JOURNAL OF THE
CHUNGCHEONG MATHEMATICAL SOCIETY
Volume **29**, No. 1, 2016
http://dx.doi.org/10.14403/jcms.2016.29.1.95

AN APPROXIMATION FOR THE QUEUE LENGTH DISTRIBUTION IN A MULTI-SERVER RETRIAL QUEUE

Jeongsim Kim*

Abstract. Multi-server queueing systems with retrials are widely used to model problems in a call center. We present an explicit formula for an approximation of the queue length distribution in a multi-server retrial queue, by using the Lerch transcendent. Accuracy of our approximation is shown in the numerical examples.

1. Introduction

Retrial queues are queueing systems in which arriving customers who find all servers occupied may retry for service again after a random amount of time. Retrial queues have been widely used to model many problems/situations in telephone systems, call centers, telecommunication networks, computer networks and computer systems, and in daily life. For an overview regarding retrial queues, refer to the surveys [9, 13, 14]. For further details, refer to the books [7, 10] and the bibliographies [3, 4, 5].

Typically a call center consists of a finite number of servers that answer customer's calls, and it can be modelled as a queueing system. In a queueing model of a call center, the customers are callers and the servers are either telephone operators or communication equipment. Queues are formed by callers who are waiting service. The call center can be described as follows: When a customer's call arrives, it will be served immediately if a server is available. However, if all servers are busy at

Received December 30, 2015; Accepted February 05, 2016.
2010 Mathematics Subject Classification: Primary 60K25.
Key words and phrases: $M/M/m$ retrial queue, queue length distribution, Lerch transcendent.
The research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2011-0011887).

the arrival time of the call, the caller/customer will join the orbit, and retry for service again after a random amount of time. It is known that in a call center environment, the impact of the customer retrial phenomenon on the overall performance, cannot be ignored, refer to [1, 2]. Therefore the multi-server retrial queue is useful in modelling of a call center. The $M/M/m$ retrial queue is the simplest and most widely used in call centers. For the applications of retrial queues to call centers, refer to [1, 2, 6].

Kim et al. [12] studied the exact tail asymptotic formula for the queue length (the number of customers in the orbit) distribution in the $M/M/m$ retrial queue. In this paper, based on the tail asymptotic formula of Kim et al. [12], we present an approximate formula for the queue length distribution in the $M/M/m$ retrial queue. This approximate formula is very simple, but very accurate. Accuracy of our approximation is shown in the numerical examples. The approximation for the queue length distribution in a single server retrial queue, can be found in Kim and Kim [11].

2. The model and preliminaries

We consider the $M/M/m$ retrial queue where customers arrive from outside the system according to a Poisson process with rate λ. The service facility consists of m identical servers, and service times are exponentially distributed with mean μ^{-1}. If there is a free server when a customer arrives from outside the system, this customer begins to be served immediately and leaves the system after the service is completed. On the other hand, any customer who finds all the servers busy upon arrival joins the orbit and then attempts service after a random amount of time. If there is a free server when a customer from the orbit attempts service, this customer receives service immediately and leaves the system after the service completion. Otherwise the customer comes back to the orbit immediately and repeats the retrial process. The inter-retrial time, i.e., the length of the time interval between two consecutive attempts made by a customer in the orbit, is exponentially distributed with mean ν^{-1}. The arrival process, the service times, and the inter-retrial times are assumed to be mutually independent. The offered load ρ is defined as $\rho = \frac{\lambda}{m\mu}$. It is assumed that $\rho < 1$ for stability of the system.

Let $N(t)$ denote the number of customers in the orbit at time t and $S(t)$ the number of busy servers at time t. Then $\{(N(t), S(t)) : t \geq 0\}$

is a Markov process with state space $\{0,1,2,\ldots\}\times\{0,1,\ldots,m\}$. Its infinitesimal transition rates $q_{(i,j)(k,l)}$ are given by

(a) for $0\le j\le m-1$,

$$q_{(i,j)(k,l)}=\begin{cases}\lambda & \text{if } (k,l)=(i,j+1),\\ j\mu & \text{if } (k,l)=(i,j-1),\\ i\nu & \text{if } (k,l)=(i-1,j+1),\\ -(\lambda+j\mu+i\nu) & \text{if } (k,l)=(i,j),\\ 0 & \text{otherwise.}\end{cases}$$

(b) for $j=m$,

$$q_{(i,m)(k,l)}=\begin{cases}\lambda & \text{if } (k,l)=(i+1,m),\\ m\mu & \text{if } (k,l)=(i,m-1),\\ -(\lambda+m\mu) & \text{if } (k,l)=(i,m),\\ 0 & \text{otherwise.}\end{cases}$$

Let N be the number of customers in the orbit at the steady state and S the number of busy servers at the steady state. Let

$$p_{ni}=\mathbb{P}(N=n,S=i),\ n=0,1,\ldots,i=0,1,\ldots,m.$$

Kim et al. [12] obtained the following exact tail asymptotic formula for p_{ni}: For $i=0,1,\ldots,m$,

$$p_{ni}\sim\frac{c}{i!}\Big(\frac{\nu}{\mu}\Big)^{i}n^{\frac{\lambda}{m\nu}-m+i}\,\rho^{n}\quad\text{as } n\to\infty, \tag{2.1}$$

where

$$\begin{aligned}c=&\frac{(m-1)!\,(1-\rho)^{\frac{\lambda}{m\nu}}}{\Gamma(\frac{\lambda}{m\nu})}\Big(\frac{\mu}{\nu}\Big)^{m-1}\mathbb{E}[(m-S)\rho^{m-S-1}]\\ &\times\exp\Big(\int_1^{\frac{1}{\rho}}\frac{\mathbb{E}[h(z,S)z^N]}{\mathbb{E}[(m-S)\rho^{m-S-1}z^N]}dz\Big),\end{aligned}$$

with

$$\begin{aligned}h(z,i)=&\frac{\lambda}{m\nu}\sum_{j=0}^{m-i-1}(\rho z)^j\,\frac{(j+1)(i-m\rho z)(1-z)+\rho(m-i)z}{\rho z^{m-i+1}}\\ &+\frac{\lambda}{m\nu}\Big(\rho^{m-i-1}\frac{(i-mz)(m-i)+i}{z}-\frac{i}{\rho z^{m-i+1}}\Big),\end{aligned}$$

and $\Gamma(\cdot)$ denoting the gamma function defined by $\Gamma(\alpha)=\int_0^\infty e^{-t}t^{\alpha-1}dt$. Here and subsequently, $f_n\sim g_n$ as $n\to\infty$ denotes $\lim_{n\to\infty}\frac{f_n}{g_n}=1$.

3. Approximation for the queue length distribution

In this section, we present an explicit formula for an approximation of the queue length distribution. Note that for every positive constant b, we have $n \sim n + b$ as $n \to \infty$. Thus (2.1) implies that

$$p_{ni} \sim \frac{c}{i!}\big(\frac{\nu}{\mu}\big)^i (n+b)^{\frac{\lambda}{m\nu}-m+i}\rho^n \quad \text{as } n \to \infty,$$

for $i = 0, 1, \ldots, m$ and $b > 0$. The constant c is very difficult to obtain in practice. Therefore, we will use an approximation $\tilde{p}_{ni}$ for p_{ni} as shown below: For positive real numbers $\tilde{b}$ and $\tilde{c}$,

$$\tilde{p}_{ni} = \frac{\tilde{c}}{i!}\big(\frac{\nu}{\mu}\big)^i (n+\tilde{b})^{\frac{\lambda}{m\nu}-m+i}\rho^n. \tag{3.1}$$

To use this approximation, we have to determine $\tilde{b}$ and $\tilde{c}$. We choose $\tilde{b}$ and $\tilde{c}$ such that $\tilde{p}_{ni}$ satisfies the following two conditions:

$$\sum_{n=0}^{\infty}\sum_{i=0}^{m} \tilde{p}_{ni} = 1, \tag{3.2}$$

$$\sum_{n=0}^{\infty}\sum_{i=0}^{m} i\mu\tilde{p}_{ni} = \lambda, \tag{3.3}$$

where (3.2) follows from the condition that the total probability is 1 and (3.3) follows from the fact that the departure rate is equal to the arrival rate at steady state. By substituting (3.1) into (3.2), we have

$$\frac{1}{\tilde{c}} = \sum_{n=0}^{\infty}\sum_{i=0}^{m} \frac{1}{i!}\big(\frac{\nu}{\mu}\big)^i (n+\tilde{b})^{\frac{\lambda}{m\nu}-m+i}\rho^n,$$

which can be rewritten as

$$\frac{1}{\tilde{c}} = \sum_{i=0}^{m} \frac{1}{i!}\big(\frac{\nu}{\mu}\big)^i \Phi\big(\rho, m-i-\frac{\lambda}{m\nu}, \tilde{b}\big), \tag{3.4}$$

where $\Phi(z, s, a)$ is the Lerch transcendent given by (see Section 1.11 of [8])

$$\Phi(z, s, a) = \sum_{n=0}^{\infty} \frac{z^n}{(n+a)^s}, \ |z| < 1, a \neq 0, -1, -2, \ldots.$$

By substituting (3.1) into (3.3), we have

$$\frac{\lambda}{\tilde{c}} = \sum_{i=1}^{m} \frac{\nu}{(i-1)!}\big(\frac{\nu}{\mu}\big)^{i-1} \Phi\big(\rho, m-i-\frac{\lambda}{m\nu}, \tilde{b}\big). \tag{3.5}$$

By eliminating $\tilde{c}$ from (3.4) and (3.5), we get

$$\sum_{i=0}^{m} \frac{\lambda - i\mu}{i!} \Big(\frac{\nu}{\mu}\Big)^i \Phi\Big(\rho, m-i-\frac{\lambda}{m\nu}, \tilde{b}\Big) = 0. \tag{3.6}$$

From this we can compute the value of $\tilde{b}$.

The above procedures can be summarized as follows: The approximation $\tilde{p}_{ni}$ for p_{ni} is given as

$$\tilde{p}_{ni} = \frac{\tilde{c}}{i!}\Big(\frac{\nu}{\mu}\Big)^i (n+\tilde{b})^{\frac{\lambda}{m\nu}-m+i}\rho^n,$$

where $\tilde{b}$ is calculated by numerically solving equation (3.6) and $\tilde{c}$ is given by (3.4).

4. Numerical examples

Numerical examples are presented to show the accuracy of the approximate formula (3.1). We consider the following three queueing models, all with retrial rate $\nu = 1$.

EXAMPLE 4.1. *(The M/M/30 retrial queue). We consider the M/M/30 retrial queue where the arrival rate is $\lambda = 1$ and the mean service time is $\mu^{-1} = 27$, and hence the offered load is $\rho = \frac{\lambda}{30\mu} = 0.9$.*

EXAMPLE 4.2. *(The M/M/100 retrial queue). We consider the M/M/100 retrial queue where the arrival rate is $\lambda = 1$ and the mean service time is $\mu^{-1} = 90$, and hence the offered load is $\rho = \frac{\lambda}{100\mu} = 0.9$.*

EXAMPLE 4.3. *(The M/M/200 retrial queue). We consider the M/M/200 retrial queue where the arrival rate is $\lambda = 1$ and the mean service time is $\mu^{-1} = 180$, and hence the offered load is $\rho = \frac{\lambda}{200\mu} = 0.9$.*

	$\tilde{b}$	$\tilde{c}$
Example 4.1	9.824789864115019e-001	1.176790359366119e-012
Example 4.2	9.930371010595211e-001	6.699231550073337e-040
Example 4.3	9.971693179242537e-001	6.052717076200984e-079

TABLE 1. The values of $\tilde{b}$ and $\tilde{c}$.

In Figs. 1-3, we plot the exact and approximate values of p_{ni} for Examples 4.1-4.3. The approximate values are obtained by using the

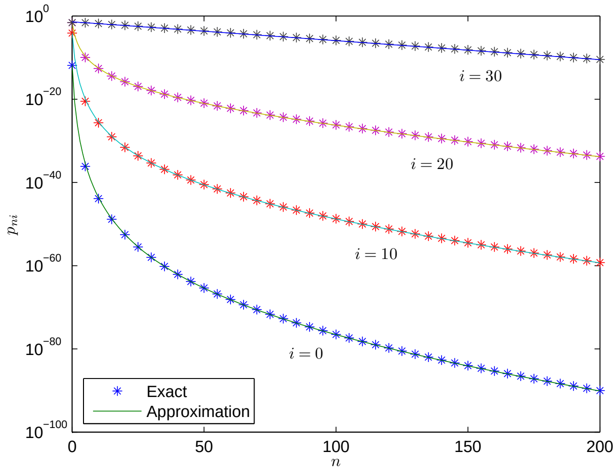

FIGURE 1. Exact and approximate values of p_{ni} for Example 4.1.

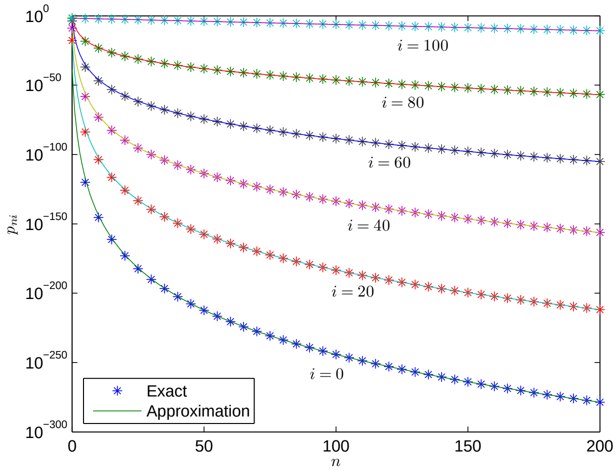

FIGURE 2. Exact and approximate values of p_{ni} for Example 4.2.

formula (3.1), along with the values of $\tilde{b}$ and $\tilde{c}$ given in Table 1. The exact values are obtained as follows: It is known that $\lim_{K\to\infty} p_{ni}^{(K)} = p_{ni}$, where $p_{ni}^{(K)}$ is the probability that there are n customers in the orbit and

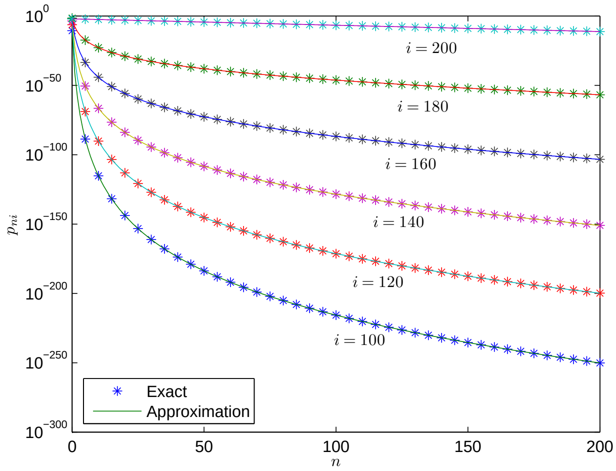

FIGURE 3. Exact and approximate values of p_{ni} for Example 4.3.

the number of busy servers is i at steady state in the M/M/m retrial queue with finite orbit capacity K. The probability p_{ni} is obtained as $p_{ni}^{(K)}$ such that $p_{ni}^{(K)}$ does not vary numerically as K increases. We note that the calculation time of the exact values rapidly increases as the number of servers becomes larger. Figs. 1-3 show that our approximation is very accurate.

References

[1] M. S. Aguir, O. Z. Akşin, F. Karaesmen, and Y. Dallery, *On the interaction between retrials and sizing of call centers*, European Journal of Operational Research **191** (2008), 398-408.

[2] M. S. Aguir, F. Karaesmen, O. Z. Akşin, and F. Chauvet, *The impact of retrials on call center performance*, OR Spectrum **26** (2004), 353-376.

[3] J. R. Artalejo, *A classified bibliography of research on retrial queues: Progress in 1990-1999*, Top **7** (1999), 187-211.

[4] J. R. Artalejo, *Accessible bibliography on retrial queues*, Math. Comput. Model. **30** (1999), 1-6.

[5] J. R. Artalejo, *Accessible bibliography on retrial queues: Progress in 2000-2009*, Math. Comput. Model. **51** (2010), 1071-1081.

[6] J. R. Artalejo, A. Economou, and A. Gómez-Corral, *Applications of maximum queue lengths to call center management*, Computers & Operations Research **34** (2007), 983-996.

[7] J. R. Artalejo and A. Gómez-Corral, *Retrial Queueing Systems*, Springer, 2008.

[8] A. Erdelyi, *Higher Transcendental Functions*, Vol. 1, McGraw-Hill, New York, 1953.
[9] G. I. Falin, *A survey of retrial queues*, Queueing Systems **7** (1990), 127-168.
[10] G. I. Falin and J. G. C. Templeton, *Retrial Queues*, Chapman & Hall, London, 1997.
[11] J. Kim and J. Kim, *An approximation for the distribution of the number of retrying customers in an M/G/1 retrial queue*, Journal of the Chungcheong Mathematical Society **27** (2014), 405-411.
[12] J. Kim, J. Kim, and B. Kim, *Tail asymptotics of the queue size distribution in the M/M/m retrial queue*, Journal of Computational and Applied Mathematics **236** (2012), 3445-3460.
[13] V. G. Kulkarni and H. M. Liang, *Retrial queues revisited, In: Frontiers in Queueing: Models and Applications in Science and Engineering (J.H. Dshalalow, ed.)*, CRC Press, Boca Raton, 1997, 19-34.
[14] T. Yang and J. G. C. Templeton, *A survey on retrial queues*, Queueing Systems **2** (1987), 201-233.

*
Department of Mathematics Education
Chungbuk National University
Cheongju 28644, Republic of Korea
E-mail: jeongsimkim@chungbuk.ac.kr