European Journal of Operational Research 239 (2014) 456–468

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Equilibrium arrival times to a queue with order penalties

Liron Ravner *

Department of Statistics and the Center for the Study of Rationality, The Hebrew University of Jerusalem, Israel

ARTICLE INFO

Article history:
Received 30 August 2013
Accepted 4 June 2014
Available online 17 June 2014

Keywords:
Queueing
Strategic arrival times to a queue
Queueing games

ABSTRACT

Suppose customers need to choose when to arrive to a congested queue with some desired service at the end, provided by a single server that operates only during a certain time interval. We study a model where the customers incur not only congestion (waiting) costs but also penalties for their index of arrival. Arriving before other customers is desirable when the value of service decreases with every admitted customer. This may be the case for example when arriving at a concert or a bus with unmarked seats or going to lunch in a busy cafeteria. We provide game theoretic analysis of such queueing systems with a given number of customers, specifically we characterize the arrival process which constitutes a symmetric Nash equilibrium.

1. Introduction

When customers are faced with the decision of when to arrive to a queueing system with some desired service at the end, the first issue to consider is avoiding congestion. This disutility is typically modelled as a waiting time cost. Such a model was first considered by Glazer and Hassin (1983). They assumed that the number of customers arriving to the queue is a Poisson random variable, which as it turns out makes the analysis easier than the deterministic case, or any other distribution. However, customers may also be interested in being served at an early time. Such an example is driving home from work, where commuters wish to avoid traffic but are not willing to stay at work until midnight in order to achieve this. This type of disutility has been modelled as a tardiness cost that increases the later one is admitted into service. Some recent research has been carried out on this model by Haviv (2013) and by Juneja and Shimkin (2013). The first considered a Poisson number of customers and studied the equilibrium properties when limiting the allowed arrival period. The latter considered a general number of customers and focused on a rigorous characterization of the Nash equilibrium, and the proof of convergence to a fluid limit. Both also presented fluid approximation models which are technically less cumbersome, and often provide insight on the discrete stochastic case.

In many queueing scenarios customers are not actually worried about tardiness, but rather about the number of customers who arrived ahead of them. This is the case in a concert or flight with unmarked seats, when there is no actual penalty for tardiness unless other customers have arrived and taken hold of the better seats. Equivalently, one can consider situations where the actual value of the service deteriorates with every service completion, for example a repair machine which depreciates with every use. Other examples can be found in Myrick Freeman and Haveman (1977) who find an optimal toll for the use of a facility in which the demand is a decreasing function of the aggregate level of facility use, and in Cicchetti and Kerry Smith (1973) who model the demand for wilderness recreation using a congestion model that users suffer disutility because of recreation of other users in the same area, and also provide empirical support for the model. This brings us to the focus of this work, which is to present a model where one's cost is not necessarily time based, but rather dependent on the number of prior arrivals. If this is the only disutility assumed in the model, then obviously all customers arrive as early as possible. However, if there are also waiting costs, customers may improve their utility by not arriving in close proximity to others, which leads to a more interesting analysis of their strategic behaviour.

In the remainder of this section we introduce the model and review some related literature. Our analysis commences in Section 2 by illustrating an example of a two customer game, and comparing it to the known results for the tardiness model. We show that the support of the symmetric equilibrium arrival distribution is infinite if there is no closing time for the server, as opposed to the finite support obtained in the tardiness model. In particular, the equilibrium distribution is uniform prior to the opening time, and exponential after it. We further show how the equilibrium is adjusted if early birds are not allowed, and when the server has a closing time. All solutions for the two customer game are explicitly derived. In Section 3 we consider a general model with any number of customers. The explicit solution is not

* Corresponding author. Address: Department of Statistics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel. Tel.: +972 25883790.

E-mail address: liron.ravner@mail.huji.ac.il

http://dx.doi.org/10.1016/j.ejor.2014.06.005
0377-2217/© 2014 Elsevier B.V. All rights reserved.

tractable for the general model, but we characterize the equilibrium properties and dynamics, accompanied by a numerical technique to compute it. We also provide an analysis of the tail behaviour of the arrival distribution, and prove that it is exponential. Numerical analysis also suggests that the tail of the arrival hazard rate equals exactly to the exponential rate found in the two customer case, regardless of the population size. We then examine how an immediate generalization can be made to a model with both order and tardiness costs. We further provide bounds on the cost incurred by any single customer in equilibrium, in the general setting. In Section 4 we present the symmetric equilibrium for a random number of customers, which follows a Poisson distribution. In Section 5 we briefly discuss the social optimization problem and explain how the existing literature relates to the model we have presented. Finally, in Section 6 we summarize the results and discuss possible extensions and future work.

Remark. Much of the "essence" of the model is captured in the two-customer analysis of Section 2. Section 3 is a more technical generalization. The main results are stated in Theorems 3.1, 3.2 and 3.3, and numerical examples for the general model are presented in subSection 3.4.

1.1. Model

Suppose that $N+1$ customers wish to obtain service. We assume that a single server provides the service according to a *First Come First Served* regime, and that service times are independent and exponentially distributed with rate μ. If multiple customers arrive at exactly the same time, then they are admitted in a uniformly randomized order. The customers incur a delay cost of α per unit of time, a tardiness cost of β per unit of time until their admittance into service and an index of arrival cost γ for every customer that has arrived before them. We denote the closing time by $T>0$, where $T=\infty$ means there is no closing time. For most of this work we assume that $\beta=0$, and analyse the model with only waiting and order costs. Where possible, we also consider $\beta>0$ for the sake of comparison and generalization.

In Juneja and Shimkin (2013), the equilibrium for a general N customer was characterized under the assumption that customers are limited to arrival distributions F such that: *"For each F, the corresponding support can locally (i.e., on any finite interval) be represented as a finite union of closed intervals and points"*. They proved that under this assumption, the equilibrium arrival profile is unique and symmetric. We focus here only on distributions that satisfy this assumption.

The symmetric equilibrium mixed strategy is defined by a cdf denoted by $F(t)$ for all $t\in\mathbb{R}$. We also denote the density function $f(t)=F'(t)$ for all t such that $F(t)$ is continuous and differentiable. We seek a cdf F such that if the other N customers arrive according to F, then the last customer is indifferent between arriving at all points of the support of F, and does not prefer any point outside of the support. The expected cost of arriving at time $t\in\mathbb{R}$ is:

$$c_F(t) = -\alpha t\mathbb{1}_{\{t<0\}} + \frac{\alpha+\beta}{\mu}\mathbb{E}Q_F(t) + \beta t\mathbb{1}_{\{t\geqslant 0\}} + \gamma\mathbb{E}A_F(t), \qquad (1.1)$$

where $Q_F(t)$ and $A_F(t)$ are the queue size and the arrival process at time t, respectively, when N customers are arriving independently according to F. The value of the arrival process at time t, $A_F(t)$, is in fact the index of arrival of the last customer to arrive up until time t. In the following sections we will simply denote these processes by $Q(t)$ and $A(t)$, although their distribution is always determined by F. Note that $\mathbb{E}A_F(t)=NF(t)$, but the expected queue size depends on both arrivals and departures, and typically does not have an explicit form as a function of F.

Remark. The majority of our analysis assumes that the size of N is common knowledge. It is important to note however, that all results may be generalized to any prior distribution on N in a fairly straightforward manner. The special case of the Poisson distribution has simplifying properties which we shall elaborate on in Section 4.

1.2. Preliminary analysis of the index cost model

Suppose $\beta=0$ and $\alpha,\gamma>0$, i.e. customers only incur waiting and index costs. This special case of the model has several unique equilibrium properties which will be used throughout our analysis. We state these properties in the following two lemmata and their subsequent corollary.

The cost function in (1.1) can now be rewritten:

$$c(t) = -\alpha t\mathbb{1}_{\{t<0\}} + \frac{\alpha}{\mu}\mathbb{E}Q(t) + \gamma\mathbb{E}A(t). \qquad (1.2)$$

Lemma 1.1. *There exists no symmetric equilibrium arrival profile such that for some finite time t_b, all customers have arrived with probability one; $F(t_b)=1$. Furthermore, the expected cost in equilibrium is at most γ.*

Proof. We assume that there exists such an equilibrium arrival profile, and show that this leads to a contradiction. Any customer can achieve the cost $\gamma+\epsilon$ for any $\epsilon>0$ by arriving at a very large $t>t_b$. This is because the probability that the server is still busy approaches zero when $t\to\infty$. Therefore, the expected cost, denoted by c_e, in this equilibrium is at most γ. This cost is constant on all of the support, specifically at time t_b:

$$c_e = c(t_b) = \frac{\alpha}{\mu}\mathbb{E}Q(t) + \gamma > \gamma, \qquad (1.3)$$

thus contradicting the previous argument, that $c_e\leqslant\gamma$. □

Lemma 1.2. *There can be no holes in a symmetric equilibrium arrival profile. In other words, there exists no time t such that $F(t)>F(t-)$, where $F(t-)=\lim_{s\uparrow t}F(s)$ is the limit from the left of the cdf at point t (the point of upward discontinuity).*

Proof. Assume there exists a time t such that $F(t)>F(t-)$. The left limit from the left of the cost function (1.2) at t is:

$$c(t-) = -\alpha t\mathbb{1}_{\{t<0\}} + \frac{\alpha}{\mu}\mathbb{E}Q(t-) + \gamma F(t-).$$

The expected queue size can only have upward jumps, i.e. $\mathbb{E}Q(s-)\leqslant\mathbb{E}Q(s)$ for any time s (see for example Lemma 2 in Juneja & Shimkin (2013)). Therefore, we can conclude that $c(t-)<c(t)$, which contradicts the equilibrium assumption. □

Corollary 1.3. *The support of the equilibrium distribution F can be represented as an interval $[t_a,\infty)$, for some finite and negative t_a.*

Note that $t_a>-\infty$ because $\lim_{t\to-\infty}c(t)=\infty$ for any F, and in Lemma 1.1 we established that the equilibrium cost is at most γ.

1.3. Related literature

In 1969 Vickrey published his seminal paper *"Congestion Theory and Transport Investment"* (Vickrey, 1969), which presented a fluid model for congestion dynamics. In particular, equilibrium arrival dynamics where characterized for a bottleneck model. This model was studied and developed in various directions in the following