Technical Research Report

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

Stochastic Monotonicity of the Output Process of Parallel Queues

by P. Tsoucas

TR 89-56

STOCHASTIC MONOTONICITY OF THE OUTPUT PROCESS OF PARALLEL QUEUES

by

Pantelis Tsoucas

Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

This paper considers the output process of a system of $K \cdot/M/1$ queues in parallel with Bernoulli routing of jobs upon arrival. It is shown that the output process is stochastically increasing as the routing probabilities approach $1/K$ in a certain sense. The proof crucially depends on the fact that the absolute value of a simple random walk is a time-homogeneous birth and death process.

1. Introduction

Jobs arrive at a service system at fixed instants $(a_n)_{n=1}^{\infty}$. The service system consists of two $\cdot/M/1$ queues in parallel, i.e., there is one server in each queue and the service times are exponential. Let the service rates in both queues be equal to μ and assume that the system is initially empty. Upon arrival, a job is routed to queue 1 with probability $p \geq 1/2$ and to queue 2 with probability $q \stackrel{\text{def}}{=} 1-p$. Denote by (D_t^i) the departure process of queue i $(i = 1, 2)$ and set $D_t = D_t^1 + D_t^2$. The aim of this paper is to show

Theorem 1.1: (D_t) is stochastically decreasing with increasing p.

The fact that choosing equal routing probabilities in a system of identical $M/G/1$ parallel queues in order to maximize the throughput in steady state is well known. See Gün and Jean-Marie [GJ] and references therein. In [GJ] the result is shown to hold in the case where a resequencing buffer is added at the output of the system. Our Theorem 1.1 continues to hold in this case too.

This problem was posed to the author by Professor G. Shantikumar as a "dual" of the problem considered in Lehtonen [Le]. There, service rates μ_1 and μ_2 $(\mu_1 \geq \mu_2)$, subject to the constraint $\mu_1 + \mu_2 = \mu$, are allocated to two $\cdot/M/1$ queues in tandem. It was shown that the departure process of the second queue for an initially empty system is stochastically decreasing with increasing μ_1. An alternate proof of that was given in [TW], using properties of an embedded random walk. Another "dual" version of the problem in [Le] is the following. Service rates μ_1 and μ_2 $(\mu_1 \geq \mu_2)$, subject to the constraint $\mu_1 + \mu_2 = \mu$, are allocated to two $\cdot/M/1$ queues in parallel. Arriving jobs are routed to each node with equal probability. It can be shown by arguments similar to the ones employed in Section 3 that the departure process of the system is stochastically decreasing with increasing μ. Partial results on this problem are reported in [Be].

Let z_t^i be the number of jobs at queue i $(i = 1, 2)$ at time $t \geq 0$. Consider the quantities

$$x_t \stackrel{\text{def}}{=} z_t^1 - z_t^2, \tag{1.1}$$

$$\eta_t \stackrel{\text{def}}{=} z_t^1 + z_t^2. \tag{1.2}$$

Our proof relies on the observation that the distribution of x_t given $y_s \stackrel{\text{def}}{=} |x_s|$, $s \leq t$, and η_s, $s \leq t$, depends only on y_t. In turn, this is implied by the fact that the absolute value of a simple random walk is a Markov process.

In Section 2 some properties of the absolute value of a simple random walk are presented. Theorem 1.1 is proved in Section 3, and a generalization of it is stated and proved in Section 4.

2. The absolute value of a simple random walk

Let $(\xi_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables where

$$\xi_i = \begin{cases} +1, & \text{w.p. } p \\ -1, & \text{w.p. } q = 1-p., \end{cases}$$

and assume that $p > q$.

Set $X_0 = 0$, $X_n = \xi_1 + \ldots + \xi_n$ for $n \geq 1$ and let $Y_n \stackrel{\text{def}}{=} |X_n|$. We first present the well known result of Pitman that (Y_n) is a Markov chain.

Rogers and Pitman [RP] give a sufficient condition for a function of a Markov process to be Markov. For a discrete Markov chain the result specializes to the following. (See condition (b) of Theorem 2 in [RP].)

Fact 2.1: Let (U_n) be a Markov chain on a set S_U and let $f : S_U \to S_V$ be a function onto a set S_V. Then, the process $V_n \stackrel{\text{def}}{=} f(U_n)$ is a Markov chain on S_V if the following condition is satisfied for $n \geq 1$ and all $v_n \in S_V$, $u_{n+1} \in S_U$.

$$P\{U_{n+1} = u_{n+1} | V_{n+1} = f(u_{n+1})\} P\{V_{n+1} = f(u_{n+1}) | V_n = v_n\} = \\ = \sum_{u_n \in f^{-1}(v_n)} P\{U_{n+1} = u_{n+1} | U_n = u_n\} P\{U_n = u_n | V_n = v_n\}. \quad (2.1)$$

Proof: It suffices to show that (2.1) implies that

$$P\{U_n = u_n | V_n = v_n,\ V_{n-1} = v_{n-1}, \ldots, V_{n-s} = v_{n-s}\} = P\{U_n = u_n | V_n = v_n\}, \quad (2.2)$$

for all $u_n \in S_U$, $v_{n-i} \in S_V$, $i = 0, \ldots, s$. Condition (2.2) implies that the process (Y_n) is Markov (see Kelly [Ke]). □

For the processes (X_n) and (Y_n) defined above, a simple calculation gives

$$\Lambda(p,k) \stackrel{\text{def}}{=} P\{X_n = k | Y_n = k\} = \begin{cases} 1, & k = 0, \\ (p^k) / (p^k + q^k), & k = 1, 2, \ldots \end{cases} \quad (2.3)$$

$$Q_{k,k+1}(p) \stackrel{\text{def}}{=} P\{Y_{n+1} = k+1 | Y_n = k\} = \begin{cases} 1, & k = 0 \\ (p^{k+1} + q^{k+1}) / (p^k + q^k), & k = 1, 2, \ldots \end{cases} \quad (2.4)$$

From (2.3) and (2.4) condition (2.1) is seen to be satisfied. It follows that process (Y_n) is a time-homogeneous birth and death process with transition probabilities given by (2.4).

Remark 2.1: Fact 2.1 implies the stronger result that process (Y_n) is Markov if the distribution of X_0 is $\Lambda(p, Y_0)$.

As will be shown in Section 3, the stochastic ordering results are implied by the following lemma which is a result of easy calculations.

Lemma 2.1: One has the monotonicity properties:

(a) $Q_{k,k+1}(p)$ is an increasing function of p for $k = 0, 1, \ldots$.

(b) $Q_{k,k+1}(p) \le Q_{k+1,k+2}(p)$, for all $0 \le p \le 1$, $k = 1, 2, \ldots$.

3. Proof of the main result

We now consider the process (z_t^1, z_t^2) defined in Section 1. By (S_t^i) denote the virtual service process of queue i ($i = 1, 2$) and set $S_t \stackrel{\text{def}}{=} S_t^1 + S_t^2$. Let (σ_n) be the set of points of process (S_t) and set $(\tau_n) = (a_n) \cup (\sigma_n)$. In what follows we condition on (σ_n) and only consider the processes (z_t^1, z_t^2), (x_t, η_t) and (y_t) at instants (τ_n). They are denoted by (z_n^1, z_n^2), (x_n, η_n) and (y_n).

The probability distribution of the conditional transitions of $(y)_n$ and a result corresponding to relation (2.2) are the key to the proof of Theorem 1.1. They are given next.

Lemma 3.1: For $k \ge 0$, $n \ge 0$, and some $l \ge 0$, one has

(a) $P\{y_{n+1} = k+1 | (y_m, \eta_m),\ m \le n; y_n = k;\ \tau_{n+1} = a_l\} = Q_{k,k+1}(p)$,

(b)

$$P\{y_{n+1} = k+1 | (y_m, \eta_m),\ m \le n;\ y_n = k;\ \tau_{n+1} = \sigma_l\} = \begin{cases} 1, & k = 0 \\ 1/2, & \eta_n > k > 0 \\ 0, & \eta_n = k. \end{cases}$$

(c) $P\{x_n = k | (y_m, \eta_m),\ m \le n;\ y_n = k\} = \Lambda(p, k)$.

Proof: It is done by induction on n simultaneously on Parts (a)-(c). At each step of the induction Part (c) follows from Parts (a) and (b). □

In the same queueing system increase the routing probability p to $p' \in (p, 1]$ and denote by (x'_n, η'_n) and (y'_n) the processes corresponding to (x_n, η_n) and (y_n).

Lemma 3.2: Processes (y_n, η_n) and (y'_n, η'_n) can be constructed on the same sample space such that

$$(y_n, \eta_n) \le (y'_n, \eta'_n), \; n = 0, 1, \ldots, \text{ a.s.} \tag{3.1}$$

Proof: Making use of Lemma 2.1 one shows that (3.1) is satisfied in the "naive" coupling construction. The details are as follows. At $\tau_1 = a_1$, $(y_1, \eta_1) = (y'_1, \eta'_1) = (1, 1)$. Assume that (3.1) is satisfied in some construction of the processes (y_n, η_n) and (y'_n, η'_n) up to some $n > 1$. Then, if $\tau_{n+1} = a_l$, it follows that $\eta_{n+1} = \eta_n + 1$, $\eta'_{n+1} = \eta'_n + 1$. There are two cases. If $y_n > 0$ from Lemma 3.1 and Lemma 2.1 (a) and (b) one picks y'_{n+1}, y_{n+1} such that $y_{n+1} - y_n \le y'_{n+1} - y'_n$. If however $y_n = 0$ then $y_{n+1} = 1$ and (3.1) is not violated if $y'_{n+1}$is chosen according to (2.4). It is easy to see how to finish the construction such that (3.1) is satisfied in the remaining case when $\tau_{n+1} = \sigma_l$. □

It now remains to note that processes (x_n, η_n) and (x'_n, η'_n) can be easily constructed from the processes (y_n, η_n) and (y'_n, η'_n), respectively, of Lemma 3.3. Specifically, the process (x_n, η_n) is completely determined between successive visits of (y_n) to 0, given the first transition of x_n after such a visit. If $y_n = 0$ and $\tau_{n+1} = a_l$, then $x_{n+1} = 1$ (respectively -1) with probability p (respectively q). If $y_n = 0$ and $\tau_{n+1} = \sigma_l$, then $x_{n+1} = 1$ (respectively -1) with probability 1/2 (respectively 1/2). Process (x'_n, η'_n), is constructed similarly. Moreover, in this construction one has $D_t \ge D'_t$, $t \ge 0$, almost surely. This establishes Theorem 1.1.

4. A generalization

Theorem 1.1 can be generalized by considering $K > 2 \cdot/M/1$ queues in parallel fed again by an input stream arriving at instants (a_n). An arriving job is routed to queue i with probability p_i, $i = 1, \ldots, K$, and $\sum_{i=1}^{K} p_i = 1$. Such a vector $\mathbf{p}$ is called a *routing vector*. A result can be formulated and proved to the effect that the total departure process from the system, denoted by (D_t), stochastically increases as the p_i's become more homogeneous. The situation is entirely similar to Section 4 in Lehtonen [Le] and the same approach is followed here. This section is included mainly for completeness.

Definitions: For a routing vector $\mathbf{p}$ denote by $p_{[1]} \ge \cdots \ge p_{[K]}$ the decreasing rearrangement of the coordinates. A routing vector $\mathbf{p}'$ is said to *majorize* vector $\mathbf{p}$ if $\sum_{i=1}^{k} p_{[i]} \le \sum_{i=1}^{k} p'_{[i]}$, $k = 1, \ldots, K-1$. Then, one writes $\mathbf{p} \prec \mathbf{p}'$. (See Marshall and Olkin [MO].)

By (D'_t) denote the departure process from the system when the arrival process is (a_n) and the routing vector is $\mathbf{p}'$.

Theorem 4.1: If $\mathbf{p}' \succ \mathbf{p}$, then $(D'_t) \le_{st} (D_t)$ ($\le_{st}$ denotes stochastic ordering.)

Proof: For a routing vector $\mathbf{q}$ define

$$T(\mathbf{q}) \stackrel{\text{def}}{=} (q_1, \ldots, \lambda q_i + (1-\lambda) q_j, \ldots, (1-\lambda) q_i + \lambda q_j, \ldots, q_K). \tag{4.1}$$

Then, since $\mathbf{p}' > \mathbf{p}$, $\mathbf{p}$ can be obtained from $\mathbf{p}'$ be a finite number of successive transformations of the type (4.1) (see Marshall and Olkin [MO], p. 21). It therefore suffices to show that $(D'_t) \le_{st} (D_t)$ when $\mathbf{p} = (p'_1, \ldots, \lambda p'_i + (1-\lambda) p'_j, \ldots, (1-\lambda) p'_i + \lambda p'_j, \ldots, p'_K)$. But this is a corollary of Theorem 1.1. □

Acknowledgement: I would like to thank Professor J. Pitman for pointing out reference [RP].

5. References

[Be] Benchekroun, M. (1989) MS Thesis. Dept. of Electrical Engineering, Cornell University.

[GJ] Gün, L. and Jean-Marie, A. (1989) Resequencing in parallel queues with Bernoulli loading. Preprint.

[Ke] Kelly, F.P. (1982) Markovian functions of a Markov chain. *Sankhya* Vol. 44, Series A, Pt. 3, 372-379.

[Le] Lehtonen, T. (1986) On the ordering of tandem queues with exponential servers. *J. Appl. Prob.* **23**, 115-129.

[MO] Marshall, A.W. and Olkin, I. (1979) *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York.

[RP] Rogers, L.C.G. and Pitman, J.W. (1981) Markov funtions. *Ann. Prob.* 9, 4, 573-582.

[TW] Tsoucas, P. and Walrand, J. (1987) On the interchangeability and stochastic ordering of $\cdot/M/1$ queues in tandem. *Adv. Appl. Prob.* 19, 515-520.