CHAPTER 2

CONFIDENCE LIMITS FOR EXPECTED WAITING TIME OF TWO QUEUEING MODELS

A modified version of this chapter has been published in ORION, Vol. 20, No. 1, 2004

2.1.1 Introduction

Once one is armed with a queueing model of a system, one which is described by equations which emulate the relevant birth-death process, parametric estimation is one of the essential tools to understand the random phenomena using stochastic models. Whenever systems are fully observable in terms of their basic random components such as inter arrival times and service times, standard parametric estimation techniques of statistical theory are quite appropriate. Most of the studies of several queueing models are confined to only obtaining expressions for transient or stationary (steady state) solutions and do not consider the associated inference problems. Recently, Bhat [41] has provided an overview of methods available for estimation, when the information is restricted to the number of entities in the system at certain discrete points in time. Narayan Bhat has also described how maximum likelihood estimation (MLE) is applied directly to the underlying Markov chain in the queue length process in $M|G|1$ and $GI|M|1$. An attempt is made in this chapter to obtain MLE, a consistent asymptotically normal estimator (CAN) and asymptotic confidence limits for the expected waiting time per entity in $M|M|1|\infty$ and $M|M|1|N$ queues. These two models and the expected waiting time per entity for each model are explained briefly.

2.1.2 Description of Systems

Model I The $(\boldsymbol{M}|\boldsymbol{M}|\boldsymbol{1}):(\boldsymbol{FCFS}|\infty|\infty)$ queue

It can be readily seen that (Taha [3]) the difference-differential equations governing $M|M|1$ are given by

$$p_n'(t) = \lambda p_{n-1}(t) - (\lambda + \mu) p_n(t) + \mu p_{n+1}(t), \qquad n = 1,2,3,... \tag{2.1.1}$$

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t), \qquad n = 0 \tag{2.1.2}$$

As $t \to \infty$, the steady state solution can be proved to exist, when $\lambda < \mu$. Assuming that $p_n'(t) \to 0$ and $p_n(t) \to p_n$ as $t \to \infty$, for $n = 0,1,2,...$, it is clear that

$$-\lambda p_0 + \mu p_1 = 0, \qquad n = 0 \tag{2.1.3}$$

$$\lambda p_{n-1} - (\lambda + \mu) p_n + \mu p_{n+1} = 0, \qquad n = 1,2,3,... \tag{2.1.4}$$

Solving these difference-differential equations,

$$p_n = (1 - \rho)\rho^n, \qquad n = 0,1,2,... \tag{2.1.5}$$

where $\rho = \dfrac{\lambda}{\mu} < 1$.

Clearly (2.1.5) corresponds to the probability mass function of the Geometric distribution. The expected waiting time per entity in the queue is given by

$$ {}_1W_Q = \frac{\lambda}{\mu(\mu - \lambda)} . \tag{2.1.6} $$

Model II The $(\boldsymbol{M}|\boldsymbol{M}|\boldsymbol{1}):(\boldsymbol{GD}|\boldsymbol{N}|\infty)$ queue

The model is essentially the same as Model I, except that the maximum number of entities in the system is limited to N (maximum queue length is N-1) (Taha [3]). The steady state equations for the model are given by

$$ -\rho p_0 + p_1 = 0, \qquad n = 0 \tag{2.1.7} $$

$$ \rho p_{n-1} - (\rho + 1) p_n + p_{n+1} = 0, \qquad n = 1,2,3,\ldots, N-1 \tag{2.1.8} $$

$$ \rho p_{N-1} - p_N = 0, \qquad n = N \tag{2.1.9} $$

The solution of the above difference-differential equations is given by

$$ p_n = \frac{(1-\rho)}{(1-\rho^{N+1})}\rho^n, \qquad n = 0,1,2,\ldots, N \tag{2.1.10} $$

The expected number of entities in the system is given by

$$ L_s = \frac{\rho\{1-(N+1)\rho^N + N\rho^{N+1}\}}{(1-\rho)(1-\rho^{N+1})}, \qquad \rho \neq 1 \tag{2.1.11} $$

Since the queue length is limited and some entities are lost, it is necessary to compute the effective arrival rate λ_{eff}, which is given by

$\lambda_{eff} = \lambda(1 - p_N)$.

The expected number of entities in the queue L_Q is

$$L_Q = L_s - \frac{\lambda_{eff}}{\mu}$$
$$= \frac{\rho^2\left[1 - N\rho^{N-1} + (N-1)\rho^N\right]}{(1-\rho)(1-\rho^{N+1})} \quad .$$

(2.1.12)

Hence the expected waiting time per entity in the queue is given by

$${}_2W_Q = \frac{L_Q}{\lambda_{eff}}$$
$$= \frac{\lambda\left[(\mu^N - \lambda^N) - N\lambda^{N-1}(\mu - \lambda)\right]}{\mu(\mu - \lambda)(\mu^N - \lambda^N)} \quad .$$

(2.1.13)

2.1.3 The ML and CAN estimators for expected waiting time

2.1.3.1 The ML Estimator

Considering $X_{i1}, X_{i2}, ..., X_{in}$ (with $i = 1,2$ representing Models I and II) to be random samples of size n, each randomly drawn from different exponential inter arrival time populations with the parameter λ. and letting $Y_{i1}, Y_{i2}, ..., Y_{in}$ (with $i = 1,2$ representing Models I and II) be random samples of size n, each drawn from different exponential service time populations with the parameter μ, it follows that $\mathrm{E}(\overline{X}_i) = \frac{1}{\lambda}$ and $\mathrm{E}(\overline{Y}_i) = \frac{1}{\mu}$, where $\overline{X}_i$ and $\overline{Y}_i$, $i = 1,2$, are the sample means of inter arrival times and service times respectively corresponding to Models I and II. Further $\overline{X}_i$ and $\overline{Y}_i$ (with $i = 1,2$ representing Models I and II) are the MLEs of $\frac{1}{\lambda}$ and $\frac{1}{\mu}$ respectively. Let $\theta_1 = \frac{1}{\lambda}$ and $\theta_2 = \frac{1}{\mu}$ respectively.

Model I

The average waiting time per entity in the queue given in (2.1.6) reduces to

$$ {}_1W_Q = \frac{\theta_2^2}{(\theta_1 - \theta_2)} \tag{2.1.14} $$

and hence the MLE of W_Q is given by

$$ {}_1\hat{W}_Q = \frac{\overline{Y}_1^{\,2}}{(\overline{X}_1 - \overline{Y}_1)} \tag{2.1.15} $$

Model II

The average waiting time per entity in the queue given in (2.1.13) reduces to

$$ {}_2W_Q = \frac{\theta_2^2[(\theta_1^N - \theta_2^N) + N\theta_2^{N-1}(\theta_2 - \theta_1)]}{(\theta_2 - \theta_1)(\theta_2^N - \theta_1^N)} $$

(2.1.16)

and hence the MLE of W_Q is given by

$$ {}_2\hat{W}_Q = \frac{\overline{Y}_2^2[(\overline{X}_2^N - \overline{Y}_2^N) + N.\overline{Y}_2^{N-1}(\overline{Y}_2 - \overline{X}_2)]}{(\overline{Y}_2 - \overline{X}_2)(\overline{Y}_2^N - \overline{X}_2^N)} $$

(2.1.17)

It may be noted that ${}_i\hat{W}_Q$ given in (2.1.15) and (2.1.17) are real valued functions in $\overline{X}_i$ and $\overline{Y}_i$, $i = 1,2$, which are also differentiable. The following application of the multivariate central limit theorem may be considered (Rao [42]).

2.1.3.2 An application of the multivariate central limit theorem

Suppose $T_1', T_2', T_3',...$ are independent and identically distributed k-dimensional random variables such that

$$ T_n' = (T_{1n}, T_{2n}, T_{3n},..., T_{kn}), \qquad n = 1,2,3... $$

having the first and second order moments $\mathrm{E}(T_n) = \mu$ and $Var(T_n) = \Sigma$. The sequence of random variables may be defined as

$$ \overline{T_n'} = (\overline{T}_{1n}, \overline{T}_{2n}, \overline{T}_{3n},..., \overline{T}_{kn}), \qquad n = 1,2,3... $$

where $\overline{T}_{in} = \frac{\sum_{j=1}^{n} T_{ij}}{n}, \qquad i = 1,2,...,k$ and $j = 1,2,...,n$

Then, $\sqrt{n}(\overline{T}_n - \mu) \xrightarrow{d} N(0,\Sigma)$ as $n \to \infty$

2.1.3.3 The CAN Estimator

Model I

By applying the multivariate central limit theorem given to (2.1.15), it readily follows that

$$\sqrt{n}[(\overline{X}_1, \overline{Y}_1) - (\theta_1, \theta_2)] \xrightarrow{d} N(0,\Sigma)$$

as $n \to \infty$, where the dispersion matrix $\Sigma = ((\sigma_{ij}))$ is given by

$$\Sigma = diag(\theta_1^2, \theta_2^2)$$

From (Rao [42]), it follows that

$\sqrt{n}({}_1\hat{W}_Q - {}_1W_Q) \xrightarrow{d} N(0, {}_1\sigma^2(\theta))$, as $n \to \infty$, where $\theta = (\theta_1, \theta_2)$ and

$$\begin{aligned} {}_1\sigma^2(\theta) &= \sum_{i=1}^{2} \left(\frac{\partial {}_1W_Q}{\partial \theta_i} \right)^2 . \sigma_{ii} \\ &= \frac{\theta_2^2[\theta_1^2 + \theta_2^2(2\theta_1 - \theta_2)^2]}{(\theta_1 - \theta_2)^4} \end{aligned} \qquad (2.1.18)$$

Hence, ${}_1\hat{W}_Q$ is a CAN estimator of ${}_1W_Q$. There are several methods for generating CAN estimators and the Method of Moments and the Method of Maximum likelihood are commonly used to generate such estimators (Sinha [43]).

Model II

As in Model I,

$\sqrt{n}({}_2\hat{W}_Q - {}_2W_Q) \xrightarrow{d} N(0, {}_2\sigma^2(\theta))$, as $n \to \infty$, where $\theta = (\theta_1, \theta_2)$, ${}_2W_Q$ and ${}_2\hat{W}_Q$ are given by (4.16) and (4.17) respectively. Further, ${}_2\sigma^2(\theta)$ is computed from the partial derivatives $\left(\frac{\partial\, {}_2W_Q}{\partial \theta_i}\right), i = 1,2$ as in Model I. Thus ${}_2\hat{W}_Q$ is a CAN estimator of ${}_2W_Q$.

2.1.4 Confidence limits for the expected waiting time

Let ${}_i\sigma^2(\hat{\theta})$ be the estimator of ${}_i\sigma^2(\theta)$ (with $i = 1,2$ representing Models I and II) obtained by replacing θ by a consistent estimator ${}_i\hat{\theta}$ namely ${}_i\hat{\theta} = (\overline{X}_i, \overline{Y}_i)$, $i = 1,2$. Let ${}_i\hat{\sigma}^2 = {}_i\sigma^2(\hat{\theta})$. Since ${}_i\sigma^2(\theta)$ is a continuous function of ϑ, ${}_i\hat{\sigma}^2$ is a consistent estimator of ${}_i\sigma^2(\theta)$, i.e., ${}_i\hat{\sigma}^2 \xrightarrow{P} {}_i\sigma^2(\theta)$ as $n \to \infty$, $i = 1,2$. By the Slutsky theorem

$$\frac{\sqrt{n}({}_i\hat{W}_Q - W_Q)}{{}_i\hat{\sigma}} \xrightarrow{d} N(0,1)$$

i.e., $\Pr\left[-k_{\frac{\alpha}{2}} < \frac{\sqrt{n}({}_i\hat{W}_Q - {}_iW_Q)}{{}_i\hat{\sigma}} < k_{\frac{\alpha}{2}}\right] = (1-\alpha)$

where $k_{\frac{\alpha}{2}}$ is obtained from Normal tables. Hence, a $100(1-\alpha)\%$ asymptotic confidence interval for ${}_iW_Q$ is given by

$${}_i\hat{W}_Q \pm k_{\frac{\alpha}{2}} \cdot \frac{{}_i\hat{\sigma}}{\sqrt{n}}, \qquad i = 1,2 \tag{2.1.19}$$

Numerical Results

Table 2

Confidence limits for M/M/1/ ∞ : FCFS with 99% confidence interval and sample size of 20

λ \ μ	0.04	0.06	0.08	0.1
0.01	(8.289842877:8.376823789)	(3.316926722:3.349739944)	(1.777150118:1.794278453)	(1.105858754:1.116363468)
0.02	(24.83223528:25.16776472)	(8.286718407:8.37994826)	(4.144904413:4.18842892)	(2.487402722:2.512597278)
0.03	(74.16104529:75.83895471)	(16.55477386:16.77855947)	(7.456343866:7.543656134)	(4.262400695:4.309027877)

Table 3

Confidence limits for M/M/1/N: FCFS with 99% confidence interval and sample size of 20

	λ \ μ	0.04	0.06	0.08	0.1
N=10	**0.01**	(8.331269275:8.334920554)	(3.331916366:3.334744788)	(1.784518941:1.786909398)	(1.110057009:1.112165194)
	0.02	(24.75248955:24.7587519)	(8.328275966:8.332745576)	(4.164721818:4.168373097)	(2.498408635:2.501570885)
	0.03	(60.07542472:60.08759256)	(16.50061597:16.50687832)	(7.490679322:7.495571505)	(4.283054105:4.287193479)
	λ \ μ	**0.04**	**0.06**	**0.08**	**0.1**
N=20	**0.01**	(8.331507591:8.335159075)	(3.33191912:3.334747547)	(1.784519057:1.786909514)	(1.110057019:1.112165204)
	0.02	(24.99636102:25.0026853)	(8.33109717:8.335569306)	(4.164840925:4.168492408)	(2.498418861:2.501581139)
	0.03	(73.40387298:73.41482651)	(16.66318663:16.66951092)	(7.497549755:7.502448733)	(4.283644082:4.287784475)
	λ \ μ	**0.04**	**0.06**	**0.08**	**0.1**
N=40	**0.01**	(8.331507591:8.335159075)	(3.33191912:3.334747547)	(1.784519057:1.786909514)	(1.110057019:1.112165204)
	0.02	(24.99683772:25.00316228)	(8.331097265:8.335569401)	(4.164840925:4.168492409)	(2.498418861:2.501581139)
	0.03	(74.98446701:74.99541962)	(16.66350439:16.66982894)	(7.49755051:7.50244949)	(4.283644089:4.287784482)

As is to be expected, W_q is an increasing function of λ, and a decreasing function of μ, for both M/M/1/∞ and M/M/1/N queueing systems [See Tables 2&3].

2.2 Statistical analysis for a tandem queue with blocking

A maximum likelihood estimator (MLE), a consistent asymptotically normal (CAN) estimator and asymptotic confidence limits for the expected service time per customer in the system in a two station tandem queue with zero queue capacity and with blocking are obtained.

2.2.1 Introduction

Many studies of queueing models are confined to obtaining expressions for transient or stationary (steady state) solutions and do not consider the associated statistical inference problems. Parametric estimation is one of the essential tools to understand random phenomena using stochastical models. Analysis of queueing systems in this context has not received due attention. Whenever the systems are fully observable in terms of their basic random components such as inter-arrival times and service times, standard parametric techniques of statistical theory are quite appropriate. Recently Bhat [41] has provided an overview of methods available for estimation, when the information is restricted to the number of entities in the system at some discrete point in time. Bhat has also described how maximum likelihood estimation is applied directly to the underlying Markov chain in the queue length process in M/G/1 and GI/M/1 queues. Yadavalli *et al* [44] have obtained asymptotic confidence limits for the expected waiting time per customer in the queues of $M/M/1/\infty$ and M/M/1/N. Further, Yadavalli *et al* [45] have extended the same results to c parallel servers ($c \geq 1$).

Generally speaking, the queueing models assume that each service channel consists of only one station. Situations do exist, where each service channel may consist of several stations in series. In this situation, an entity must successively pass through all the stations before completing service. Such situations are known as queues in series or tandem queues. Examples of such situations are as follows:

a) In a manufacturing process, units must pass through a series of service channels (work stations), where each service channel performs a given task or job.

b) In a University registration process, each registrant must pass through a series of counters such as advisor, departmental chairman (Head of the Department), Cashier etc.

c) In a clinical physical examination procedure, a patient goes through a series of stages such as laboratory tests, Electro Cardio Graph, Chest X-ray etc.

In all these model structures, it is not only sufficient to know how many persons are in the system but also where they are.

An attempt is made in this paper to study a two station tandem queue with blocking in detail, Taha [3]. An MLE, CAN and asymptotic confidence limits are obtained for the expected service time per entity in the system.

2.2.2 System description and assumptions

Consider a simplified single channel queueing system consisting of two series stations as below:

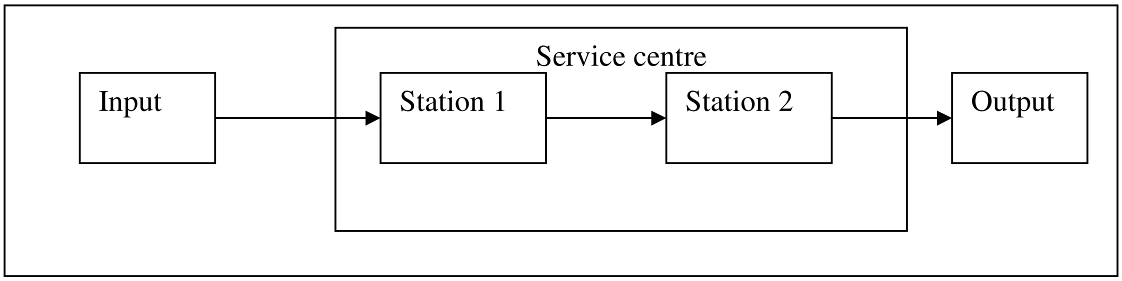


Fig. 2.2.1 System configuration

An entity arriving for service must pass through station 1 and station 2 before completion of service. The precise assumptions of the model are as follows:

(i) Arrivals occur according to a Poisson distribution with a mean rate λ.

(ii) Service times at each station are exponentially distributed with a service rate μ.

(iii) Queues are not permitted ahead of station 1 or station 2.

(iv) Each station is either free or busy.

(v) Station 1 is said to be blocked when the entity in station 1 completes service before station 2 becomes free. In such a case the entity cannot wait between the stations, since this is not allowed.

2.2.3 Analysis of the system

Let the symbols 0,1 and b represent free, busy or blocked states of a station. Let $X(t)$ and $Y(t)$ respectively denote the states of station 1 and station 2 and the vector process $Z(t) = \{(X(t), Y(t)), t \geq 0\}$ with state space

$$E = \{(0,0), (0,1), (1,0), (1,1), (b,1)\}, \qquad (2.2.3.1)$$

the state of the system at time *t.* Since the inter-arrival and service times are exponential, it follows that the process *Z(t)* is a Markov process with the infinitesimal generator given by

$$
\begin{array}{c|ccccc}
E & (0,0) & (0,1) & (1,0) & (1,1) & (b,1) \\
\hline
(0,0) & -\lambda & 0 & \lambda & 0 & 0 \\
(0,1) & \mu & -(\lambda+\mu) & 0 & \lambda & 0 \\
(1,0) & 0 & \mu & -\mu & 0 & 0 \\
(1,1) & 0 & 0 & \mu & -2\mu & \mu \\
(b,1) & 0 & \mu & 0 & 0 & -\mu
\end{array}
\tag{2.2.3.2}
$$

Let $p_{ij}(t) = p[Z(t) = (i, j)], \forall (i, j) \in E$ represent the probability that the system is in state (i, j) at time t with the initial condition $p_{00}(0) = 1$. From the infinitesimal generator given in (2.2.3.2), the following system of differential-difference equations is obtained:

$$\frac{dp_{00}(t)}{dt} = -\lambda p_{00}(t) + \mu p_{01}(t) \tag{2.2.3.3}$$

$$\frac{dp_{01}(t)}{dt} = -(\lambda + \mu)p_{01}(t) + \mu p_{10}(t) + \mu p_{b1}(t) \tag{2.2.3.4}$$

$$\frac{dp_{10}(t)}{dt} = \lambda p_{00}(t) - \mu p_{10}(t) + \mu p_{11}(t) \tag{2.2.3.5}$$

$$\frac{dp_{11}(t)}{dt} = \lambda p_{01}(t) - 2\mu p_{11}(t) \tag{2.2.3.6}$$

$$\frac{dp_{b1}(t)}{dt} = \mu p_{11}(t) - \mu p_{b1}(t) \tag{2.2.3.7}$$

2.2.3.1 Transient Solution

Solving the system of equation (2.2.3.3)-(2.2.3.7) along with the equation $\sum_{(i,j)\in E} p_{ij}(t)=1$ and using Laplace transforms, it is evident that:

$$p_{00}(t)=\frac{2\mu^2}{(3\lambda^2+4\lambda\mu+2\mu^2)}+\lambda\mu^2\sum_{i=1}^{3}\frac{(\alpha_i+2\mu)}{\alpha_i(\alpha_i+\lambda)\prod_{\substack{j=i\\ j\neq 1}}^{3}(\alpha_i-\alpha_j)}e^{\alpha_i t}$$

(2.2.3.8)

$$p_{01}(t)=\frac{2\lambda\mu}{(3\lambda^2+4\lambda\mu+2\mu^2)}+\lambda\mu\sum_{i=1}^{3}\frac{(\alpha_i+2\mu)}{\alpha_i\prod_{\substack{j=i\\ j\neq 1}}^{3}(\alpha_i-\alpha_j)}e^{\alpha_i t}$$

(2.2.3.9)

$$p_{10}(t)=\frac{\lambda(\lambda+2\mu)}{(3\lambda^2+4\lambda\mu+2\mu^2)}+\lambda^2\mu^2\sum_{i=1}^{3}\frac{(2\alpha_i+\lambda+2\mu)}{\alpha_i(\alpha_i+\lambda)(\alpha_i+\mu)}*\frac{e^{\alpha_i t}}{\prod_{\substack{j=i\\ j\neq 1}}^{3}(\alpha_i-\alpha_j)}+\frac{\lambda}{(2\lambda-\mu)}e^{\mu t}$$

(2.2.3.10)

$$p_{11}(t)=\frac{\lambda^2}{(3\lambda^2+4\lambda\mu+2\mu^2)}+\lambda^2\mu\sum_{i=1}^{3}\frac{1}{\alpha_i\prod_{\substack{j=i\\ j\neq 1}}^{3}(\alpha_i-\alpha_j)}e^{\alpha_i t}$$

(2.2.3.11)

$$p_{b1}(t)=\frac{\lambda^2}{(3\lambda^2+4\lambda\mu+2\mu^2)}+\lambda^2\mu^2\sum_{i=1}^{3}\frac{1}{\alpha_i(\alpha_i+\mu)\prod_{\substack{j=i\\ j\neq 1}}^{3}(\alpha_i-\alpha_j)}e^{\mu t}+\frac{\lambda}{(\mu-2\lambda)}e^{-\mu t}$$

(2.2.3.12)

where α_1,α_2 and α_3 are the roots of

$$s^3+(2\lambda+4\mu)s^2+(\lambda^2+7\lambda\mu+5\mu^2)s+\mu(3\lambda^2+4\lambda\mu+2\mu^2)=0$$

2.2.3.2 The Steady state solution

Since the stationary behaviour of the system is to be modelled, let $\lim_{t\to\infty} p_{ij}(t) = p_{ij}$. Let $\underline{p} = (p_{00}, p_{01}, p_{10}, p_{11}, p_{bi})$ be the stationary distribution corresponding to the Markov process $z(t)$. It readily follows from (2.2.3.8)-(2.2.3.12) that

$$p_{00}(t) = \frac{2\mu^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \tag{2.2.3.13}$$

$$p_{01}(t) = \frac{2\lambda\mu}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \tag{2.2.3.14}$$

$$p_{10}(t) = \frac{\lambda(\lambda + 2\mu)}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \tag{2.2.3.15}$$

$$p_{11}(t) = \frac{\lambda^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \tag{2.2.3.16}$$

$$p_{b1}(t) = \frac{\lambda^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \tag{2.2.3.17}$$

It may be noted that the solution given in (2.2.3.13)-(2.2.3.17) is in agreement with Taha [3] with $\rho = \frac{\lambda}{\mu}$

2.2.3.3 Expected service time per entity in the system

The expected number of entities in the system is given by

$$\begin{aligned} L_s &= \sum_{n=0}^{\infty} np_n \\ &= (p_{01} + p_{10}) + 2(p_{11} + p_{b1}) \\ &= \frac{\lambda(5\lambda + 4\mu)}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \end{aligned}$$

(2.2.3.18)

The probability that an entity will enter station 1 is

$$\begin{aligned} &(p_{00} + p_{01}) \\ &= \frac{2\mu(\lambda + \mu)}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \end{aligned}$$

(2.2.3.19)

W_s represents the expected service time per entity in the system since queues are allowed and is given by

$$W_s = \frac{L_s}{\lambda_{eff}} = \frac{L_s}{\lambda(p_{00} + p_{01})} = \frac{(5\lambda + 4\mu)}{2\mu(\lambda + \mu)}$$

(2.2.3.20)

In the next section, the maximum likelihood and consistent asymptotically normal estimators for the expected service time per entity in the system are obtained.

2.2.4 MLE and CAN estimator for the expected service time per entity in the system

2.2.4.1 The ML estimator

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be random samples of size n, each drawn from exponential inter-arrival time and exponential service time populations with parameters λ and μ respectively. It is clear that $\mathrm{E}(\overline{X}) = \dfrac{1}{\lambda}$ and $\mathrm{E}(Y) = \dfrac{1}{\mu}$, where $\overline{X}_i$ and $\overline{Y}_i$ are the sample means of inter-arrival times and service time respectively.

It can be shown that $\overline{X}$ and $\overline{Y}$ are MLEs of $\frac{1}{\lambda}$ and $\frac{1}{\mu}$ respectively.

Let $\theta_1 = \frac{1}{\lambda}$ and $\theta_2 = \frac{1}{\mu}$. The average service time per customer in the system given in (2.2.3.20) reduces to

$$W_s = \frac{\theta_2 (4\theta_1 + 5\theta_2)}{2(\theta_1 + \theta_2)} \tag{2.2.4.1}$$

and hence the MLE of W_s is given by

$$\hat{W}_s = \frac{\overline{Y}(4\overline{X} + 5\overline{Y})}{2(\overline{Y} + \overline{X})} \tag{2.2.4.2}$$

It may be noted that $\hat{W}_s$ given in (2.2.4.2) is a real valued function in $\overline{X}$ and $\overline{Y}$, which are also differentiable. Consider the following application of the multivariate central limit theorem. See Rao [42].

2.2.4.2 An application of the multivariate central limit theorem

Suppose $T_1', T_2', T_3',\ldots$ are independent and identically distributed k-dimensional random variables such that

$$T_n' = (T_{1n}, T_{2n}, T_{3n}, \ldots, T_{kn}), \qquad n = 1,2,3\ldots$$

having the first and second order moments $\mathrm{E}(T_n) = \mu$ and $Var(T_n) = \Sigma$. The sequence of random variables may be defined as

$$\overline{T_n'} = (\overline{T}_{1n}, \overline{T}_{2n}, \overline{T}_{3n}, \ldots, \overline{T}_{kn}), \qquad n = 1,2,3\ldots$$

where $\overline{T}_{in} = \frac{\sum_{j=1}^{n} T_{ij}}{n}$, $\quad i = 1,2,\ldots,k$ and $j = 1,2,\ldots,n$

Then, $\sqrt{n}(\overline{T}_n - \mu) \xrightarrow{d} N(0,\Sigma)$ as $n \to \infty$

2.2.4.3 The CAN Estimator

By applying the multivariate central limit theorem to (2.2.4.2), it readily follows that

$$\sqrt{n}[(\overline{X},\overline{Y}) - (\theta_1,\theta_2)] \xrightarrow{d} N(0,\Sigma)$$

as $n \to \infty$, where the dispersion matrix $\Sigma = ((\sigma_{ij}))$ is given by

$$\Sigma = diag(\theta_1^2,\theta_2^2)$$

Again from Rao [42] it follows that

$\sqrt{n}({}_1\hat{W}_s - {}_1W_s) \xrightarrow{d} N(0, {}_1\sigma^2(\theta))$, as $n \to \infty$, where $\theta = (\theta_1,\theta_2)$ and

$$\sigma^2(\theta) = \sum_{i=1}^{2} \left(\frac{\partial_1 W_s}{\partial \theta_i} \right)^2 .\sigma_{ii}$$

$$= \frac{\vartheta_2^2[\theta_1^2\theta_2^2 + (4\vartheta_1^2 + 10\theta_1\theta_2 + 5\vartheta_2^2)^2]}{4(\vartheta_1 + \vartheta_2)^4}$$

Thus, $\hat{W}_s$ is a CAN estimator of $\hat{W}_s$. There are several methods for generation of CAN estimators and the Method of Moments and the Method of Maximum likelihood are commonly used to generate such estimators. See Sinha [43].

2.2.4.4 Confidence limits for the expected waiting time

Let $\sigma^2(\hat{\theta})$ be the estimator of $\sigma^2(\theta)$ obtained by replacing θ by a consistent estimator $\hat{\theta}$ namely. Let $\hat{\sigma}^2 = \sigma^2(\hat{\theta})$. Since $\sigma^2(\theta)$ is a continuous function of θ, $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2(\theta)$, i.e., $\hat{\sigma}^2 \xrightarrow{P} \sigma^2(\theta)$ as $n \to \infty$, $i = 1,2$. By the Slutsky theorem

$$\sqrt{n}(\hat{W}_s - W_s) \xrightarrow{d} N(0,1)$$

i.e., $$\Pr\left[-k_{\frac{\alpha}{2}} < \frac{\sqrt{n}(\hat{W}_s - W_s)}{{}_i\hat{\sigma}} < k_{\frac{\alpha}{2}}\right] = (1-\alpha)$$

where $k_{\frac{\alpha}{2}}$ is obtained from Normal tables. Hence, a $100(1-\alpha)\%$ asymptotic confidence interval for W_s is given by

$$\hat{W}_s \pm k_{\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}} \qquad (2.2.5.1)$$

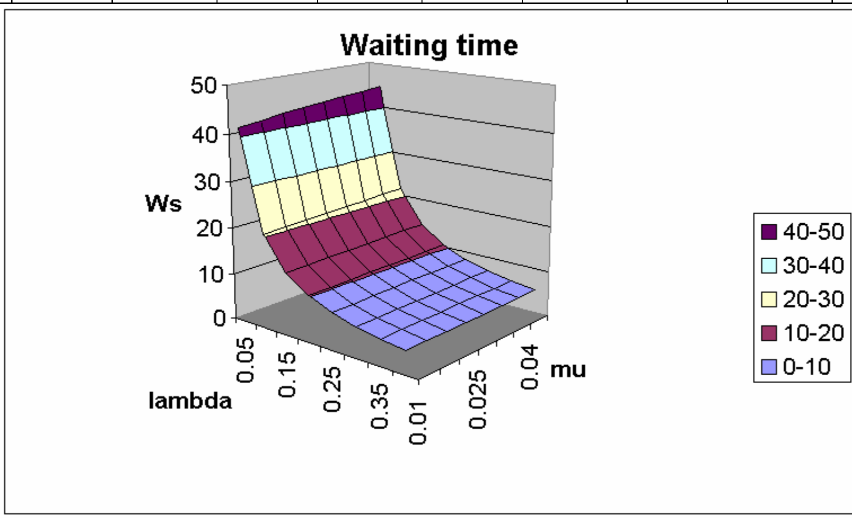
As is to be expected, W_q is an increasing function of λ, and a decreasing function of μ, for a tandem queue with blocking. The numerical illustration of the confidence interval of this model (tandem queues) is shown in Table 4.

Numerical Results

Table 4

Confidence limits for a tandem queue with blocking: 99% confidence interval and sample size of 20

μ / λ	0.05		0.1		0.15		0.2		0.25		0.3		0.35		0.4	
	LCL	**UCL**	**LCL**	**UCL**	**LCL**	**UCL**	**LCL**	**UCL**	**LCL**	**UCL**	**LCL**	**UCL**	**LCL**	**UCL**	**LCL**	**UCL**
0.01	41.57	41.76	20.41	20.50	13.51	13.57	10.10	10.14	8.06	8.10	6.71	6.74	5.74	5.77	5.02	5.04
0.015	42.21	42.41	20.60	20.70	13.61	13.67	10.15	10.20	8.09	8.13	6.73	6.76	5.76	5.79	5.03	5.06
0.02	42.76	42.96	20.79	20.88	13.69	13.76	10.20	10.25	8.13	8.17	6.76	6.79	5.78	5.80	5.05	5.07
0.025	43.23	43.44	20.95	21.05	13.78	13.84	10.25	10.30	8.16	8.20	6.78	6.81	5.80	5.82	5.06	5.09
0.03	43.65	43.85	21.10	21.20	13.86	13.92	10.30	10.35	8.20	8.23	6.80	6.83	5.81	5.84	5.08	5.10
0.035	44.01	44.22	21.25	21.35	13.93	14.00	10.35	10.40	8.23	8.26	6.83	6.86	5.83	5.86	5.09	5.11
0.04	44.34	44.55	21.38	21.48	14.00	14.07	10.39	10.44	8.26	8.29	6.85	6.88	5.85	5.87	5.10	5.13
0.045	44.63	44.84	21.50	21.60	14.07	14.14	10.43	10.48	8.29	8.32	6.87	6.90	5.86	5.89	5.11	5.14



Graph illustrating W_q as a function of λ and μ.