# Equilibrium joining probabilities for an M/G/1 queue

Yoav Kerner[*]

## Abstract

We study the customers' Nash equilibrium behavior in a single server observable queue with a Poisson arrival process and general service times. Each customer takes one decision: to join or not to join. Furthermore, he takes it upon arrival and no future regrets. The customers are homogenous with respect to their waiting cost linear function and with the reward associated with service completion. The cost of joining depends of others' behavior, and therefore a strategic game is formed here. A full recursive algorithm for computing the (possibly mixed) Nash equilibrium strategy is presented. Its output is queue dependent joining probabilities. We demonstrate that depending on the service distribution, this equilibrium is sometimes unique while at other times it is not. Also, the 'follow the crowd' phenomenon holds in times, as is the case regarding the 'avoid the crowd' phenomenon.

## 1 Introduction

Consider an M/G/1 first-come first-served model. By that we mean that there exists a single server who serves customers at the order of their arrival. Service times follow an arbitrary distribution function $G$ and the arrival process is Poisson with rate denoted by $\lambda$. Upon arriving to the system and observing the number of customers there, the arrival assesses his waiting

[*]EURANDOM, Eindhoven University of Technology, Eindhoven, The Netherlands. E-mail: kerner@eurandom.tue.nl .

time. This is in aim to decide whether to join the queue or to leave for good. The trade off is between the reward associated with receiving service (assumed here homogenous across customers and is denoted by $V$) and the expected waiting cost (assumed here linear with time and with a homogenous across customers slope denoted by $C$). This cost is applicable also while in service. Once a customer joins, he may not leave. Note that in some applications the latter is endogenous. For example, in a case where a significantly high entrance fee is payed upon joining.

The waiting time of one who joins is composed of two parts which by no means are independent. The first part is the remaining service time of the one found to receive service upon the arrival of our tagged customer. The second part is due to future service times of those he finds in queue upon his arrival. The contribution of the second component to the mean waiting time is straightforward: If $n-1$ such customers are found in the queue, then their contribution to the mean waiting time is $(n-1)\bar{x}$ where $\bar{x}$ denotes the mean service time. Assessing the mean remaining service time of the one currently in service is more complicated. The reason behind that is two-fold. First, as said, this value if a function of $n$, $n \geq 0$, the total number of customers found in the system upon arrival. So even in the supposedly simple case where all join, leaving the analysis as in a standard M/G/1 model, dealing with the remaining service time is not trivial. Yet, this analysis appears in the literature. See [8],[3],[7] or [1]. Second, our model gets another twist. Since customers can balk after observing the queue length, the remaining service time is, at least in principle, a function of the behavior of all those who arrive earlier, including those who balked.

A pure strategy in our model is a prescription which says for which queue lengths to join and for which to balk. Allowing randomization (mixing), a typical strategy is a set of probabilities $p_n$, $n \geq 0$, such that a customer who faces $n$ customers upon arrival, joins with probability $p_n$ and balks for good with probability $1 - p_n$. Without loss of generality, we assume a reward of zero in the latter case. As we will see below, the use of a mixed strategy is quite natural here. Our goal is to develop a *symmetric Nash equilibrium* strategy profile. By that we mean a strategy which if followed by all, an individual's best response, is to follow it too. From now when we refer to equilibrium, we mean a symmetric Nash equilibrium. In order to avoid trivialities we assume that $V > C\bar{x}$, implying that $p_0 = 1$. Also, for $n$ large enough so that $V < Cn\bar{x}$, $p_n = 0$.

2

We next explain, by the use of an example borrowed from [2], why one's assessment on the remaining service time depends on the behavior of other customers. Suppose service times follow a Bernoulli distribution, namely they are either zero or one. Suppose a tagged customer arrives to an empty queue but observes a customer receiving service. In case of a large value of $p_1$ used by others, the tagged customer can assess, from the fact that the queue is empty, that his arrival took place at an earlier stage of the current service (and hence much is left) in comparison with the case where they all used a small value for $p_1$. Thus, the higher $p_1$ is, the less appealing it is to the tagged customer to join. This phenomenon is coined 'avoid the crowd'. See more on it in [5]. Thus, it is possible (depending on $\lambda$, $C$ and $V$), that if all use $p_1 = 0$ ($p_1 = 1$, respectively), then one's best response is to join, i.e. $p_1 = 1$ (balk, i.e. $p_1 = 0$, respectively). Hence, under equilibrium, $p_1$ is a fraction between zero or one, i.e. one would mixed between joining and balking. For some other service distributions the opposite is the case: The higher $p_1$ is, the more appealing it is for an individual to join. This is the 'follow the crowd' phenomenon. In this case (again, depending on the model's parameters), when all use $p_1 = 1$ ($p_1 = 0$, respectively) one should join (balk, respectively) too. This situation leads to a multiple equilibria. We later give an example for such a case.

Above we gave an example under which a $p_1$ which is strictly between zero and one, is part of an equilibrium profile. The same is the possibility regarding $p_n$ for any $n \geq 1$. Thus, the first thing needed here towards the construction of an equilibrium, is the computation of the mean remaining service times for any queue length under any mixed strategy profile. This was achieved in [7]. Specifically, let $\bar{r}_n$, be the mean remaining service time given $n$ customers in the system, $n \geq 1$, (applicable, by PASTA, at arrival times too). Then, $\bar{r}_n$ is a function of $P = (p_1, p_2, ....)$, the probabilities used by others. As it turns out, $\bar{r}_n$ is a function only of $p_1, \ldots, p_n$. An explanation for the latter is as follows. First, the only information about the past is reflected into the distribution of the remaining service time is information about events and actions taken by other customers since the beginning of current service. Second, since obviously there were no departure since the current service started, there were never more than $n$ customers in the system since then. Thus, the behavior of those who find more than $n$ upon arrival is irrelevant to the distribution the remaining service time given that there are $n$ customers in the system. In [7], a recursive expression for $\bar{r}_n$ is

given. See (3) and (4) below. Thus, the equilibrium probabilities $p_n$, $n \geq 1$, will too be derived recursively. Yet, as $\bar{r}_n$ is a function also of $p_n$, computing $p_n$, given $p_i$, $1 \leq i \leq n - 1$, requires a fixed point analysis. In particular, multiple equilibria are a possibility. Details on how this is done are given in Section 2.

The next question we address is if we can characterize some of the cases where the equilibrium is unique and others where it is not. Details are given in Section 3. Here we highlight our main findings. Under some distributions of service times, the mean remaining service time is decreasing with the age of that service (DMRL, for decreasing mean residual life, an acronym used mainly in reliability models). Others are IMRL (where 'I' stands for 'increasing'), while the rest are of course none of the above. We show in Section 3 that in the DMRL case there exists a unique equilibrium strategy (which can prescribed mixing for some queue lengths). The intuition behind this uniqueness is exemplified upon in the case of Bernoulli distribution for service. Indeed, in the DMRL case, the smaller is the joining probability used by all, the higher is one's assessment regarding the age of service he finds upon arrival, which in turn implies a smaller remaining service time, making him more inclined to join. This is a typical situation in which equilibrium is unique. Moreover, as we show via an example, that although counter-intuitive, it is possible that under equilibrium $p_n < p_{n+1}$ for some values of $n$. An explanation is as follows. True, one more $\bar{x}$ is added to the mean waiting time when comparing a system with $n + 1$ customers with that of $n$ but the information that one more has joined (let alone arrived), may change upwards the prior regarding the age of service. This in turn, in the DMRL case, reduces the prior regarding the remaining service time of the one currently in service, making joining at the case of $n + 1$ more appealing.

The opposite is the case under IMRL distributions. Here too, a high joining probability implies that the arrival takes place at an early stages of service. But now it means a lower remaining service time, making joining appealing. In other words, the more who join the queue, the more it is appealing for an individual to do so. This is the follow the crowd phenomenon. A unique equilibrium is not ruled out here but, as we show in Section 3, it is possible to have three equilibria. Two among them are pure, namely they subscribe joining or balking while the third is mixed. We conclude with some examples in Section 4.

We end this introduction with a short literature review. The decision prob-

lem of joining or not a queue after observing its queue length goes back to [9] who solved the M/M/1 case. There due the memoryless property of the service distribution, customers do not interact via assessing the remaining service time, making the equilibrium policy simple: Join if and and only if the number in the system is smaller than $n_e$, where $n_e = \lfloor V/(C\bar{x}) \rfloor$. In [2] an observable $M/G/1$ queue was presented where service times follow a Bernoulli distribution and customers decide whether or not to join the queue after observing its length. There it is shown that for selected values for the $V$ and $C$ parameters, and for a sufficiently large arrival rate, there exists an equilibrium strategy profile of the "delay threshold policy" type: Join with some probability $p$, $0 < p < 1$, when the queue is empty (i.e. there is only one customer in the system, and he is being served), join with probability one when the queue length is below a certain threshold, and balk when it is greater than this threshold. We consider the same model but remove the assumption of Bernoulli distribution of service times. A related model is dealt with in [6]. There, only partial information is assumed: in case that the server is busy, the arrival is informed if the queue is empty or if at least one is standing there. Thus, a strategy is stated by two values $p$ and $q$, where $p$ ($q$, respectively) is the probability of joining in case of an empty (not empty, respectively) queue. All equilibria were derived there. Indeed, under equilibrium, $p$ and our $p_1$ coincide (and it comes with all the featured described above). Interestingly, for any equilibrium $p$, there exists a unique equilibrium $q$ as the 'avoid the crowd' phenomenon prevails under this information regardless of the service distribution.

## 2  The model and main results

We consider a first-come first-served $M/G/1$ queue. An arriving customer observes the queue length and decides whether or not to join it (mixing is allowed). After taking the decision upon arrival, there are no further regrets. That is, if a customer balks, he never returns and if he joins, he can not leave before his service completion. We assume that the customers are homogenous in the sense that each one of them has the same value for receiving service and the same waiting cost parameters. Some notation is introduced: $\lambda$ is the arrival rate. $G$ is the service time distribution function. $\bar{x}$ and $\overline{x^2}$ are the first and the second moment of the service time, respectively. To avoid trivialities,

we assume that $C\bar{x} < V$. $G^*(s) = \int_{x=0}^{\infty} e^{-sx} dG(x)$ is the Laplace-Stieltjes transform (LST) associated with the service time distribution. As mentioned above, the decision variables in our model are the joining probabilities. Let $p_n$ be the probability that an arriving customer who observed $n$ customers in the system upon his arrival joins the queue, $n \geq 0$. For shortening, we denote $P_n = (p_1, \ldots, p_n)$ and $P = P_\infty$. The random variables defined below and their expected values are functions of these decision variables under steady-state conditions. $L$ and $L_a$ are the number of customers in the system at arbitrary instants and at arrival instants, respectively. $R$ is the remaining service time of the customer in service. $\bar{r}_n(P) = \mathrm{E}_P(R|L_a = n)$, under joining probabilities $P$. It was shown in [7] that the LST of the conditional residual service obeys the following recursive formulas:

$$F_n^*(s) = \frac{\lambda p_n}{s - \lambda p_n} \left( G^*(\lambda p_n) \frac{1 - F_{n-1}^*(s)}{1 - F_{n-1}^*(\lambda p_n)} - G^*(s) \right) \quad , \quad n \geq 2 \qquad (1)$$

with the initial condition

$$F_1^*(s) = \frac{\lambda p_1}{\lambda p_1 - s} \frac{G^*(s) - G^*(\lambda p_1)}{1 - G^*(\lambda p_1)} \qquad (2)$$

where $F_n^*(s) = \mathrm{E}(e^{-sR}|L_a = n)$, $n \geq 1$. Likewise,

$$\bar{r}_n(P) = \frac{G^*(\lambda p_n)}{1 - F_{n-1}^*(\lambda p_n)} \bar{r}_{n-1}(P) - \frac{1}{\lambda p_n} + \bar{x} \quad , \quad n \geq 2 \qquad (3)$$

with the initial value

$$\bar{r}_1(P) = \frac{\bar{x}}{1 - G^*(\lambda p_1)} - \frac{1}{\lambda p_1} \quad . \qquad (4)$$

**Remark 2.1** We note that $F_n^*(s)$ is a function of $p_1, ..., p_n$, as it is a function of $p_n$ and of $F_{n-1}^*(\cdot)$, while $F_1(s)$ is a function of $p_1$. Also, $\bar{r}_n(P) = \bar{r}_n(P_n)$.

When an arrival faces $n$ customers in the system, given a set of joining probabilities $P$, his expected waiting time (which includes his own service time) is $n\bar{x} + \bar{r}_n(P)$. In this model, a symmetric Nash equilibrium joining probabilities is a set $P^e = (p_1^e, p_2^e, \ldots)$ such that, if all join with these probabilities,

then one does not have a better response than following these probabilities oneself. Hence, $P^e$ satisfies:

$$p_n^e \in \arg \max_{0 \le p \le 1} p \left( V - C \left( r_n(P^e) + n\bar{x} \right) \right) \quad , \quad n \ge 1 \ . \qquad (5)$$

Following Remark 2.1, we learn that the equations in (5) can, at least in principle, be solved recursively, initiating with $n = 1$. Specifically, note that the objective function given in (5) is linear in the decision variable $p$. Therefore, the maximizer is:

$$\begin{cases} 1 & V > C \left( \bar{r}_n(P_{n-1}^e, 1) + n\bar{x} \right) \\[2ex] p & V = C \left( \bar{r}_n(P_n^e) + n\bar{x} \right) \qquad , \quad \text{every } p \in [0, 1] \\[2ex] 0 & V < C \left( \bar{r}_n(P_{n-1}^e, 0) + n\bar{x} \right) \end{cases}$$

where $p$ here means any $p \in [0, 1]$, since if $V = C(\bar{r}_n(P_n) + n\bar{x})$, then the individual's utility is zero regardless whether he joins or not, and hence he is indifferent between joining and balking, or in fact, any mixing between the two. It is clear that for $n$ large enough, for example, $n \ge \frac{V}{C\bar{x}}$, $p_n^e = 0$.

The next question is how to compute $P^e$, and in particular, how to find the initial value of $p_1^e$. We deal with this issue next. The following proposition appears in [6].

**Proposition 2.1** *Assume $L_a = 1$. Then, at least one of the following cases occurs:*

**Case 1.** *If*

$$\bar{x} + \max_{0 \le p \le 1} \left\{ \frac{\bar{x}}{1 - G^*(\lambda p)} - \frac{1}{\lambda p} \right\} \le \frac{V}{C}$$

*then joining is a dominant strategy.*[1]

---

[1] A strategy is said to be a dominant strategy, if it is one's best response for any strategy selected by all others. Of course, if there exists a dominant strategy, no other equilibrium exists.

**Case 2.** *If*

$$\bar{x} + \min_{0 \le p \le 1} \left\{ \frac{\bar{x}}{1 - G^*(\lambda p)} - \frac{1}{\lambda p} \right\} \ge \frac{V}{C}$$

*then not joining is a dominant strategy.*

**Case 3.** *If a dominant strategy does not exist, then at least one of the following holds:*

**3a.** *If* $\bar{x} + \frac{\bar{x}}{1 - G^*(\lambda)} - \frac{1}{\lambda} \ge \frac{V}{C}$ *then joining with probability one is an equilibrium.*

**3b.** *If* $\bar{x} + \frac{\overline{x^2}}{2\bar{x}} \le \frac{V}{C}$ *then joining with probability zero is an equilibrium.* [2]

**3c.** *There exists a (not necessarily unique) mixed Nash equilibrium* $p_1^e$, $0 < p_1^e < 1$, *where* $\bar{x} + \frac{\bar{x}}{1 - G^*(\lambda p_1^e)} - \frac{1}{\lambda p_1^e} = \frac{V}{C}$. *Moreover, any* $p_1^e$ *which solves this equation is an equilibrium.*

We would like to point out that in proposition 2.1, Case 1 implies Case 3a, that and Case 2 implies Case 3b, while ruling out the other cases.

**Proof.** If Case 1 (Case 2, respectively) holds, than the individual reward from joining is positive (negative, respectively) regardless of others' behavior. Hence, it is a dominant strategy for him/her to join (not to join, respectively). If the inequality in Case 3a (3b, respectively) holds, and all join (do not join, respectively), then the individual's net gain from joining is positive (negative, respectively). Therefore, one's best response is to join (not to join, respectively) and hence $p_1^e = 1$ ($p_1^e = 0$, respectively). If there exists $p_1^e$ as defined in Case 3c, and if all those who find one customer in the system upon their arrivals join with probability $p_1^e$, then the individual's net gain from joining is zero. Hence, this individual has no better response against $p_1^e$ than itelsf. Therefore, $p_1^e$ is a Nash equilibrium. Finally, as $\bar{r}_1(p)$ is continuous function in $p$, if neither Case 1 nor Case 2 holds, then Case 3c must hold for some (not necessarily unique) $p_1^e$, $0 < p_1^e < 1$.[3] $\square$

---

[2]If no one joins when $L_a = 1$ then an empty queue does not give any information about the residual service time. Also, $\lim_{p \to 0} \frac{\bar{x}}{1 - G^*(\lambda p)} - \frac{1}{\lambda p} = \frac{\overline{x^2}}{2\bar{x}}$ (a fact that can be shown analytically by applying L'Hospital rule on (4)).

[3]This of course does not rule out the possibility, for example, that all cases 3a,3b and 3c hold at the same time, leading to a multiple equilibria.

**Remark 2.2** As we demonstrate later, the existence of mixed Nash equilibrium strategies, does not rule out the possibility of the existence of pure Nash equilibrium strategies.

Once $P_{n-1}^e$, $n > 1$, is in hand (no matter if it is unique or not), the value(s) of $p_n^e$ can be determined. Details are given next.

**Proposition 2.2** *For any equilibrium probabilities $P_{n-1}^e$, the (not necessarily unique) equilibrium joining probability $p_n^e$ is determined by the not necessarily mutually exclusive (but comprehensive) cases:*

**Case 1.** *If $n\bar{x} + \bar{r}_n(P_{n-1}^e, 1) \leq \frac{V}{C}$, then $p_n^e = 1$ is an equilibrium.*

**Case 2.** *If $n\bar{x} + \bar{r}_n(P_{n-1}^e, 0) \geq \frac{V}{C}$, then $p_n^e = 0$ is an equilibrium.*

**Case 3.** *For any $p$, $0 < p < 1$, such that $n\bar{x} + \bar{r}_n(P_{n-1}^e, p) = \frac{V}{C}$, $p_n^e = p$ is an equilibrium.*

**Proof.** The same as the proof for Proposition 2.1. □

We would like to point out that for a profile $P$ in which $p_n = 0$ for some $n \geq 1$, we don't analyze the Nash equilibrium behavior for $L_a = m$ for any $m > n$. This is because of if $p_n = 0$, then the number of customers never exceeds $n$. Thus, the problem of finding $p_m$ for $n > m$ is not well-defined.

# 3 Special cases and the uniqueness issue

**Definition 3.1** A nonnegative random variable is said to be with *decreasing (increasing, respectively) mean residual life* (DMRL (IMRL, respectively)) if $\mathrm{E}(X - t | X > t)$ is monotone decreasing (increasing, respectively) with $t$.

The DMRL property looks quite natural, as it says that the longer is the past life time, the shorter is the expected future life time. Known examples for distributions with the DMRL property are the uniform distribution, and the Erlang distribution. The IMRL property looks counter intuitive, as it says that the longer is the past life time, the longer is the expected future life time. Mixtures of exponential distributions, as in Example 2 in the next section, are good examples for an IMRL distribution. In particular, assume that the service time is exponential with a small mean value with some probability

and exponential with a relatively large mean value with the complementary probability. In this case, the larger is the past service time, the larger is the posterior probability that the service time is has the large mean value. This implies that the mean residual service time is increasing with the past service time. Another known example for IMRL distribution is the Pareto distribution, in which $P(X > t) = (\beta/\beta + t)^a$, for some $a > 1, \beta > 0$. In this distribution, $E(X - t|X > t) = (\beta + t)/(a - 1)$ which is increasing with $t$. The exponential distribution, in which the past and future are independent, is both DMRL and IMRL. Moreover, it is the only one which is both DMRL and IMRL. The following lemma is now clear.

**Lemma 3.1** *If the service distribution is DMRL (IMRL, respectively) then for any $n \geq 1$ and for any $P_{n-1} \in (0, 1]^{n-1}$, $\bar{r}_n(P_{n-1}, p)$ is increasing (decreasing, respectively) with $p$.*

**Proof.** Observe by a sample path argument, that for any $n \geq 1$ and for any fixed $P_{n-1} \in (0, 1]^{n-1}$, when $L_a = n$, the completed service time of the customer who is currently in service is stochastically decreasing with $p$. Hence, if the service distribution $G$ is with $DMRL$ ($IMRL$, respectively), then the expected residual service time is decreasing (increasing, respectively) with the completed service time and hence increasing (decreasing, respectively) with $p$. $\square$

Lemma 3.1 leads to following propositions:

**Proposition 3.1** *If the service time distribution is DMRL, then the Nash equilibrium profile is unique. Moreover, it is defined recursively by $p_0^e = 1$ and for $n \geq 1$,*

$$
p_n^e = \begin{cases} 1 & C(\bar{r}_n(P_{n-1}^e, 1) + n\bar{x}) \leq V \\[2mm] 0 & C(\bar{r}_n(P_{n-1}^e, 0) + n\bar{x}) \geq V \\[2mm] \beta & C(\bar{r}_n(P_{n-1}^e, \beta) + n\bar{x}) = V \end{cases}
$$

*as long as $p_{n-1} > 0$.*

**Proof.** Since the conditional expected residual service time is increasing with $p$ (see Lemma 3.1) and continuous (see(3)), then exactly one of the three cases itemized in Proposition 2.2 holds. Moreover, if both Case 1 and

Case 2 there do not hold, then there exists a unique $\beta$, $0 < \beta < 1$, which satisfies the equality in Case 3c. □

We note that any structure of the unique solution is possible. In particular, the parameters of the model can be selected such that for some $n$, $p_n^e < p_{n+1}^e$. This phenomenon may appear in cases where the service time distribution has a small mean value, but a small fraction of customers have a relatively long service time. As a queue is observed, it is likely that the customer in service is one with a long service time. On one hand the existence of another customer in the queue does not carry with itself high marginal cost (as the marginal mean service time is small). On the other hand, a longer queue indicates that more time elapsed since the beginning of the current service. The difference between the mean value of the residual service time is larger than the mean service time of the extra customer. Thus, it makes joining more appealing for a larger queue length. Example 1 in the next section demonstrates this phenomenon.

**Proposition 3.2** *If the service distribution is IMRL, then a pure Nash equilibrium exists. Moreover, it is defined recursively by the following exhaustive and mutually conclusive cases (where the third case implies that multiple equilibria may exist):*

1. *If $C(\bar{r}_n(P_{n-1}^e, 1) + n\bar{x}) > V$, then $p_n^e = 0$ is an equilibrium. Moreover, it is unique.*

2. *If $C(\bar{r}_n(P_{n-1}^e, 0) + n\bar{x}) < V$, then $p_e = 1$ is an equilibrium. Moreover, it is unique.*

3. *If neither case 1 nor case 2 holds, then there exists $\beta$, $0 < \beta < 1$, such that $C(\bar{r}_n(P_{n-1}^e, \beta) + n\bar{x}) = V$, and three equilibria exist: $p_n^e = 0, p_n^e = 1$ and $p_n^e = \beta$. as long as $p_{n-1} > 0$.*

**Proof.** If Case 1 (Case 2, respectively) holds, then for any $p$, $0 \leq p \leq 1$, $C(\bar{r}_n(P_{n-1}^e, p) + n\bar{x}) > V$ ($C(\bar{r}_n(P_{n-1}^e, p) + n\bar{x}) < V$, respectively). Hence, the individual's best response against any $p$ is to join with probability zero (one, respectively). If neither Case 1 nor Case 2 holds, then due to the continuity of $\bar{r}_n(P)$ in $p_n$ (since we can see in (3) that $p_n$ appears inside the argument of LST's), then Case 3 holds. In this case, since $\bar{r}_n(P_{n-1}^e, p) + n\bar{x}$ is decreasing in $p$ (see Lemma 3.1), we get all of the following:

- $C(\bar{r}_n(P_{n-1}^e, 0) + n\bar{x}) > V$ and hence if all join with probability zero, then the individual's unique best response is to do so too.

- $C(\bar{r}_n(P_{n-1}^e, 1) + n\bar{x}) < V$ and hence if all join with probability one, then the individual's unique best response is to do so too.

- If all join with probability $\beta$, then the individual's gain from joining is zero and hence he has no better response than to join with probability $\beta$ as well. □

# 4 Examples

In this section we demonstrate, by numerical examples, some of the phenomena which were pointed out earlier.

**Example 1: Zero-one service times**
In this example, we present a counter-intuitive result which exemplifies a related point considered in [2]. In this example $p_1^e < p_2^e$. In other words, under equilibrium, a customer who finds two customers in the system upon his arrival, is more likely to join the system than the one who finds only one customer. Assume $G(x) = \epsilon 1_{\{x \geq 1\}} + (1 - \epsilon)1_{\{x \geq 0\}}$, i.e. the length of service is zero with probability $1 - \epsilon$ and is one with probability $\epsilon$. Note that the residual service time given $n \geq 1$ customers in the system is not a function of $\epsilon$. Fix $\epsilon$ to be small enough, say $\epsilon = 0.01$. If an arrival finds a customer in service and an empty queue, the expected residual service time is $\bar{r}_1(p_1) = \frac{1}{1-e^{-\lambda p_1}} - \frac{1}{\lambda p_1}$ (see (4)). Let $C = 1$ and $V = 0.7$. Note that the expected waiting time, service inclusive, (which equals $\bar{r}_1(p_1) + \epsilon$) is a function of $\lambda$ and $p_1$ only through their product $\lambda p_1$. If $0 < p_1^e, p_2^e < 1$, then $E(W_1) = E(W_2) = 0.7$. Solving Equation (4), we get that $\lambda p_1^e = 2.51$. After inserting this value of $\lambda p_1^e$ in Equations (2) and (3) we get that $\lambda p_2^e = 2.59$. Since in this example the residual service time is increasing, we get that the pair $(p_1^e, p_2^e)$ is as follows:

1. If $\lambda \leq 2.51$ then $p_1^e = p_2^e = 1$.

2. If $2.51 < \lambda \leq 2.59$ then $0 < p_1^e < 1, p_2^e = 1$.

3. If $\lambda > 2.59$ then $0 < p_1^e < p_2^e < 1$.

In summary, for any $\lambda$, $p_1^e \leq p_2^e$, with strict inequality if $\lambda > 2.51$.

**Example 2: A mixture of exponential distributions**
In this example, we demonstrate the case where three equilibria $p_1$ exist. Moreover, one of these three is a mixed strategy while the the other two are pure. Let $G(x) = 1 - \frac{e^{-x} - e^{-2x}}{x}$. This IMRL distribution can be looked at as a mixture of exponential distributions whose parameter is generated from a uniform distribution on the interval $[1, 2]$. For this distribution, $\bar{x} = \log 2$ and hence $\bar{r}_1(p_1) = \frac{\log 2}{\lambda p_1 (\log(2 + \lambda p_1) - \log(1 + \lambda p_1))} - \frac{1}{\lambda p_1}$. Fix the parameters $V = 2.81$, $\lambda = 1$ and $C = 1$. Using equations (1)-(4) we get the following: $W_1(0) = 1.414$ and hence $p_1^e = 1$ [4]. After inserting $p_1^e = 1$, we get that $W_2(1, 0) = 2.137$ and hence $p_2^e = 1$. Inserting further $p_2^e = 1$ we get that $W_3(1, 1, 0) = 2.827$ and $W_3(1, 1, 1) = 2.804$. Hence, as stated in Case 3 in Proposition 3.2, a multiple equilibria exists: $p_3^e = 1$, $p_3^e = 0$ and $p_3^e = 0.654$, where 0.654 is the solution $\beta$, of the equation $C(3\bar{x} + r_3(1, 1, \beta)) = V$.

<div align="center">

**Acknowledgement.**

</div>

# References

[1] Adan, I. and M. Haviv, "Conditional ages and residual service times in the M/G/1 queue," *Stochastic Models,* 25, 110-128 (2009).

[2] Altman E. and R. Hassin, "Non-threshold equilibrium for customers joining an M/G/1 Queue," *Proceeding of the second international simposioum on Dynamic Games,* Saint-Petersburg, Russia, (2002).

[3] Boxma, O.J., "Joint distribution of sojourn time and queue length in the *M/G/1* queue with (in)finite capacity," *European Journal of Operational Research,* 16, 246-256 (1984).

---

[4]According to Proposition 2.1, joining is a dominant strategy here.

[4] Edelson, N.M. and K. Hildebrand, "Congestion tolls for the Poisson queueing processes," *Econometrica*, 43, 81-92 (1973).

[5] Hassin, R. and M. Haviv, *To queue or not to queue: Equilibrium behavior in ueueing system*, Kluwer's International Series, Boston (2003).

[6] Haviv, M. and Y. Kerner, "On balking from en empty queue," *Queueing Systems: Theory and Applications* 55, 239-249 (2007).

[7] Kerner, Y., "The conditional distribution of the residual service time in the $M_n/G/1$ queue," *Stochastic Models* 24, 364-375 (2008).

[8] Mandelbaum, A. and U. Yechiali, "The conditional residual service time in M/G/1 queue," unpublished manuscript, (1979). Also at http://www.math.tau.ac.il/uriy/Publications.html.

[9] Naor, P., "The regulation of queue size by levying tolls," *Econometrica*, 37, 15-24 (1969).