Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü Dergisi
Cilt 9, Sayı: 2, 2007

Cost Function Estimation in Multiserver Queueing Systems with Impatient Customers and Application on Supermarket

Çağın KANDEMİR ÇAVAŞ[*] Özlem EGE ORUÇ[**]

Abstract: Queue forms when the demand exceeds the supply for a system. It is faced with the customers who leave without waiting or who wait a little bit because of longevity of the queue. The system that takes into account this type of customers is called "Multiserver Queueing Systems for Impatient Customers".

In this paper, it is investigated multiserver queueing systems with impatient customers by the two queue models M/M/c+M and M/M/c+D on supermarket. The first aim of the study is to identifiy the relationship between different patience times and cost functions of the two models. The second aim is to investigate effects of patience time's distributions on cost function and optimal number of servers. This study is the preliminary on the queueing systems's investigations in literature of Turkey. Consequently, it was developed the significance of the patience times in order to find performance measures of the multiserver queueing systems with impatient customers. The results were confirmed with an application by using MATHEMATICA 5.0 for observed real data.

Keywords: Cost function, impatient customers, multiserver queueing systems, optimal number of servers, patience times.

Sabırsız Müşteriler için Çok Servis Birimli Kuyruk Sistemlerinde Maliyet Fonksiyonu Tahmini ve Süpermarket Üzerine Uygulama

Öz: Bir servis için sınırlı arzın, fazla talebin sözkonusu olması kuyruklara neden olur. Kuyruk uzun olduğu zaman, hiç beklemeden veya bir süre bekleyip servisin yavaşlığından sıkılarak kuyruktan ayrılan müşterilerle karşılaşılır. Bu tür müşterileri dikkate alan sisteme "Sabırsız Müşteriler için Çok Servis Birimli Kuyruk Sistemi" adı verilir.

Bu makalede, sabırsız müşterili çok servis birimli kuyruk sistemleri M/M/c+M ve M/M/c+D modelleri ile bir süpermarket üzerinde incelenmiştir. Bu çalışma boyunca, bu iki modelin sahip olduğu farklı sabırlılık süreleri ile maliyet fonksiyonu arasındaki ilişkinin belirlenmesi amaçlanmıştır. Daha sonraki diğer

[*] DEÜ Fen-Edebiyat Fakültesi, İstatistik Bölümü, Araştırma Görevlisi.
[**] DEÜ Fen-Edebiyat Fakültesi, İstatistik Bölümü, Öğretim Üyesi, Yard.Doç.Dr.

bir amaç ise sabırlılık süresi dağılımlarının maliyet fonksiyonu ve en uygun servis birimi sayısı üzerindeki etkisini incelemektir. Bu çalışma Türkiye literatüründe kuyruk sistemleri üzerine yapılan araştırmalar içinde bir ilki temsil etmektedir.

Sonuç olarak, Sabırsız müşterili çok kanallı kuyruk modellerinin performans ölçülerini bulmak için sabırlılık süresinin önemi ortaya çıkarılmıştır. Gözlenen gerçek veriler için MATHEMATICA 5.0 yazılımı kullanılarak yapılan bir uygulama ile bulduğumuz sonuçlar desteklenmiştir

Anahtar sözcükler: Maliyet fonksiyonu, sabırsız müşteriler, çok servis birimli kuyruk sistemleri, en uygun servis birimi sayısı, sabırlılık süreleri.

1. Introduction

Queue is an often case in our daily life. It is waited for a service such as paying the bills, buying a movie ticket, passing from a motorway... etc. The queues form, when the demand for a service exceeds its supply.

Queue affects adversely the important parameters of mankind such as cost and time. Theoretically, enough servers could be provided to handle all arriving customers without delaying. However, in reality, although top management of the business may sympathize with the desire to improve service quality, it is clear that the increase in the service capacity requires extra financial resources. On the other hand, a reduction in the service capacity results in a significant increase in the cost associated with waiting (Mendelson, 1985). In addition, whether the customer is patient or not is very determinative on cost. Impatience of the customers causes a loss and a cost for the system. Several studies have been done in order to minimize the number of loss customers and the cost by guidance of the analytic framework.

In literature, the first queueing model with impatient telephone switchboard customers is investigated by Palm (1937). Barrer (1957) studied multiserver queue with impatience. Baccelli and Hebuterne (1981) investigated some special cases involving impatience with a general distribution function of practical interest for modelling telecommunication system.

In this paper, it is aimed to explain the cost funtion of the multiserver queueing systems with impatient customers by using the probabilistic concepts. The real queues data taken from İzmir Pehlivanoğlu Market in Şirinyer are examined whether the results deduced theoretically are confirmed in reality.

2. Impatient Customers

In many service systems, customers wait for service for a limited time, then they leave the system if service has not processed. This type of customer is known as impatient customer in the litterature.

In literature, the first queueing model with impatient telephone switchboard customers is investigated by Palm (1937). Barrer (1957) studies multiserver queue with impatience customers and he determines the loss probability in M/M/c+D queue. Boxma and Waal (1993) obtain the loss probability for impatient customers in queue models by considering the number of servers that minimizes a certain cost function. Boots and Tijms (1999) develop another simple and insightful formula for the loss probability in multiserver queueing system with impatient customers. Zohar, Mandelbaum and Shimkin (2002) characterize an equilibrium point for the customers's adaptation of their patience to their service expectations and then model adaptive customer behaviour in multiserver queueing model for impatient customers.

3. Erlang Loss Probability

"In a delay system each customer finding no free server upon arrival waits until a server becomes available, whereas in a loss system each customer finding no free server upon arrival are lost and have no influence on the system." (Tijms, 2003, p.194).

The probability p_i gives the probability of *i* number of occupied servers. Since $p_i = (\lambda/\mu)^i \; p_0/i!$ for $i = 1,\ldots,c$, the following result can be obtained as follows:

$$p_i = \frac{(\lambda/\mu)^i/i!}{\sum_{k=0}^{c}(\lambda/\mu)^k/k!} \text{ where } i = 0,1,\ldots,c \qquad (1)$$

The loss probability is generally denoted by π and is given by

$$\pi = \frac{probability\ that\ all\ c\ servers\ are\ busy}{total\ \ probability\ that\ an\ arriving\ customer\ \ find\ \ free\ server}$$

or it can be written as follows:

$$\pi = \frac{(\lambda/\mu)^c / c!}{\sum_{k=0}^{c} (\lambda/\mu)^k / k!} \tag{2}$$

Equation (2) is called the Erlang loss formula in the litterature.

4. Loss Probability For Impatient Customers

The loss probability P_{loss} is formulated that as $P_{loss} = \frac{(1-\rho)P\{W_q^{(\infty)} > \tau\}}{1-\rho P\{W_q^{(\infty)} > \tau\}}$ by Tijms and Boots (1999) where the notations are respectively expressed as following; each arriving customer enters the queueing system, but after a fixed time, $\tau > 0$, leave the system if the service has not begun. The probability of customers whose waiting time in queue does not exceed t, $t \geq 0$ as denoted as $W_q^{(\infty)}(t)$. $W_q^{(\infty)}(t)$ is distributed as the stationary waiting time of a customer in the standart M/G/c queue can be written as $P\{W_q^{(\infty)} \leq t\} = W_q^{(\infty)}(t)$ where $t \geq 0$.

Suppose that customers arrive at a service facility of a multiserver queueing system according to a Poisson process with rate λ. Service requests of successive customers are independent, identically distributed (i.i.d) with exponential distribution with mean μ^{-1} at each of the c servers. The service discipline is first-come-first-served.

Waiting customers may abandon the queue at any time before admitted to service. Potential abandonments times of individual customers are assumed independent and identically distributed, according to a probability distribution $G(\cdot)$ over the nonnegative real line. G referred to the patience distribution function with mean γ. Suppose that $\overline{G} = 1 - G$ denote the survival function; thus $\overline{G}$ is the probability that a waiting customer will not abandon within t

time units. The arrival, service and patience processes are independent stochastic processes.

M/M/c queue with general patience time distribution will be denoted as M/M/c+G in queueing systems for impatient customers.

The time that a nonabandoning customer would have to wait until admitted to service is called virtual offered waiting time, or offered wait, and denoted as V. V will be also the unfinished work of the server, and it will only be modified by successful customers. Let the distribution of V is the same for all customers. (Zohar et al, 2002)

The M/M/c+G queue model is given below was investigated by Baccelli and Hebuterne (1981).

$N(t)=n$, for the number of customers at time t equals n and $0 \le n \le c-1$.

$N(t)=L$, for the number of customers at time t equals c or exceeds c.

V is strictly positive when $N(t)=L$, and it equals zero elsewhere. The density of V is denoted as $v(x)$.

$$P_j = \lim_{t\to\infty} P\{N(t)=j, V=0\} \quad \text{where } j=0,\ldots,c-1 \tag{3}$$

Hence

$$v(x) = \lim_{t\to\infty}\lim_{dx\to 0} P\{N(t)=L, x \le V \le x+dx\}/dx \tag{4}$$

From the Chapman-Kolmogorov equations P_j and $v(0)$ are expressed that

$$P_j = \frac{\rho^j}{j!} P_0 \quad \text{where } j=0,\ldots,c-1 \tag{5}$$

$$v(0) = \lambda P_{c-1} \tag{6}$$

Under the stability condition $c\mu > \lambda\overline{G}(\infty)$, $v(x)$ is given as in equation (23) by Bacelli&Hebuterne (1981).

$$v(x) = v(0)\exp(J(t)) \text{ where } t>0 \tag{7}$$

where the expression of $J(t)$ in the equation (7) is expressed that:

$$J(t)) = -\int_0^t (c\mu - \lambda\overline{G}(s))ds \tag{8}$$

As it is mentioned before, P_j denote the stationary probability for exactly *j* occupied servers. Therefore the normalization condition becomes as follows:

$$\sum_{j=0}^{c-1} P_j + \int_0^{\infty} v(x)dx = 1 \tag{9}$$

where

$$P_j = \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} P_0 \tag{10}$$

Then the distribution function of $v(x)$ is expressed as follows:

$$v(x) = \frac{exp(J(t))}{\frac{K_c}{\lambda} + \int_0^{\infty} exp(J(s))ds} \tag{11}$$

where

$$K_c = \sum_{j=0}^{c-1} \frac{(c-1)!}{j!} \left(\frac{\lambda}{\mu}\right)^{j-c+1} \tag{12}$$

(Zohar et al, 2002)

The loss probability π is

$$\pi = \int_0^{\infty} G(x)v(x)dx \tag{13}$$

$$\pi = \left(1 - \frac{c}{\rho}\right)\left(1 - \sum_{j=0}^{c-1} P_j\right) + P_{c-1} \tag{14}$$

4.1. Loss Probability For M/M/c+D and M/M/c+M Model

Throughout the previous formula of the loss probability (14), the loss probability, π and the probability of all free servers for M/M/c+D is defined as follows:

$$P_0 = \left[\sum_{k=0}^{c-1} \frac{\rho^k}{k!} + \frac{\rho^c}{(\rho - c)c!}\left\{\rho e^{(\lambda - c\mu)\gamma} - c\right\}\right]^{-1} \tag{15}$$

$$\pi = P_0 \frac{\rho^c}{c!} exp[(\lambda - c\mu)\gamma] \tag{16}$$

and also the probability of all free servers and the loss probability for M/M/c+M is defined as follows:

$$P_0 = \left[\sum_{k=0}^{c-1} \frac{\rho^k}{k!} + \frac{\rho^c}{c!}\left(1 + \frac{\rho/c}{1+\alpha} + \frac{(\rho/c)^2}{(1+\alpha)(1+2\alpha)} + \ldots\right)\right]^{-1} \tag{17}$$

$$\pi = P_0 \frac{\rho^{c-1}}{(c-1)!}\left[1 + (\rho/c - 1)\left[\frac{\rho/c}{1+\alpha} + \frac{(\rho/c)^2}{(1+\alpha)(1+2\alpha)} + \ldots\right]\right] \tag{18}$$

5. Cost Model Of The Queueing System

The study of queues determines the measures of performance of queueing systems, including the average waiting time and the average queue length. This information is then used by managers to decide on an appropriate level of service for the facility. It must be identified optimal service levels to reach the minimum cost (Taha, 1995).

Cost models help to balance two conflicting costs that are:

1. Cost of offering the service
2. Cost of delay in offering the service (customer waiting time)

The two costs conflict because an increase in one causes reduction in the other and vice versa as one can see from the Figure 1.

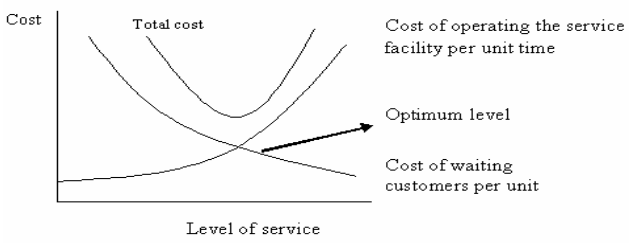

Figure 1. The costs versus level of service

Let $x=c$ represents the service level, the cost model can be expressed as

$$E(C_T(x)) = E(C_O(x)) + E(C_W(x)) \tag{19}$$

where

$E(C_T(x))$ = Expected total cost per unit time
$E(C_O(x))$ = Expected cost of operating the facility per unit time
$E(C_W(x))$ = Expected cost of waiting per unit time

At the same framework of this section, the cost models can be generated in different queueing systems. (Taha, 1995)

5.1. Cost Model of the Queueing System with Impatient Customers

Queues eventually causes to the costs (Mandelson, 1985). The basic task of the service management is to offer the best quality service to the customers under the minimum cost. Then, in litterature, there are studies in order to find optimal number of servers in taking into account all parameters that affect the cost. One of these parameters can be patience time's distribution for the systems with impatient customers. Its cost models now help to balance two conflicting costs that are

1. Cost of offering the service
2. Cost of loss customers because of their impatience.

Thus the equation (19) for the queueing systems with impatient customer is rewritten as follows:

$$E(C_T(x)) = E(C_O(x)) + E(C_L(x)) \tag{20}$$

where $E(C_L(x))$ is the expected cost of loss customers.

If we consider M/M/c+M model, in the paper of Boxma and Waal (1993), the following cost equation is defined. According to this paper, the operational costs for a server in the queue are d_1 per unit time. For any impatient customer that abandons the queue, an external server is hired, for only the service of that customer. The cost per unit time for such server is denoted as d_2. If the internal queue has *c* servers, the average cost $g(c)$ is given by

$$g(c) = d_1 c + \lambda \beta^* d_2 \pi(c) \tag{21}$$

where β^* is the mean service time of the external server, $\pi(c)$ is the loss probability as a function of *c*.

The importance of the patience time's distribution for the cost function is stressed on the following section with the real data taken from Pehlivanoğlu Market.

6. Application

6.1. Data

Since it is available to observe queue dynamic, Pehlivanoğlu Market is chosen for the application.

Pehlivanoğlu Market in Şirinyer was opened on 7th March 1990. Its service time is between 08:30 and 21:00. Application was applied in this market in November 2004. There were two activated servers in this market. According to our observation, the number of customers that arrived to system was 25 persons and it was 12 customers who were served in half an hour.

Mean patience time of customers which have to wait for receiving service from system was classified in increase order as follows; 1, 5, 15, 30 minutes.

After interviewing with executive manager of Şirinyer Pehlivanoğlu, the cost for each server was defined as 168 € for month which was just about the cashier's minimal salary. On the other hand, in the case that to open extra server for only impatient customers for not losing these customers, extra server will be able to serve just for 4 customers in half an hour. The cost of this new server

can be more than those of two servers, because the salary of the substitude cashier is 195 € for month.

The cost function value is found versus the number of servers under different patience times for the two queue models M/M/c+D and M/M/c+M. And both M/M/c+M and M/M/c+M queue models are compared with each other.

6.2. Cost Values of the M/M/c+D Queue Model

According to the results of the application based on these given data, when the M/M/c+D queue model is applied for the impatient customers in the system, the Figure 2 indicates that if there patient customers in system who are willing to wait half an hour, the number of loss customer is inverse proportional with the number of servers and as a result, the probability of loss customers converges to zero. In this case, the cost value increases related to increased number of servers. On the other hand, when the new server opens for the impatient customers in the system, in other words, the number of servers becomes 3 it is noticed that the cost decreases and its value is just about 660 €.

The optimal number of servers should be specified as 3. If the number of servers is not increased to 3, in this case, since the loss is too much, the cost is almost 730 €. One can see from that from the Figure 2.

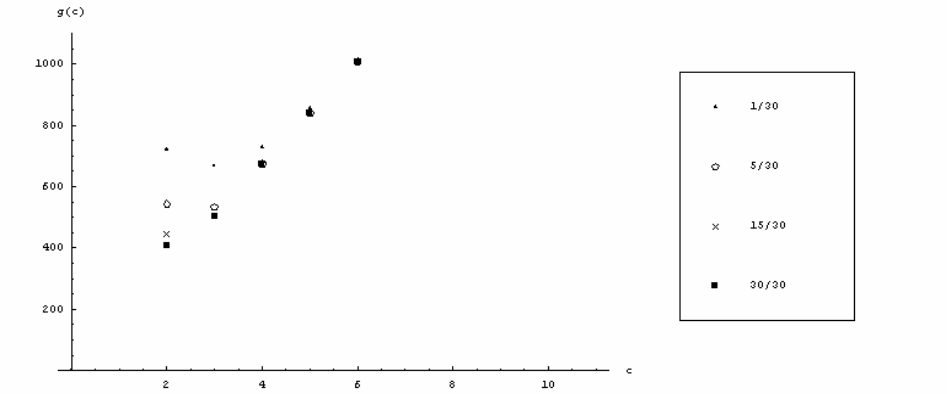

Figure 2. Cost values for different patience times in M/M/c+D queue model

The decreasing 70 € in the cost is a positive progress for the business. Then it is recommended managers to increase the number of servers up to 3.

6.3. Cost Values of the M/M/c+M Queue Model

When these data is applied to the other queue model M/M/c+M, it can be seen from the Figure 3 that the mean patience times do not have any effects in identifying the optimal number of servers and the cost value increases with the increase of the number of servers.

The cost of impatient customer that brings on the Pehlivanoğlu Market is more than the patient customers's cost. It can be agreed with the Figure 3 that the cost of a customer who is willing to wait for the service for 1 minute is 740 €. On the other hand, the cost of a customer who is willing to wait for 30 minutes in queue to receive service is about 500 €. When the number of servers in system is less than 6, the cost corresponding to each of mean patience times varies. However, when the number of servers is 6 or exceeds 6, the number of loss customer decreases. Therefore, the probability of loss customers converges to zero. And also it is considered that the difference of the patience times of the customers is important on the cost value. The reason of this case can be theoritically explained as follows: In the M/M/c+M queue model, the loss probability $\pi(c)$ is a decreasing function of c and this goes to 0 if c goes to infinity. Thus, it may be concluded that the time average costs are asymptotically equal to $d_1 c$ as c goes to infinity. In this application, after the point 6, there is no effect of loss probability on the cost function. Accordingly, the cost function becomes $g(c) = d_1 c$.

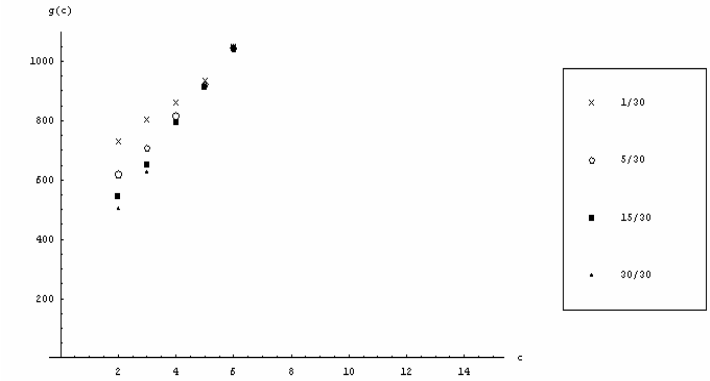

Figure 3. Cost values for different patience times in M/M/c+M queue model

6.4. Cost Values of the M/M/c+D and M/M/c+M Queue Models

In interview with customers, it is determined that they are willing to wait only 15 minutes. In this way, the other side of the study, when the patience time is about 15 minutes, it is investigated the cost values of both M/M/c+D and M/M/c+M queue models in the period of the half an hour.

As a result, when the number of servers stays the same as 2, the cost of M/M/c+M queue model is about 549 €, the cost of M/M/c+D queue model is about 449 €. It can be observed that if the number of server is increased, the cost value increases for the both queue models. When the number of servers increases up to 3, the cost value also increased significantly compared to its previous value and its cost is 650 € in the M/M/c+M queue model.

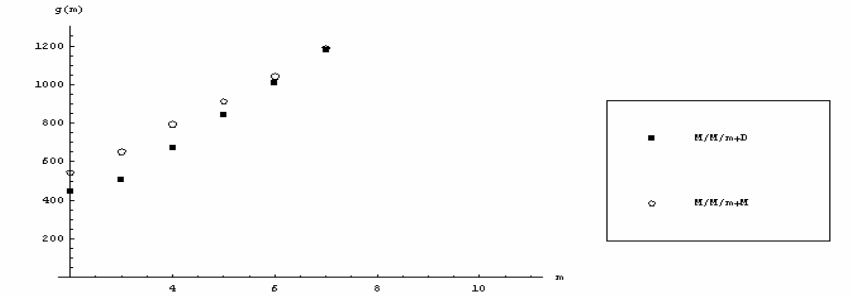

Figure 4. Cost values for both M/M/c+D and M/M/c+M queue models

7. Conclusion

The aim of this paper is to prove the importance of the patience times and its distribution function in order to identify the cost value corresponding to the optimal number of servers.

The probabilistic model of the cost function is applied with the real data taken in Pehlivanoğlu Market by using MATHEMATICA 5.0.

The exponentially distributed patience times has no effect on identifying the optimal number of servers inasmuch as the cost value increase with the increased number of servers as in Figure 3. However if the distribution of the patience times is deterministic, the optimal number of servers corresponding to the minimum cost value can be identified for the impatient customers. Thus, the cost can be shown in Figure 2.

As in Figure 2 and 3, the general result for these two models is that impatient customers cause more cost than the patient customers.

As in Figure 4, the cost value of M/M/c+D queue model is less than M/M/c+M queue model where the loss probability is the distinctive parameters between the two models. In addition, after a definite number of servers, there is no effect of loss probability on the cost function for both queue models M/M/c+D and M/M/c+M.

Clearly, there is no need to open a new server if patient customers take place in the system. However, according to our application results, the advantageous decision for the Pehlivanoğlu Market is to open a new server for the impatient customers. The cost is minimum when the number of servers is 3; thereby this value 3 is optimal number of servers. However, the executive managers of the Pehlivanoğlu Market will make the last decision.

References

Bacelli, F. & Hebuterne, G., (1981). On queues with impatient customers. Performance'81, 94, 159-179.

Barrer, D.Y., (1957). Queueing with impatient customers and ordered service. Operations Research, 5, 650-656.

Boots, N.K. & Tijms, H., (1999). A multiserver queueing system with impatient customers. Management Science, 45, 444-448.

Boxma, O.J. & De Waal, P.R., (1993). Multiserver queues with impatient customers. Centrum Voor Wiskunde en Informatica, BS-R9319, 1-20.

Mendelson, H., (1985). Pricing Computer Services: Queueing Effects. Communication of ACM. 28, 312-321.

Palm, C. (1937). Etude des delais d'attende. Ericsson Technics, 5, 37-56.

Taha, H.A., (1995). Operations Research an Introduction(6^{th} ed.). Prentice Hall.

Tijms, H.C., (2003). A First Course in Stochastic Models. Great Britain. Wiley.

Zohar, E., Mandelbaum, A., Shimkin, N., (2002). Adaptive behavior of impatient customers in tele-queues: Theory and emprical Support. Management Science, 48, 566-583.