a  $1 \times M$  vector such that  $[\mathbf{y}(k)]_i = P(L = k, J = i)$  for  $(k, i) \in \mathbb{Z}_+ \times \mathbb{M}$ , where  $L$  and  $J$  denote generic random variables for the number of customers in the system and the state of the background Markov chain, respectively, in steady state. It is known that  $\mathbf{y} := (\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2), \dots)$  is the stationary probability vector of the following transition probability matrix of M/G/1 type (Takine 2000):

$$\mathbf{T}_{M/G/1} := \begin{pmatrix} \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \mathbf{P}(3) & \cdots \\ \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \mathbf{P}(3) & \cdots \\ \mathbf{O} & \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \cdots \\ \mathbf{O} & \mathbf{O} & \mathbf{P}(0) & \mathbf{P}(1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (4.3)$$

where  $\mathbf{P}(k)$  ( $k \in \mathbb{Z}_+$ ) denotes an  $M \times M$  matrix such that

$$\widehat{\mathbf{P}}(z) := \sum_{k=0}^{\infty} z^k \mathbf{P}(k) = \int_0^{\infty} \exp\{(\mathbf{C} + \widehat{\mathbf{D}}(z))x\} dH(x). \quad (4.4)$$

It is easy to see that  $\mathbf{T}_{M/G/1}$  is equivalent to  $\mathbf{T}$  in (2.1) with

$$\mathbf{A}(k) = \begin{cases} \mathbf{P}(k+1), & k \geq -1, \\ \mathbf{O}, & k \leq -2, \end{cases} \quad \mathbf{B}(k) = \begin{cases} \mathbf{P}(k), & k \in \mathbb{Z}_+, \\ \mathbf{P}(0), & k = -1, \\ \mathbf{O}, & k \leq -2. \end{cases} \quad (4.5)$$

Note here that (4.2), (4.4) and  $\rho = \lambda h$  yield

$$\varpi \sum_{k=1}^{\infty} k \mathbf{P}(k) \mathbf{e} = \varpi \widehat{\mathbf{P}}'(1) \mathbf{e} = \varpi \sum_{k=1}^{\infty} k \mathbf{D}(k) \mathbf{e} \cdot \int_0^{\infty} x dH(x) = \lambda h = \rho. \quad (4.6)$$

We now define  $\mathbf{P}_e(k)$  ( $k \in \mathbb{Z}_+$ ) as an  $M \times M$  matrix such that

$$\widehat{\mathbf{P}}_e(z) := \sum_{k=0}^{\infty} z^k \mathbf{P}_e(k) = \int_0^{\infty} \exp\{(\mathbf{C} + \widehat{\mathbf{D}}(z))x\} dH_e(x), \quad (4.7)$$

where  $H_e$  is the equilibrium distribution of the service time distribution  $H$ . We then have the following lemma:

**Lemma 4.1**

$$\overline{\mathbf{P}}(k) \mathbf{e} = h \cdot \mathbf{P}_e * \overline{\mathbf{D}}(k) \mathbf{e}, \quad k \in \mathbb{Z}_+. \quad (4.8)$$

*Proof.* Post-multiplying both sides of (4.7) by  $-C - \widehat{D}(z)$  and integrating the right hand side by parts yield

$$\widehat{P}_e(z)(-C - \widehat{D}(z)) = h^{-1}(\mathbf{I} - \widehat{P}(z)), \quad |z| < 1. \quad (4.9)$$

It follows from (4.9) and  $-Ce = De = \widehat{D}(1)e$  that

$$\widehat{P}_e(z) \frac{\widehat{D}(1)e - \widehat{D}(z)e}{1-z} = h^{-1} \frac{e - \widehat{P}(z)e}{1-z}, \quad |z| < 1. \quad (4.10)$$

Note here that

$$\sum_{k=0}^{\infty} z^k \overline{D}(k)e = \frac{\widehat{D}(1)e - \widehat{D}(z)e}{1-z}, \quad \sum_{k=0}^{\infty} z^k \overline{P}(k)e = \frac{e - \widehat{P}(z)e}{1-z}.$$

Substituting these equations into (4.10), we have

$$\widehat{P}_e(z) \sum_{k=0}^{\infty} z^k \overline{D}(k)e = h^{-1} \sum_{k=0}^{\infty} z^k \overline{P}(k)e,$$

and thus

$$\overline{P}(k)e = h \cdot \sum_{l=0}^k P_e(l) \overline{D}(k-l)e, \quad k \in \mathbb{Z}_+,$$

which shows that (4.8) holds.  $\square$

## 4.2 Asymptotic formulas for the queue length

In this subsection, we present some subexponential asymptotic formulas for the stationary queue length distribution of the BMAP/GI/1 queue. For this purpose, we use the following result:

**Corollary 4.1** *Suppose that there exists some random variable  $Y$  in  $\mathbb{Z}_+$  such that  $Y \in \mathcal{S}$  and*

$$\lim_{k \rightarrow \infty} \frac{\overline{\overline{P}}(k)e}{\mathbb{P}(Y > k)} = \mathbf{c} \geq \mathbf{0}, \neq \mathbf{0}. \quad (4.11)$$

*We then have*

$$\overline{\mathbf{y}}(k) \stackrel{k}{\sim} \frac{\varpi \mathbf{c}}{1-\rho} \varpi \cdot \mathbb{P}(Y > k). \quad (4.12)$$

*Proof.* Recall that  $\mathbf{T}_{M/G/1}$  in (4.3) is equivalent to  $\mathbf{T}$  in (2.1) with block matrices  $\mathbf{A}(k)$  and  $\mathbf{B}(k)$  ( $k \in \mathbb{Z}$ ) satisfying (4.5). Recall also that  $\varpi$  is the stationary probability vector of  $\mathbf{C} + \mathbf{D}$ . Thus (4.4) implies that  $\varpi$  satisfies  $\varpi \widehat{P}(1) = \varpi$  and corresponds to the stationary probability vector  $\pi$  of  $\mathbf{A} = \sum_{k \in \mathbb{Z}} \mathbf{A}(k)$ . Combining these facts with (4.5), (4.6) and (4.11), we have

$$\begin{aligned} \overline{\overline{\mathbf{A}}}(k)e &\stackrel{k}{\sim} \overline{\overline{\mathbf{B}}}(k)e \stackrel{k}{\sim} \mathbf{c} \cdot \mathbb{P}(Y > k), \\ \sigma &= \varpi \sum_{k=0}^{\infty} (k-1) P(k)e = \rho - 1. \end{aligned}$$

Therefore (4.12) follows from Theorem 3.1 and  $[\mathbf{y}(0)]_i + [\overline{\mathbf{y}}(0)]_i = \mathbb{P}(J = i) = \varpi_i$  ( $i \in \mathbb{M}$ ).  $\square$

In the following, we consider three cases: (i) the service time distribution is light-tailed; (ii) second-order long-tailed; and (iii) consistently varying.

### 4.2.1 Light-tailed service time

Let  $G$  denote a random variable in  $\mathbb{Z}_+$  such that  $\mathbb{P}(G = 0) = 0$  and

$$\mathbb{P}(G = k) = \frac{\varpi \mathbf{D}(k) \mathbf{e}}{\lambda_G}, \quad k \in \mathbb{N}, \quad (4.13)$$

where  $\lambda_G$  is the arrival rate of batches, i.e.,  $\lambda_G = \varpi \mathbf{D} \mathbf{e}$ . From the definition of  $G$ , we have  $\mathbb{E}[G] = \lambda/\lambda_G$  and thus

$$\mathbb{P}(G_{\text{de}} > k) = \frac{\varpi \overline{\mathbf{D}}(k) \mathbf{e}}{\lambda}, \quad k \in \mathbb{Z}_+. \quad (4.14)$$

We now make the following assumption:

**Assumption 4.1** There exists some  $\tilde{\mathbf{d}}_G \geq \mathbf{0}, \neq \mathbf{0}$  such that

$$\lim_{k \rightarrow \infty} \frac{\overline{\mathbf{D}}(k) \mathbf{e}}{\mathbb{P}(G_{\text{de}} > k)} = \tilde{\mathbf{d}}_G. \quad (4.15)$$

**Theorem 4.1** Suppose that  $H$  is light-tailed, i.e.,  $\int_0^\infty e^{\delta x} dH(x) < \infty$  for some  $\delta > 0$ . Further if Assumption 4.1 holds and  $G_{\text{de}} \in \mathcal{S}$ , then

$$\overline{\mathbf{P}}(k) \mathbf{e} \stackrel{k}{\sim} h \widehat{\mathbf{P}}_e(1) \tilde{\mathbf{d}}_G \cdot \mathbb{P}(G_{\text{de}} > k), \quad (4.16)$$

and

$$\mathbb{P}(L > k, J = i) \stackrel{k}{\sim} \frac{\rho}{1 - \rho} \varpi_i \cdot \mathbb{P}(G_{\text{de}} > k). \quad (4.17)$$

*Proof.* It follows from (4.7) and  $\varpi(\mathbf{C} + \mathbf{D}) = \mathbf{0}$  that

$$\varpi \widehat{\mathbf{P}}_e(1) = \varpi, \quad (4.18)$$

and from (4.14) and (4.15) that

$$\varpi \tilde{\mathbf{d}}_G = \lambda. \quad (4.19)$$

Thus if (4.16) holds, then (4.18), (4.19) and Corollary 4.1 yield

$$\overline{\mathbf{y}}(k) \stackrel{k}{\sim} \frac{\rho}{1 - \rho} \varpi \cdot \mathbb{P}(G_{\text{de}} > k),$$

which shows that (4.17) holds.

In what follows, we prove (4.16). Let  $\mathbf{A}(k)$  ( $k \in \mathbb{Z}_+$ ) denote

$$\mathbf{A}(k) = \begin{cases} \mathbf{I} + \theta^{-1} \mathbf{C}, & k = 0, \\ \theta^{-1} \mathbf{D}(k), & k \in \mathbb{N}, \end{cases} \quad (4.20)$$

where  $\theta = \max_{j \in \mathbb{M}} |[\mathbf{C}]_{j,j}|$ . We then rewrite (4.7) as

$$\sum_{k=0}^{\infty} z^k \mathbf{P}_e(k) = \int_0^\infty \sum_{n=0}^{\infty} e^{-\theta x} \frac{(\theta x)^n}{n!} dH_e(x) \left[ \sum_{k=0}^{\infty} z^k \mathbf{A}(k) \right]^n,$$

which implies that

$$\overline{P}_e(k) = \int_0^\infty \sum_{n=1}^{\infty} e^{-\theta x} \frac{(\theta x)^n}{n!} dH_e(x) \overline{\mathbf{A}}^{*n}(k), \quad k \in \mathbb{Z}_+. \quad (4.21)$$

According to Corollary 3.3 in Sigman (1999),  $G_{\text{de}} \in \mathcal{S} \subset \mathcal{L}$  implies  $\mathbb{P}(G > k) = o(\mathbb{P}(G_{\text{de}} > k))$ . It thus follows from (4.13), (4.14), (4.20) and  $\varpi > \mathbf{0}$  that for  $i \in \mathbb{M}$ ,

$$\begin{aligned} [\overline{\mathbf{A}}(k)\mathbf{e}]_i &= \frac{\lambda_G [\overline{\mathbf{D}}(k)\mathbf{e}]_i}{\theta \lambda_G} \leq \frac{\lambda_G \varpi \overline{\mathbf{D}}(k)\mathbf{e}}{\theta \varpi_i \lambda_G} \\ &= \frac{\lambda_G}{\theta \varpi_i} \mathbb{P}(G > k) = o(\mathbb{P}(G_{\text{de}} > k)). \end{aligned} \quad (4.22)$$

Using this and Proposition 2.4, we obtain

$$\overline{\mathbf{A}}^{*n}(k) = o(\mathbb{P}(G_{\text{de}} > k)), \quad n \in \mathbb{N}. \quad (4.23)$$

Note here that  $H$  is light-tailed if and only if  $H_e$  is light-tailed. Therefore similarly to the proof of Lemma 3.5 in Masuyama et al. (2009), we can readily prove from (4.21) and (4.23) that

$$\overline{P}_e(k) = o(\mathbb{P}(G_{\text{de}} > k)). \quad (4.24)$$

As a result, we obtain (4.16) by applying Proposition 2.4 to (4.8) and using (4.15) and (4.24).  $\square$

Masuyama et al. (2009) present a similar result:

**Proposition 4.1 (Masuyama et al. 2009, Theorem 3.2)** *Suppose that (i)  $H$  is light-tailed; and (ii) there exists some  $\tilde{\mathbf{D}} \geq \mathbf{O}, \neq \mathbf{O}$  such that  $\overline{\mathbf{D}}(k) \stackrel{k}{\sim} \tilde{\mathbf{D}}\mathbb{P}(G > k)$ . Further if  $G \in \mathcal{S}$  and  $G_{\text{de}} \in \mathcal{S}$ , then (4.17) holds.*

Theorem 4.1 shows that the condition  $G \in \mathcal{S}$  in Proposition 4.1 is not necessary for the subexponential asymptotic formula (4.17). In addition, condition (ii) of Proposition 4.1 implies Assumption 4.1 whereas its converse does not. This fact is confirmed similarly to the comparison of Theorem 3.1 and Proposition 3.1 in Section 3. As a result, the conditions of Proposition 4.1 are more restrictive than those of Theorem 4.1.

#### 4.2.2 Second-order long-tailed service time

**Theorem 4.2** *Suppose that (i)  $H_e \in \mathcal{L}^\mu$  for some  $\mu \geq 2$ ; and (ii)  $\sum_{k=1}^{\infty} e^{Q(k)} \mathbf{D}(k) < \infty$  for some cumulative hazard function  $Q \in \mathcal{SC}$  such that  $x^{1/\mu} = O(Q(x))$ . We then have*

$$\overline{P}_e(k) \stackrel{k}{\sim} e^{\varpi} \cdot \overline{H}_e(k/\lambda). \quad (4.25)$$

In addition, if (iii)  $H_e \in \mathcal{S}$ , then

$$\overline{\overline{P}}(k)\mathbf{e} \stackrel{k}{\sim} \rho \mathbf{e} \cdot \overline{H}_e(k/\lambda), \quad (4.26)$$

and

$$\mathbb{P}(L > k, J = i) \stackrel{k}{\sim} \frac{\rho}{1 - \rho} \varpi_i \cdot \overline{H}_e(k/\lambda). \quad (4.27)$$

**Remark 4.1** Condition (i) implies that  $\overline{H}_e(x) = \exp\{-o(x^{1/\mu})\}$  (see Proposition 2.3 (ii)). Further condition (ii) implies that  $\overline{D}(k) = o(\exp\{-\delta k^{1/\mu}\})$  for some  $\delta > 0$ . Thus  $\overline{D}(k) = o(\overline{H}_e(k))$ .

*Proof of Theorem 4.2.* Let  $T$  denote a nonnegative random variable distributed with  $H_e$  independently of BMAP  $\{\mathbf{C}, \mathbf{D}(1), \mathbf{D}(2), \dots\}$ . We can readily obtain

$$\mathbb{P}(N(T) > k \mid J(0) = i) \stackrel{k}{\sim} \mathbb{P}(T > k/\lambda), \quad i \in \mathbb{M}, \quad (4.28)$$

by following the proof of Lemma 3.1 in Masuyama et al. (2009) and using Corollary B.1 instead of Lemma 2.1 in Masuyama et al. (2009). Further similarly to the proof of Lemma 3.2 in Masuyama et al. (2009), we can prove from (4.28) that

$$\mathbb{P}(N(T) > k, J(T) = j \mid J(0) = i) \stackrel{k}{\sim} \varpi_j \mathbb{P}(T > k/\lambda), \quad i, j \in \mathbb{M},$$

which shows that (4.25) holds.

Next we prove (4.26). According to Remark 4.1,  $\overline{D}(k) = o(\exp\{-\delta k^{1/\mu}\})$  for some  $\delta > 0$ , which implies that

$$\begin{aligned} \overline{D}(k) &\leq o(\exp\{-(\delta/2)k^{1/\mu}\}) \sum_{l=k+1}^{\infty} \exp\{-(\delta/2)l^{1/\mu}\} \\ &= o(\exp\{-(\delta/2)k^{1/\mu}\}). \end{aligned}$$

Thus since  $\overline{H}_e(k/\lambda) = \exp\{-o(k^{1/\mu})\}$  (see Remark 4.1), we obtain

$$\overline{D}(k) = o(\overline{H}_e(k/\lambda)). \quad (4.29)$$

Applying Proposition 2.4 to (4.8) and using (4.25) and (4.29) yield

$$\overline{P}(k)e \stackrel{k}{\sim} h e \varpi \sum_{k=0}^{\infty} \overline{D}(k)e \cdot \overline{H}_e(k/\lambda) = \rho e \cdot \overline{H}_e(k/\lambda),$$

where the last equality is due to (4.2) and  $\rho = \lambda h$ . Therefore we have (4.26).

Finally, from (4.26) and Corollary 4.1, we have

$$\overline{y}(k) \stackrel{k}{\sim} \frac{\rho}{1 - \rho} \varpi \cdot \overline{H}_e(k/\lambda),$$

which shows that (4.27) holds. □

We now compare Theorem 4.2 with a similar result presented in Masuyama et al. (2009), which is as follows:

**Proposition 4.2 (Masuyama et al. 2009, Theorem 3.1)** *If (i)  $H \in \mathcal{L}^2$  and  $H_e \in \mathcal{S}$ ; and (ii)  $\sum_{k=1}^{\infty} e^{\phi\sqrt{k}} \mathbf{D}(k) < \infty$  for some  $\phi > 0$ , then (4.27) holds.*

Note that if  $H \in \mathcal{L}^2$ , then  $H_e \in \mathcal{L}^2$  (see Lemma A.2 in Masuyama et al. 2009). Note also that  $H_e \in \mathcal{L}^2$  if and only if  $H_e \in \mathcal{L}^\mu$  for some  $\mu \geq 2$  (see Proposition 2.3 (i)). Thus conditions (i) and (iii) of Theorem 4.2 are weaker than condition (i) of Proposition 4.2. Further if  $Q(x) = \phi\sqrt{x}$ , then condition (ii) of Theorem 4.2 is reduced to condition (ii) of Proposition 4.2. As a result, Theorem 4.2 is a more general result than Proposition 4.2.

Actually, Asmussen et al. (1999) consider an M/GI/1 queue with arrival rate  $\lambda$  and service time distribution  $H$ , and the authors prove that if  $H_e \in \mathcal{L}^2 \cap \mathcal{S}$ ,

$$\mathbb{P}(L > k) \stackrel{k}{\sim} \frac{\rho}{1 - \rho} \overline{H}_e(k/\lambda).$$

Theorem 4.2 includes this result as a special case whereas Proposition 4.2 does not.

### 4.2.3 Consistently varying service time

**Theorem 4.3** *Suppose that (i)  $H_e \in \mathcal{C}$  and  $\int_0^\infty \overline{H}_e(x) dx < \infty$  and (ii)  $\overline{\mathbf{D}}(k) = o(\overline{H}_e(k))$ . We then have (4.25). Further if (iii) there exists some finite  $\tilde{\mathbf{d}}_H \geq \mathbf{0}$  such that  $\overline{\mathbf{D}}(k) e \stackrel{k}{\sim} \overline{H}_e(k/\lambda) \tilde{\mathbf{d}}_H$ , then*

$$\overline{\mathbf{P}}(k) e \stackrel{k}{\sim} \left( \rho e + h \widehat{\mathbf{P}}_e(1) \tilde{\mathbf{d}}_H \right) \overline{H}_e(k/\lambda), \quad (4.30)$$

and

$$\mathbb{P}(L > k, J = i) \stackrel{k}{\sim} \frac{\rho + h \varpi \tilde{\mathbf{d}}_H}{1 - \rho} \varpi_i \cdot \overline{H}_e(k/\lambda). \quad (4.31)$$

*Proof.* As in the proof of Theorem 4.2, let  $T$  denote a nonnegative random variable distributed with  $H_e$  independently of BMAP  $\{\mathbf{C}, \mathbf{D}(1), \mathbf{D}(2), \dots\}$ . It is easy to see that the conditions of Proposition B.2 are satisfied. Using Proposition B.2, we can obtain (4.28) and thus (4.25) in the same way as the proof of Theorem 4.2, where we do not require condition (iii).

In addition, applying Proposition 2.4 to (4.8) and using (4.25) and condition (iii), we obtain

$$\begin{aligned} \overline{\mathbf{P}}(k) e &\stackrel{k}{\sim} h \left( e \varpi \sum_{k=0}^{\infty} \overline{\mathbf{D}}(k) e + \widehat{\mathbf{P}}_e(1) \tilde{\mathbf{d}}_H \right) \overline{H}_e(k/\lambda) \\ &= \left( \rho e + h \widehat{\mathbf{P}}_e(1) \tilde{\mathbf{d}}_H \right) \overline{H}_e(k/\lambda), \end{aligned}$$

where the last equality follows from (4.2) and  $\rho = \lambda h$ . Therefore we have (4.30). Combining (4.30), (4.18) and Corollary 4.1 yields

$$\overline{\mathbf{y}}(k) \stackrel{k}{\sim} \frac{\rho + h \varpi \tilde{\mathbf{d}}_H}{1 - \rho} \varpi \cdot \overline{H}_e(k/\lambda),$$

which leads to (4.31). □

Suppose  $\tilde{\mathbf{d}}_H = \mathbf{0}$ . It then follows that asymptotic formula (4.31) in Theorem 4.3 has the same expression as (4.27) in Theorem 4.2. The two theorems assume that  $\overline{\mathbf{D}}(k) = o(\overline{H}_e(k))$  (see Remark 4.1 and condition (ii) of Theorem 4.3) and thus that the service time distribution has a dominant impact on the tail of the stationary queue length distribution.

Conversely, the following theorem assumes, as with Theorem 4.1, that the batch size distribution has a dominant impact on the tail of the stationary queue length distribution.

**Theorem 4.4** *Suppose that conditions (i) and (ii) of Theorem 4.3 are satisfied. Further suppose that Assumption 4.1 holds for  $G_{\text{de}} \in \mathcal{S}$  such that  $\overline{H}_e(k/\lambda) = o(\mathbb{P}(G_{\text{de}} > k))$ . We then have (4.16) and thus (4.17).*

*Proof.* As shown in the proof of Theorem 4.3, the asymptotics (4.25) holds under conditions (i) and (ii) of Theorem 4.3. From (4.25) and  $\overline{H}_e(k/\lambda) = o(\mathbb{P}(G_{\text{de}} > k))$ , we have (4.24), i.e.,  $\overline{P}_e(k) = o(\mathbb{P}(G_{\text{de}} > k))$ . The rest of the proof is the same as that of Theorem 4.1.  $\square$

## 5 Application to MAP/GI<sup>(a,b)</sup>/1 Queue

In this section, we apply our main result to a single-server queue with Markovian arrivals and the  $(a, b)$ -bulk-service rule, which is denoted by MAP/GI<sup>(a,b)</sup>/1 queue (Singh et al. 2013).

### 5.1 Model description

We assume that the arrival process is a Markovian arrival process (MAP), which is a special case of the BMAP  $\{\mathbf{C}, \mathbf{D}(1), \mathbf{D}(2), \dots\}$  (introduced in Section 4) such that  $\mathbf{D}(k) = \mathbf{O}$  for all  $k \geq 2$ . For convenience, we use the symbols defined for the BMAP in Section 4, though we denote, for simplicity,  $\mathbf{D}(1)$  by  $\mathbf{D}$ . Thus the MAP is characterized by  $\{\mathbf{C}, \mathbf{D}\}$ . As with Section 4, we assume that  $\mathbf{C} + \mathbf{D}$  is irreducible and that the arrival rate  $\lambda$  is strictly positive, i.e.,

$$\lambda = \boldsymbol{\varpi} \mathbf{D} \mathbf{e} > 0, \quad (5.1)$$

where  $\boldsymbol{\varpi}$  is the unique stationary probability vector of  $\mathbf{C} + \mathbf{D}$ .

We also assume that the server works according to the  $(a, b)$ -bulk-service rule (Singh et al. 2013). To explain the  $(a, b)$ -bulk-service rule, we suppose that  $l$  customers are waiting in the queue at the completion of a service. The  $(a, b)$ -bulk-service rule is as follows:

- (i) If  $0 \leq l < a$ , the server keeps idle until the queue length is equal to the lower threshold  $a$  and then starts serving all the  $a$  customers when the queue length reaches  $a$ ; and
- (ii) If  $l \geq a$ , the server immediately starts serving  $\min(l, b)$  customers in the queue and makes the other  $l - b$  customers (if any) be in the queue.

The service times are assumed to be independent of the number of customers in service and i.i.d. according to distribution function  $H$  with mean  $h \in (0, \infty)$  and  $H(0) = 0$ . We assume that the offered load  $\rho = \lambda h$  satisfies

$$\rho < b, \quad (5.2)$$

under which the system is stable (Loynes 1962).

It should be noted that since  $\mathbf{D}(k) = \mathbf{O}$  for all  $k \geq 2$ , (4.4) and (4.7) are reduced to

$$\widehat{\mathbf{P}}(z) = \int_0^\infty \exp\{(\mathbf{C} + z\mathbf{D})x\}dH(x), \quad (5.3)$$

$$\widehat{\mathbf{P}}_e(z) = \int_0^\infty \exp\{(\mathbf{C} + z\mathbf{D})x\}dH_e(x). \quad (5.4)$$

In addition, since  $\overline{\mathbf{D}}(0) = \mathbf{D}$  and  $\overline{\mathbf{D}}(k) = \mathbf{O}$  for all  $k \in \mathbb{N}$ , it follows from Lemma 4.1 that

$$\overline{\mathbf{P}}(k)\mathbf{e} = h \cdot \mathbf{P}_e(k)\mathbf{D}\mathbf{e}, \quad k \in \mathbb{Z}_+. \quad (5.5)$$

## 5.2 Queue length process

Let  $L^{(a,b)}(t)$  ( $t \geq 0$ ) denote the total number of customers in the system at time  $t$ . Let  $J(t)$  ( $t \geq 0$ ) denote the state of the background Markov chain at time  $t$ . Let  $0 = t_0 \leq t_1 \leq t_2 \leq \dots$  denote time points at each of which a service is completed.

Let  $L_n^{(a,b)}$  and  $J_n$  ( $n \in \mathbb{Z}_+$ ) denote

$$L_n^{(a,b)} = \lim_{\varepsilon \downarrow 0} L^{(a,b)}(t_n + \varepsilon), \quad J_n = \lim_{\varepsilon \downarrow 0} J(t_n + \varepsilon).$$

Thus  $L_n^{(a,b)}$  and  $J_n$  denote the number of customers in the queue and the state of the background Markov chain, respectively, immediately after the completion of the  $n$ th service. It follows (Singh et al. 2013) that  $\{(L_n^{(a,b)}, J_n); n \in \mathbb{N}_+\}$  is a discrete-time Markov chain with state space  $\mathbb{Z}_+ \times \mathbb{M}$ , whose transition probability matrix  $\mathbf{T}_+^{(a,b)}$  is given by

$$\mathbf{T}_+^{(a,b)} = \begin{pmatrix} \mathbf{P}_0(0) & \mathbf{P}_0(1) & \mathbf{P}_0(2) & \cdots & \mathbf{P}_0(a) & \cdots & \mathbf{P}_0(b) & \cdots \\ \mathbf{P}_1(0) & \mathbf{P}_1(1) & \mathbf{P}_1(2) & \cdots & \mathbf{P}_1(a) & \cdots & \mathbf{P}_1(b) & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\ \mathbf{P}_{a-1}(0) & \mathbf{P}_{a-1}(1) & \mathbf{P}_{a-1}(2) & \cdots & \mathbf{P}_{a-1}(a) & \cdots & \mathbf{P}_{a-1}(b) & \cdots \\ \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \cdots & \mathbf{P}(a) & \cdots & \mathbf{P}(b) & \cdots \\ \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \cdots & \mathbf{P}(a) & \cdots & \mathbf{P}(b) & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\ \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \cdots & \mathbf{P}(a) & \cdots & \mathbf{P}(b) & \cdots \\ \mathbf{O} & \mathbf{P}(0) & \mathbf{P}(1) & \cdots & \mathbf{P}(a-1) & \cdots & \mathbf{P}(b-1) & \cdots \\ \mathbf{O} & \mathbf{O} & \mathbf{P}(0) & \cdots & \mathbf{P}(a-2) & \cdots & \mathbf{P}(b-2) & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \end{pmatrix}, \quad (5.6)$$

where

$$\mathbf{P}_l(k) = [(-\mathbf{C})^{-1}\mathbf{D}]^{a-l} \mathbf{P}(k), \quad l = 0, 1, \dots, a-1, k \in \mathbb{Z}_+. \quad (5.7)$$

Under the stability condition (5.2), the Markov chain  $\{(L_n^{(a,b)}, J_n); n \in \mathbb{N}_+\}$  and thus  $\mathbf{T}_+^{(a,b)}$  have the unique stationary distribution. Let  $\mathbf{y}(k)$  denote a  $1 \times M$  vector such that

$$[\mathbf{y}_+^{(a,b)}(k)]_i = \lim_{n \rightarrow \infty} \mathbf{P}(L_n^{(a,b)} = k, J_n = i), \quad (k, i) \in \mathbb{Z}_+ \times \mathbb{M}.$$

It should be noted that the stochastic process  $\{(L^{(a,b)}(t), J(t)); t \geq 0\}$  is a semi-regenerative process with the embedded Markov renewal process  $\{(L_n^{(a,b)}, J_n, t_n); n \in \mathbb{Z}_+\}$  (Çinlar 1975, Chapter 10). Note also that  $\{(L_n^{(a,b)}, J_n, t_n); n \in \mathbb{Z}_+\}$  is aperiodic because the arrival process is Markovian (Çinlar 1975, Chapter 10, Definition 2.22). Further the mean regenerative cycle (mean inter-departure time) is given by

$$\begin{aligned} \eta &:= \sum_{k=0}^{\infty} \sum_{i \in \mathbb{M}} [\mathbf{y}_+^{(a,b)}(k)]_i \cdot \mathbf{E}[t_1 \mid L_0^{(a,b)} = k, J_0 = i] \\ &= h + \sum_{k=0}^{a-1} \mathbf{y}_+^{(a,b)}(k) \cdot (-1) \lim_{s \downarrow 0} \frac{d}{ds} \left[ \int_0^{\infty} e^{-sx} \exp\{\mathbf{C}x\} dx \mathbf{D} \right]^{a-k} \mathbf{e} \\ &= h - \sum_{k=0}^{a-1} \mathbf{y}_+^{(a,b)}(k) \lim_{s \downarrow 0} \frac{d}{ds} [(s\mathbf{I} - \mathbf{C})^{-1} \mathbf{D}]^{a-k} \mathbf{e} \\ &= h + \sum_{k=0}^{a-1} \mathbf{y}_+^{(a,b)}(k) \sum_{l=0}^{a-k-1} [(-\mathbf{C})^{-1} \mathbf{D}]^l (-\mathbf{C})^{-2} \mathbf{D} \mathbf{e} \\ &= h + \sum_{k=0}^{a-1} \mathbf{y}_+^{(a,b)}(k) \sum_{l=0}^{a-k-1} [(-\mathbf{C})^{-1} \mathbf{D}]^l (-\mathbf{C})^{-1} \mathbf{e}. \end{aligned} \quad (5.8)$$

According to Theorem 6.12 in Chapter 10 of Çinlar (1975), we have for  $(k, j) \in \mathbb{Z}_+ \times \mathbb{M}$ ,

$$\begin{aligned} &[\mathbf{y}^{(a,b)}(k)]_j \\ &:= \lim_{t \rightarrow \infty} \mathbf{P}(L^{(a,b)}(t) = k, J(t) = j) \\ &= \frac{1}{\eta} \sum_{k=0}^{\infty} \sum_{i \in \mathbb{M}} [\mathbf{y}_+^{(a,b)}(l)]_i \cdot \int_0^{\infty} \mathbf{P}_{(l,i)}(L^{(a,b)}(x) = k, J(x) = j, t_1 > x) dx, \end{aligned} \quad (5.9)$$

where  $\mathbf{P}_{(l,i)}(\cdot) = \mathbf{P}(\cdot \mid L^{(a,b)}(0) = l, J(0) = i)$ .

We now define  $\mathbf{P}(t, k)$  ( $t \geq 0, k \in \mathbb{Z}_+$ ) as an  $M \times M$  matrix such that

$$[\mathbf{P}(t, k)]_{i,j} = \mathbf{P}(N(t) = k, J(t) = j \mid J(0) = i), \quad i, j \in \mathbb{M}.$$

It then follows from (5.9) that

$$\mathbf{y}^{(a,b)}(k) = \frac{1}{\eta} \sum_{l=0}^k \mathbf{y}_+^{(a,b)}(l) [(-\mathbf{C})^{-1}\mathbf{D}]^{k-l} (-\mathbf{C})^{-1}, \quad 0 \leq k \leq a-1, \quad (5.10)$$

$$\mathbf{y}^{(a,b)}(a) = \frac{1}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) [(-\mathbf{C})^{-1}\mathbf{D}]^{a-l} \int_0^\infty \mathbf{P}(x,0)\overline{H}(x)dx, \quad (5.11)$$

$$\begin{aligned} \mathbf{y}^{(a,b)}(k) &= \frac{1}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) [(-\mathbf{C})^{-1}\mathbf{D}]^{a-l} \int_0^\infty \mathbf{P}(x, k-a)\overline{H}(x)dx \\ &\quad + \frac{1}{\eta} \sum_{l=a+1}^k \mathbf{y}_+^{(a,b)}(l) \int_0^\infty \mathbf{P}(x, k-l)\overline{H}(x)dx, \quad k \geq a+1. \end{aligned} \quad (5.12)$$

Note here that  $H'_e(x) = h^{-1}\overline{H}(x)$  for  $x \geq 0$ . Note also that

$$\int_0^\infty \mathbf{P}(x, k)H'_e(x)dx = \int_0^\infty \mathbf{P}(x, k)dH_e(x) = \mathbf{P}_e(k),$$

where the last equality is due to (4.7). Thus (5.11) and (5.12) can be rewritten as

$$\mathbf{y}^{(a,b)}(a) = \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) [(-\mathbf{C})^{-1}\mathbf{D}]^{a-l} \mathbf{P}_e(0), \quad (5.13)$$

$$\begin{aligned} \mathbf{y}^{(a,b)}(k) &= \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) [(-\mathbf{C})^{-1}\mathbf{D}]^{a-l} \mathbf{P}_e(k-a) \\ &\quad + \frac{h}{\eta} \sum_{l=a+1}^k \mathbf{y}_+^{(a,b)}(l) \mathbf{P}_e(k-l), \quad k \geq a+1. \end{aligned} \quad (5.14)$$

### 5.3 Asymptotic formulas for the queue length

Let  $\mathbf{S}(0)$  denote a  $bM \times bM$  matrix such that

$$\mathbf{S}(0) = \begin{pmatrix} \mathbf{P}_0(0) & \mathbf{P}_0(1) & \mathbf{P}_0(2) & \cdots & \mathbf{P}_0(a) & \cdots & \mathbf{P}_0(b-1) \\ \mathbf{P}_1(0) & \mathbf{P}_1(1) & \mathbf{P}_1(2) & \cdots & \mathbf{P}_1(a) & \cdots & \mathbf{P}_1(b-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{P}_{a-1}(0) & \mathbf{P}_{a-1}(1) & \mathbf{P}_{a-1}(2) & \cdots & \mathbf{P}_{a-1}(a) & \cdots & \mathbf{P}_{a-1}(b-1) \\ \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \cdots & \mathbf{P}(a) & \cdots & \mathbf{P}(b-1) \\ \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \cdots & \mathbf{P}(a) & \cdots & \mathbf{P}(b-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \cdots & \mathbf{P}(a) & \cdots & \mathbf{P}(b-1) \end{pmatrix}, \quad (5.15)$$

and let  $\mathbf{S}(k)$  ( $k \in \mathbb{N}$ ) denote a  $bM \times M$  matrix such that

$$\mathbf{S}(k) = \begin{pmatrix} \mathbf{P}_0(k+b-1) \\ \mathbf{P}_1(k+b-1) \\ \vdots \\ \mathbf{P}_{a-1}(k+b-1) \\ \mathbf{P}(k+b-1) \\ \mathbf{P}(k+b-1) \\ \vdots \\ \mathbf{P}(k+b-1) \end{pmatrix}. \quad (5.16)$$

Further let  $\mathbf{S}(-k)$  ( $k = 1, 2, \dots, b$ ) denote an  $M \times bM$  matrix such that

$$\mathbf{S}(-k) = \left( \overbrace{\mathbf{O}, \mathbf{O}, \dots, \mathbf{O}}^{k-1}, \mathbf{P}(0), \mathbf{P}(1), \dots, \mathbf{P}(b-k) \right). \quad (5.17)$$

We then rewrite (5.6) as

$$\mathbf{T}_+^{(a,b)} = \left( \begin{array}{c|cccc} \mathbf{S}(0) & \mathbf{S}(1) & \mathbf{S}(2) & \mathbf{S}(3) & \dots \\ \hline \mathbf{S}(-1) & \mathbf{P}(b) & \mathbf{P}(b+1) & \mathbf{P}(b+2) & \dots \\ \mathbf{S}(-2) & \mathbf{P}(b-1) & \mathbf{P}(b) & \mathbf{P}(b+1) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathbf{S}(-b) & \mathbf{P}(1) & \mathbf{P}(2) & \mathbf{P}(3) & \dots \\ \mathbf{O} & \mathbf{P}(0) & \mathbf{P}(1) & \mathbf{P}(2) & \dots \\ \mathbf{O} & \mathbf{O} & \mathbf{P}(0) & \mathbf{P}(1) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right), \quad (5.18)$$

which is a GI/G/1-type Markov chain without disasters.

**Lemma 5.1** *Suppose that the arrival process is the MAP  $\{\mathbf{C}, \mathbf{D}\}$ , i.e., a BMAP characterized by  $\{\mathbf{C}, \mathbf{D}(1), \mathbf{D}(2), \dots\}$  such that  $\mathbf{D}(k) = \mathbf{O}$  for all  $k \geq 2$ . If  $H_e \in \mathcal{L}^2$ , then*

$$\overline{\mathbf{P}}_e(k) \stackrel{k}{\sim} e\varpi \cdot \overline{H}_e(k/\lambda), \quad (5.19)$$

$$\overline{\overline{\mathbf{P}}}(k) \stackrel{k}{\sim} \rho e \cdot \overline{H}_e(k/\lambda), \quad (5.20)$$

$$\overline{\overline{\mathbf{S}}}(k) \stackrel{k}{\sim} \rho e \cdot \overline{H}_e(k/\lambda). \quad (5.21)$$

*Proof.* Since conditions (i) and (ii) of Theorem 4.2 are satisfied, the asymptotic equation (5.19) hold. Substituting (5.19) into (5.5) and using (5.1) and  $\rho = \lambda h$  yield

$$\overline{\overline{\mathbf{P}}}(k)e \stackrel{k}{\sim} h e \varpi \mathbf{D} e \cdot \overline{H}_e(k/\lambda) = \rho e \cdot \overline{H}_e(k/\lambda),$$

which shows that (5.20) holds. Further applying (5.20) to (5.7) and using  $(-\mathbf{C})^{-1} \mathbf{D} e = e$ , we obtain for  $l = 0, 1, \dots, b-1$ ,

$$\overline{\overline{\mathbf{P}}}_l(k)e \stackrel{k}{\sim} [(-\mathbf{C})^{-1} \mathbf{D}]^{a-l} \rho e \cdot \overline{H}_e(k/\lambda) = \rho e \cdot \overline{H}_e(k/\lambda).$$

Finally, incorporating this and (5.20) into (5.16) yields (5.21).  $\square$

**Theorem 5.1** *If  $H_e \in \mathcal{L}^2 \cap \mathcal{S}$ , then*

$$\bar{\mathbf{y}}_+^{(a,b)}(k) \stackrel{k}{\sim} \frac{\rho}{b-\rho} \boldsymbol{\varpi} \cdot \bar{H}_e(k/\lambda), \quad (5.22)$$

$$\bar{\mathbf{y}}^{(a,b)}(k) \stackrel{k}{\sim} \frac{h}{\eta} \frac{b}{b-\rho} \boldsymbol{\varpi} \cdot \bar{H}_e(k/\lambda). \quad (5.23)$$

*Proof.* Note that  $\mathbf{T}_+^{(a,b)}$  in (5.18) is equivalent to  $\mathbf{T}$  in (2.1) with

$$\mathbf{A}(k) = \begin{cases} \mathbf{P}(k+b), & k \geq -b, \\ \mathbf{O}, & k \leq -b-1, \end{cases} \quad \mathbf{B}(k) = \begin{cases} \mathbf{S}(k), & k \geq -b, \\ \mathbf{O}, & k \leq -b-1. \end{cases} \quad (5.24)$$

It then follows from (4.6) and (5.2) that

$$\boldsymbol{\varpi} \sum_{k \in \mathbb{Z}} k \mathbf{A}(k) \mathbf{e} = \boldsymbol{\varpi} \sum_{k=-b}^{\infty} k \mathbf{P}(k+b) \mathbf{e} = \rho - b < 0.$$

It also follows from (5.20), (5.21) and (5.24) that

$$\bar{\bar{\mathbf{A}}}(k) \mathbf{e} \stackrel{k}{\sim} \rho \mathbf{e} \cdot \bar{H}_e(k/\lambda), \quad \bar{\bar{\mathbf{B}}}(k) \mathbf{e} \stackrel{k}{\sim} \rho \mathbf{e} \cdot \bar{H}_e(k/\lambda),$$

where the dimensions of  $\bar{\bar{\mathbf{A}}}(k) \mathbf{e}$  and  $\bar{\bar{\mathbf{B}}}(k) \mathbf{e}$  are different each other. Combining these results and Theorem 3.1 yields

$$\bar{\mathbf{y}}_+^{(a,b)}(k) \stackrel{k}{\sim} \frac{\rho \sum_{k=0}^{\infty} \mathbf{y}_+^{(a,b)}(k) \mathbf{e}}{b-\rho} \boldsymbol{\varpi} \cdot \bar{H}_e(k/\lambda) = \frac{\rho}{b-\rho} \boldsymbol{\varpi} \cdot \bar{H}_e(k/\lambda),$$

which shows that (5.22) holds.

Next we prove (5.23). From (5.14), we have for  $k \geq a$ ,

$$\begin{aligned} \bar{\mathbf{y}}^{(a,b)}(k) &= \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) [(-\mathbf{C})^{-1} \mathbf{D}]^{a-l} \bar{\mathbf{P}}_e(k-a) \\ &\quad + \frac{h}{\eta} \overline{\mathbf{y}_+^{(a,b)}} * \mathbf{P}_e(k) - \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) \bar{\mathbf{P}}_e(k-l). \end{aligned}$$

Applying (5.19), (5.22) and Proposition 2.4 to the above equation and using the long-tailed property of  $H_e$ , we obtain

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\bar{\mathbf{y}}^{(a,b)}(k)}{\bar{H}_e(k/\lambda)} &= \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) [(-\mathbf{C})^{-1} \mathbf{D}]^{a-l} \mathbf{e} \boldsymbol{\varpi} - \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) \mathbf{e} \boldsymbol{\varpi} \\ &\quad + \frac{h}{\eta} \left[ \sum_{l=0}^{\infty} \mathbf{y}_+^{(a,b)}(l) \mathbf{e} \boldsymbol{\varpi} + \frac{\rho}{b-\rho} \boldsymbol{\varpi} \sum_{l=0}^{\infty} \mathbf{P}_e(l) \right] \\ &= \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) \mathbf{e} \boldsymbol{\varpi} - \frac{h}{\eta} \sum_{l=0}^a \mathbf{y}_+^{(a,b)}(l) \mathbf{e} \boldsymbol{\varpi} + \frac{h}{\eta} \left[ \boldsymbol{\varpi} + \frac{\rho}{b-\rho} \boldsymbol{\varpi} \right] \\ &= \frac{h}{\eta} \frac{b}{b-\rho} \boldsymbol{\varpi}, \end{aligned}$$

where the second equality follows from

$$(-\mathbf{C})^{-1}\mathbf{D}\mathbf{e} = \mathbf{e}, \quad \boldsymbol{\varpi} \sum_{l=0}^{\infty} \mathbf{P}_e(l) = \boldsymbol{\varpi}, \quad \sum_{l=0}^{\infty} \mathbf{y}_+^{(a,b)}(l)\mathbf{e} = 1.$$

The proof is completed.  $\square$

**Remark 5.1** Suppose  $a = b = 1$ . It then follows that the MAP/G<sup>(a,b)</sup>/1 queue is reduced to the standard MAP/GI/1 queue, which is a special case of the BMAP/GI/1 queue. Further from (5.8) and (5.10), we have

$$1 = \frac{h}{\eta} + \frac{\mathbf{y}_+^{(1,1)}(0)(-\mathbf{C})^{-1}\mathbf{e}}{\eta} = \frac{h}{\eta} + \mathbf{y}^{(1,1)}(0)\mathbf{e} = \frac{h}{\eta} + 1 - \rho, \quad (5.25)$$

where the last equality holds because  $\mathbf{y}^{(1,1)}(0)\mathbf{e} = 1 - \rho$  (due to Little's law). The equation (5.25) yields  $h/\eta = \rho$ . Substituting this into (5.23), we have

$$\bar{\mathbf{y}}^{(1,1)}(k) \stackrel{k}{\sim} \frac{\rho}{1 - \rho} \boldsymbol{\varpi} \cdot \bar{H}_e(k/\lambda),$$

which is consistent with (4.27) in Theorem 4.2.

## A Proofs

### A.1 Proof of Lemma 3.1

We prove (3.7) only. The proof of (3.8) is omitted because it is similar to that of (3.7).

According to Proposition 2.2, we fix  $\varepsilon > 0$  arbitrarily and  $m_* := m_*(\varepsilon)$  such that for all  $m \geq m_*$  and  $l = 0, 1, \dots, \tau - 1$ ,

$$\mathbf{e}(\tau\psi - \varepsilon\mathbf{e}^t) \leq \sum_{l=0}^{\tau-1} \mathbf{L}(\lfloor m/\tau \rfloor \tau + l) \leq \mathbf{e}(\tau\psi + \varepsilon\mathbf{e}^t). \quad (\text{A.1})$$

Further since  $\mathbf{L}(m) \leq \mathbf{e}\mathbf{e}^t$  for all  $m \in \mathbb{N}$ , it follows from (3.2) and  $Y \in \mathcal{L}$  that

$$\limsup_{k \rightarrow \infty} \sum_{m=1}^{m_*-1} \frac{\bar{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)} \leq \sum_{m=1}^{m_*-1} \limsup_{k \rightarrow \infty} \frac{\bar{\bar{\mathbf{A}}}(k+m-1)\mathbf{e}\mathbf{e}^t - \bar{\bar{\mathbf{A}}}(k+m)\mathbf{e}\mathbf{e}^t}{\mathbf{P}(Y > k)} = \mathbf{O},$$

and thus

$$\lim_{k \rightarrow \infty} \sum_{m=1}^{\infty} \frac{\bar{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)} = \lim_{k \rightarrow \infty} \sum_{m=m_*}^{\infty} \frac{\bar{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)}. \quad (\text{A.2})$$

To prove (3.7) it suffices to show that for any fixed  $\varepsilon > 0$ ,

$$\limsup_{k \rightarrow \infty} \sum_{m=m_*}^{\infty} \frac{\bar{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)} \leq \mathbf{c}_A(\boldsymbol{\psi} + \varepsilon \mathbf{e}^t/\tau), \quad (\text{A.3})$$

$$\liminf_{k \rightarrow \infty} \sum_{m=m_*}^{\infty} \frac{\bar{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)} \geq \mathbf{c}_A(\boldsymbol{\psi} - \varepsilon \mathbf{e}^t/\tau). \quad (\text{A.4})$$

Indeed, letting  $\varepsilon \downarrow 0$  in (A.3) and (A.4) we obtain

$$\lim_{k \rightarrow \infty} \sum_{m=m_*}^{\infty} \frac{\bar{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)} = \mathbf{c}_A \boldsymbol{\psi} = \frac{\mathbf{c}_A \boldsymbol{\pi}(\mathbf{I} - \mathbf{R})(\mathbf{I} - \boldsymbol{\Phi}(0))}{-\boldsymbol{\sigma}},$$

where the second equality is due to (2.6). Substituting the obtained equation into (A.2), we have (3.7).

We first prove (A.3). By definition,  $\{\bar{\mathbf{A}}(k); k \in \mathbb{Z}_+\}$  is nonincreasing. We thus obtain

$$\begin{aligned} \sum_{m=m_*}^{\infty} \bar{\mathbf{A}}(k+m)\mathbf{L}(m) &\leq \sum_{n=\lfloor m_*/\tau \rfloor}^{\infty} \sum_{l=0}^{\tau-1} \bar{\mathbf{A}}(k+n\tau+l)\mathbf{L}(n\tau+l) \\ &\leq \sum_{n=\lfloor m_*/\tau \rfloor}^{\infty} \bar{\mathbf{A}}(k+n\tau) \sum_{l=0}^{\tau-1} \mathbf{L}(n\tau+l) \\ &\leq \sum_{n=\lfloor m_*/\tau \rfloor}^{\infty} \frac{1}{\tau} \sum_{i=0}^{\tau-1} \bar{\mathbf{A}}(k+n\tau-i) \cdot \sum_{l=0}^{\tau-1} \mathbf{L}(n\tau+l). \end{aligned}$$

Substituting (A.1) into the above inequality yields

$$\begin{aligned} \sum_{m=m_*}^{\infty} \frac{\bar{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)} &\leq \sum_{n=\lfloor m_*/\tau \rfloor}^{\infty} \sum_{i=0}^{\tau-1} \frac{\bar{\mathbf{A}}(k+n\tau-i)\mathbf{e}}{\mathbf{P}(Y > k)} (\boldsymbol{\psi} + \varepsilon \mathbf{e}^t/\tau). \\ &= \frac{\bar{\mathbf{A}}(k + \lfloor m_*/\tau \rfloor \tau - \tau)\mathbf{e}}{\mathbf{P}(Y > k)} (\boldsymbol{\psi} + \varepsilon \mathbf{e}^t/\tau). \end{aligned} \quad (\text{A.5})$$

From (A.5), (3.2) and  $Y \in \mathcal{L}$ , we have (A.3).

Next we prove (A.4). Since  $\{\bar{\mathbf{A}}(k)\}$  is nonincreasing, we have

$$\begin{aligned} \sum_{m=m_*}^{\infty} \bar{\mathbf{A}}(k+m)\mathbf{L}(m) &\geq \sum_{n=\lceil m_*/\tau \rceil}^{\infty} \sum_{l=0}^{\tau-1} \bar{\mathbf{A}}(k+n\tau+l)\mathbf{L}(n\tau+l) \\ &\geq \sum_{n=\lceil m_*/\tau \rceil}^{\infty} \bar{\mathbf{A}}(k+n\tau+\tau+1) \sum_{l=0}^{\tau-1} \mathbf{L}(n\tau+l) \\ &\geq \sum_{n=\lceil m_*/\tau \rceil}^{\infty} \frac{1}{\tau} \sum_{i=1}^{\tau} \bar{\mathbf{A}}(k+n\tau+\tau+i) \cdot \sum_{l=0}^{\tau-1} \mathbf{L}(n\tau+l). \end{aligned}$$

Combining this with (A.1) yields

$$\begin{aligned} \sum_{m=m_*}^{\infty} \frac{\overline{\mathbf{A}}(k+m)\mathbf{L}(m)}{\mathbf{P}(Y > k)} &\geq \sum_{n=\lceil m_*/\tau \rceil + 1}^{\infty} \sum_{i=1}^{\tau} \frac{\overline{\mathbf{A}}(k+n\tau+i)\mathbf{e}}{\mathbf{P}(Y > k)} (\boldsymbol{\psi} - \varepsilon \mathbf{e}^t/\tau). \\ &= \frac{\overline{\mathbf{A}}(k + \lceil m_*/\tau \rceil \tau + \tau)\mathbf{e}}{\mathbf{P}(Y > k)} (\boldsymbol{\psi} - \varepsilon \mathbf{e}^t/\tau). \end{aligned}$$

Therefore similarly to (A.3), we can obtain (A.4).  $\square$

## B Cumulative process sampled at heavy-tailed random times

This section summarizes some of the results presented in Masuyama (2013), which are used in Sections 4 and 5.

Let  $\{B(t); t \geq 0\}$  denote a stochastic process on  $(-\infty, \infty)$ , where  $|B(0)| < \infty$  with probability one (w.p.1). We assume that there exist regenerative points  $0 \leq \tau_0 < \tau_1 < \tau_2 < \dots$  such that  $\{B(t + \tau_n) - B(\tau_n); t \geq 0\}$  ( $n \in \mathbb{Z}_+$ ) is independent of  $\{B(u); 0 \leq u < \tau_n\}$  and is stochastically equivalent to  $\{B(t + \tau_0) - B(\tau_0); t \geq 0\}$ . The process  $\{B(t); t \geq 0\}$  is called (*regenerative*) *cumulative process*, which is introduced by Smith (1955).

Let  $\Delta\tau_0 = \tau_0$  and  $\Delta\tau_n = \tau_n - \tau_{n-1}$  for  $n \in \mathbb{N}$ . Let

$$\Delta B_n = \begin{cases} B(\tau_0), & n = 0, \\ B(\tau_n) - B(\tau_{n-1}), & n \in \mathbb{N}, \end{cases} \quad \Delta B_n^* = \begin{cases} \sup_{0 \leq t \leq \tau_0} \max(B(t), 0), & n = 0, \\ \sup_{\tau_{n-1} \leq t \leq \tau_n} B(t) - B(\tau_{n-1}), & n \in \mathbb{N}. \end{cases}$$

It is easy to see that  $\Delta B_n^* \geq \Delta B_n$  for  $n \in \mathbb{Z}_+$  and that  $\{\Delta\tau_n; n \in \mathbb{N}\}$  (resp.  $\{\Delta B_n; n \in \mathbb{N}\}$  and  $\{\Delta B_n^*; n \in \mathbb{N}\}$ ) is a sequence of i.i.d. random variables, which is independent of  $\Delta\tau_0$  (resp.  $\Delta B_0$  and  $\Delta B_0^*$ ).

**Remark B.1** The counting process  $\{N(t); t \geq 0\}$  of BMAP  $\{\mathbf{C}, \mathbf{D}(1), \mathbf{D}(2), \dots\}$  is a cumulative process such that regenerative points are hitting times to any fixed background state and the regenerative cycle follows a phase-type distribution (see equations (3.3)–(3.5) in Masuyama 2013).

We now assume that

$$\begin{aligned} \mathbf{P}(0 \leq \Delta\tau_n < \infty) &= \mathbf{P}(0 \leq \Delta B_n^* < \infty) = 1 \quad (n = 0, 1), \\ \mathbf{E}[|\Delta B_1|] < \infty, \quad 0 < \mathbf{E}[\Delta\tau_1] < \infty, \quad b := \mathbf{E}[\Delta B_1]/\mathbf{E}[\Delta\tau_1] > 0. \end{aligned}$$

We then obtain the following results.

**Proposition B.1 (Masuyama 2013, Theorem 3.3)** *Suppose that  $T$  is a nonnegative random variable independent of  $\{B(t); t \geq 0\}$ . Further suppose that (i)  $T \in \mathcal{L}^\mu$  for some  $\mu \geq 2$ ; (ii)  $\mathbf{E}[(\Delta\tau_1)^2] < \infty$  and  $\mathbf{E}[(\Delta B_1)^2] < \infty$ ; and (iii)  $\mathbf{E}[\exp\{Q(\Delta B_n^*)\}] < \infty$  ( $n = 0, 1$ ) for some cumulative hazard function  $Q \in \mathcal{SC}$  such that  $x^{1/\mu} = O(Q(x))$ . We then have  $\mathbf{P}(B(T) > bx) \stackrel{x}{\sim} \mathbf{P}(T > x)$ .*

**Corollary B.1** *Suppose that  $T$  is a nonnegative random variable independent of  $\{(N(t), J(t)); t \geq 0\}$ , where  $\{N(t)\}$  and  $\{J(t)\}$  denote the counting process and the background Markov chain, respectively, of BMAP  $\{C, D(1), D(2), \dots\}$  introduced in subsection 4.1. Suppose that (i)  $T \in \mathcal{L}^\mu$  for some  $\mu \geq 2$ ; and (ii)  $\sum_{k=1}^{\infty} \exp\{Q(k)\} D(k) < \infty$  ( $n = 0, 1$ ) for some cumulative hazard function  $Q \in \mathcal{SC}$  such that  $x^{1/\mu} = O(Q(x))$ . We then have  $P(N(T) > k) \stackrel{k}{\sim} P(T > k/\lambda)$ .*

*Proof.* It suffices to prove that conditions (i)–(iii) of Proposition B.1 are satisfied. For this purpose, fix  $B(t) = N(t)$  for  $t \geq 0$ . Since the regenerative cycle follows a phase-type distribution (see Remark B.1), we have  $E[(\Delta\tau_1)^2] < \infty$ . Further since  $\{B(t) = N(t); t \geq 0\}$  is nondecreasing, we have  $\Delta B_n^* = \Delta B_n$  for all  $n \in \mathbb{Z}_+$ . Therefore it follows from the renewal reward theorem (see, e.g., Wolff 1989, Chapter 2, Theorem 2) that

$$\frac{E[\Delta B_1^*]}{E[\Delta\tau_1]} = \lambda \in (0, \infty), \quad \frac{E[\exp\{Q(\Delta B_1^*)\}]}{E[\Delta\tau_1]} = \pi \sum_{k=1}^{\infty} \exp\{Q(k)\} D(k) e < \infty,$$

which lead to  $E[\exp\{Q(\Delta B_1^*)\}] < \infty$  and thus  $E[(\Delta B_1)^2] < \infty$ .

It remains to prove  $E[\exp\{Q(\Delta B_0^*)\}] < \infty$ . Let  $i_0$  denote the background state at regenerative points, i.e.,  $J(\tau_n) = i_0$  for all  $n \in \mathbb{Z}_+$ . Suppose that there exists some  $i \in \mathbb{M}$  such that

$$E[\exp\{Q(N(\tau_0))\}] \cdot \mathbb{1}(J(\tau_0) = i_0) \mid J(0) = i] = \infty, \quad (\text{B.1})$$

where  $\tau_0 = \inf\{t \geq 0; J(t) = i_0\}$ . Let  $T_i^{\geq \tau_0} = \inf\{t \geq \tau_0; J(t) = i\}$ . Since the background Markov chain is irreducible, we have

$$P(T_i^{\geq \tau_0} < \tau_1 \mid J(\tau_0) = i_0) > 0, \quad (\text{B.2})$$

where  $\tau_1 = \inf\{t \geq \tau_0; J(t) = i_0\}$ . It follows from  $\Delta B_1^* = N(\tau_1) - N(\tau_0)$ , (B.1) and (B.2) that

$$\begin{aligned} E[\exp\{Q(\Delta B_1^*)\}] &= E[\exp\{Q(N(\tau_1) - N(\tau_0))\}] \\ &\geq P(T_i^{\geq \tau_0} < \tau_1 \mid J(\tau_0) = i_0) \\ &\quad \times E[\exp\{Q(N(\tau_1) - N(T_i^{\geq \tau_0}))\} \mid J(T_i^{\geq \tau_0}) = i, T_i^{\geq \tau_0} < \tau_1] \\ &= P(T_i^{\geq \tau_0} < \tau_1 \mid J(\tau_0) = i_0) \\ &\quad \times E[\exp\{Q(N(\tau_0))\} \mid J(0) = i] = \infty, \end{aligned}$$

which is inconsistent with  $E[\exp\{Q(\Delta B_1^*)\}] < \infty$ . Thus (B.1) is not true. As a result, for any  $i \in \mathbb{M}$ , we have  $E[\exp\{Q(N(\tau_0))\}] \cdot \mathbb{1}(J(\tau_0) = i_0) \mid J(0) = i] = \infty$ , which implies that  $E[\exp\{Q(\Delta B_0^*)\}] < \infty$ .  $\square$

A similar result is presented in Masuyama (2013).

**Proposition B.2 (Masuyama 2013, Corollary 3.1)** *Suppose that  $T$  is a nonnegative random variable independent of  $\{(N(t), J(t)); t \geq 0\}$ , where  $\{N(t)\}$  and  $\{J(t)\}$  denote the counting*

process and the background Markov chain, respectively, of BMAP  $\{C, D(1), D(2), \dots\}$  introduced in subsection 4.1. Suppose that (i)  $T \in \mathcal{C}$ ; (ii)  $E[T] < \infty$ ; and (iii)  $\overline{D}(k) = o(P(T > k))$ . We then have  $P(N(T) > k) \stackrel{k}{\sim} P(T > k/\lambda)$ .

## Acknowledgments

Research of the author was supported in part by Grant-in-Aid for Young Scientists (B) of Japan Society for the Promotion of Science under Grant No. 24710165.

## References

- Aleškevičienė, A., Leipus, R., & Šiaulyš, J. (2008). Tail behavior of random sums under consistent variation with applications to the compound renewal risk model. *Extremes*, *11*(3), 261–279.
- Asmussen, S. (2003). *Applied Probability and Queues* (2nd ed.). New York: Springer.
- Asmussen, S., & Møller J. R. (1999). Tail asymptotics for M/G/1 type queueing processes with subexponential increments. *Queueing Systems*, *33*(1–3), 153–176.
- Asmussen, S., Klüppelberg, C., & Sigman, K. (1999). Sampling at subexponential times, with queueing applications. *Stochastic Processes and their Applications*, *79*(2), 265–286.
- Asmussen, S., Foss, S., & Korshunov, D. (2003). Asymptotics for sums of random variables with local subexponential behaviour. *Journal of Theoretical Probability*, *16*(2), 489–518.
- Çınlar, E. (1975). *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall.
- Goldie, C. M., & Klüppelberg, C. (1998). Subexponential distributions. In R. J. Adler, R. E. Feldman, & M. S. Taqqu (Eds), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (pp. 435–459). Boston: Birkhäuser.
- He, Q. M. (2014). *Fundamentals of Matrix-Analytic Methods*. New York: Springer.
- Jelenković, P. R., & Lazar, A. A. (1998). Subexponential asymptotics of a Markov-modulated random walk with queueing applications. *Journal of Applied Probability*, *35*(2), 325–347.
- Kim, B., & Kim, J. (2012). A note on the subexponential asymptotics of the stationary distribution of M/G/1 type Markov chains. *European Journal of Operational Research*, *220*(1), 132–134.
- Kimura, T., Daikoku, K., Masuyama, H., & Takahashi, Y. (2010). Light-tailed asymptotics of stationary tail probability vectors of Markov chains of M/G/1 type. *Stochastic Models*, *26*(4), 505–548.

- Kimura, T., Masuyama, H., & Takahashi, Y. (2013). Subexponential asymptotics of the stationary distributions of GI/G/1-type Markov chains. *Stochastic Models*, 29(2), 190–239.
- Li, Q. L., & Zhao, Y. Q. (2005). Heavy-tailed asymptotics of stationary probability vectors of Markov chains of GI/G/1 type. *Advances in Applied Probability*, 37(2), 482–509.
- Loynes, R. M. (1962). The stability of a queue with non-independent inter-arrival and service times. *Mathematical Proceedings of the Cambridge Philosophical Society*, 58(3), 497–520.
- Lucantoni, D. M. (1991). New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7(1), 1–46.
- Masuyama, H. (2011). Subexponential asymptotics of the stationary distributions of M/G/1-type Markov chains. *European Journal of Operational Research*, 213(3), 509–516.
- Masuyama, H. (2013). Tail asymptotics for cumulative processes sampled at heavy-tailed random times with applications to queueing models in Markovian environments. *Journal of the Operations Research Society of Japan*, 56(4), 257–308.
- Masuyama, H., Liu, B., & Takine, T. (2009). Subexponential asymptotics of the BMAP/GI/1 queue. *Journal of the Operations Research Society of Japan*, 52(4), 377–401.
- Shneer, V. V. (2006). Estimates for interval probabilities of the sums of random variables with locally subexponential distributions. *Siberian Mathematical Journal*, 47(4), 779–786.
- Sigman, K. (1999). Appendix: A primer on heavy-tailed distributions. *Queueing Systems*, 33(1–3), 261–275.
- Singh, G., Gupta, U. C., & Chaudhry, M. L. (2013). Computational analysis of bulk service queue with Markovian arrival process: MAP/R<sup>(a,b)</sup>/1 queue. *Opsearch*, 50(4), 582–603.
- Smith, W. L. (1955). Regenerative stochastic processes. *Proceedings of the Royal Society of London, Series A*, 232(1188), 6–31.
- Takine, T. (2000). A new recursion for the queue length distribution in the stationary BMAP/G/1 queue. *Stochastic Models*, 16(2), 335–341.
- Takine, T. (2004). Geometric and subexponential asymptotics of Markov chains of M/G/1 type. *Mathematics of Operations Research*, 29(3), 624–648.
- Wolff, R. W. (1989). *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, NJ: Prentice Hall.