PROBABILISTIC PROOF OF THE INTERCHANGEABILITY OF ./M/1 QUEUES IN SERIES

V. ANANTHARAM

School of Electrical Engineering, Phillips Hall, Cornell University, Ithaca, NY 14853, U.S.A.

Received: 14 April 1987
Revised: 15 August 1987

Abstract

Given a finite number of empty ./M/1 queues, let customers arrive according to an arbitrary arrival process and be served at each queue exactly once, in some fixed order. The process of departing customers from the network has the same law, whatever the order in which the queues are visited. This remarkable result, due to R. Weber [4], is given a simple probabilistic proof.

Keywords

Departure process, filtering theory, insensitivity, M/M/1 queues, point processes, queueing theory, tandem queues.

1. Introduction

Given a finite number of empty ./M/1 queues, let customers arrive according to an arbitrary arrival process, and be served at each queue exactly once, in some fixed order, i.e., once they finish service at the first server they joint the queue at the next, and so on till they leave the system. By an arrival process we mean the counting process associated with an a.s. strictly increasing sequence of times

$$0 \leqslant A_1 \leqslant A_2 \leqslant \dots$$

and a sequence of a.s. finite, positive integer valued random variables

$$N_1, N_2, \dots$$

giving the number of arrivals at the respective times. Note that the process may be explosive. In [4], R. Weber proved the remarkable fact that whatever the order in which the queues are met, the law of the departure process is the same. Here we give a simple probabilistic proof of this result.

© J.C. Baltzer A.G. Scientific Publishing Company

Consider an initially empty tandem connection of two exponential servers in series, and a deterministic sequence of arrivals at times $0 < t_1 < t_2 < \cdots$, the number of customers arriving at t_k being n_k. Each arriving customer joins the first queue, moves to the second queue after being served, and finally leaves the system after being served in the second queue. We let $0 < T_1 < T_2 < \ldots$ be the successive departure times from the system and (D_t) the associated counting process. Let X_t denote the number of customers in the first queue at time t and Y_t the number in the second. Let P denote the probability law of the process (X_t, Y_t) when the service rates are μ_1 in the first server and μ_2 in the second, and $\tilde{P}$ the law when the service rates are μ_2 in the first server and μ_1 in the second.

To establish Weber's result, it suffices to show that the law of the departure process (D_t) is the same under P and $\tilde{P}$. Indeed, it follows that, for an arbitrary arrival process, the law of the departure process from an initially empty connection of two exponential servers in series is independent of the order in which the servers are met. If a finite number of empty exponential servers are in series, for an arbitrary arrival process, we may interchange a pair of successive servers without changing the law of the departure process. Note that any permutation of the servers can be generated by such interchanges.

2. Main result and proof

Let (G_t) be the filtration generated by (D_t). Let (h_t) be the right continuous version of the stochastic intensity of (D_t) under P, i.e., the unique right continuous process adapted to (G_t) such that $(D_t - \int_0^t h_s \, \mathrm{d}s)$ is a (P, G_t) local martingale. We define $(\tilde{h}_t)$ similarly, but with respect to $(\tilde{P}, G_t)$. Since a point process is uniquely determined by its intensity, to show that (D_t) has the same law under P and $\tilde{P}$, it suffices to show that, for every $t \geqslant 0$,

$$h_t = \tilde{h}_t \text{ a.s.}$$

The term "almost surely" is unambiguous, because P and $\tilde{P}$ are mutually absolutely continuous. Let

$$\xi_t(x, y) \triangleq P[X_t = x, Y_t = y \mid G_t],$$

and

$$\tilde{\xi}_t(x, y) \triangleq \tilde{P}[X_t = x, Y_t = y \mid G_t].$$

If right continuous versions can be chosen for ξ_t and $\tilde{\xi}_t$, then, for each $t \geqslant 0$,

$$h_t = \mu_2 P[Y_t \geqslant 1 \mid G_t] = \mu_2 \sum_{y \geqslant 1} \xi_t(x, y), \tag{2.1a}$$

and

$$\tilde{h}_t = \mu_1 \tilde{P}[Y_t \geqslant 1 \mid G_t] = \mu_1 \sum_{y \geqslant 1} \tilde{\xi}_t(x, y). \tag{2.1b}$$

Consider the Markov process $((X_t, Y_t),\ t \in [t_k, t_{k+1}))$. The filtering theory, [1], allows us to write the following equations for the evolution of a right continuous version of ξ_t, (vector notation for $\xi_t(x, y)$), on $[t_k, t_{k+1})$):

$$\dot{\xi}_t = \xi_t Q_D + \xi_t h_t \text{ for } t \neq T_j,\ j = 1, 2, \ldots, \tag{2.2a}$$

$$\xi_t = \frac{\xi_{t\_} Q^D}{h_{t\_}} \text{ when } t = T_j \text{ for some } j = 1, 2, \ldots. \tag{2.2b}$$

Here

$$Q_D(\alpha, \beta) \triangleq Q(\alpha, \beta) 1(\beta \neq \alpha + (0, -1)),$$
$$Q^D(\alpha, \beta) \triangleq Q(\alpha, \beta) 1(\beta = \alpha + (0, -1)),$$

and Q is the infinitesimal generator of $((X_t, Y_t),\ t \in [t_k, t_{k+1}))$ under P. Similarly, equations parallel to (2.2a) and (2.2b) can be written for $\tilde{\xi}_t$ in terms of $\tilde{Q}$, the infinitesimal generator of $((X_t, Y_t),\ t \in [t_k, t_{k+1}))$ under $\tilde{P}$.,

A heuristic understanding of (2.2a) can be obtained by writing

$$\begin{aligned} \xi_{t+\epsilon}(\beta) 1(D_{t+\epsilon} = D_t) &= P[(X_{t+\epsilon}, Y_{t+\epsilon}) = \beta \mid G_{t+\epsilon}] 1(D_{t+\epsilon} = D_t) \\ &= P[(X_{t+\epsilon}, Y_{t+\epsilon}) = \beta \mid G_t,\ D_{t+\epsilon} = D_t] 1(D_{t+\epsilon} = D_t) \end{aligned}$$

So

$$\begin{aligned} \xi_{t+\epsilon}(\beta) &= \frac{\sum_\alpha P[(X_{t+\epsilon}, Y_{t+\epsilon}) = \beta,\ (X_t, Y_t) = \alpha,\ D_{t+\epsilon} = D_t \mid G_t]}{P[D_{t+\epsilon} = D_t \mid G_t]} \\ &= (1 - h_t \epsilon)^{-1} \sum_\alpha \xi_t(\alpha) Q_D(\alpha, \beta) \epsilon + o(\epsilon) \end{aligned}$$

on $\{D_{t+\epsilon} = D_t\}$, from which (2.2a) follows upon letting $\epsilon \to 0$. (2.2b) can be heuristically understood as giving ξ_t by putting $\xi_{t\_}$ through the transition probability matrix at a jump corresponding to a departure. (This transition matrix is Q^D normalized by its row sums.)

For $n \geqslant 0$, $N \geqslant n$, let

$$S_t(n, N) \triangleq \sum_{\substack{x+y=N \\ y \geqslant n}} \xi_t(x, y). \tag{2.3}$$

Similarly define $\tilde{S}_t(n, N)$. The crucial observation is the following

Claim: For any $n \geqslant 0$, $N \geqslant n$, we have

$$S_t(n, N) = (\mu_1/\mu_2)^n \tilde{S}_t(n, N)$$

for all $t \geqslant 0$, a.s.

REMARK

To motivate the claim, note that if Weber's result is true, the total number of customers in the system is the same in law, so we must have $S_t(0, N) = \tilde{S}_t(0, N)$. Further, we must have $S_t(1, N) = (\mu_1/\mu_2)\tilde{S}_t(1, N)$ to get the same intensity of departures in the two situations. Finally, the claimed identity holds for the corresponding partial sums of the stationary distribution when each system is stable with the same Poisson input.

Proof

First note that, for $k = 1, 2, \ldots,$

claim holds upto $t_{k^-} \Rightarrow$ claim holds upto t_k.

Indeed, the arrivals at t_k do not change the number of customers in the tail queue, and there are a.s. no jumps at t_k. Clearly the claim holds upto t_1. Thus it suffices to prove that

claim holds at $t_k \Rightarrow$ claim holds upto $(t_{k+1})_-$.

Writing (2.2a) explicitly gives

$$\dot{\xi}_t(x, y) = \mu_1 \xi_t(x+1, y-1)1(y>0) - \mu_1 \xi_t(x, y)1(x>0) \\ - \mu_2 \xi_t(x, y)1(y>0) + h_t \xi_t(x, y). \tag{2.4}$$

From (2.3) and (2.4) and elementary manipulations, we get the following:

$$\dot{S}_t(n, N) = \mu_1 S_t(n-1, N) - (\mu_1 + \mu_2) S_t(n, N) + h_t S_t(n, N) \\ (n > 0,\ N \geqslant n) \tag{2.5a}$$

$$\dot{S}_t(0, N) = -\mu_2 S_t(1, N) + h_t S_t(0, N) \quad (N \geqslant 1) \tag{2.5b}$$

$$\dot{S}_t(0, 0) = h_t S_t(0, 0). \tag{2.5c}$$

Equations (2.5 a–c) hold for $t \in [t_k, t_{k+1})$, $t \neq T_j$, $j = 1, 2, \ldots$. At the departure times of the tail queue we have, from (2.1b) and (2.2),

$$S_t(n, N) = \frac{S_{t^-}(n+1, N+1)\mu_2}{h_{t^-}}. \tag{2.6}$$

Now, we define the functions

$$\tau_t(n, N) \triangleq (\mu_2/\mu_1)^n S_t(n, N), \tag{2.7}$$

and

$$\eta_t \triangleq \mu_1 \sum_{N \geqslant 1} \tau_t(1, N). \tag{2.8}$$

Using the definitions (2.7) and (2.8), it is then easy to see that η_t and $\tau_t(n, N)$ satisfy the following equations, for $t \in [t_k, t_{k+1})$, $t \neq T_j$, $j = 1, 2, \ldots,$
From (2.5a), for $n \geqslant 0$, $N \geqslant n$,

$$\dot{\tau}_t(n, N) = \mu_2 \tau_t(n-1, N) - (\mu_1 + \mu_2)\tau_t(n, N) + \eta_t \tau_t(n, N). \tag{2.9a}$$

From (2.5b), for $N \geqslant 1$,

$$\dot{\tau}_t(0, N) = -\mu_1 \tau_t(1, N) + \eta_t \tau_t(0, N). \tag{2.9b}$$

From eq. (2.5c),

$$\dot{\tau}_t(0, 0) = \eta_t \tau_t(0, 0). \tag{2.9c}$$

Further, from (2.6), if $t = T_j$ for some $j = 1, 2, \ldots, n \geqslant 0$, $N \geqslant n$,

$$\tau_t(n, N) = \frac{\tau_{t-}(n+1, N+1)\mu_1}{\eta_{t-}}. \tag{2.10}$$

But (2.8), (2.9 a–c) and (2.10) are also satisfied by $\tilde{h}_t$ and $\tilde{S}_t(n, N)$ for $t \in [t_k, t_{k+1})$, [replacing η_t and $\tau_t(n, N)$ respectively], as can be seen by writing equations parallel to (2.5 a–c) and (2.6), under $\tilde{P}$. This establishes the claim. □

COROLLARY

$\tilde{h}_t = h_t \forall t \geqslant 0$ a.s.

Proof

This follows from (2.1a) and (2.1b) and the claim. □

This establishes Weber's result. The probabilistic nature of our reasoning is partly hidden in the proof of the filtering formulas, [1], a reading of which is suggested to convince the reader that the proof is not merely a sleight of hand with differential equations.

Since the preparation of the original version of this paper, two other proofs of Weber's result, using entirely different techniques, have appeared in the literature [2,3].

Acknowledgement

I thank Prof. J. Walrand for bringing R. Weber's paper, [4], to my attention.

References

[1] J. Walrand, Filtering formulas and the ./M/1 queue in a quasireversible network, Stochastics 6 (1981) 1–22.
[2] T. Lehtonen, On the ordering of tandem queues with exponential servers, Jour. Appl. Prob., 23 (1986) 115–129.
[3] P. Tsoucas and J. Walrand, A note on the interchangeability of ./M/1 queues in tandem, Preprint (1986).
[4] R. Weber, The interchangeability of ./M/1 queues in series, Jour. Appl. Prob., 16 (1979) 690–695.