RAIRO Operations Research
RAIRO Oper. Res. **38** (2004) 227–241
DOI: 10.1051/ro:2004021

MAXIMUM LIKELIHOOD ESTIMATES AND CONFIDENCE INTERVALS OF AN M/M/R/N QUEUE WITH BALKING AND HETEROGENEOUS SERVERS

Kuo-Hsiung Wang[1], Sheau-Chyi Chen[1] and Jau-Chuan Ke[2]

Communicated by Shunji Osaki

Abstract. This paper considers an M/M/R/N queue with heterogeneous servers in which customers balk (do not enter) with a constant probability $(1-b)$. We develop the maximum likelihood estimates of the parameters for the M/M/R/N queue with balking and heterogeneous servers. This is a generalization of the M/M/2 queue with heterogeneous servers (without balking), and the M/M/2/N queue with balking and heterogeneous servers in the literature. We also develop the confidence interval formula for the parameter ρ, the probability of empty system P_0, and the expected number of customers in the system $E[N]$, of an M/M/R/N queue with balking and heterogeneous servers. The effects of varying b, N, and R on the confidence intervals of P_0 and $E[N]$ are also investigated.

Keywords. Balk, confidence interval, heterogeneous servers, maximum likelihood estimate, queue.

Received October 24, 2003. Revised January 15, 2004.

[1] Department of Applied Mathematics National Chung-Hsing University Taichung, 402, Taiwan, R.O.C.; e-mail: khwang@amath.nchu.edu.tw

[2] Department of Statistics National Taichung Institute of Technology Taichung 404, Taiwan, R.O.C.

1. INTRODUCTION

Several researchers have investigated point estimation, hypothesis testing, and confidence interval in queueing problems. This paper deals with both point estimation and confidence interval of an M/M/R/N queue with balking and heterogeneous servers. We consider the behavior of impatient customers which upon arrival may or may not enter the queue for service depending on the number of customers in the system.

It is assumed that customers arrive according to a Poisson process with rate λ. A customer which on arrival finds n customers in the system, either enters the queue with probability b_n or balks with probability $1-b_n$. The balking probabilities are defined as: $1-b_n = 0$, when $n < R$, and $1-b_n = 1-b$, when $n \geq R$, where b denotes the probability of a customer entering the queue. The service times follow exponential distribution with R unequal rates μ_i, $(i = 1, 2, \ldots R)$, where $\mu_1 > \mu_2 > \cdots > \mu_R$. We assume that arriving customers at the servers form a single waiting line and are served in the order of their arrivals. It is further assumed that each server may serve only one customer at a time, and that the service is independent of the arrival of the customers. If all servers are busy, then the first customer must wait until any one server is available. If all servers are idle, the first customer in the waiting line goes to the fastest server. On the other hand, if part of the servers are idle, the first customer goes to the faster server.

The statistical inference in queueing problems are rarely found in the literature and the work of related problems in the past mainly concentrates on only one server or two servers. The pioneering paper in parameter estimation problem was first proposed by Clarke [4], who developed maximum likelihood estimates for the arrival and service parameters of an M/M/1 queue. Lilliefors [10] examined the confidence intervals for the M/M/1, M/E$_k$/1 and M/M/2 queues. For a G/G/1 queue, Basawa and Prabhu [3] studied moment estimates as well as maximum likelihood estimates. Maximum likelihood estimates and confidence intervals in an M/M/2 queue with heterogeneous servers were derived by Dave and Shah [5] and Jain and Templeton [9], respectively. The confidence interval estimation of a single server queue with random arrivals and balking is investigated by Rubin and Robson [12]. The hypothesis testing and simultaneous confidence intervals of the estimators for an M/E$_k$/1 queue are analyzed by Jain [8]. Abou-El-Ata and Hariri [1] developed point estimation and confidence intervals of the truncated M/M/2/N queue with balking and heterogeneous servers. Basawa *et al.* [2] used the waiting time data to deal with the maximum likelihood estimation for a single server queue. An overview of literature on the statistical analysis of several queueing systems was provided by Dshalalow [6]. Rodrigues and Leite [11] applied Bayesian analysis to study the confidence intervals of an M/M/1 queue. Recently, Huang and Brill [7] derived the minimum variance unbiased estimator and the maximum likelihood estimator of a collection of n independent M/G/c/c queues.

The main purpose of this paper is fourfold. Firstly we develop the maximum likelihood estimates and confidence intervals of the M/M/R/N queue with balking and heterogeneous servers. Secondly we show that it generalizes either the M/M/2 queue with heterogeneous servers which is studied by Dave and Shah [5] and Jain and Templeton [9], or the M/M/2/N queue with balking and heterogeneous servers which is studied by Abou-El-Ata and Hariri [1]. Thirdly we provide the confidence interval formula for the parameter ρ, the probability of empty system P_0, and the expected number of customers in the system $E[N]$, of an M/M/R/N queue with balking and heterogeneous servers. Finally we study the effects of varying b, N, and R on the confidence intervals of P_0 and $E[N]$.

2. M/M/R/N queue with balking and heterogeneous servers

For the M/M/R/N queue with balking and heterogeneous servers, we develop the maximum likelihood estimates of the arrival rate λ, and the R unequal service rates μ_i, $(i = 1, 2, \dots R)$, where $\mu_1 > \mu_2 > \cdots > \mu_R$.

The mean arrival rate λ_n and the mean service rate θ_n for the M/M/R/N queue with balking and heterogeneous servers are as follows:

$$\lambda_n = \begin{cases} \lambda & n = 0, 1, 2, \dots, R-1, \\ \lambda b & n = R, R+1, R+2, \dots, N, \\ 0 & \text{otherwise.} \end{cases}$$

$$\theta_n = \begin{cases} \sum_{i=1}^{n} \mu_i & n = 1, 2, 3, \dots, R-1, \\ \mu & n = R, R+1, R+2, \dots, N, \\ 0 & \text{otherwise.} \end{cases}$$

In steady-state, the following notations are used.

$P_0 \equiv$ probability that there are no customers in the system,

$P_n \equiv$ probability that there are n customers in the system,

where $n = 1, 2, 3, \dots, N$.

The steady-state equations for P_n are given by:

$$\lambda P_0 = \mu_1 P_1, \tag{1}$$

$$\left(\lambda + \sum_{i=1}^{n} \mu_i\right) P_n = \lambda P_{n-1} + \left(\sum_{i=1}^{n+1} \mu_i\right) P_{n+1}, \quad 1 \le n \le R-1 \tag{2}$$

$$(\lambda b + \mu) P_R = \lambda P_{R-1} + \mu P_{R+1}, \tag{3}$$

$$(\lambda b + \mu) P_n = \lambda b P_{n-1} + \mu P_{n+1}, \quad R+1 \le n \le N-1, \tag{4}$$

$$\mu P_N = \lambda b P_{N-1}, \tag{5}$$

where $\mu = \sum_{i=1}^{R} \mu_i$.

Using the general birth and death results given by

$$P_n = \prod_{j=1}^{n} \frac{\lambda_{j-1}}{\theta_j} P_0, \tag{6}$$

or solving (1)–(5) recursively, we obtain

$$P_n = \frac{\lambda^n}{\prod_{j=1}^{n} \sum_{i=1}^{j} \mu_i} P_0, \quad n = 1, 2, \ldots, R, \tag{7}$$

$$P_n = \frac{\lambda^n}{\prod_{j=1}^{R} \sum_{i=1}^{j} \mu_i} \left(\frac{b}{\mu}\right)^{n-R} P_0, \quad n = R+1, R+2, \ldots, N, \tag{8}$$

and

$$P_0 = \left[1 + \sum_{n=1}^{R} \frac{\lambda^n}{\prod_{j=1}^{n} \sum_{i=1}^{j} \mu_i} + \sum_{n=R+1}^{N} \frac{\lambda^n}{\prod_{j=1}^{R} \sum_{i=1}^{j} \mu_i} \left(\frac{b}{\mu}\right)^{n-R}\right]^{-1}. \tag{9}$$

Let $E[N]$ denote the expected number of customers in the system. From (7)–(9), we finally obtain

$$E[N] = \sum_{n=1}^{R} nP_n + \sum_{n=R+1}^{N} nP_n = \left[\sum_{n=1}^{R} \frac{n\lambda^n}{\prod_{j=1}^{n} \sum_{i=1}^{j} \mu_i} + \sum_{n=R+1}^{N} \frac{n\lambda^n}{\prod_{j=1}^{R} \sum_{i=1}^{j} \mu_i} \left(\frac{b}{\mu}\right)^{n-R}\right] P_0. \tag{10}$$

2.1. Likelihood function

Let m_0 be the initial number of customers in the queue at time $t = 0$. The queue is being observed for a fixed amount of time T, where T is sufficiently large to obtain enough number of observations. During T, we assume that there are N_a number of arrivals to the queue and N_d number of departures from the queue. During T, we observe the following:

$T_e \equiv$ amount of time during which all servers are idle;
$T_{B_1} \equiv$ amount of time during which only the fastest server is busy;
$T_{B_n} \equiv$ amount of time during which n faster servers are busy where $n = 2, 3, \ldots, R-1$;
$T_{B_R} \equiv$ amount of time during which all servers are busy;
$N_e \equiv$ number of arrivals to an empty queue when all servers are idle (transitions E_0 to E_1);

$N_{B_1} \equiv$ number of arrivals to a partially busy queue when only the fastest server is busy (transitions E_1 to E_2);

$N_{B_n} \equiv$ number of arrivals to a partially busy queue when n servers are busy (transitions E_n to E_{n+1}) where $n = 2, 3, \ldots, R-1$;

$N_{B_R} \equiv$ number of arrivals to a completely busy queue when all servers are busy (transitions E_n to E_{n+1}) where $n = R, R+1, \ldots, N-1$;

$N_{D_1} \equiv$ number of departures from a partially busy queue when only the fastest server is busy (transitions E_1 to E_0);

$N_{D_n} \equiv$ number of departures from a partially busy queue when n faster servers are busy (transitions E_n to E_{n-1}) where $n = 2, 3, \ldots, R$;

$N_{D_R} \equiv$ number of departures from a completely busy queue when all servers are busy (transitions E_n to E_{n-1}) where $n = R+1, R+2, \ldots, N$.

It is clear that

$$T = T_e + \sum_{n=1}^{R} T_{B_n},$$

$$N_a = N_e + \sum_{n=1}^{R} N_{B_n},$$

$$N_d = \sum_{n=1}^{R} N_{D_n}.$$

We observe that the time axis can be decomposed into three random sequences of intervals: the idle intervals consisting of all times when the queue size is empty, the busy intervals consisting of all times when the queue size is one or more and less than R, and the busy intervals consisting of all times when the queue size is R or more. As analyzed in Clarke's paper [4], the corresponding likelihood function in this queueing system can be obtained using the following three basic components:

(i) the probability density function of N_e transitions (E_0 to E_1) occurring during time T_e is given by $\lambda^{N_e} \mathrm{e}^{-\lambda T_e}$;

(ii) the probability density function of N_{B_n} transitions (E_n to E_{n+1}, $1 \leq n \leq R-1$) occurring and N_{D_n} transitions (E_n to E_{n-1}, $1 \leq n \leq R-1$) occurring during time T_{B_n} is given by $\left(\lambda^{N_{B_n}} \mathrm{e}^{-\lambda T_{B_n}}\right)\left[\left(\sum_{i=1}^{n} \mu_i\right)^{N_{D_n}} \mathrm{e}^{-\left(\sum_{i=1}^{n} \mu_i\right) T_{B_n}}\right]$;

(iii) the probability density function of N_{B_R} transitions (E_n to E_{n+1}, $R \leq n \leq N-1$) occurring and N_{D_R} transitions (E_n to E_{n-1}, $R \leq n \leq N-1$) occurring during time T_{B_R} is given by $\left[\left(\lambda b\right)^{N_{B_R}} \mathrm{e}^{-(\lambda b) T_{B_R}}\right]\left(\mu^{N_{D_R}} \mathrm{e}^{-\mu T_{B_R}}\right)$.

Following Dave and Shah [5], the log-likelihood function is given by

$$\begin{aligned}\ln L = \ln L(\lambda, b, \mu_1, \mu_2, \ldots, \mu_R) &= \ln P_{m_0} + N_a \ln\lambda + N_{B_R} \ln b - \lambda T - \lambda(b-1)T_{B_R} \\ &+ \sum_{n=1}^{R-1} N_{D_n} \ln\left(\sum_{i=1}^{n} \mu_i\right) + N_{D_R} \ln\mu - \sum_{n=1}^{R-1}\sum_{i=1}^{n} \mu_i T_{B_n} - \mu T_{B_R}, \quad (11)\end{aligned}$$

where P_{m_0} denotes the steady-state probability that there are m_0 customers in the system at time $t = 0$.

2.2. Maximum likelihood estimates

Since the queue is in steady-state, the initial queue length m_0 may be omitted. After some algebraic manipulations in (11), we obtain the maximum likelihood estimates of λ, and μ_n $(n = 1, 2, \ldots, R)$ as follows:

$$\hat{\lambda} = \frac{N_a}{T + (b-1)T_{B_R}}, \quad (12)$$

$$\hat{\mu}_1 = \frac{N_{D_1}}{T_{B_1}}, \quad (13)$$

$$\hat{\mu}_n = \frac{N_{D_n}}{T_{B_n}} - \frac{N_{D_{n-1}}}{T_{B_{n-1}}}, \quad 2 \le n \le R. \quad (14)$$

From (13)–(14), we get

$$\hat{\mu} = \sum_{n=1}^{R} \hat{\mu}_n = \frac{N_{D_R}}{T_{B_R}}\cdot \quad (15)$$

Thus, we get the maximum likelihood estimate of the traffic intensity ρ

$$\hat{\rho} = \frac{\hat{\lambda}}{\hat{\mu}} = \frac{N_a T_{B_R}}{N_{D_R}\left[T + (b-1)T_{B_R}\right]}\cdot \quad (16)$$

2.3. Special cases

Case 1. The maximum likelihood estimates of the M/M/2/N queue with balking and heterogeneous servers are obtained by setting $R = 2$.

Case 2. The maximum likelihood estimates of the M/M/2/N queue with heterogeneous servers are obtained by setting $R = 2$ and $b = 1$.

Case 3. The maximum likelihood estimates of the M/M/2 queue with heterogeneous servers are obtained by letting $N \to \infty$, and setting $R = 2$ and $b = 1$.

3. Confidence interval formula for ρ, P_0 and $E[N]$

We will develop the confidence interval formula for ρ, P_0 and $E[N]$ of an M/M/R/N queue with balking and heterogeneous servers, where the traffic intensity $\rho = \lambda/\mu = \lambda/\sum_{i=1}^{R} \mu_i$. We also study the effects of changing b, N, and R on the confidence intervals of P_0 and $E[N]$. We first demonstrate the following results.

For a simple birth-death process to an M/M/R/N queueing system with balking and heterogeneous servers, we obtain from Appendix that

$$E[N] = -\rho \frac{\partial \ln P_0}{\partial \rho} \geq 0, \tag{17}$$

and the variance

$$\mathrm{Var}[N] = \rho \frac{\partial E[N]}{\partial \rho} \geq 0. \tag{18}$$

Next, following the results of Lilliefors [10], we get the $(1-\alpha) \times 100\%$ lower and upper confidence limits ρ_L and ρ_U of ρ as follows

$$\rho_L = \hat{\rho} F_{1-\alpha/2}(2N_a, 2N_d), \tag{19}$$

and

$$\rho_U = \hat{\rho} F_{\alpha/2}(2N_a, 2N_d), \tag{20}$$

where $\hat{\rho}$ is given by (16).

3.1. Confidence interval for P_0

One can easily see from (17) that P_0 is a monotonic decreasing function of ρ. Therefore, the $(1-\alpha) \times 100\%$ lower and upper confidence limits, P_{0L} and P_{0U} of P_0 can be achieved through (9) and (19)–(20). That is,

$$P_{0L} = P_0|_{\rho=\rho_U}, \tag{21}$$

and

$$P_{0U} = P_0|_{\rho=\rho_L}. \tag{22}$$

For the convenience of numerical experiments, we let $\mu_1 = \mu_2 = \cdots = \mu_R = \mu$. Thus (9) becomes

$$P_0 = \left[1 + \sum_{n=1}^{R} \frac{\rho^n}{n!} + \sum_{n=R+1}^{N} \frac{\rho^n}{R!} \left(\frac{b}{R}\right)^{n-R}\right]^{-1}, \tag{23}$$

where $\rho = \lambda/\mu$.

We assume that $\hat{\rho} = 1.5$ and $N_a = N_d = m$. The following three cases are analyzed graphically to study the effects of various parameters on the confidence interval of P_0.

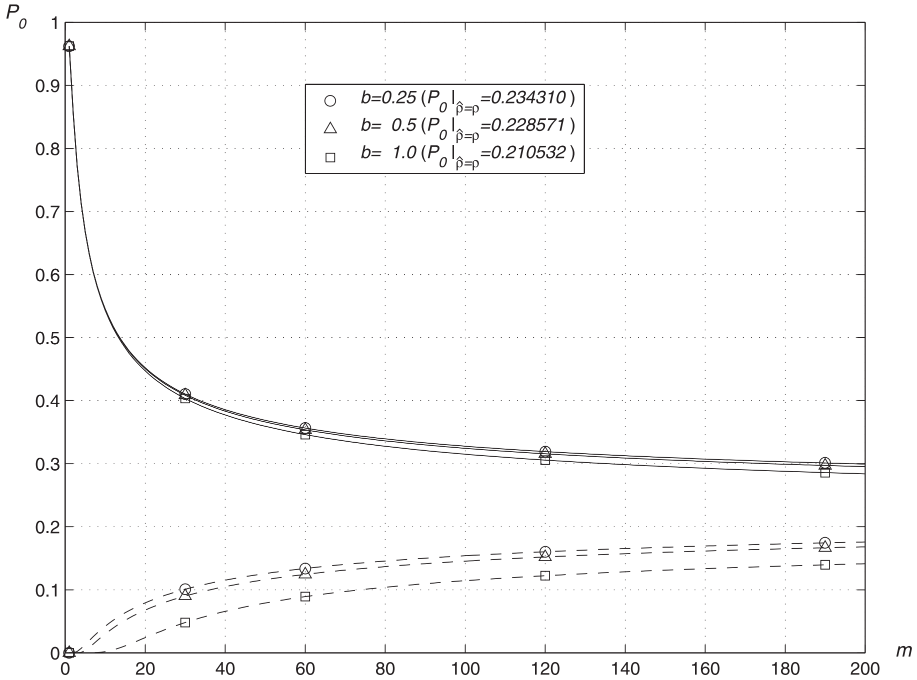

FIGURE 1. The 95% upper and lower confidence limits of P_0 for $R = 3$ and $N = 15$.

Case 1. $R = 3$, $N = 15$, choose different values of $b = 0.25$, 0.5, 1.0.
Case 2. $R = 3$, $b = 0.5$, select various values of $N = 5$, 10, 15.
Case 3. $N = 15$, $b = 0.5$, choose different values of $R = 3$, 5, 8.

The 95% lower and upper confidence limits, P_{0L} and P_{0U} of P_0 are shown in Figures 1–3 for cases 1–3, respectively. Figures 1–3 show the effects of varying b, N, and R, respectively. One observes from Figures 1–3 that (i) at a level of significance equal to 0.05, $P_0|_{\rho=\hat{\rho}}$ lies between the two confidence limits P_{0L} and P_{0U} for all cases; (ii) for given R and N, the confidence interval bands increase as b increases; (iii) for given R and b, the tracks of the confidence interval nearly stay the same when N varies from 5 to 15; and (iv) for given N and b, the branches of the confidence interval are almost no difference when R changes from 3 to 8. Intuitively, this seems too insensitive to changes in R and N. It to be noted that the numerical results show that the length of the confidence interval is getting tight if m is getting large.

3.2. CONFIDENCE INTERVAL FOR $E[N]$

It follows from (18) that $E[N]$ is a monotonic increasing function of ρ. We obtain the $(1 - \alpha) \times 100\%$ lower and upper confidence limits, $E[N]_L$ and $E[N]_U$

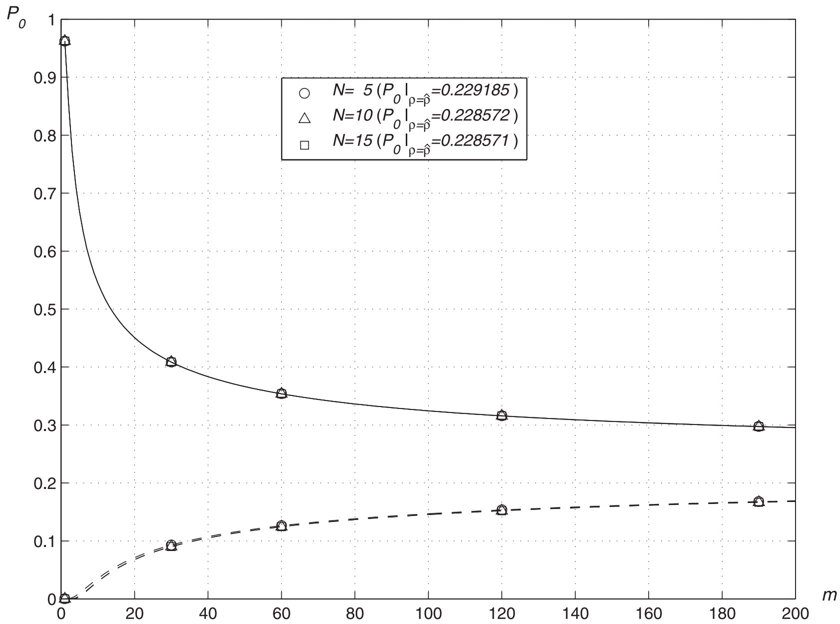

FIGURE 2. The 95% upper and lower confidence limits of P_0 for $R = 3$ and $b = 0.5$.

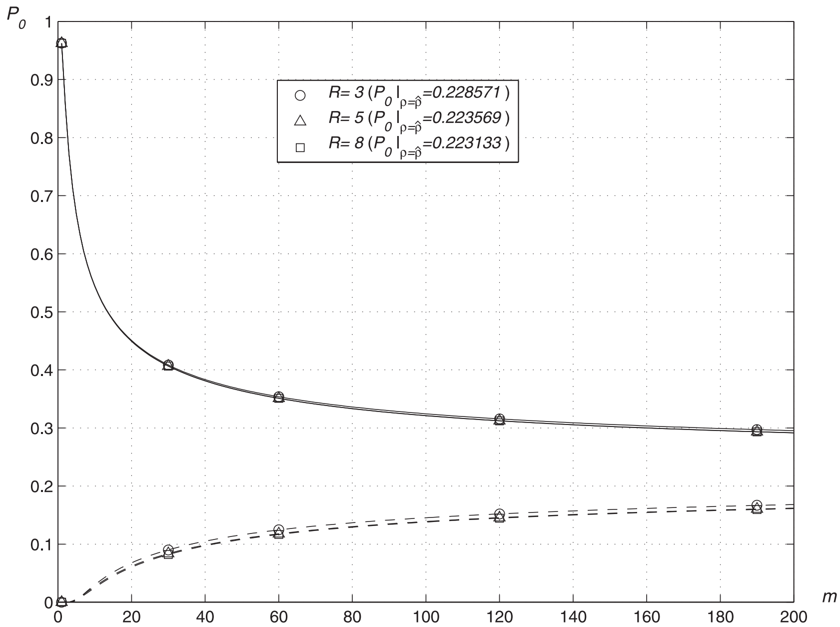

FIGURE 3. The 95% upper and lower confidence limits of P_0 for $N = 15$ and $b = 0.5$.

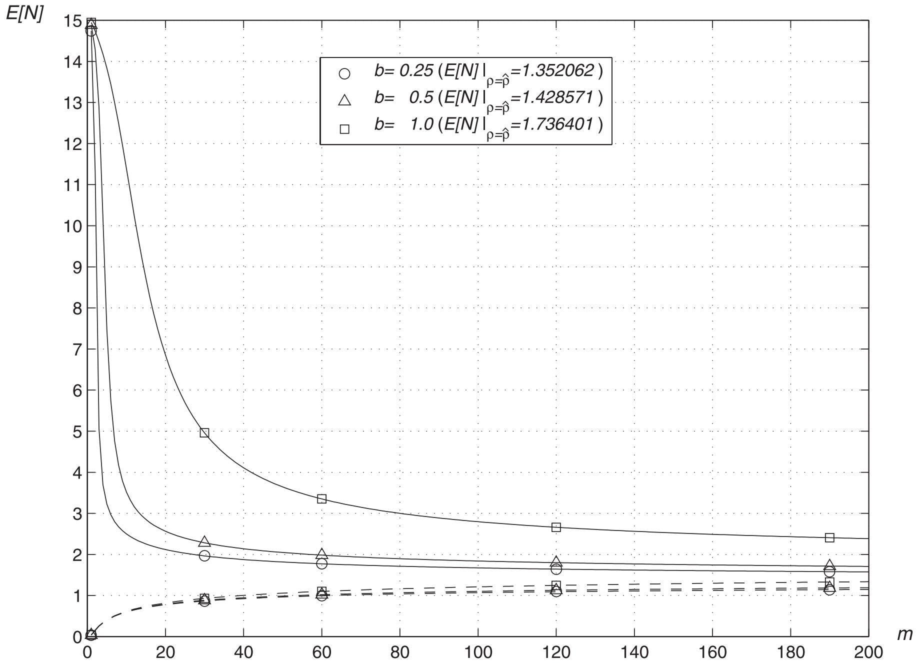

FIGURE 4. The 95% upper and lower confidence limits of E[N] for $R = 3$ and $N = 15$.

of $E[N]$, through (10) and (19)–(20). That is,

$$E[N]_L = E[N]|_{\rho=\rho_L}, \tag{24}$$

and

$$E[N]_U = E[N]|_{\rho=\rho_U}. \tag{25}$$

As in Section 3.1, we let $\mu_1 = \mu_2 = \cdots = \mu_R = \mu$. Thus (10) becomes

$$E[N] = \left[\sum_{n=1}^{R} \frac{n\rho^n}{n!} + \sum_{n=R+1}^{N} \frac{n\rho^n}{R!}\left(\frac{b}{R}\right)^{n-R}\right] P_0, \tag{26}$$

where $\rho = \lambda/\mu$.

We assume $\hat{\rho} = 1.5$ and $N_a = N_d = m$. The following cases are analyzed graphically to study the effects of various parameters on the confidence interval of $E[N]$.

Case 1. $R = 3$, $N = 15$, choose different values of $b = 0.25, 0.5, 1.0$.
Case 2. $R = 3$, $b = 0.5$, select various values of $N = 5, 10, 15$.
Case 3. $N = 15$, $b = 0.5$, choose different values of $R = 3, 5, 8$.

The 95% lower and upper confidence limits, $E[N]_L$ and $E[N]_U$ of $E[N]$ are shown in Figures 4–6 for cases 1–3, respectively. Figures 4–6 show the effects of

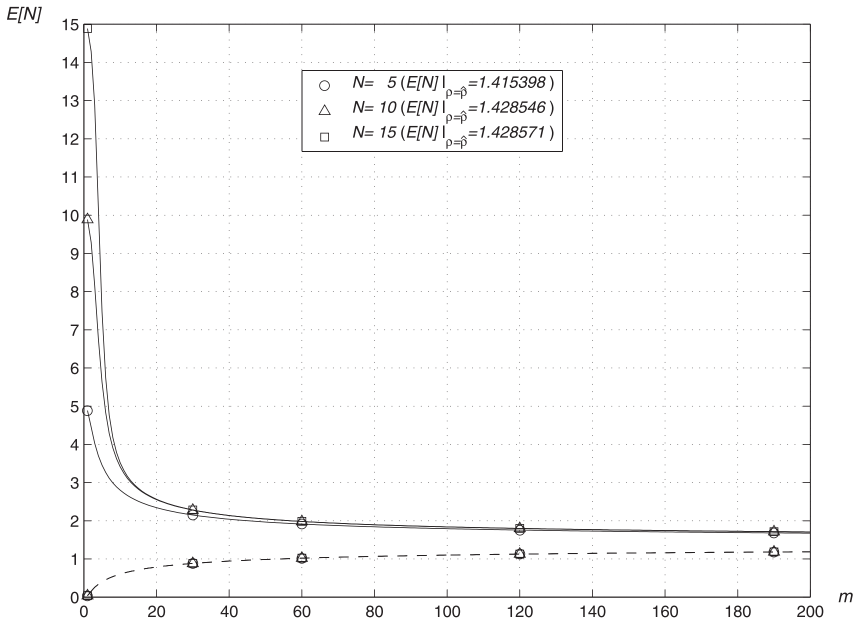

FIGURE 5. The 95% upper and lower confidence limits of E[N] for $R = 3$ and $b = 0.5$.

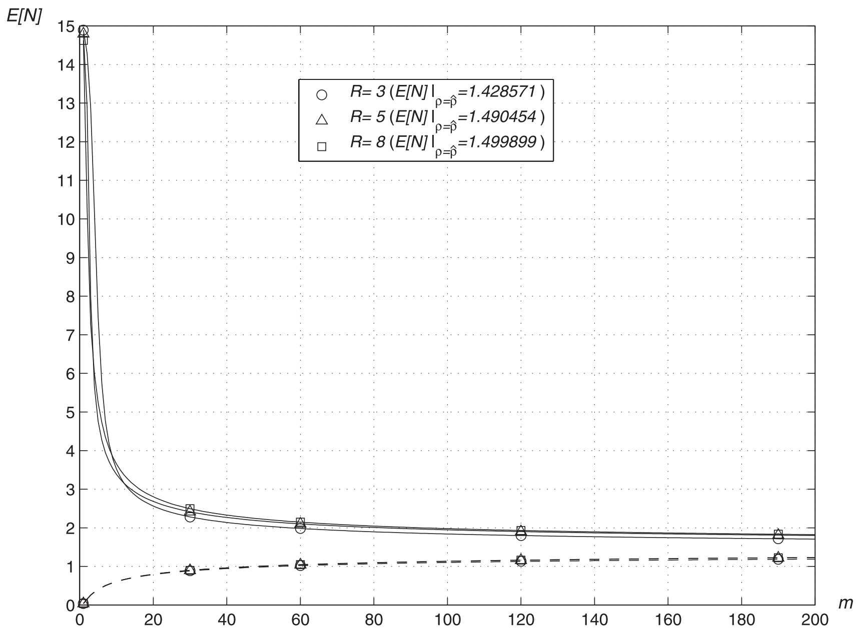

FIGURE 6. The 95% upper and lower confidence limits of E[N] for $N = 15$ and $b = 0.5$.

varying b, N, and R, respectively. We observes from Figures 4–6 that at a level of significance equal to 0.05, $E[N]|_{\rho=\hat{\rho}}$ lies between the two confidence limits $E[N]_L$ and $E[N]_U$ for all cases. The reader can be referred to Figures 4–6 to achieve similar conclusions as listed above for those of P_0.

Example

A library rents three Xerox machines located in reader service suite. Assume users (readers) that arrive at the machines and choose the faster machine to perform some copy tasks if the three machines are available. The management wants to know the long-run fraction of time the server is busy. He observes that 60 users arrived at the machines for performing copy tasks during a period of four hour. The amount of time is thirty minutes during which three machines are busy. There are 15 users to complete their copy tasks when three machines are busy. Therefore, the management gets $T = 4$, $T_{B_R} = 0.5$, $N_{D_R} = 15$, and $N_a = 60$. If $b = 1$ (without balking), then

$$\hat{\rho} = \frac{N_a T_{B_R}}{N_{D_R} T} = 0.5.$$

The system capacity $N = 100$ is considered. At a level of significance equal to 0.05, $P_0|_{\rho=\hat{\rho}} = 0.2105$ lies between the two confidence limits $P_{0L} = 0.0886$ and $P_{0U} = 0.3462$. Furthermore, at a level of significance equal to 0.05, $E[N]|_{\rho=\hat{\rho}} = 1.7368$ lies between the two confidence limits $E[N]_L = 1.1015$ and $E[N]_U = 3.4534$.

On the other hand, if $b = 0.5$ (balking), then

$$\hat{\rho} = \frac{N_a T_{B_R}}{N_{D_R}\left[T + (b-1)T_{B_R}\right]} \approx 1.846.$$

At a level of significance equal to 0.05, $P_0|_{\rho=\hat{\rho}} = 0.1649$ lies between the two confidence limits $P_{0L} = 0.0790$ and $P_{0U} = 0.2799$. Furthermore, at a level of significance equal to 0.05, $E[N]|_{\rho=\hat{\rho}} = 1.7265$ lies between the two confidence limits $E[N]_L = 1.2414$ and $E[N]_U = 2.4120$.

4. Conclusions

In this paper, the maximum likelihood estimates of the parameters for the M/M/R/N queue with balking and heterogeneous servers are developed. We have shown that it generalizes either the M/M/2 queue with heterogeneous servers, or the M/M/2/N queue with balking and heterogeneous servers. The confidence interval formula for ρ, P_0 and $E[N]$ of an M/M/R queue with balking and heterogeneous servers are developed. We also provide the numerical results to show the effects of varying b, N, and R.

Appendix

Derivations of (17) and (18)

Taking the logarithm of (9) and differentiating it with respect to λ, we finally get

$$\frac{\partial \mathrm{ln} P_0}{\partial \lambda} = -\left[\sum_{n=1}^{R} \frac{n\lambda^{n-1}}{\prod_{j=1}^{n} \sum_{i=1}^{j} \mu_i} + \sum_{n=R+1}^{N} \frac{n\lambda^{n-1}}{\prod_{j=1}^{R} \sum_{i=1}^{j} \mu_i} \left(\frac{b}{\mu}\right)^{n-R}\right] P_0. \tag{A.1}$$

Multiplying (A.1) by $-\lambda$ and using (10), we obtain

$$E[N] = -\lambda \frac{\partial \mathrm{ln} P_0}{\partial \lambda}. \tag{A.2}$$

Since

$$\frac{\partial \mathrm{ln} P_0}{\partial \rho} = \frac{\frac{\partial \mathrm{ln} P_0}{\partial \lambda}}{\frac{\partial \rho}{\partial \lambda}},$$

it follows that

$$\frac{\partial ln P_0}{\partial \rho} = \mu \frac{\partial \mathrm{ln} P_0}{\partial \lambda} \cdot$$

Thus we obtain

$$\rho \frac{\partial \mathrm{ln} P_0}{\partial \rho} = \lambda \frac{\partial \mathrm{ln} P_0}{\partial \lambda} \cdot \tag{A.3}$$

From (A.2) and (A.3), it implies that

$$E[N] = -\rho \frac{\partial \mathrm{ln} P_0}{\partial \rho} \geq 0, \tag{A.4}$$

which is the derivation of (17).

Using the chain rule, we obtain

$$\frac{\partial P_0}{\partial \lambda} = \frac{\frac{\partial \mathrm{ln} P_0}{\partial \lambda}}{\frac{\partial \mathrm{ln} P_0}{\partial P_0}} = P_0 \frac{\partial \mathrm{ln} P_0}{\partial \lambda} \cdot \tag{A.5}$$

From (A.5), we finally get

$$\lambda\frac{\partial E[N]}{\partial\lambda} = \lambda\left\{\left[\sum_{n=1}^{R}\frac{n^2\lambda^{n-1}}{\prod_{j=1}^{n}\sum_{i=1}^{j}\mu_i} + \sum_{n=R+1}^{N}\frac{n^2\lambda^{n-1}}{\prod_{j=1}^{R}\sum_{i=1}^{j}\mu_i}\left(\frac{b}{\mu}\right)^{n-R}\right]P_0 + \left[\sum_{n=1}^{R}\frac{n\lambda^{n}}{\prod_{j=1}^{n}\sum_{i=1}^{j}\mu_i} + \sum_{n=R+1}^{N}\frac{n\lambda^{n}}{\prod_{j=1}^{R}\sum_{i=1}^{j}\mu_i}\left(\frac{b}{\mu}\right)^{n-R}\right]P_0\frac{\partial \ln P_0}{\partial\lambda}\right\}. \quad (A.6)$$

It yields from (A.2) that

$$\begin{aligned}\lambda\frac{\partial E[N]}{\partial\lambda} &= \left[\sum_{n=1}^{R}\frac{n^2\lambda^{n}}{\prod_{j=1}^{n}\sum_{i=1}^{j}\mu_i} + \sum_{n=R+1}^{N}\frac{n^2\lambda^{n}}{\prod_{j=1}^{R}\sum_{i=1}^{j}\mu_i}\left(\frac{b}{\mu}\right)^{n-R}\right]P_0 - (E[N])^2\\ &= E[N^2] - (E[N])^2\\ &= \mathrm{Var}[N]. \end{aligned} \quad (A.7)$$

Thus

$$\mathrm{Var}[N] = \lambda\frac{\partial E[N]}{\partial\lambda} = \lambda\frac{\partial\rho}{\partial\lambda}\frac{\partial E[N]}{\partial\rho} = \rho\frac{\partial E[N]}{\partial\rho} \geq 0, \quad (A.8)$$

which is the derivation of (18).

References

[1] M.O. Abou-E1-Ata and A.M.A. Hariri, Point estimation and confidence intervals of the M/M/2/N queue with balking and heterogeneity. *Amer. J. Math. Manage. Sci.* **15** (1995) 35–55.

[2] I.V. Basawa, U.N. Bhat and R. Lund, Maximum likelihood estimation for single server queues from waiting time data. *Queue. Syst. Theory Appl.* **24** (1996) 155–167.

[3] I.V. Basawa and N.U. Prabhu, Estimation in single server queues. *Naval Res. Logist. Quarterly* **28** (1981) 475–487.

[4] A.B. Clarke, Maximum likelihood estimates in a simple queue. *Ann. Math. Statist.* **28** (1957) 1036–1040.

[5] U. Dave and Y.K. Shah, Maximum likelihood estimates in an M/M/2 queue with heterogeneous servers. *J. Oper. Res. Soc.* **31** (1980) 423–426.

[6] J.H. Dshalalow, *Frontiers in queueing: Models and Applications in Science and Engineering.* CRC Press, Inc. (1997).

[7] M.L. Huang and P. Brill, On estimation in M/G/c/c queues. *Internat. Trans. Oper. Res.* **8** (2001) 647–657.

[8] S. Jain, Estimation in M/E_k /1 queueing systems. *Comm. Statist. Theory Methods* **20** (1991) 1871–1879.

[9] S. Jain and J.G.C. Templeton, Confidence interval for M/M/2 queue with heterogeneous servers. *Oper. Res. Lett.* **10** (1991) 99–101.

[10] H.W. Lilliefors, Some confidence intervals for queues. *Oper. Res.* **14** (1966) 723–727.

[11] J. Rodrigues and J. G. Leite, A note on Bayesian analysis in M/M/1 queues derived from confidence intervals. *Statistics* **31** (1998) 35–42.

[12] G. Rubin and D.S. Robson, A single server queue with random arrivals and balking: confidence interval estimation. *Queue. Syst. Theory Appl.* **7** (1990) 283–306.