Journal of the Korean
Statistical Society
Vol. 24, No. 1, 1995

Asymptotic Distributions of Maximum Queue Lengths for M/G/1 and GI/M/1 Systems

YouSung Park [1]

ABSTRACT

In this paper, we investigate the asymptotic distributions of maximum queue length for M/G/1 and GI/M/1 systems which are positive recurrent. It is well known that for any positive recurrent queueing systems, the distributions of their maxima linearly normalized do not have non-degenerate limits. We show, however, that by concerning an array of queueing processes limiting behaviors of these maximum queue lengths can be established under certain conditions.

KEYWORDS : Maximum queue length, M/G/1 and GI/M/1 systems, Positive recurrent processes

1. INTRODUCTION

Serfozo(1988a) and McCormick and Park(1992) found the limiting distribution of maximum queue length in positive recurrent M/M/s systems. They also showed by simulation results that their limiting distributions are very accurate. We present a generalization of McCormick and Park's results in M/M/s

[1]Department of Statistics, Korea University, Seoul, 136-701, Korea

system to allow for either a general service distribution or a general interarrival time distribution. Consider the M/G/1 case, where in Kendall's notation this refers to a queueing system with exponential interarrival time, general service distribution and one server. Only the M/M/s systems are Markovian. Although the queue length process $\{X(t), t \geq 0\}$ for the M/G/1 process is not Markovian, it possesses the regenerative behavior critical to our analysis. Let $\{Y_n, n \geq 1\}$ be an *i.i.d.* sequence and $S_n = \sum_{k=1}^{n} Y_n, n \geq 1$. We say that a process $\{\zeta_t, t \geq 0\}$ is regenerative if the cycles

$$\begin{aligned} C_1 &= \{\zeta_t, 0 \leq t < S_1\},\ C_2 = \{\zeta_{S_1+t}, 0 \leq t < S_2 - S_1\}, \\ &\cdots\ C_k = \{\zeta_{S_{k-1}+t}, 0 \leq t < S_k - S_{k-1}\} \end{aligned}$$

are independent and have the same distribution. The M/G/1 queue length process $\{X_t, t \geq 0\}$ with $X(0) \equiv 0$ is seen to be regenerative by taking

$$S_1 = \min\{t \geq 0;\ X(t^-) \neq 0 = X(t)\}$$

and

$$S_n = \min\{t > S_{n-1};\ X(t^-) \neq 0 = X(t)\},\ n \geq 2.$$

(see Asmussen, p.187).

Unlike the M/M/s case where an explicit form of the distribution of the maximum queue length in a busy cycle is available(cf. Serfozo(1988a)), in the M/G/1 and GI/M/1 cases the distributions are expressed only as integral forms. More precisely, for the M/G/1 queue length process with service time distribution $B(t)$ possessing Laplace transform $\gamma(z) = \int_0^\infty e^{-zt} dB(t)$, if $F(m) = P\{\max_{0 \leq t \leq S_1} X(t) \leq m\}$ then

$$1 - F(m) = \int_c (1-z) z^{-m} \phi(z)^{-1} dz \Big/ \int_c z^{-m} \phi(z)^{-1} dz$$

where $\phi(z) = \gamma(\lambda(1-z)) - z$, λ denotes the arrival rate, and c is any circle in the complex plane centered at zero with radius less than one. Despite the complicated form of F, Serfozo(1988b) noticed that it may be expressed in a form suitable for asymptotic analysis. With this preliminary step accomplished, he could then obtain the asymptotic behavior of maximum queue length over a large number of cycles which are themselves random variables and hence we can not observe them in advance. Our results in this paper determine approximations to the distribution of maximum queue length over a long period of time. Thus by our results one could evaluate the possibility that the queue length of a system overflows a critical value in a fixed duration(say 100 units of time), for example, the probability that the number of vehicles waiting at a tollgate is greater than a maimum value which the tollgate manages without traffic jam, while Serfozo's result applies to a period of unknown duration, e.g., 20 busy cycles.

2. EXTREME VALUES IN M/G/1

Suppose that $\{X_n(t)\}$ is a sequence of M/G/1 queueing processes, i.e., exponential interarrival time distribution with mean λ_n^{-1} and service time distribution $B_n(t)$ with mean μ_n^{-1} such that $\rho_n = \dfrac{\lambda_n}{\mu_n} < 1$ and $\rho_n \to 1$ as $n \to \infty$. These modification from the original M/G/1 queueing processses $\{X(t)\}$ to $\{X_n(t)\}$, consequently restrictions on the rate parameters λ_n and μ_n, makes us to tackle the critical obstacle that is nonconvergence of $\max_{0\leq s\leq t} X(s)$ for positive recurrent queueing systems as discussed by McCormick and Park(1992). We assume that the Laplace transform $\gamma_n(z) = \int_0^\infty e^{-zt}dB_n(t)$, for z complex, is analytic on the region $Re(z) > -\bar{x}_n$, where $-\infty < -\bar{x}_n < 0$ and that $\gamma_n(x) \to \infty$ for real

$x \downarrow -\bar{x}_n$. Serfozo(1988b) showed that $\phi_n(z)$ has exactly one zero on the real interval $(1, 1 + \bar{x}_n/\lambda_n)$.

We let r_n be this zero. Furthermore we assume that $B_n(t/\lambda_n) \to B_0(t)$ for each continuity point t of a distribution function B_0 to guarantee that $\gamma_n(\lambda_n z) \to \gamma_0(z)$ and $\phi_n(z) = \gamma_n(\lambda_n(1-z)) - z \to \phi_0(z)$ uniformly on compact sets which ensures $r_n \to 1$ and that there is an $\eta > 1$ and N such that $\phi_n(z)$ for $n \geq N$ has no zeros other than at r_n on the annulus $1 < |z| < r_n < \eta$.

Define for $j \geq 1$

$$S_j^n = \min\{t; X_n(t^-) \neq 0 = X_n(t) \text{ and } t > S_{j-1}^n\} \text{ with } S_0^n = 0$$

and $N_t^n = \max\{j;\, S_j^n \leq t\}$.

That is, $S_j^n - S_{j-1}^n$ is the jth busy cycle of M/G/1 sequence $\{X_n(t)\}$ and N_t^n is the number of busy cycles of the sequence $\{X_n(t)\}$ up to time t. Then we find from properties of S_j^n discussed in Section 1 and from Gross and Harris(1985) that

$$E(S_1^n) = E(S_j^n - S_{j-1}^n) = \frac{1}{(1-\rho_n)\lambda_n}$$

and

$$Var(S_1^n) = Var(S_j^n - S_{j-1}^n) = \frac{\beta_{2n}\lambda_n^2 + 1 - 3\rho_n + 2\rho_n^2}{(1-\rho_n)^3\lambda_n^2} \tag{2.1}$$

where β_{2n} is the second moment of $B_n(t)$.

Lemma 2.1. Let $\{X_n(t)\}$ be a sequence of M/G/1 queueing processes satisfying $\beta_{2n}\lambda_n^2 < \infty$, $n(1-\rho_n) \to \infty$ and for some constant $0 < q < \infty$, $\lambda_n(1-\rho_n) \to q$ as $n \to \infty$. Then

$$\frac{N_t^{[t]}}{t} \xrightarrow{p} q.$$

where [ ] is the integer operator.

Proof. Observe that for any given $\epsilon > 0$

$$\begin{aligned}
&P[N_t^{[t]} > (1+\epsilon)tq] \\
&= P[S^{[t]}_{[(1+\epsilon)tq]} \le t] \\
&= P[S^{[t]}_{[(1+\epsilon)tq]} - E(S^{[t]}_{[(1+\epsilon)tq]}) \le t - E(S^{[t]}_{[(1+\epsilon)tq]})] \\
&= P[S^{[t]}_{[(1+\epsilon)tq]} - E(S^{[t]}_{[(1+\epsilon)tq]}) \le t - [(1+\epsilon)tq](\lambda_{[t]}(1-\rho_{[t]}))^{-1}]
\end{aligned} \tag{2.2}$$

from (2.1).
For sufficiently large t we bound (2.2) by

$$P[S^{[t]}_{[(1+\epsilon)tq]} - E(S^{[t]}_{[(1+\epsilon)tq]}) \le -\frac{1}{2}\epsilon t] \tag{2.3}$$

since $\lambda_{[t]}(1-\rho_{[t]}) \to q$ as $t \to \infty$.
Thus (2.3) has an upper bound as

$$\frac{4(\beta_{2[t]}\lambda_{[t]}^2 + 1 - 3\rho_{[t]} + 2\rho_{[t]}^2)(1+\epsilon)q}{\epsilon^2 t(1-\rho_{[t]})^3\lambda_{[t]}^2} = o(1) \tag{2.4}$$

from the assumptions of $\beta_{2n}\lambda_{[t]}^2 = O(n)$ and $n(1-\rho_n) \to \infty$.
By similar fashion, one can show that

$$P[N_t^{[t]} < (1-\epsilon)tq] = o(1) \text{ as } t \to \infty \tag{2.5}$$

Thus combining (2.4) and (2.5) we have the lemma.
We now determine the asymptotic behavior of $M_t^{[t]} = \max_{0\leq s\leq t} X_{[t]}(s)$ where $X_n(t)$ is defined as before.

Theorem 2.2. Suppose that $\{X_n(t)\}$ is a sequence of M/G/1 queue length processes with $X(0) \equiv 0$ which satisfy the following assumptions:

(i) $B_n(t/\lambda_n) \to B_0(t)$ for each continuity point t of B_0,

(ii) $\beta_{2n}\lambda_n^2$ converges,

(iii) for some constants $0 < c, q < \infty$

$$n(1-\rho_n)(r_n - 1) \to c \text{ and } \lambda_n(1-\rho_n) \to q,$$

(iv) $\liminf_{n\to\infty} \frac{\ln(1-\rho_n)/\phi_n'(r_n)}{\ln r_n} > -\infty.$

Then

$$P[\frac{M_t^{[t]} - a_t}{b_t} \leq x] \to \exp(-\frac{cq}{e^x}),\ -\infty < x < \infty \tag{2.6}$$

where $a_t = -(\ln r_{[t]})^{-1}\ln\phi_{[t]}'(r_{[t]})$ and $b_t = (\ln r_{[t]})^{-1}$.

Proof. By τ_i we denote the moment of the ith departure from the queueing system after $t = 0$. Define $Z_n(i) = X_n(\tau_i)$, $i = 0, 1, \cdots$, then it is well known that $\{Z_n(i), i = 0, 1, \cdots\}$ is a discrete time parameter Markov chain with stationary transition probabilities. Now, take the instants of inition of idle periods as regeneration points, i.e., S_j^n defined in Section 1 and 2. Moreover $\max_{S_{j-1}^n \leq \tau_i \leq S_j^n} Z_n(i) = \max_{S_{j-1}^n \leq t \leq S_j^n} X_n(t)$. And the sequence $\{\max_{S_{j-1}^n\leq\tau_i\leq S_j^n} Z_n(i), j \geq 1\}$ are *i.i.d.* by the strong Markov properties. Thus, to show (2.6) it is enough to consider M/G/1 queueing processes through the process $\{Z_n(i), i = 0, 1, \cdots\}$. Since Lemma 2.1 holds by assumptions (ii) and (iii), it suffices to show by Theorem 2.1 of McCormick and Park(1992) that

$$n(1 - F_n(a_n + b_n x)) \to ce^{-x} \tag{2.7}$$

where $F_n(\cdot)$ is the distribution function of maximum queue length over a single busy cycle. We know from Cohen(1982) that

$$F_n([x]) = 1 - (2\pi i)^{-1} \int_{c_n^1} (1 - z) z^{-[x]} \phi_n(z)^{-1} dz / (2\pi i)^{-1} \int_{c_n^1} z^{-[x]} \phi_n(z)^{-1} dz$$

where c_n^1 is any circle in the z-plane with center at origin and radius less than 1 and the complex integral are over c_n^1 in the counter-clockwise direction. Then, as given in Serfozo(1988b)

$$n(1 - F_n(a_n + b_n x)) = (1 + \delta_n)/[\{(\xi_n - 1)/n(r_n - 1)\} + E_n].$$

where

$$\begin{aligned}
\xi_n &= r_n^{[a_n + b_n x]} \phi_n'(r_n)/(1 - \rho_n) \\
\delta_n &= \xi_n (1 - \rho_n)\{n(1 - r_n)\}^{-1} (2\pi i)^{-1} \int_{c_n} (1 - z) z^{[a_n + b_n x]} \phi_n(z)^{-1} dz \\
E_n &= \xi_n (1 - \rho_n)\{n(1 - r_n)\}^{-1} (2\pi i)^{-1} \int_{c_n} z^{[a_n + b_n x]} \phi_n(z)^{-1} dz
\end{aligned}$$

and c_n is a circle in the z-plane with center at the origin and radius greater than r_n.
The assertion in (2.7) follows since

$$\begin{aligned}
\xi_n &= r_n^{[a_n + b_n x]} \phi_n' / (1 - \rho_n) \\
&= \exp[\{-(\ln r_n)^{-1} \ln \phi_n'(r_n) + \frac{x}{\ln r_n} + O(1)\} \ln r_n \\
&\quad + \ln \phi_n'(r_n) - \ln(1 - \rho_n)] \\
&= (1 - \rho_n)^{-1} \exp(x + o(1))
\end{aligned}$$

so that

$$\frac{\xi_n - 1}{n(r_n - 1)} \to \frac{e^x}{c} \text{ by assumption (iii).}$$

One can check by assumptions (i) and (ii) that δ_n and E_n converge to zero(cf. Serfozo(1988b)). This completes the proof.

Remark 2.3. (i) The asymptotic results of (2.6) can be applied to M/G/1 queueing process $\{X(t)\}$ without subscript n when only one process among the sequence $\{X_n(t)\}$ is considered. More precisely, simply view the first n rates for the result pertaining to a sequence of queues as constant(i.e, $\lambda_k = \lambda$, $\mu_k = \mu$, $1 \le k \le n$), then estimate the value of the limit c and q by the n-th term in the sequence, that is, take

$$c = c_n = n(1-\rho)(r-1) \text{ and } q = q_n = \lambda(1-\rho).$$

Moreover, since we drop the subscript n the conditions (i), (ii) and (iv) of Theorem 2.2 always are true for moderate large but finite t.
Thus (2.6) suggests the approximation

$$P[M_t \le x] \simeq \exp(-t\lambda(1-\rho)^2(1-r)e^{-x\ln r - \ln \phi'(r)})$$

where $M_t = \max_{0\le s\le t} X(s)$, $X(t)$ being M/G/1 queueing process.
Although the approximation was motivated for ρ near 1, it should be satisfactory over all range of $0 < \rho < 1$ since the conditions (2.8) are true even for small ρ when we drop the artificial subscript n as shown in M/M/s case by McCormick and Park(1992).
(ii) Theorem 2.2 may hold when the initial state $X_n(0)$ has a distribution function π on $\{0, 1, 2, \cdots\}$.

Corollary 2.4. The conclusion of Theorem 2.2 holds under assumptions $\rho_n \to 1$, $n(1-\rho_n)(r_n - 1) \to c$ and $\lambda_n(1-\rho_n) \to q$, $0 < c, q < \infty$ when

(i) $B_n(t) = \begin{cases} 1 & \text{if } t \ge \mu_n^{-1} \\ 0 & \text{if } t < \mu_n^{-1} \end{cases}$
or
(ii) each $B_n(t)$ is a gamma distribution with scale parameter σ_n and order k.

Proof. (i) An easy check shows that

$$\phi_n(z) \to \exp(-(1-z)) - z, \ \gamma(\lambda_n z) \to e^{-z}$$

and $r_n \to 1$ as $\rho_n \to 1$.
Using $\phi'_n(z) = \rho_n(\phi_n(z) + z) - 1$ and $\phi'_n(r_n) = \rho_n r_n - 1$, we have

$$\lim_{n\to\infty} \frac{\ln[(1-\rho_n)\phi'_n(r_n)]}{\ln r_n} \to -1.$$

(ii) Clearly, $\mu_n = \frac{\sigma_n}{k}$ and thus $\mu_n^2 \int_0^\infty t^2 dB_n(t) = (k+1)/k$.
Furthermore, it can be shown that

$$\phi_n(z) \to \{1 + \frac{1}{k}(1-z)\}^{-k}, \ \gamma_n(\lambda_n z) \to \gamma_0(z) = (1 + k^{-1}z)^{-k}.$$

3. EXTREME VALUES IN GI/M/1

By the same reason as in the case of M/G/1, suppose that $\{X_n(t)\}$ is a sequence of GI/M/1 queueing processes. Let $A_n(t)$ be the interarrival time distribution with mean λ_n^{-1} and μ_n^{-1} indicating the mean of the exponential service times such that $\frac{\lambda_n}{\mu_n} = \rho_n < 1$ and $\rho_n \to 1$ as $n \to \infty$. Define a transform $\gamma_n(z,a)$ of $A_n(t)$ to be

$$\gamma_n(z,a) = \int_0^\infty e^{-(a+z)t} dA_n(t) \text{ for } Re(z) > -\bar{x}_n, \ Re(a) \geq 0.$$

where $-\infty \leq -\bar{x}_n < 0$.
Let $\psi_n(z,a) = \gamma_n(\mu_n(1-z),a) - z$ for $Re(z) < 1 + \bar{x}_n/\mu_n$ and let $r_n(a)$ be roots of $\psi_n(z,a) = 0$. Then as given by Serfozo(1988b), $\psi_n(z,0)$ has exactly two zeros on $(0, 1+\bar{x}_n/\mu_n)$ at 1 and at some $r_n(0) < 1$. Since from Cohen(1982)

$$\int_0^\infty e^{-at} dD_n(t) = \frac{1 - r_n(a)}{\mu_n^{-1}a - 1 - r_n(a)}, \ Re(a) \geq 0$$

where $D_n(t)$ is the distribution function of the busy period of GI/M/1 queue, it is easily found that

$$E(S_j^n - S_{j-1}^n) = \frac{\rho_n + 1 - r_n(0)}{\lambda_n(1 - r_n(0))},$$

$$Var(S_j^n - S_{j-1}^n) = \frac{\mu_n^{-2} - 2r_n'(0)\mu_n^{-1} + \sigma_n^2(1-r_n(0))^2}{(1-r_n(0))^2}$$

where $\rho_n = \dfrac{\lambda_n}{\mu_n}$ and σ_n^2 is the variance of $A_n(t)$.

Theorem 3.1. Suppose that $\{X_n(t)\}$ is a sequence of GI/M/1 queue length processes satisfying

(i) $A_n(t/\mu_n) \xrightarrow{d} A_0(t)$ and $\rho_n \to 1$ where A_0 is a distribution function,

(ii) $r_n'(0)$ converges and $\sigma_n^2 = o(n)$,

(iii) for some constants $0 < c, q < \infty$

$$n(1-r_n(0))(1-\rho_n) \to c \text{ and } \lambda_n(1-r_n(0)) \to q.$$

(iv) $\lim_{n\to\infty} \inf a_n$ is finite.

Then

$$P[\frac{M_t^{[t]} - a_t}{b_t} \le x] \to \exp(-\frac{cq}{e^x}), \quad -\infty < x < \infty$$

where

$$a_n = (\ln r_n)^{-1}\{\ln(1-r_n) + \ln(-(1-\rho_n)/\psi_n'(r_n,0))\},$$

$$b_n = -(\ln r_n)^{-1} \text{ and } r_n = r_n(0) \tag{3.1}$$

Proof. By the same argument of Lemma 2.1 and assumptions (ii) and (iii), one can show that

$$\frac{N_t^{[t]}}{t} \xrightarrow{p} q.$$

Note that assumption (i) ensures $r_n(0) \to 1$ as $n \to \infty$.

Let $F_n(\cdot)$ be the distribution function of the maximum queue length over a single busy cycle in GI/M/1 queues. It is known by Cohen(1982) that

$$1 - F_n([x]) = 2\pi i / \int_c z^{-[x]}\psi_n(z,0)^{-1}dz$$

where c is a circle in the z-plane with center at the origin and radius less than r_n. Now, $n[1 - F_n(a_n + b_n x)]$ can be rewritten by Serfozo(1988b) as follows:

$$n[1 - F_n(a_n + b_n x)] = \{(\zeta_n - \rho_n)/(n(1 - \rho_n)) + E_n\}^{-1} \tag{3.2}$$

where

$$\zeta_n = -r_n^{-[a_n + b_n x]}(1 - \rho_n)\psi_n'(r_n, 0),$$

$$E_n = (2\pi i n)^{-1} \int_{c_n} z^{-[a_n + b_n x]} \psi_n(z, 0)^{-1} dz,$$

and c_n is a circle in the z-plane with center out the origin and radius greater than 1.

It is straightforward to show by a_n, b_n and r_n defined in (3.1) and assumptions (i), (iii) and (iv) that for sufficiently large n,

$$\zeta_n = (1 - r_n)^{-1} e^{x + o(1)} \text{ and } E_n = o(1) \tag{3.3}$$

Hence (3.2), (3.3) and letting the embedded process $\{Y_n(i) = X_n(t_i), i = 0, 1, 2, \cdots\}$ where t_i is the moment of the i-th arrival to the system yield the conclusion.

Remark 3.2. Analogues of Remark 2.3 and Corollary 2.4 apply to Theorem 3.1.

REFERENCES

(1) Asmussen, S.(1987). *Applied Probability and Queues*, John Wiley, New York.

(2) Cohen, J.W.(1982). *The Single Server Queue*, North Holland, Amsterdam.

(3) Gross, D. and Harris, C.M.(1985). *Fundamentals of Queueing Theory*, 2nd ed., John Wiley & Sons, Inc..

(4) McCormick, W.P. and Park, Y.S.(1992). Approximating the distribution of the maximum queue length for M/M/s queues. *Queueing and Related Models* (ed. Bhat, U.N. and Basawa, I.V.), 240-261, Oxford Press.

(5) Serfozo, R.F.(1988a). Extreme values of birth and death processes and queues. *Stochastic Processes and Their Applications*, **27**, 291-306.

(6) Serfozo, R.F.(1988b). Extreme values of queue lengths in M/G/1 and GI/M/1 systems. *Mathematics of Operations Research*, **13**, 349-359.