Adv. Appl. Prob. **21**, 481–484 (1989)
Printed in N. Ireland

AN EXTREMAL PROPERTY OF FIFO DISCIPLINE IN *G/IFR/1* QUEUES

TETSUJI HIRAYAMA,* *University of Electro-communications, Tokyo*
MASAAKI KIJIMA,** *Tokyo Institute of Technology*

Abstract

In a recent paper by Shanthikumar and Sumita (1987), it is conjectured that the ergodic sojourn time of a customer in $G/IFR/1$ queues is minimized by FIFO (first in, first out) discipline in the sense of increasing and convex ordering. This paper shows that their conjecture is true. In fact, FIFO discipline minimizes $E[\sum_{k=1}^{N} f(\tau_{\theta(k)} - a_{\theta(k)})]$ for any non-decreasing and convex function f, where N is the (constant) number of arrivals, $\theta(k)$ is the customer identity of the kth departing customer, and a_n and τ_n denote the arriving and departing times of the nth customer, respectively.

IFR SERVICE TIME; SINGLE-SERVER QUEUE; DYNAMIC SCHEDULING; INTERCHANGE ARGUMENT

1. Introduction and summary

Consider a single-server queue having a sequence of random arrival epochs $(A_n)_{n=1}^{N}$ and an i.i.d. (independently and identically distributed) random sequence $(S_n)_{n=1}^{N}$ of service times where N is the (constant) number of arrivals. Suppose that customers are served according to one of the work-conserving service disciplines which are independent of actual service times. Denote by SD the class of such service disciplines. SD includes pre-emptive (resume) service disciplines. In a recent paper by Shanthikumar and Sumita [8] it is conjectured that the ergodic sojourn time is minimized by FIFO (first in, first out) service discipline, in the sense of increasing and convex ordering $\leqq_{c\uparrow}$, if the service time is IFR (increasing failure rate). In their notation,

$$G/IFR/1; T_{\mathrm{FIFO}} \leqq_{c\uparrow} T_\pi, \quad \text{for all} \quad \pi \in SD, \tag{1.1}$$

where T_π is the ergodic sojourn time when discipline π is used. Indeed, they have showed (1.1) for Erlang service time distributions of any order. This conjecture seems to stem from Jackson [3] and Schrage [7] (see also Kleinrock [4], p. 146).

The main purpose of this paper is to prove a stronger result than (1.1) for a discrete-time model of dynamic scheduling problems associated with $G/IFR/1$ queues. Since continuous IFR distributions are absolutely continuous except possibly at the right endpoint (see e.g. [1]), a continuous-time model can be well approximated by our discrete-time model using an ordinary discretization of time. A simple limiting argument then proves (1.1) for the continuous-time model except for a pathological arrival process.

Received 13 June 1988; revision received 10 October 1988.

* Postal address: Department of Communications and Systems Engineering, University of Electro-communications, Chofugaoka, Chofu-shi, Tokyo 182, Japan.

** Postal address: Department of Information Sciences, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan.

2. Main results

Suppose that N customers arrive only at discrete time epochs, say $t = 1, 2, \cdots$. Let $A = (a_n)_{n=1}^N$ be the arbitrary arrival process which is statistically independent of service times. It is assumed that customers are appropriately numbered so that a_n is the arrival epoch of customer n. Each customer requires a positive integer-valued service time with finite mean. Let S_n be the service time of customer n. Define $\mu(t) = \Pr[S_n = t + 1 \mid S_n > t]$. $\mu(t)$ is often called the failure rate function of S_n. The service time is said to be IFR if $\mu(t)$ is non-decreasing in t, where $\mu(t)$ is appropriately defined for t such that $\Pr[S_n > t] = 0$. If this is the case, the single-server queue is denoted by $G/IFR/1$ when there are infinite waiting rooms.

Let $\pi \in SD$ be any scheduling policy and denote by τ_n the departure epoch of customer n. Of course, τ_n is dependent on policy π and service times S_n. Let $\theta(k)$ denote the customer identity who departs the system at the kth departure epoch. For any non-decreasing and convex function f, define

$$V_N(\pi) = E_\pi\left[\sum_{k=1}^N f(\tau_{\theta(k)} - a_{\theta(k)})\right], \qquad \pi \in SD. \tag{2.1}$$

Then the next theorem holds

Theorem 1. Suppose that the system is initially empty. If the generic service time is IFR, then FIFO discipline minimizes $V_N(\pi)$ for all non-decreasing and convex functions over all $\pi \in SD$.

We note that $\tau_{\theta(k)} - a_{\theta(k)}$ is the sojourn time of customer $\theta(k)$. If the system is ergodic, all customers eventually leave the system. Thus, for sufficiently large N, $V_N(\pi)/N$ closely approximates $E[f(T_\pi)]$, if it exists. This observation leads to (1.1) for the discrete-time model. The transform of the result to the continuous-time model can be done in such a way as described in the end of Section 1.

Remark 1. In the previous paper [2], we showed that the expected sojourn time in $G/IFR/1$ queues with two classes of customers is minimized by a simple index rule. As a consequence of that result, it has been seen there that FIFO discipline minimizes the expected sojourn time in $G/IFR/1$ queues (in fact, this result holds for $G/DMRL/1$ queues).

To prove Theorem 1, the next lemma is the key.

Lemma 1. Let $\pi \in SD$. For each non-decreasing function g, define

$$J_k(\pi) = E_\pi[g(\tau_{\theta(k)})], \qquad k = 1, 2, \cdots. \tag{2.2}$$

Then the policy that, at each time epoch, selects a customer with the largest expended service time minimizes $J_k(\pi)$ for all $k = 1, 2, \cdots$, and all non-decreasing functions g.

Proof. Denote the policy stated in the lemma by π_F. We prove that π_F minimizes $J_k^L(\pi) = E_\pi[g(\tau_{\theta(k)} \wedge L)]$ for all g, k and L over $\pi \in SD$ where we denote $\min\{x, y\}$ as $x \wedge y$. Then by letting $L \to \infty$, we can obtain the desired result by the monotone convergence theorem [6]. For $L = 1$, π_F obviously minimizes $J_k^1(\pi)$. Hence we prove it for L, assuming that it holds for $l = 1, \cdots, L - 1$. Without any loss of generality, we assume that, at $t = 1$, there are n customers in the system and $t_1 \geqq t_2 \geqq \cdots \geqq t_n$ where t_k denotes the expended service time of customer k. Let π_m be the policy that, at $t = 1$, selects customer m and then follow the policy π_F after $t = 1$. If the inequality $J_k^L(\pi_F) \leqq J_k^L(\pi_m)$ is shown to hold for all k, $m = 1, \cdots, n$, and g, the principle of optimality of dynamic programming [5] guarantees the lemma (note that the function g is bounded below). To this end, we should consider the following two cases.

Case 1. $t_{m-1} > t_m$. In this case, under policy π_m, customers 1 through n receive service in the order of their numbers.

Case 2. $t_{m-1} = t_m$. Suppose $t_l \geqq t_m + 1 \geqq t_{l+1} \geqq \cdots \geqq t_{m-1}$, where $m - 2 \geqq l \geqq 0$ with $t_0 = \infty$.

Then, under policy π_m, customer m completes his service before customers $l+1$ through $m-1$ are admitted to service.

In both cases, note that customers who arrive at the system after time 1 are not admitted to service before customers 1 through n complete their service. This follows from the definitions of π_F and π_m.

Let r_k be the remaining service time of customer k at $t=1$, and define $R_k = \sum_{i=1}^k r_i$. For policy π_F, one easily sees that $J_k^L(\pi_F) = E[g(R_k \wedge L)]$. Let $r_r^* = r_k - 1$ given that $r_k > 1$. Then, since S_k is IFR, it follows that $E[g(r_k^*)] \leqq E[g(r_k)]$ for any non-decreasing function g (we write this by $r_k^* \leqq_d r_k$).

Case 1. If $k \geqq m$, it is seen by the definition of π_m that $J_k^L(\pi_m) = E[g(R_k \wedge L)]$. For $k < m$, by conditioning on r_m, we have

(2.3) $J_k^L(\pi_m \mid r_m = 1) = E[g((1+R_{k-1}) \wedge L)]$ and $J_k^L(\pi_m \mid r_m > 1) = E[g(1+R_k) \wedge L)]$.

It follows that

(2.4) $J_k^L(\pi_m) = \mu(t_m)E[g((1+R_{k-1}) \wedge L)] + (1-\mu(t_m))E[g((1+R_k) \wedge L)]$.

By conditioning on r_k $(k<m)$, one also has

(2.5) $J_k^L(\pi_F) = \mu(t_k)E[g((1+R_{k-1}) \wedge L)] + (1-\mu(t_k))E[g((1+r_k^* + R_{k-1}) \wedge L)]$.

Note that $r_k^* \leqq_d r_k$. Thus, $E[g((1+R_k) \wedge L)] \geqq E[g((1+r_k^*+R_{k-1}) \wedge L)]$ so that, since $\mu(t_k) \geqq \mu(t_m)$, (2.4) and (2.5) imply $J_k^L(\pi_m) \geqq J_k^L(\pi_F)$, which is desired.

Case 2. For $l<k<m$, by conditioning on r_m, we have

$$J_k^L(\pi_m \mid r_m = 1) = E[g((1+R_{k-1}) \wedge L)] \text{ and } J_k^L(\pi_m \mid r_m > 1) = E[g((1+r_m^*+R_{k-1}) \wedge L)].$$

Hence,

(2.6) $J_k^L(\pi_m) = \mu(t_m)E[g((1+R_{k-1}) \wedge L)] + (1-\mu(t_m))E[g((1+r_m^*+R_{k-1}) \wedge L)]$.

Since $t_k \geqq t_m$ so that $r_k^* \leqq_d r_m^*$ as well as $\mu(t_k) \geqq \mu(t_m)$, (2.6) together with (2.5) shows $J_k^L(\pi_F) \leqq J_k^L(\pi_m)$. For $k \leqq l$, one has (2.3). Thus, the same arguments applying (2.4) and (2.5) as in Case 1 conclude the desired result. If $k \geqq m$, it is trivial that $J_k^L(\pi_F) = J_k^L(\pi_m)$. This completes the proof of the lemma.

It should be noted that $n=0$ or $t_1 = \cdots = t_n = 0$, i.e. the system is initially empty, the optimal policy coincides with FIFO discipline.

We are now in a position to prove Theorem 1.

Proof of Theorem 1. We now assume that customers are numbered in the order of arrival. If customers arrive simultaneously, these customers are appropriately numbered so that $0 \leqq a_1 \leqq a_2 \leqq \cdots$. Since f is convex, it is easily seen that

$$\sum_{k=1}^N f(\tau_{\theta(k)} - a_{\theta(k)}) \geqq \sum_{k=1}^N f(\tau_{\theta(k)} - a_k)$$

with probability 1. It follows that, for any $\pi \in SD$,

$$\begin{aligned} V_N(\pi) &= E_\pi\left[\sum_{k=1}^N F(\tau_{\theta(k)} - a_{\theta(k)})\right] \\ &\geqq E_\pi\left[\sum_{k=1}^N f(\tau_{\theta(k)} - a_k)\right] \\ &\geqq E_{\pi_F}\left[\sum_{k=1}^N f(\tau_{\theta(k)} - a_k)\right] \\ &= E_{\pi_F}\left[\sum_{k=1}^N f(\tau_{\theta(k)} - a_{\theta(k)})\right] = V_N(\text{FIFO}). \end{aligned}$$

The second inequality above is due to Lemma 1, since f is non-decreasing. Therefore, the theorem follows.

Acknowledgement

The authors wish to thank the referee and Dr M. Ohnishi for pointing out an error contained in the original manuscript.

References

[1] BARLOW, R. E. AND PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York.

[2] HIRAYAMA, T., KIJIMA, M. and NISHIMURA, S. (1989). Further results for dynamic scheduling of multiclass $G/G/1$ queues. *J. Appl. Prob.* **26**(3).

[3] JACKSON, J. R. (1961) Queues with dynamic priority discipline. *Management Sci.* **8,** 18–34.

[4] KLEINROCK, L. (1976) *Queueing Systems,* Vol. 2. Wiley, New York.

[5] ROSS, S. (1983) *Introduction to Stochastic Dynamic Programming*. Academic Press, New York.

[6] ROYDEN, H. (1968) *Real Analysis*. Macmillan, New York.

[7] SCHRAGE, L. (1974) Optimal Scheduling Disciplines for a Single Machine under Various Degrees of Information. Working Paper, Graduate School of Business, University of Chicago.

[8] SHANTHIKUMAR, J. G. and SUMITA, U. (1987) Convex ordering of sojourn times in single-server queues: extremal properties of FIFO and LIFO service disciplines. *J. Appl. Prob.* **24,** 737–748.