

First and Second Derivative Estimators for Closed Jackson-Like Queueing Networks Using Perturbation Analysis Techniques¹

GANG BAO

Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003

CHRISTOS G. CASSANDRAS

Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003

MICHAEL A. ZAZANIS

Dept. of Industrial Engineering and Operations Research, University of Massachusetts, Amherst, MA 01003

Received January 21, 1994; Revised March 3, 1996; Accepted June 10, 1996

Abstract. We consider a closed Jackson—like queueing network with arbitrary service time distributions and derive an unbiased second derivative estimator of the throughput over N customers served at some node with respect to a parameter of the service distribution at that node. Our approach is based on observing a single sample path of this system, and evaluating all second-order effects on interdeparture times as a result of the parameter perturbation. We then define an estimator as a conditional expectation over appropriate observable quantities, as in Smoothed Perturbation Analysis (SPA). This process recovers the first derivative estimator along the way (which can also be derived using other techniques), and gives new insights into event order change phenomena which are of higher order, and on the type of sample path information we need to condition on for higher-order derivative estimation. Despite the complexity of the analysis, the final algorithm we obtain is relatively simple. Our estimators can be used in conjunction with other techniques to obtain rational approximations of the entire throughput response surface as a function of system parameters.

Keywords: Queueing Network, gradient estimation, perturbation analysis, response surface estimation, rational approximation.

1. Introduction

In dealing with stochastic Discrete Event Dynamic Systems (DEDS), we are often faced with situations where no functional relationship between design or control parameters and performance metrics of interest is available. Still, by observing a single sample path of such a system (in a simulation or in an actual operating environment) it is often possible to efficiently estimate gradients of performance metrics with respect to various parameters. This can be accomplished through techniques such as Perturbation Analysis (PA) (Ho and Cao 1991, Suri 1989, Glasserman 1991) and the Likelihood Ratio (LR) methodology (Glynn 1987, Reiman and Weiss 1989, Rubinstein 1986). These techniques provide an alternative to costly (sometimes infeasible in real-time) simulation where sensitivity estimation requires multiple sample path generations. In addition, they can often be integrated into gradient-based optimization algorithms (e.g. Cassandras, Abidi, and Towsley (1990), Chong and Ramadge (1992)) for problems of considerable complexity.

Techniques such as perturbation analysis can generally be used to estimate not only the first, but also higher-order derivatives of performance metrics with respect to some parameters (see Brémaud and Vázquez-Abad (1992), Zazanis and Suri (1994), Fu and Hu (1997a), Bao, Cassandras, and Zazanis (1996)). Although second derivative estimators become difficult to analyze and are harder to implement in practice (compared to first derivative estimators), there are two recent developments that have provided renewed motivation for deriving higher-order derivative estimates for performance metrics of DEFS. First, as first derivative estimators are used in gradient-based optimization, we are often faced with the practical problem of instabilities in the form of large oscillations (Cassandras, Abidi, and Towsley (1990), Bertsekas (1982)). To alleviate this problem, it is known that algorithms using second derivative information may be used (Bertsekas, Gafni, and Gallager 1984). The second, perhaps more important, development is the emergence of Padé approximation techniques as viable means to accurately estimate the entire response surface of a complex system with respect to some parameter. Given some function $J(\theta)$, a Padé approximant is a rational function of the form $P_L(\theta)/Q_M(\theta)$, where $P_L(\theta)$ and $Q_M(\theta)$ are appropriately selected polynomials of degree L and M respectively (see Baker 1975). The coefficients of these polynomials typically involve first and higher-order derivative information at a single point θ_0 . As was recently shown in Gong, Nananukul, and Yan (submitted), Padé approximants of performance metrics of GI/G/1 systems show remarkable accuracy using first and second derivative information alone. This opens up a range of exciting possibilities for estimating global response surfaces of more complex systems based on information extracted from a single sample path observed under a parameter setting θ_0 . Lastly, it is worth mentioning that a byproduct of sample-path-based first and second derivative estimators is the fact that they sometimes lead directly to the establishment of structural properties of a system, such as monotonicity or convexity/concavity of some performance metric with respect to some parameter (if, for example, it turns out that the sign of an unbiased such estimator is always positive/negative).

In this paper, we consider a Jackson-like closed queueing network consisting of m servers, providing service to a fixed population of n customers. By “Jackson-like queueing network” we mean that all service time distributions are arbitrary (except for some mild technical conditions), routing is Markovian (as explained in section 2), and all queues are assumed to have infinite capacity. This is an extension of our work in Bao, Cassandras, and Zazanis (1996), where our analysis was limited to a *serial* closed network. PA techniques were first applied to this type of system in Ho, Cao, and Cassandras (1983) to approximate first derivatives of the throughput. In Ho and Cao (1983) and Cao (1987) estimators for the throughput of closed Jackson queueing networks using Infinitesimal Perturbation Analysis (IPA) were derived, and extended by Cao (1990) to general service time distributions. However, IPA generally yields biased estimates if applied to second derivative estimation. The main reason is that IPA is based on limited information obtained from the observed sample path; to estimate second derivatives one needs additional information to account for second-order effects in event order changes.

The main contribution of this paper is the derivation of second derivative estimators for the throughput of closed Jackson-like networks. Our basic approach is to evaluate all second-order effects on interdeparture times in a sample path of this network as a result of

a parameter perturbation. We then proceed as in SPA (Gong and Ho 1987, Glasserman and Gong 1991) to define an estimator as a conditional expectation. We derive along the way a number of new results pointing at event order change phenomena which are of higher order. The analysis in our paper is at times tedious, and the form of the second derivative estimator appears complex. As we will show, however, its implementation turns out to be relatively simple. Further, we show that our estimators over a fixed number of events are unbiased.

The paper is organized as follows. In section 2, we set up the estimation problem and introduce some notation. In section 3, we derive an unbiased second derivative estimator for the throughput; in the process, our approach recovers the first derivative, which can also be obtained through standard IPA. We also present an algorithm for implementing our first and second derivative estimators in section 4. Some numerical examples are then presented in section 5. In section 6, we present an application to the estimation of the throughput over all parameter values, based on the analysis recently provided in Gong, Nananukul, and Yan (submitted). Finally, section 7 contains a summary and discussion of future research in this area.

2. Notation and Estimation Problem Setup

Consider an m -node, n -customer closed Jackson-like queueing network (as defined in the previous section), with single server nodes. The routing is Markovian, i.e., a customer, upon completing service at node r , moves to node q with probability p_{rq} independent of anything else (without loss of generality we will assume the routing matrix $[p_{rq}]$, $r, q = 1, \dots, m$ to be irreducible.) All queues have infinite capacity and all nodes serve customers in FCFS fashion. Let S_{ik} denote the service time of the i th customer served at node k . We assume that the service times $\{S_{ik}; i = 1, 2, \dots\}$ are an i.i.d. sequence of random variables with distribution $F_k(\cdot)$, $k = 1, \dots, m$. The sequences $\{S_{ik}; i = 1, 2, \dots\}$, $k = 1, \dots, m$, are also assumed independent.

Our objective is to estimate the first and second derivatives of the expected departure time of the N th customer served at a node, say node 1, with respect to a parameter θ of the service time distribution of one of the nodes based on observations extracted from single sample path (the “nominal sample path”). Without loss of generality we assume that $\theta \in \Theta$ is a parameter of $F_1(\cdot)$ and Θ is an interval in \mathbf{R} .

As in previous related work in this area (e.g., Glasserman (1991), Suri and Zazanis (1988)) suppose that our probability space, (Ω, \mathcal{F}, P) supports m sequences of i.i.d. random variables $\{U_{ik}; i = 1, 2, \dots, k = 1, \dots, m\}$, uniformly distributed on $[0, 1]$. Let $F_1^{-1}(u, \theta) = \inf\{x: F_1(x, \theta) > u\}$, $F_k^{-1}(u) = \inf\{x: F_k(x) > u\}$, $k = 2, \dots, m$. Thus, letting $S_{i1}(\theta) = F_1^{-1}(U_{i1}, \theta)$, $S_{ik} = F_k^{-1}(U_{ik})$, $k = 2, \dots, m$ defines a family of sample paths parameterized by θ .

We are now ready to state the assumptions under which we carry out our analysis and derive the first and second derivative estimators.

ASSUMPTION A.1 $S_{i1}(\theta)$ is an increasing function of θ , i.e. $\Delta S_{i1} \stackrel{\text{def}}{=} S_{i1}(\theta + \Delta\theta) - S_{i1}(\theta) \geq 0$ w.p.1, for $\Delta\theta \geq 0$.

This assumption simplifies the sample path analysis since it guarantees that a positive change in the parameter θ will result in positive perturbations. It can be relaxed via the approach described in §8 of Zazanis and Suri (1994).

ASSUMPTION A.2 *The derivative $\frac{\partial S_{i1}}{\partial \theta}(\theta) = \lim_{\Delta\theta \rightarrow 0} \frac{S_{i1}(\theta + \Delta\theta) - S_{i1}(\theta)}{\Delta\theta}$ exists and is a continuous function of $\theta \in \Theta$ w.p.1. Furthermore, there exist positive constants c_1, c_2 , such that*

$$\frac{\partial S_{i1}}{\partial \theta} \leq c_1 + c_2 S_{i1} \quad \text{w.p.1, for all } \theta \in \Theta .$$

The above assumption (together with the mean value theorem) implies that

$$\frac{\Delta S_{i1}}{\Delta\theta} \leq c_1 + c_2 S_{i1}. \quad (1)$$

This assumption is introduced purely for convenience in our analysis. Note that it is a condition which is easy to verify for any given distribution and is satisfied by most commonly encountered parametric distribution families. In particular it is always satisfied when θ is a scale or a location parameter (see Suri and Zazanis 1988).

ASSUMPTION A.3 *The second derivative, $\frac{\partial^2 S_{i1}}{\partial \theta^2}(\theta) = \lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta} [\frac{\partial S_{i1}}{\partial \theta}(\theta + \Delta\theta) - \frac{\partial S_{i1}}{\partial \theta}(\theta)]$, exists and is continuous for all $\theta \in \Theta$ w.p.1. Furthermore, $E|\sup_{\theta \in \Theta} \frac{\partial^2 S_{i1}}{\partial \theta^2}(\theta)| < \infty$.*

ASSUMPTION A.4 *The distributions $F_k, k = 1, \dots, m$, are absolutely continuous with density $f_k(t)$ and corresponding hazard rate $\frac{f_k(t)}{1-F_k(t)}$ bounded above by γ for all $t \geq 0$. In particular, $\frac{f_1(t, \theta)}{1-F_1(t, \theta)} \leq \gamma$ for all $\theta \in \Theta$.*

The last part of the assumption above may be relaxed, though with considerable effort. The reader is referred to Fu and Hu (1997b), Fu and Hu (1997c), and Zazanis (1995).

ASSUMPTION A.5 *The service times at all nodes have finite third moments:*

$$\max_{1 \leq k \leq m} \int_0^\infty x^3 dF_k(x) < \infty.$$

The following notation will be needed in the sample path analysis of section 3.

C_{ik} : the i th customer served at node k in the nominal path;

e_{ik} : departure event of C_{ik} in the nominal path;

S_{ik} : service time of C_{ik} in the nominal path;

A_{ik} : arrival time of C_{ik} in the nominal path;

D_{ik} : departure time of C_{ik} in the nominal path;

U_{ik} : routing indicator of the i th departure at node k in the nominal path;

where $i = 1, 2, \dots, k = 1, \dots, n$. The routing indicator is uniformly distributed over $[0, 1]$ and is used to implement the probabilistic mechanism through which the i th departure at node k is routed to node j with the given probability p_{ij} .

In this paper we focus on finite horizon performance metrics. With the above notation, the main performance metric we consider is

$$\bar{D}(\theta) = \frac{1}{N} E[D_{N1}] \quad (2)$$

which can be thought of as the mean interdeparture time. Next, we seek an expression for D_{N1} in terms of $\{S_{ik}, i = 1, 2, \dots, k = 1, \dots, n\}$. For this purpose, we proceed as in Bao, Cassandras, and Zazanis (1996). First, we say that event e_{ik} is *induced* by another event e_{jl} if e_{ik} becomes feasible at the time when e_{jl} takes place. Then, fixing a given sample path and observing the precedence in the events, we construct a set B_{ik} associated with event e_{ik} as follows:

1. $(i, k) \in B_{ik}$.
2. If e_{i_1, k_1} induces e_{ik} , then $(i_1, k_1) \in B_{ik}$.
3. For all $j = 2, 3, \dots$, if e_{i_j, k_j} induces $e_{i_{j-1}, k_{j-1}}$ and $(i_{j-1}, k_{j-1}) \in B_{ik}$, then $(i_j, k_j) \in B_{ik}$.
4. The procedure ends at the beginning of the sample path with $(i_s, k_s) \in B_{ik}$ such that $D_{i_s, k_s} = \min\{D_{jl} : (j, l) \in B_{ik}\}$.

Therefore, B_{ik} is of the form

$$B_{ik} = \{(i_s, k_s), (i_{s-1}, k_{s-1}), \dots, (i_2, k_2), (i_1, k_1), (i, k)\},$$

Since D_{ik} is the time when e_{ik} occurs, we have

$$D_{ik} = \sum_{(j,l) \in B_{ik}} S_{jl}. \quad (3)$$

Therefore,

$$E[D_{ik}] = E \left[\sum_{(j,l) \in B_{ik}} S_{jl} \right]. \quad (4)$$

Since it is known (Glasserman 1991) that the commuting condition is satisfied in a closed Jackson-like network, the interchangeability of the derivative and expectation operations is allowed to first order, that is,

$$\frac{\partial}{\partial \theta} E D_{N1} = E \left[\frac{\partial}{\partial \theta} D_{N1} \right]. \quad (5)$$

As also shown in Glasserman (1991), we have

$$\frac{\partial}{\partial \theta} D_{ik} = \sum_{(j,l) \in B_{ik}} \frac{\partial}{\partial \theta} S_{jl}. \quad (6)$$

Therefore, since only service times S_{j1} , $j = 1, 2, \dots$ are dependent on θ , we get

$$\left[\frac{\partial}{\partial \theta} E D_{N1} \right]_{\text{est.}} = \sum_{(j,l) \in B_{N1}} \frac{\partial}{\partial \theta} S_{jl} = \sum_{(j,1) \in B_{N1}} \frac{\partial}{\partial \theta} S_{j1} \quad (7)$$

which is an unbiased estimator of the first derivative of $E[D_{N1}]$. However, for the second derivative of $E[D_{N1}]$, by simply using $\sum_{(j,l) \in B_{N1}} \frac{\partial^2}{\partial \theta^2} S_{jl}$, we will generally have a biased estimator. Roughly speaking, the problem is that with θ increasing to $\theta + \Delta\theta$ several event order changes will result, and the probability of these event order changes is $O(\Delta\theta)$. With the increments themselves being $O(\Delta\theta)$, we see that the event order changes will produce effects of $O(\Delta\theta^2)$ which are not negligible as far as the second derivative is concerned.

With this observation in mind, in what follows we proceed in a way similar to the serial network case considered in Bao, Cassandras, and Zazanis (1996). As we will see, however, the fact that customers are now routed to one of several nodes and may arrive at a node from one of several other nodes introduces additional complexities which have to be handled.

3. Derivation of Estimators

When θ is increased to $\theta + \Delta\theta$, we get a perturbed sample path. We use the superscript p to denote various quantities in the perturbed path (for example, D_{ik}^p denotes the i th departure time from node k in the perturbed path).

3.1. Lindley Recursions for Perturbed Event Times

Applying Lindley's equation to both nominal and perturbed paths, we have:

$$\begin{aligned} D_{ik} &= S_{ik} + \max(D_{i-1,k}, A_{ik}) \\ D_{ik}^p &= S_{ik}^p + \max(D_{i-1,k}^p, A_{ik}^p). \end{aligned}$$

where $D_{0k} = D_{0k}^p = 0$ for all $k = 1, \dots, m$. Defining $\Delta D_{ik} = D_{ik}^p - D_{ik}$,

$$\Delta D_{ik} = \Delta S_{ik} + \max(D_{i-1,k}^p, A_{ik}^p) - \max(D_{i-1,k}, A_{ik}). \quad (8)$$

We use \hat{i} , \hat{k} , to denote the customer index and the node index respectively of C_{ik} just prior to arriving at node k in the nominal path, i.e., $C_{\hat{i}\hat{k}}$ becomes C_{ik} immediately after event $e_{\hat{i}\hat{k}}$. Stated another way, $A_{ik} = D_{\hat{i}\hat{k}}$. In general, in the perturbed path both the node from which C_{ik} comes and his index will be different and thus we will designate them by k^* , i^* respectively. In other words, in the perturbed path customer C_{ik} arrives at node k

at time $A_{ik}^p = D_{i^*k^*}^p$. We note that, for cyclic, single server networks, $\hat{k} = k^* = k - 1$ since customers arrive to node k only from node $k - 1$ directly upstream, and $\hat{i} = i^*$ (Bao, Cassandras, and Zazanis, 1996).

For consistency of notation, if there are initially n_k customers at node k , we simply set $\hat{i} = 0$ (i.e. $A_{ik} = D_{0k} = 0$) for all $i = 1, \dots, n_k$. We also note that, since the service distributions at all nodes are absolutely continuous (see Assumption **A.4**), the probability of two events occurring at the same time is zero. Denoting by $(x)^+ := \max(0, x)$ the positive part of $x \in \mathbf{R}$, we now consider two cases in (8):

Case 1. $D_{i-1,k} \geq A_{ik} = D_{i\hat{k}}$ (Thus, $D_{i_1,k_1} = D_{i-1,k}$ or e_{ik} is induced by $e_{i-1,k}$).

$$\begin{aligned} \Delta D_{ik} &= \Delta S_{ik} + \max(D_{i-1,k}^p, A_{ik}^p) - D_{i-1,k} \\ &= \Delta S_{ik} + (D_{i-1,k}^p - D_{i-1,k}) + \max(0, A_{ik}^p - D_{i-1,k}^p) \\ &= \Delta S_{ik} + \Delta D_{i-1,k} + (A_{ik}^p - D_{i-1,k}^p)^+ \\ &= \Delta S_{ik} + \Delta D_{i_1,k_1} + (D_{i^*k^*}^p - D_{i-1,k}^p)^+. \end{aligned}$$

Case 2. $D_{i-1,k} < A_{ik} = D_{i\hat{k}}$ (Thus, $D_{i_1,k_1} = A_{ik}$ or e_{ik} is induced by $e_{i\hat{k}}$).

$$\begin{aligned} \Delta D_{ik} &= \Delta S_{ik} + \max(D_{i-1,k}^p, A_{ik}^p) - A_{ik} \\ &= \Delta S_{ik} + (A_{ik}^p - A_{ik}) + \max(0, D_{i-1,k}^p - A_{ik}^p) \\ &= \Delta S_{ik} + (D_{i^*k^*}^p - D_{i\hat{k}}) + (D_{i-1,k}^p - D_{i^*k^*}^p)^+ \\ &= \Delta S_{ik} + (D_{i\hat{k}}^p - D_{i\hat{k}}) + (D_{i^*k^*}^p - D_{i\hat{k}}^p) + (D_{i-1,k}^p - D_{i^*k^*}^p)^+ \\ &= \Delta S_{ik} + \Delta D_{i_1,k_1} + (D_{i^*k^*}^p - D_{i\hat{k}}^p) + (D_{i-1,k}^p - D_{i^*k^*}^p)^+. \end{aligned}$$

Combining the above two cases we write:

$$\Delta D_{ik} = \Delta D_{i_1,k_1} + \Delta S_{ik} + \mathbf{1}(D_{i-1,k} \geq D_{i\hat{k}})W_{ik}^p + \mathbf{1}(D_{i-1,k} < D_{i\hat{k}}) \{X_{ik}^p + I_{ik}^p\}$$

where,

$$\begin{aligned} W_{ik}^p &= (D_{i^*k^*}^p - D_{i-1,k}^p)^+ \\ I_{ik}^p &= (D_{i-1,k}^p - D_{i^*k^*}^p)^+ \\ X_{ik}^p &= (D_{i^*k^*}^p - D_{i\hat{k}}^p). \end{aligned} \tag{9}$$

Similarly we define the following quantities which will be needed later on:

$$\begin{aligned} W_{ik} &= (D_{i\hat{k}} - D_{i-1,k})^+ \\ I_{ik} &= (D_{i-1,k} - D_{i\hat{k}})^+ \\ X_{ik} &= (D_{i+1,\hat{k}} - D_{i\hat{k}}). \end{aligned} \tag{10}$$

Observing that the parameter θ affects only service times at node 1, we have $\Delta S_{ik} = S_{ik}^p - S_{ik} = 0$ for $k \neq 1$, and $\Delta S_{i1} = S_{i1}^p - S_{i1} \geq 0$ (see Assumption **A.1**). Therefore,

$$\begin{aligned} \Delta D_{ik} &= \Delta D_{i_1,k_1} + \Delta S_{ik}\mathbf{1}(k=1) + \mathbf{1}(D_{i-1,k} \geq D_{i\hat{k}})W_{ik}^p \\ &\quad + \mathbf{1}(D_{i-1,k} < D_{i\hat{k}})X_{ik}^p + \mathbf{1}(D_{i-1,k} < D_{i\hat{k}})I_{ik}^p. \end{aligned} \tag{11}$$

The above equation provides an iterative expression for the perturbation of the i th departure from node k , ΔD_{ik} : the first term on the rhs is the perturbation in the departure epoch D_{i_1, k_1} which induces departure D_{ik} . The second term is the perturbation in the service time of customer C_{ik} (which is nonzero only if $k = 1$), the third term represents a perturbation that introduces an idle period at node k in the perturbed path; the fifth term represents a perturbation that causes the elimination of an idle period at node k present in the nominal path; and the fourth term represents a perturbation due to a change in the identity of the node that terminates an idle period at node k (whenever two or more nodes are supplying input to node k).

Define

$$\begin{aligned} \Delta_{ik}^1 &= \Delta S_{ik} \mathbf{1}(k = 1) + \mathbf{1}(D_{i-1, k} \geq D_{i\hat{k}}) W_{ik}^p \\ &\quad + \mathbf{1}(D_{i-1, k} < D_{i\hat{k}}) X_{ik}^p + \mathbf{1}(D_{i-1, k} < D_{i\hat{k}}) I_{ik}^p. \end{aligned} \quad (12)$$

We then have

$$\Delta D_{ik} = \Delta D_{i_1, k_1} + \Delta_{ik}^1. \quad (13)$$

By recursively applying (13) back to the beginning of the observed sample path, we have

$$\Delta D_{ik} = \sum_{(j, l) \in B_{ik}} \Delta_{jl}^1. \quad (14)$$

Define now the following three subsets of $\mathbf{N} \times \{1, 2, \dots, m\}$

$$\begin{aligned} P &= \{(j, l): l = 1\} \\ Q &= \{(j, l): D_{j-1, l} \geq D_{j\hat{l}}\} \\ R &= \{(j, l): D_{j-1, l} < D_{j\hat{l}}\} \end{aligned} \quad (15)$$

and the corresponding subsets of B_{ik}

$$\begin{aligned} P_{ik} &= P \cap B_{ik} \\ Q_{ik} &= Q \cap B_{ik} \\ R_{ik} &= R \cap B_{ik}. \end{aligned} \quad (16)$$

P_{ik} contains those events in B_{ik} that are departures from node 1 and, as we will see, it is related to the IPA part of our estimators. Q_{ik} (resp. R_{ik}) is the set of all events in B_{ik} that are arrivals to a busy (resp. idle) node.

Using the above definitions and (11), (14) yields the following expression for the change in the departure time D_{ik}

$$\Delta D_{ik} = \sum_{(j, l) \in P_{ik}} \Delta S_{jl} + \sum_{(j, l) \in Q_{ik}} W_{jl}^p + \sum_{(j, l) \in R_{ik}} X_{jl}^p + \sum_{(j, l) \in R_{ik}} I_{jl}^p \quad (17)$$

and taking expectations

$$E[\Delta D_{ik}] = E \sum_{(j, l) \in P_{ik}} \Delta S_{jl} + E \sum_{(j, l) \in Q_{ik}} W_{jl}^p + E \sum_{(j, l) \in R_{ik}} X_{jl}^p + E \sum_{(j, l) \in R_{ik}} I_{jl}^p. \quad (18)$$

In what follows, we will first concentrate on defining an appropriate characterization for each term in (18).

3.2. Defining Sample Path Characterizations

Our task now is to select an appropriate sample path characterization, which, as in Bao, Cassandras, and Zazanis (1996), is a sequence of additional conditionings, one for each term in (18). We remind the reader that U_{jl} is the routing indicator of the event (departure) e_{jl} occurring at time D_{jl} . Let us also denote by e'_{jl} the event that immediately follows e_{jl} and by U'_{jl} its associated routing indicator. Let Λ_{jl} denote the information in the sample path up to (and including) time D_{jl} (including also the value of the routing indicator U_{jl}) as well as the identity of the next event e'_{jl} and the value of its routing indicator, U'_{jl} .

Denote by $\mathcal{M}(t)$ the set of nodes that are busy at time t . The process $\{\mathcal{M}(t); t \geq 0\}$ taking values on the class of subsets of the set of nodes, $\{1, 2, \dots, m\}$, is assumed to have *right-continuous paths*. In particular, $\mathcal{M}(D_{j-1,l})$ is the set of nodes that are busy immediately after the $(j - 1)$ th departure from node l . Next, define

$$S_k^a(t) = \begin{cases} \text{age of service time of customer present at node } k \text{ at time } t & \text{if } k \in \mathcal{M}(t) \\ 0 & \text{if } k \notin \mathcal{M}(t) \end{cases},$$

$$S_k^r(t) = \begin{cases} \text{residual service time of customer present at node } k \text{ at time } t & \text{if } k \in \mathcal{M}(t) \\ 0 & \text{if } k \notin \mathcal{M}(t) \end{cases},$$

$$S_k(t) = \begin{cases} \text{total service time of customer present at node } k \text{ at time } t & \text{if } k \in \mathcal{M}(t) \\ 0 & \text{if } k \notin \mathcal{M}(t) \end{cases},$$

where, of course, $S_k(t) = S_k^a(t) + S_k^r(t)$. Hence, we define the families of processes $\{S_k^a(t); t \geq 0, k = 1, \dots, m\}$, $\{S_k^r(t); t \geq 0, k = 1, \dots, m\}$, $\{S_k(t); t \geq 0, k = 1, \dots, m\}$, and, again, we consider the *right-continuous* versions of these processes.

We now impose a sequence of conditions, z_{jl} , associated with events e_{jl} as follows: As we have seen, Λ_{jl} contains all events, event times, and routing indicators up to and including D_{jl} . In addition it includes *the identity of the next event, e'_{jl} , and its routing indicator, U'_{jl}* . In particular we point out that Λ_{jl} contains the *ages* of the service time processes of all active nodes at D_{jl} .

Now let D'_{jl} denote the time of the next event, e'_{jl} , and define the condition z_{jl} as follows:

$$z_{jl} = \{\Lambda_{jl}, \mathcal{M}(D'_{jl}), S_q^r(D'_{jl}); q \in \mathcal{M}(D'_{jl})\}. \quad (19)$$

Note that, in addition to the history of the process up to the event time and the identity of the next event, z_{jl} includes the list of active nodes at the time the next event occurs, as well as the residual service times at the time of the next event. It does *not*, however, contain the *time of the next event, D'_{jl}* . Readers familiar with the Smoothed Perturbation Analysis (SPA) methodology will appreciate the importance of carefully selecting the conditions z_{jl} : “just enough” information from the observed sample path is included in (19) to allow us to “smooth out” discontinuities that prevent a PA estimator from being unbiased. The justification for the specific choices made here will become clear in the analysis that follows.

Remark. In a simulation setting, conditioning on information such as the identity of the next event or residual service times presents no implementation problem, since this information is routinely available. In a real-time setting, on the other hand, future information is not

available. This, however, presents no problem for the implementation of the estimators we will derive; it simply requires additional memory associated with certain events. For example, if one requires residual time data for certain active nodes when event e_{jl} occurs, one should wait until all corresponding service completions take place and then proceed with whatever computation is involved associated with e_{jl} .

Returning to (18), our objective now becomes to evaluate explicitly appropriate *conditional expectations* of the terms involving I_{jl}^p , W_{jl}^p , and X_{ji}^p .

Define now the following three subsets of $\mathbf{N} \times \{1, 2, \dots, m\}$:

$$\begin{aligned} Q^* &= \{(j, l): D_{j-1, l} > D_{j\hat{l}}, D_{j-1, l} \neq \widehat{D_{j+1, l}}, \\ &\quad \text{and no events occur in } (D_{j\hat{l}}, D_{j-1, l})\}, \\ R^* &= \{(j, l): D_{j-1, l} < D_{j\hat{l}}, \text{ and no events occur in } (D_{j-1, l}, D_{j\hat{l}})\}, \\ R' &= \{(j, l): D_{j-1, l} < D_{j\hat{l}}, D_{jl} \neq \widehat{D_{j+1, l}}, \\ &\quad \text{and no events occur in } (D_{j\hat{l}}, \widehat{D_{j+1, l}})\}. \end{aligned} \quad (20)$$

In particular, we point out that $\mathbf{1}((j, l) \in Q^*) \in \Lambda_{j\hat{l}}$, $\mathbf{1}((j, l) \in R^*) \in \Lambda_{j-1, l}$, $\mathbf{1}((j, l) \in R') \in \Lambda_{j-1, l}$. We also define

$$\begin{aligned} Q_{ik}^* &= B_{ik} \cap Q^*, \\ R_{ik}^* &= B_{ik} \cap R^*, \\ R'_{ik} &= B_{ik} \cap R'. \end{aligned} \quad (21)$$

For an interpretation of these three sets, note that Q_{ik}^* is a subset of Q_{ik} and R_{ik}^* , R'_{ik} are subsets of R_{ik} . In Q_{ik}^* , we exclude events such that the $(j-1)$ th departure from node l returns to l to become the $(j+1)$ th customer, or such that the waiting period of the j th customer contains at least one other event anywhere in the system. In R_{ik}^* , we exclude events such that the idle period following the $(j-1)$ th departure at node l contains at least one other event anywhere in the system. Finally, in R'_{ik} , we exclude events such that the j th departure from node l immediately returns to l to become the $(j+1)$ th customer, or such that the interval between two successive arrivals at l following an idle period contains at least one other event anywhere in the system.

The above definitions are motivated by the following considerations: As we will see in the sequel,

$$E \left[\sum_{(j, l) \in Q_{ik}} W_{jl}^p + \sum_{(j, l) \in R_{ik}} X_{jl}^p + \sum_{(j, l) \in R_{ik}} I_{jl}^p \right] = O(\Delta\theta^2)$$

and therefore will contribute to the second derivative estimator. However, as we will see,

$$E \sum_{(j, l) \in Q_{ik} \setminus Q_{ik}^*} W_{jl}^p = o(\Delta\theta^2)$$

which means that we need not consider any events in Q_{ik} which do not belong to Q_{ik}^* for

the purpose of estimating second derivatives. Similarly,

$$E \sum_{(j,l) \in R_{ik} \setminus R'_{ik}} I_{jl}^p = o(\Delta\theta^2), \quad \text{and} \quad E \sum_{(j,l) \in R_{ik} \setminus R'_{ik}} X_{jl}^p = o(\Delta\theta^2).$$

We refer to events in B_{ik} that do not belong to Q_{ik}^* , R_{ik}^* , or R'_{ik} as *non-critical* and we will show that their contribution is indeed of order $o(\Delta\theta^2)$ in the next subsection.

3.3. Conditional Expectations for Non-Critical Events

We begin by considering all non-critical events in the induced event set B_{ik} , i.e., those that do not belong to R_{ik}^* , R'_{ik} or Q_{ik}^* . We want to show that the contribution of such events is of order $o(\Delta\theta^2)$.

LEMMA 1 *For any events e_{ik}, e_{jl} , such that $D_{ik}(\theta) < D_{jl}(\theta)$ and the interval $(D_{ik}(\theta), D_{jl}(\theta))$ contains at least one event,*

$$\frac{1}{\Delta\theta^2} P(D_{ik}(\theta + \Delta\theta) > D_{jl}(\theta + \Delta\theta) | \Lambda_{ik}) \leq V_{ik}(\Delta\theta), \quad (22)$$

where V_{ik} is a random variable such that $EV_{ik}(\Delta\theta) < \infty$.

Proof: See Appendix. ■

This lemma is similar in spirit to Glasserman and Gong (1991, Lemma 3). We condition with respect to the information available at time D_{ik} and argue that the probability that a future event in the nominal path, namely the departure D_{jl} , occurring before D_{ik} in the perturbed path is $o(\Delta\theta)$, when another event intervenes. This result is then used in the following lemma to show that the contribution of non-critical events will not play a role in the final expression for the second derivative of $E[D_{ik}]$ and the resulting estimator.

LEMMA 2 *For the sets Q, R , defined in (15) and the sets Q^*, R^*, R' , defined in (20):*

- (i) *If $(j, l) \in Q \setminus Q^*$, then $\frac{1}{\Delta\theta^2} E[W_{jl}^p | \Lambda_{j\hat{l}}] \leq K_{jl}^W(\Delta\theta)$, where $K_{jl}^W(\Delta\theta) \in \Lambda_{j\hat{l}}$, $EK_{jl}^W < \infty$, and $\lim_{\Delta\theta \rightarrow 0} K_{jl}^W(\Delta\theta) = 0$.*
- (ii) *If $(j, l) \in R \setminus R^*$, then $\frac{1}{\Delta\theta^2} E[I_{jl}^p | \Lambda_{j-1,l}] \leq K_{jl}^I(\Delta\theta)$, where $K_{jl}^I(\Delta\theta) \in \Lambda_{j-1,l}$, $EK_{jl}^I < \infty$, and $\lim_{\Delta\theta \rightarrow 0} K_{jl}^I(\Delta\theta) = 0$.*
- (iii) *If $(j, l) \in R \setminus R'$, then $\frac{1}{\Delta\theta^2} E[X_{jl}^p | \Lambda_{j\hat{l}}] \leq K_{jl}^X(\Delta\theta)$, where $K_{jl}^X(\Delta\theta) \in \Lambda_{j\hat{l}}$, $EK_{jl}^X < \infty$, and $\lim_{\Delta\theta \rightarrow 0} K_{jl}^X(\Delta\theta) = 0$.*

Proof: See Appendix. ■

3.4. Conditional Expectations for Critical Events

Next, we consider critical events in the induced event set B_{ik} . We first need to obtain some conditional densities. In particular, suppose event $e_{j-1,l}$ occurs at time $D_{j-1,l}$ and is followed by an idle period of duration I_{jl} . In the following lemma, an expression for the conditional density of I_{jl} given $z_{j-1,l}$, $g_{jl}^I(x|z_{j-1,l})$, is obtained for the case where no events occur during the idle period, or, equivalently, $e_{j\hat{l}}$ occurs immediately after $e_{j-1,l}$ (with obvious dual results for W_{jl} and X_{jl}).

LEMMA 3

- (i) If event $e_{j-1,l}$ initiates an idle period of length I_{jl} at node l and the next event to occur is $e_{j\hat{l}}$, i.e., the event terminating the idle period, then the conditional density of this idle period given $z_{j-1,l}$, $g_{jl}^I(x|z_{j-1,l})$, is

$$g_{jl}^I(x|z_{j-1,l}) = \frac{f_l(S_l^a(D_{j-1,l}) + x) [\prod_{\substack{q \in \mathcal{M}(D_{j-1,l}) \\ q \neq \hat{l}}} f_q(S_q(D_{j-1,l}) - I_{jl} + x)]}{\int_0^\infty f_l(S_l^a(D_{j-1,l}) + u) [\prod_{\substack{q \in \mathcal{M}(D_{j-1,l}) \\ q \neq \hat{l}}} f_q(S_q(D_{j-1,l}) - I_{jl} + u)] du}. \quad (23)$$

- (ii) If customer C_{jl} upon arrival to node l at time $D_{j\hat{l}}$ finds the server busy and the next event to occur is $e_{j-1,l}$, $D_{j-1,l} \neq \widehat{D_{j+1,\hat{l}}}$ then the conditional density of this customer's waiting time given $z_{j\hat{l}}$, $g_{jl}^W(x|z_{j\hat{l}})$, is

$$g_{jl}^W(x|z_{j\hat{l}}) = \frac{f_l(S_l^a(D_{j\hat{l}}) + x) [\prod_{\substack{q \in \mathcal{M}(D_{j\hat{l}}) \\ q \neq \hat{l}}} f_q(S_q(D_{j\hat{l}}) - W_{jl} + x)]}{\int_0^\infty f_l(S_l^a(D_{j\hat{l}}) + u) [\prod_{\substack{q \in \mathcal{M}(D_{j\hat{l}}) \\ q \neq \hat{l}}} f_q(S_q(D_{j\hat{l}}) - W_{jl} + u)] du}. \quad (24)$$

- (iii) If event $e_{j\hat{l}}$ initiates a busy period at node l and the next event to occur is $e_{\widehat{j+1,\hat{l}}}$, i.e., the second arrival event at node l with inter-event time of X_{jl} , then the conditional density of this period given $z_{j\hat{l}}$, $g_{jl}^X(x|z_{j\hat{l}})$, is

$$g_{jl}^X(x|z_{j\hat{l}}) = \frac{f_l(S_l^a(D_{j\hat{l}}) + x) [\prod_{\substack{q \in \mathcal{M}(D_{j\hat{l}}) \\ q \neq \hat{l}}} f_q(S_q(D_{j\hat{l}}) - X_{jl} + x)]}{\int_0^\infty f_l(S_l^a(D_{j\hat{l}}) + u) [\prod_{\substack{q \in \mathcal{M}(D_{j\hat{l}}) \\ q \neq \hat{l}}} f_q(S_q(D_{j\hat{l}}) - X_{jl} + u)] du}. \quad (25)$$

Proof: See Appendix. ■

We now consider the conditional expectation $E[I_{jl}^D | z_{j-1,l}]$. Unlike Lemma 1, however, the condition here is $z_{j-1,l}$, not just $\Lambda_{j-1,l}$. In the next lemma we show that the contribution of critical events is no longer of order $o(\Delta\theta^2)$, but instead it depends on the conditional density function $g_{jl}^I(\cdot | z_{j-1,l})$ and a quantity Y_{jl}^I defined next. Let S_{jl}^a denote the age of the

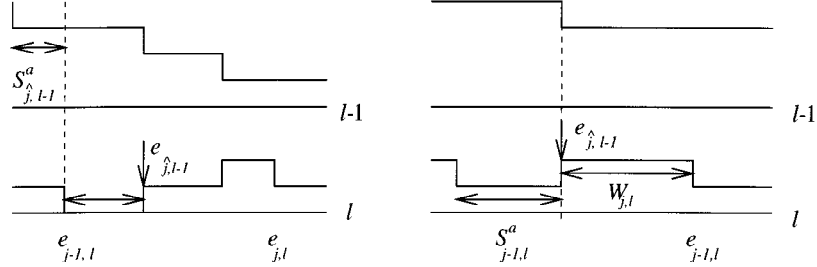


Figure 1. Illustrating the definitions of $S_{j\hat{l},l}^a$ and $S_{j-1,l}^a$.

service time of customer $C_{j\hat{l}}$ at the time $e_{j-1,l}$ occurs. In other words, using the definition of a service time age, we simply set $S_{j\hat{l}}^a = S_{j\hat{l}}^a(D_{j-1,l})$.

As illustrated in Figure 1 (where, for simplicity, we have set $\hat{l} = l - 1$), this corresponds to any event $e_{j\hat{l}}$ which is the first departure in a busy period of node l initiated by $e_{j\hat{l}}$ following an idle period of length $I_{j\hat{l}}$. The crucial observation here is that $S_{j\hat{l}}^a$ belongs to $z_{j-1,l}$, the condition imposed in $E[I_{j\hat{l}}^p | z_{j-1,l}]$. Let us then define

$$Y_{j\hat{l}}^I = \sum_{(q,r) \in P_{j-1,l}} \frac{\partial}{\partial \theta} S_{qr} - \sum_{\substack{(q,r) \in P_{j\hat{l}} \\ (q,r) \neq (j,\hat{l})}} \frac{\partial}{\partial \theta} S_{qr} - \frac{\partial}{\partial \theta} S_{j\hat{l}}^a. \quad (26)$$

Similarly, we use $S_{j-1,l}^a$ to denote the age of the service time of customer $C_{j-1,l}$ at time $D_{j\hat{l}}$, i.e. when $e_{j\hat{l}}$ occurs (see Figure 1). Accordingly, we define

$$Y_{j\hat{l}}^W = \sum_{(q,r) \in P_{j\hat{l}}} \frac{\partial}{\partial \theta} S_{qr} - \sum_{\substack{(q,r) \in P_{j-1,l} \\ (q,r) \neq (j-1,l)}} \frac{\partial}{\partial \theta} S_{qr} - \frac{\partial}{\partial \theta} S_{j-1,l}^a. \quad (27)$$

Finally, we use $S_{j+1,\hat{l}}^a$ to denote the age of the service time of customer $C_{j\hat{l}}$ at time $D_{j+1,\hat{l}}$ occurs (see Figure 1). Accordingly, we define

$$Y_{j\hat{l}}^X = \sum_{(q,r) \in P_{j\hat{l}}} \frac{\partial}{\partial \theta} S_{qr} - \sum_{\substack{(q,r) \in P_{j+1,\hat{l}} \\ (q,r) \neq (j+1,\hat{l})}} \frac{\partial}{\partial \theta} S_{qr} - \frac{\partial}{\partial \theta} S_{j+1,\hat{l}}^a. \quad (28)$$

LEMMA 4 For the sets Q^* , R^* , R' defined in (20)

(i) if $(j, l) \in R^*$, then

$$\lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[I_{j\hat{l}}^p | z_{j-1,l}] = \frac{1}{2} g_{j\hat{l}}^I(0 | z_{j-1,l}) (Y_{j\hat{l}}^I)^+. \quad (29)$$

(ii) if $(j, l) \in Q^*$, then

$$\lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[W_{j\hat{l}}^p | z_{j\hat{l}}] = \frac{1}{2} g_{j\hat{l}}^W(0 | z_{j\hat{l}}) (Y_{j\hat{l}}^W)^+. \quad (30)$$

(iii) if $(j, l) \in R'$, then

$$\lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[X_{jl}^p | z_{j\hat{l}}] = -\frac{1}{2} g_{jl}^W(0 | z_{j\hat{l}}) ([Y_{jl}^X]^+)^2. \quad (31)$$

Proof: See Appendix. ■

3.5. Second Derivative Estimators using Conditional Expectations

Returning to (18)

$$E[\Delta D_{ik}] = E \sum_{(j,l) \in P_{ik}} \Delta S_{jl} + E \sum_{(j,l) \in Q_{ik}} W_{jl}^p + E \sum_{(j,l) \in R_{ik}} I_{jl}^p + E \sum_{(j,l) \in R_{ik}} X_{jl}^p. \quad (32)$$

Combining the results from the previous sections, we will now replace the rhs above by terms involving conditional expectations. This leads to our main result, Theorem 1 below.

First, looking at (29) and (30) in Lemma 4, let us set

$$Z_{jl}^I = \frac{1}{2} g_{jl}^I(0 | z_{j-1,l}) \left([Y_{jl}^I]^+ \right)^2, \quad (33)$$

$$Z_{jl}^W = \frac{1}{2} g_{jl}^W(0 | z_{j\hat{l}}) \left([Y_{jl}^W]^+ \right)^2, \quad (34)$$

$$Z_{jl}^X = \frac{1}{2} g_{jl}^X(0 | z_{j\hat{l}}) \left([Y_{jl}^X]^+ \right)^2. \quad (35)$$

We are now ready to state our main result.

THEOREM 1 With Z_{jl}^I , Z_{jl}^W and Z_{jl}^X as defined above,

$$\begin{aligned} E[\Delta D_{ik}] &= E \left[\sum_{(j,l) \in B_{ik}} \Delta S_{jl} \mathbf{1}(l=1) \right] \\ &\quad + E \left[\sum_{(j,l) \in R_{ik}^*} Z_{jl}^I + \sum_{(j,l) \in Q_{ik}^*} Z_{jl}^W - \sum_{(j,l) \in R_{ik}'} Z_{jl}^X \right] \Delta\theta^2 + o(\Delta\theta^2). \end{aligned} \quad (36)$$

Proof: Recalling (12) and (14), we have

$$\begin{aligned} \Delta D_{ik} &= \sum_{(j,l) \in B_{ik}} \Delta S_{jl} \mathbf{1}(l=1) + \sum_{(j,l) \in B_{ik}} W_{jl}^p \mathbf{1}(D_{j-1,l} > D_{j\hat{l}}) \\ &\quad + \sum_{(j,l) \in B_{ik}} X_{jl}^p \mathbf{1}(D_{j-1,l} < D_{j\hat{l}}) + \sum_{(j,l) \in B_{ik}} I_{jl}^p \mathbf{1}(D_{j-1,l} < D_{j\hat{l}}) \end{aligned}$$

Comparing the right-hand-side above with the right-hand-side of (36), note that the first

summations are identical. Let us next consider the remaining three sums above. The last one can be written as

$$\sum_{(j,l) \in B_{ik}} I_{jl}^p \mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) = \sum_{l=1}^m \sum_{j=1}^{\infty} I_{jl}^p \mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) \mathbf{1}[(j,l) \in B_{ik}].$$

Taking expectations and using Fubini's theorem we obtain

$$E \sum_{(j,l) \in B_{ik}} I_{jl}^p \mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) = \sum_{l=1}^m \sum_{j=1}^{\infty} E[I_{jl}^p \mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) \mathbf{1}((j,l) \in B_{ik})]. \quad (37)$$

Define the set

$$\Gamma = \{(j,l): \quad j = 1, 2, \dots, l = 1, \dots, m, \quad (38)$$

$$\text{and no events occur in } (\min\{D_{j-1,l}, D_{\hat{j}l}\}, \max\{D_{j-1,l}, D_{\hat{j}l}\})\}.$$

Recalling the definitions of the sets R_{ik} and R_{ik}^* we have

$$\mathbf{1}((j,l) \in B_{ik}) \mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) = \mathbf{1}((j,l) \in R_{ik}), \quad (39)$$

$$\mathbf{1}((j,l) \in R_{ik}) \mathbf{1}((j,l) \in \Gamma) = \mathbf{1}((j,l) \in R_{ik}^*). \quad (40)$$

We next note that the information in $\Lambda_{j-1,l}$ is enough to determine whether $D_{j-1,l} < D_{\hat{j}l}$, and whether (j,l) belongs to Γ or not, that is

$$\mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) \in \Lambda_{j-1,l} \subset \mathcal{Z}_{j-1,l}, \quad (41)$$

and

$$\mathbf{1}((j,l) \in \Gamma) \in \Lambda_{j-1,l} \subset \mathcal{Z}_{j-1,l}. \quad (42)$$

The above two remarks are crucial in what follows. (41) is simply due to the fact that $\Lambda_{j-1,l}$ includes the whole history of the process up to time $D_{j-1,l}$, so that by that time we obviously know whether $D_{\hat{j}l}$ has occurred (in which case the inequality is satisfied) or not. To check (42) we need to recall that $\Lambda_{j-1,l}$ contains the identity of the next event: If $D_{j-1,l} < D_{\hat{j}l}$, observe that no events occur in $(D_{j-1,l}, D_{\hat{j}l})$ iff $e_{\hat{j}l}$ is the next event. If on the other hand $D_{j-1,l} > D_{\hat{j}l}$ then we can tell whether any events occurred in $(D_{\hat{j}l}, D_{j-1,l})$ or not since this time interval clearly belongs to the past history of the process, $\Lambda_{j-1,l}$.

With these observations in mind, let us now return to the rhs of (37) and examine a typical term in the double summation. Noting that $\mathbf{1}((j,l) \in B_{ik}) = \mathbf{1}((j,l) \in B_{ik} \cap \Gamma) + \mathbf{1}((j,l) \in B_{ik} \cap \Gamma^c)$, we have

$$E \left[I_{jl}^p \mathbf{1}((j,l) \in B_{ik}) \mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) \right] = E \left[\mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) \right. \\ \left. \times I_{jl}^p \mathbf{1}((j,l) \in B_{ik} \cap \Gamma) \right] \\ + E \left[\mathbf{1}(D_{j-1,l} < D_{\hat{j}l}) \right. \\ \left. \times I_{jl}^p \mathbf{1}((j,l) \in B_{ik} \cap \Gamma^c) \right].$$

Next, take conditional expectations on the rhs of the above display, conditioning on $z_{j-1,l}$ for the first term above and on $\Lambda_{j-1,l} \subset z_{j-1,l}$ for the second term:

$$\begin{aligned}
& E \left[I_{jl}^p \mathbf{1}((j, l) \in B_{ik}) \mathbf{1}(D_{j-1,l} < D_{j\hat{l}}) \right] \\
&= E \left[E[\mathbf{1}(D_{j-1,l} < D_{j\hat{l}}) I_{jl}^p \mathbf{1}((j, l) \in B_{ik} \cap \Gamma) \mid z_{j-1,l}] \right] \\
&\quad + E \left[E[\mathbf{1}(D_{j-1,l} < D_{j\hat{l}}) I_{jl}^p \mathbf{1}((j, l) \in B_{ik} \cap \Gamma^c) \mid \Lambda_{j-1,l}] \right] \\
&= E \left[\mathbf{1}(D_{j-1,l} < D_{j\hat{l}}) E[I_{jl}^p \mathbf{1}((j, l) \in B_{ik} \cap \Gamma) \mid z_{j-1,l}] \right] \\
&\quad + E \left[\mathbf{1}(D_{j-1,l} < D_{j\hat{l}}) E[I_{jl}^p \mathbf{1}((j, l) \in B_{ik} \cap \Gamma^c) \mid \Lambda_{j-1,l}] \right], \tag{43}
\end{aligned}$$

where the last step above follows from (41).

Examine now the two terms on the rhs of (43) separately, starting with the second term. First, note that $I_{jl}^p \geq 0$ w.p.1. Hence the conditional expectation in the second term is dominated by $E[I_{jl}^p \mid \Lambda_{j-1,l}]$. By virtue of Lemma 2 it follows that the second term is of order $o(\Delta\theta^2)$.

Turning our attention to the first term, we next argue that the random variables $\mathbf{1}((j, l) \in B_{ik} \cap \Gamma)$ and $I_{jl}^p \mathbf{1}((j, l) \in \Gamma)$ are conditionally independent given $z_{j-1,l}$. Indeed, recalling the definition of I_{jl}^p in (9), we see that it clearly depends only on events that have occurred up to time $D_{j\hat{l}}$. Furthermore, note that, on Γ , the event immediately after $e_{j-1,l}$ is $e_{j\hat{l}}$. Therefore, by the definition in (19), $z_{j-1,l}$ contains the history of the process up to time $D_{j-1,l}$, as well as the identity of the next event, i.e., $e_{j\hat{l}}$, and hence the number of customers at each node at time $D_{j\hat{l}}$. In addition, the residual service times at the active nodes at time $D_{j\hat{l}}$ are also part of $z_{j-1,l}$. In other words, the complete state information at time $D_{j\hat{l}}$ (though not the time $D_{j\hat{l}}$ itself) is part of $z_{j-1,l}$, provided that $(j, l) \in \Gamma$.

On the other hand, $(j, l) \in B_{ik}$ is a statement about the future evolution of the system (after time $D_{j\hat{l}}$). It is completely determined by the segment of the sample path of the system in the interval (D_{jl}, D_{ik}) , and depends on the succession of events but not on ‘‘absolute time’’; in particular, it is independent of $D_{j\hat{l}}$. Thus, it follows from the Strong Markov Property that, on $(j, l) \in \Gamma$, the future evolution of the network (and in particular $\mathbf{1}(j, l) \in B_{ik}$) is conditionally independent of the past (and in particular I_{jl}^p) given the ‘‘present’’ state, i.e. the state at time $D_{j\hat{l}}$.

Since $z_{j-1,l}$ contains the complete state of the network at time $D_{j\hat{l}}$, we see that $\mathbf{1}((j, l) \in B_{ik} \cap \Gamma)$ and $I_{jl}^p \mathbf{1}((j, l) \in \Gamma)$ are conditionally independent given $z_{j-1,l}$. Consequently,

$$\begin{aligned}
E[I_{jl}^p \mathbf{1}((j, l) \in B_{ik} \cap \Gamma) \mid z_{j-1,l}] &= E[\mathbf{1}((j, l) \in \Gamma) I_{jl}^p \mathbf{1}((j, l) \in \Gamma) \mid z_{j-1,l}] \\
&\quad \times E[\mathbf{1}((j, l) \in B_{ik}) \mid z_{j-1,l}] \\
&= \mathbf{1}((j, l) \in \Gamma) E[I_{jl}^p \mid z_{j-1,l}] \\
&\quad \times E[\mathbf{1}((j, l) \in B_{ik}) \mid z_{j-1,l}],
\end{aligned}$$

where, in the last equation we have used (42).

From Lemma 4, using the definition of Z_{jl}^l in (33), we have

$$E[I_{jl}^p \mid z_{j-1,l}] = Z_{jl}^l \Delta\theta^2 + o(\Delta\theta^2) \quad \text{on } \{(j, l) \in \Gamma\},$$

where Z_{jl}^l belongs to $z_{j-1,l}$. This in turn yields

$$E \left[I_{jl}^p \mathbf{1}((j, l) \in B_{ik} \cap \Gamma) | z_{j-1,l} \right] = \mathbf{1}((j, l) \in \Gamma) E \left[\mathbf{1}((j, l) \in B_{ik}) | z_{j-1,l} \right] Z_{jl}^l \Delta\theta^2 + o(\Delta\theta^2).$$

Therefore, returning to (37) and combining all of the above results,

$$\begin{aligned} & E \sum_{(j,l) \in B_{ik}} I_{jl}^p \mathbf{1}(D_{j-1,l} < D_{\hat{j}}) \\ &= \sum_{l=1}^m \sum_{j=1}^{\infty} E \left[E \left[\mathbf{1}((j, l) \in B_{ik}) | z_{j-1,l} \right] \mathbf{1}((j, l) \in \Gamma) Z_{jl}^l \mathbf{1}(D_{j-1,l} < D_{\hat{j}}) \right] \Delta\theta^2 \\ &\quad + o(\Delta\theta^2) \\ &= \sum_{l=1}^m \sum_{j=1}^{\infty} E \left[E \left[Z_{jl}^l \mathbf{1}((j, l) \in B_{ik} \cap \Gamma) \mathbf{1}(D_{j-1,l} < D_{\hat{j}}) | z_{j-1,l} \right] \right] \Delta\theta^2 + o(\Delta\theta^2) \\ &= \Delta\theta^2 E \sum_{l=1}^m \sum_{j=1}^{\infty} Z_{jl}^l \mathbf{1}((j, l) \in B_{ik} \cap \Gamma) \mathbf{1}(D_{j-1,l} < D_{\hat{j}}) + o(\Delta\theta^2) \\ &= \Delta\theta^2 E \sum_{(j,l) \in R_{ik}^*} Z_{jl}^l + o(\Delta\theta^2). \end{aligned}$$

Similarly, we can show that

$$\begin{aligned} E \sum_{(j,l) \in B_{ik}} W_{jl}^p \mathbf{1}(D_{j-1,l} > D_{\hat{j}}) &= E \sum_{(j,l) \in Q_{ik}^*} Z_{jl}^W \Delta\theta^2 + o(\Delta\theta^2), \\ E \sum_{(j,l) \in B_{ik}} X_{jl}^p \mathbf{1}(D_{j-1,l} < D_{\hat{j}}) &= -E \sum_{(j,l) \in R_{ik}'} Z_{jl}^X \Delta\theta^2 + o(\Delta\theta^2), \end{aligned}$$

which completes the proof. \blacksquare

The desired first and second derivative estimators can be readily obtained from the above theorem and the following

LEMMA 5

$$E \left[\sum_{(j,1) \in P_{ik}} \Delta S_{j1} \right] = E \left[\sum_{(j,1) \in P_{ik}} \frac{\partial}{\partial \theta} S_{j1} \right] \Delta\theta + \frac{1}{2} E \left[\sum_{(j,1) \in P_{ik}} \frac{\partial^2}{\partial \theta^2} S_{j1} \right] \Delta\theta^2 + o(\Delta\theta).$$

Proof: See Appendix. \blacksquare

Using Lemma 5, (36) can be written as

$$\begin{aligned}
E[\Delta D_{ik}] &= E \left[\sum_{(j,l) \in P_{ik}} \frac{\partial}{\partial \theta} S_{jl} \mathbf{1}(l=1) \right] \Delta \theta \\
&+ \frac{1}{2} E \left[\sum_{(j,l) \in P_{ik}} \frac{\partial^2}{\partial \theta^2} S_{jl} \mathbf{1}(l=1) \right. \\
&\quad + \sum_{(j,l) \in Q_{ik}^*} g_{jl}^W(0|z_{j\hat{l}}) \left([Y_{jl}^W]^+ \right)^2 + \sum_{(j,l) \in R_{ik}^*} g_{jl}^I(0|z_{j-1,l}) \left([Y_{jl}^I]^+ \right)^2 \\
&\quad \left. - \sum_{(j,l) \in R'_{ik}} g_{jl}^X(0|z_{j\hat{l}}) \left([Y_{jl}^X]^+ \right)^2 \right] \Delta \theta^2 + o(\Delta \theta^2). \quad (44)
\end{aligned}$$

From the expression above, letting $(i, k) = (N, 1)$, we obtain the following estimator for the first-order derivative of \bar{D}

$$\left[\frac{\partial}{\partial \theta} \bar{D} \right]_{\text{est.}} = \frac{1}{N} \sum_{(j,1) \in P_{N1}} \frac{\partial}{\partial \theta} S_{j1}. \quad (45)$$

Thus, we have recovered the standard IPA first derivative estimator for this type of network (e.g. Cao 1990).

Similarly, our second derivative estimator of \bar{D} is given by

$$\begin{aligned}
\left[\frac{\partial^2}{\partial \theta^2} \bar{D} \right]_{\text{est.}} &= \frac{1}{N} \left\{ \sum_{(j,1) \in P_{N1}} \frac{\partial^2}{\partial \theta^2} S_{j1} \right. \\
&\quad + \sum_{(j,l) \in Q_{N1}^*} g_{jl}^W(0|z_{j\hat{l}}) \left(\left[\sum_{(q,1) \in P_{j\hat{l}}} \frac{\partial}{\partial \theta} S_{q1} \right. \right. \\
&\quad \quad \quad \left. \left. - \sum_{\substack{(q,1) \in P_{j-1,l} \\ (q,1) \neq (j-1,l)}} \frac{\partial}{\partial \theta} S_{q1} - \frac{\partial}{\partial \theta} S_{j-1,l}^a \right]^+ \right)^2 \\
&\quad + \sum_{(j,l) \in R_{N1}^*} g_{jl}^I(0|z_{j-1,l}) \left(\left[\sum_{(q,1) \in P_{j-1,l}} \frac{\partial}{\partial \theta} S_{q1} \right. \right. \\
&\quad \quad \quad \left. \left. - \sum_{\substack{(q,1) \in P_{j\hat{l}} \\ (q,1) \neq (j\hat{l})}} \frac{\partial}{\partial \theta} S_{q1} - \frac{\partial}{\partial \theta} S_{j\hat{l}}^a \right]^+ \right)^2
\end{aligned}$$

$$\begin{aligned}
& - \sum_{(j,l) \in R_{N_1}^*} g_{jl}^X(0|z_{j\hat{l}}) \left(\left[\sum_{(q,1) \in P_{j\hat{l}}} \frac{\partial}{\partial \theta} S_{q1} - \sum_{\substack{(q,1) \in P_{j+1\hat{l}} \\ (q,1) \neq (j+1, \hat{l})}} \frac{\partial}{\partial \theta} S_{q1} \right. \right. \\
& \left. \left. - \frac{\partial}{\partial \theta} S_{j+1, \hat{l}}^a \right]^+ \right)^2 \Bigg\}. \tag{46}
\end{aligned}$$

It readily follows from Theorem 1 that the two derivative estimators above are indeed unbiased. It is worth pointing out that in the second, third and fourth terms of (45) above not all (j, l) belonging to the critical even sets $Q_{N_1}^*$ and $R_{N_1}^*$ contribute to the estimator. This is because the corresponding differences in the $[\cdot]^+$ in these terms may be negative.

An explicit algorithm for implementing the second-order derivative estimator above is provided in the next section.

4. The Estimation Algorithm

Although the expression for the second derivative estimator in (45) is rather complicated, we will present in this section an algorithm for implementing both first and second derivative estimators which is quite simple.

We begin with the first derivative estimator in (45). Defining

$$L_1(i, k) = \sum_{(j,1) \in P_{ik}} \frac{\partial}{\partial \theta} S_{j1}, \tag{47}$$

we have:

$$\left[\frac{\partial}{\partial \theta} \bar{D} \right]_{\text{est.}} = \frac{1}{N} L_1(N, 1).$$

Now, let e_{i_1, k_1} be the event that induces e_{ik} . From the definition of B_{ik} in section 2, we know that

$$B_{ik} = B_{i_1, k_1} \cup \{(i, k)\}; \quad P_{ik} = P_{i_1, k_1} \cup \{(i, k)\mathbf{1}(k=1)\},$$

where we agree to let $\{(i, k)\mathbf{1}(k=1)\}$ be the empty set if $\mathbf{1}(k=1) = 0$. Therefore, we have the following iterative scheme for obtaining $L_1(i, k)$ from $L_1(i_1, k_1)$:

$$L_1(i, k) = \sum_{(j,1) \in P_{i_1, k_1} \cup \{(i,k)\mathbf{1}(k=1)\}} \frac{\partial}{\partial \theta} S_{j1} = L_1(i_1, k_1) + \mathbf{1}(k=1) \frac{\partial}{\partial \theta} S_{ik}. \tag{48}$$

We can obtain a similar iterative scheme for the second derivative. Define

$$L_2(i, k) = N \left[\frac{\partial^2}{\partial \theta^2} \bar{D}(\theta) \right]_{\text{est.}}, \tag{49}$$

and

$$L_1^a(i, k) = \sum_{\substack{(j,1) \in P_{ik} \\ (j,1) \neq (i,k)}} \frac{\partial}{\partial \theta} S_{j1} + \frac{\partial}{\partial \theta} S_{ik}^a = L_1(i_1, k_1) + \mathbf{1}(k=1) \frac{\partial}{\partial \theta} S_{ik}^a. \tag{50}$$

Then, from the definition of $L_1(i, k)$ in (47) and also (45), we have

$$\begin{aligned}
L_2(i, k) &= \sum_{(j,1) \in P_{ik}} \frac{\partial^2}{\partial \theta^2} S_{j1} \\
&\quad + \sum_{(j,l) \in Q_{ik}^*} g_{jl}^W(0|z_{j\hat{l}}) \left([L_1(\hat{l}) - L_1^a(j-1, l)]^+ \right)^2 \\
&\quad + \sum_{(j,l) \in R_{ik}^*} g_{jl}^I(0|z_{j-1,l}) \left([L_1(j-1, l) - L_1^a(\hat{l})]^+ \right)^2 \\
&\quad - \sum_{(j,l) \in R'_{ik}} g_{jl}^X(0|z_{j\hat{l}}) \left([L_1(\hat{l}) - L_1^a(\widehat{j+1}, \hat{l})]^+ \right)^2 \\
&= L_2(i_1, k_1) + \mathbf{1}(k=1) \frac{\partial^2}{\partial \theta^2} S_{i1} \\
&\quad + \mathbf{1}[(i, k) \in Q_{ik}^*] g_{i,k}^W(0|z_{i,\hat{k}}) \left([L_1(\hat{i}, \hat{k}) - L_1^a(i-1, k)]^+ \right)^2 \\
&\quad + \mathbf{1}[(i, k) \in R_{ik}^*] g_{i,k}^I(0|z_{i-1,k}) \left([L_1(i-1, k) - L_1^a(\hat{i}, \hat{k})]^+ \right)^2 \\
&\quad - \mathbf{1}[(i, k) \in R'_{ik}] g_{i,k}^X(0|z_{i\hat{k}}) \left([L_1(\hat{i}\hat{k}) - L_1^a(\widehat{i+1}, \hat{k})]^+ \right)^2 \tag{51}
\end{aligned}$$

The last expression above corresponds to the following four cases:

Case 1. $D_{i-1,k} > D_{i,\hat{k}}$ and no event occurs during the waiting time of customer C_{ik} i.e. during the time interval $(D_{i,\hat{k}}, D_{i-1,k})$.

It follows from the definition of Q_{ik}^* and R_{ik}^* that (51) yields:

$$\begin{aligned}
L_2(i, k) &= L_2(i_1, k_1) + \mathbf{1}(k=1) \frac{\partial^2}{\partial \theta^2} S_{i1} \\
&\quad + g_{i,k-1}^W(0|z_{i,\hat{k}}) \left([L_1(\hat{i}, \hat{k}) - L_1^a(i-1, k)]^+ \right)^2.
\end{aligned}$$

Case 2. $D_{i-1,k} \leq D_{i,\hat{k}}$ and no event occurs during the idle period at node k , i.e. during the time interval $(D_{i-1,k}, D_{i,\hat{k}})$. Then

$$\begin{aligned}
L_2(i, k) &= L_2(i_1, k_1) + \mathbf{1}(k=1) \frac{\partial^2}{\partial \theta^2} S_{i1} \\
&\quad + g_{i-1,k}^I(0|z_{i-1,k}) \left([L_1(i-1, k) - L_1^a(\hat{i}, \hat{k})]^+ \right)^2.
\end{aligned}$$

Case 3. $D_{i-1,k} \leq D_{\hat{i},\hat{k}}$ and no event occurs between the arrival of C_{ik} and $C_{i+1,k}$, i.e. during the time interval $(D_{\hat{i},\hat{k}}, \widehat{D_{i+1,k}})$. Then

$$L_2(i, k) = L_2(i_1, k_1) + \mathbf{1}(k=1) \frac{\partial^2}{\partial \theta^2} S_{i1} - g_{ik}^X(0|z_{i\hat{k}}) \left(\left[L_1(\hat{i}\hat{k}) - L_1^a(\widehat{i+1}, \hat{k}) \right]^+ \right)^2.$$

Case 4. Otherwise,

$$L_2(i, k) = L_2(i_1, k_1) + \mathbf{1}(k=1) \frac{\partial^2}{\partial \theta^2} S_{i1}.$$

Using (48) and the four cases in (51), we have the following:

First and Second Derivative Estimation Algorithm

1. Initialize: $L_1(1, k) := 0$, $L_1^a(1, k) := 0$, $L_2(1, k) := 0$, for $k = 1, \dots, n$.
2. At each (departure) event e_{ik} induced by e_{i_1, k_1} :

- 2.1. $L_1(i, k) := L_1(i_1, k_1)$, $L_2(i, k) := L_2(i_1, k_1)$.

- 2.2. If $k = 1$,

$$L_1(i, k) := L_1(i, k) + \frac{\partial}{\partial \theta} S_{i1},$$

$$L_1^a(i, k) := L_1^a(i, k) + \frac{\partial}{\partial \theta} S_{i1}^a,$$

$$L_2(i, k) := L_2(i, k) + \frac{\partial^2}{\partial \theta^2} S_{i1}.$$

- 2.3. If $D_{i-1,k} > D_{\hat{i},\hat{k}}$ and no event has occurred in $(D_{\hat{i},\hat{k}}, D_{i-1,k})$ (i.e. if C_{ik} had to wait and no event occurred in the network during his waiting time),

$$L_2(i, k) := L_2(i, k) + g_{ik}^W(0|z_{i\hat{k}}) \left(\left[L_1(\hat{i}, \hat{k}) - L_1^a(i-1, k) \right]^+ \right)^2.$$

- 2.4. If $D_{i-1,k} \leq D_{\hat{i},\hat{k}}$ and no event has occurred in $(D_{i-1,k}, D_{\hat{i},\hat{k}})$ (i.e. if C_{ik} did not have to wait and no event occurred in the network during the idle period in node k preceding his arrival),

$$L_2(i, k) := L_2(i, k) + g_{ik}^I(0|z_{i-1,k}) \left(\left[L_1(i-1, k) - L_1^a(\hat{i}, \hat{k}) \right]^+ \right)^2.$$

- 2.5. If $D_{i-1,k} \leq D_{i\hat{k}}$ and no event has occurred in $(D_{i\hat{k}}, \widehat{D}_{i+1,k})$ (i.e. if C_{ik} started a busy period and no event occurred before $C_{i+1,k}$ arrived),

$$L_2(i, k) := L_2(i, k) - g_{ik}^X(0|z_{i\hat{k}}) \left(\left[L_1(i\hat{k}) - L_1^a(i+1, \hat{k}) \right]^+ \right)^2.$$

3. If N events have occurred at node 1, stop and set:

$$L_1(N, 1) := \frac{1}{N} L_1(N, 1), \quad L_2(N, 1) := \frac{1}{N} L_2(N, 1),$$

otherwise return to step 2 for the next observed event.

Intuitively, step 2.2. calculates first derivatives as well as the part of the second derivative due to $\frac{\partial^2}{\partial \theta^2} S_{i1}$. Step 2.3 accounts for the contribution to the second derivative of changes in the order of events that create idle periods, while step 2.4 accounts for the disappearance of idle periods. There is however a third type of event order change that contributes to the second derivative in this system: a customer who initiates a busy period at a given node in the nominal sample path may be overtaken by another customer (coming from a different node) who now initiates the busy period instead, in the perturbed path. This effect is accounted for by step 2.5. (This last scenario is clearly not possible in a single-server, cyclic network and as a result this step was not necessary in Bao, Cassandras, and Zazanis (submitted).) Note that in each of steps 2.3–2.5 the contribution to the second derivative consists of two parts: the actual event time change given in terms of the $L_1(\cdot)$ and $L_2(\cdot)$ expressions, and a conditional probability as evaluated in Lemma 3.

In checking the validity of conditions at each step, it is necessary to know the identity of the next event. Furthermore, to carry out the computations involving the conditional densities, we need to know the residual service times at the *time of the next event*. When implementing the algorithm this is accomplished by letting the second derivative calculations “lag behind” the simulation until the necessary information becomes available (see also the Remark in section 3.2). The calculations of $g_{ik}^W(0|z_{i\hat{k}})$, $g_{ik}^I(0|z_{i-1,k})$ and $g_{ik}^X(0|z_{i\hat{k}})$ whenever required in the algorithm for a particular (i, k) , are carried out by using (23),(24) and (25). Observe, however, that the expressions for these conditional densities (and in particular the necessary integrations) are performed *off line* before the simulation is carried out. The calculation of the first and second derivatives of the service times $S_{i1}(\theta)$ is done using standard techniques (e.g. see §2 of Zazanis and Suri (1994)).

Finally, when the algorithm stops we have:

$$L_1(N, 1) = \left[\frac{\partial \bar{D}(\theta)}{\partial \theta} \right]_{\text{est.}}, \quad L_2(N, 1) = \left[\frac{\partial^2 \bar{D}(\theta)}{\partial \theta^2} \right]_{\text{est.}}.$$

Finally, the reader may notice that calculations of first and second order derivatives at all possible nodes are performed in parallel since, from the definition of B_{ik} , each calculation of $L_2(i, k)$ depends on the value $L_2(j, l)$ at some other node in steps 2.3–2.5 where e_{jl} is a neighboring event of e_{ik} and l could be any node in the network.

5. Numerical Results

The numerical results in this section illustrate the performance of our algorithms. The first two experiments refer to systems with exponential service times where the decision parameter $\theta = 1/\mu$ is the mean service time of one of the nodes. The simulation horizon was sufficiently long to permit comparison with the analytical results available for steady state performance criteria in this case and the experiments indicate that the unbiasedness of the proposed estimators for finite horizon simulations is likely to hold in the steady state as well. We also present an experiment with non exponential (H_2) service times at all nodes and compare our estimators with “brute-force” simulation results. In the following examples, the quantities with the subscript ‘*est*’ represent estimates which are compared to the corresponding analytical results or brute force simulation results (designated by the subscript ‘*b*’). Finally, $\bar{D}'(\theta)$ and $\bar{D}''(\theta)$ represent the first and second derivatives with respect to parameter θ . Each time, we have 25 runs. The results for different N are shown below (under a 95% confidence interval):

Case 1. Exponential system with 3 nodes and 3 customers, $\mu_1 = \mu_2 = \mu_3 = 1.0$ and routing probabilities:

$$P = \begin{pmatrix} 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$$

N	D	$D_{est.}$	$D'(\theta)$	$D'(\theta)_{est.}$	$D''(\theta)$	$D''(\theta)_{est.}$
node 1						
10^4	1.56335	1.56317 ± 0.00287	0.59145	0.59037 ± 0.00312	0.37752	0.37618 ± 0.04841
10^5	1.56335	1.55584 ± 0.00120	0.59145	0.58976 ± 0.00097	0.37752	0.37247 ± 0.01176
10^6	1.56335	1.56304 ± 0.00026	0.59145	0.59117 ± 0.00024	0.37752	0.37707 ± 0.00307
node 2						
10^4	1.56335	1.56921 ± 0.00478	0.59145	0.59247 ± 0.00439	0.37752	0.37770 ± 0.04855
10^5	1.56335	1.56230 ± 0.00084	0.59145	0.59218 ± 0.00123	0.37752	0.37401 ± 0.01181
10^6	1.56335	1.56368 ± 0.00037	0.59145	0.59141 ± 0.00030	0.37752	0.37722 ± 0.00309
node 3						
10^4	1.95419	1.92756 ± 0.00543	0.73932	0.72765 ± 0.00457	0.47190	0.46401 ± 0.05952
10^5	1.95419	1.93489 ± 0.00130	0.73932	0.73340 ± 0.00125	0.47190	0.46325 ± 0.01466
10^6	1.95419	1.95336 ± 0.00036	0.73932	0.73880 ± 0.00031	0.47190	0.47123 ± 0.00385

Case 2. Exponential system with 3 nodes and 3 customers, $\mu_1 = 1$, $\mu_2 = 5.0$, $\mu_3 = 0.6$, and routing probabilities:

$$P = \begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.2 & 0.6 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$$

N	D	$D_{\text{est.}}$	$D'(\theta)$	$D'(\theta)_{\text{est.}}$	$D''(\theta)$	$D''(\theta)_{\text{est.}}$
node 1						
10^4	3.96462	3.92519 ± 0.01432	0.17430	0.17133 ± 0.00339	0.20851	0.17765 ± 0.03411
10^5	3.96462	3.92842 ± 0.00362	0.17430	0.17566 ± 0.00098	0.20851	0.20207 ± 0.01327
10^6	3.96462	3.96447 ± 0.00094	0.17430	0.17444 ± 0.00033	0.20851	0.20606 ± 0.00434
node 2						
10^4	5.13068	5.11731 ± 0.01853	0.22556	0.22293 ± 0.00477	0.26734	0.23205 ± 0.04463
10^5	5.13068	5.10633 ± 0.00174	0.22556	0.22828 ± 0.00130	0.26734	0.26269 ± 0.01723
10^6	5.13068	5.13383 ± 0.00050	0.22556	0.22589 ± 0.00045	0.26734	0.26685 ± 0.00565
node 3						
10^4	3.63423	3.56614 ± 0.01368	0.15977	0.15542 ± 0.00327	0.18937	0.16144 ± 0.03100
10^5	3.63423	3.59194 ± 0.00323	0.15977	0.16058 ± 0.00091	0.18937	0.18475 ± 0.01207
10^6	3.63423	3.63544 ± 0.00075	0.15977	0.15996 ± 0.00032	0.18937	0.18896 ± 0.00398

Case 3. H_2 system with 2 nodes and 2 customers

In this system, we have two nodes and two customers. The service time at each node is H_2 with service time distributions $f_1(x) = \alpha\mu_1 e^{-\mu_1 x} + (1 - \alpha)\mu_2 e^{-\mu_2 x}$ and $f_2(x) = \beta\lambda_1 e^{-\lambda_1 x} + (1 - \beta)\lambda_2 e^{-\lambda_2 x}$ respectively. Here, we chose $\theta = \frac{1}{\mu_1}$ and $\Delta\theta = 0.002$. The parameters are

α	μ_1	μ_2	β	λ_1	λ_2
0.6	1.0	2.0	0.4	1.0	2.0

and routing probabilities:

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{pmatrix}$$

N	D_b	$D'(\theta)_b$	$D'(\theta)_{\text{est.}}$	$D''(\theta)_b$	$D''(\theta)_{\text{est.}}$
10^4	2.68345 ± 0.00216	0.34882 ± 0.00238	0.36091 ± 0.00124	0.13396 ± 0.01463	0.13530 ± 0.00155
10^5	2.62246 ± 0.00063	0.35272 ± 0.00059	0.35793 ± 0.00031	0.12928 ± 0.00587	0.12848 ± 0.00044
10^6	2.60669 ± 0.00026	0.35666 ± 0.00020	0.35818 ± 0.00011	0.13056 ± 0.00169	0.13121 ± 0.00014

6. Padé Approximation of Interdeparture Time through its Derivatives

As mentioned in the introduction, first and second derivatives can be used to obtain an approximation of entire response curves of interest. Using the Padé approximation approach, we find that the approximated response curve is very close to the real response curve over a large range, which means we can very confidently use the approximated curve for many purposes such as optimization of the system performance. The benefit here is obvious, since we need only one simulation run plus some additional calculation for first and second derivative estimation, in order to obtain the response curve, instead of multiple simulation runs.

In the following, we will generate the response curve for the expected interdeparture time (hence, the throughput as well) by a Padé approximation based on the results of the previous sections. Since the throughput is the inverse of the expected interdeparture time,

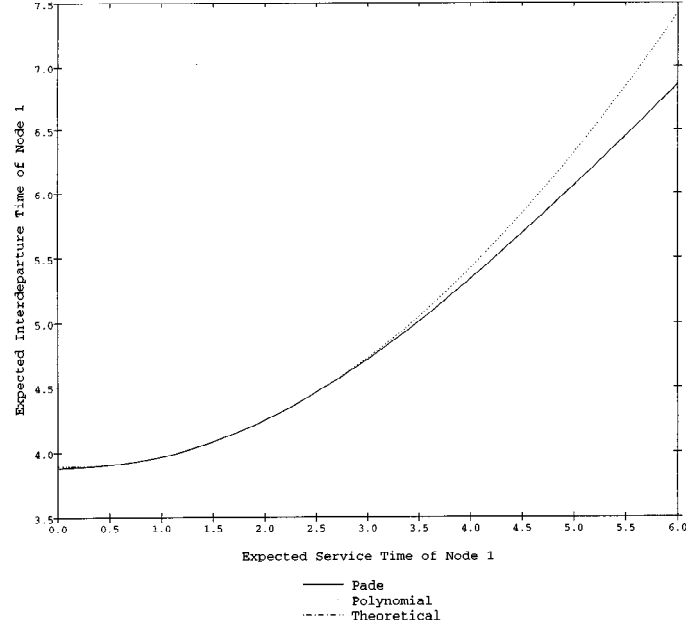


Figure 2. Padé approximation of throughput for a three-node three-customer system.

we actually obtain the response curve of the throughput in the mean time. Using results from Gong, Nananukul, and Yan (submitted), we have

$$\begin{aligned}
 [\bar{D}(\theta)]_{\text{approx.}} &= \theta + [2(\bar{D}(\theta_1)_{\text{est.}} - \theta_1)^3] \{2(\bar{D}(\theta_1)_{\text{est.}} - \theta_1)^2 \\
 &\quad - 2(\bar{D}(\theta_1)_{\text{est.}} - \theta_1)(\bar{D}'(\theta_1)_{\text{est.}} - 1)(\theta - \theta_1) \\
 &\quad + [2(\bar{D}'(\theta_1)_{\text{est.}} - 1)^2 \\
 &\quad - (\bar{D}(\theta_1)_{\text{est.}} - \theta_1)(\bar{D}''(\theta_1)_{\text{est.}})](\theta - \theta_1)^2\}^{-1}
 \end{aligned} \tag{52}$$

which is the equation we need for the Padé approximation. In the following example, we simulate a 3-node 3-customer system with $N = 10^6$. The service time at each node is exponentially distributed with service rate: $\mu_1 = 1.0$, $\mu_2 = 5.0$, and $\mu_3 = 0.6$ respectively. The routing probability (arbitrarily chosen) is as follows:

$$P = \begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.2 & 0.6 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$$

Our results are shown in Figure 2. The curves marked “Polynomial” and “Padé” represent a polynomial and a Padé approximation respectively. The curve marked “Theoretical” is obtained from the analytical expression for the throughput of this Markovian system. One can see the “Padé” and “Theoretical” curves are virtually indistinguishable. The accuracy

of “Polynomial”, on the other hand, is limited to a small range of values around the nominal point $\theta_0 = 1$.

7. Conclusions and Future Work

We have considered a Jackson-like closed queueing network with arbitrary service time distributions and derived an unbiased second derivative estimator of the throughput over N customers served at some node with respect to a parameter of the service distribution at that node. Our approach is based on observing a single sample path of this system, and evaluating all second-order effects on interdeparture times as a result of the parameter perturbation. We then define an estimator as a conditional expectation over appropriate observable quantities, as in Smoothed Perturbation Analysis (SPA). Along the way, we have also recovered the first derivative estimator of the throughput, which can also be derived using other techniques (e.g. Cao 1990). Our results can be easily extended to the second derivative of the mean delay of customers between any two points in the network proceeding as in Bao, Cassandras, and Zazanis (submitted) or other performance measures of interest.

The analysis of higher-order event order changes has given us some new insights regarding the type of sample path information we need to condition on in order to estimate higher-order performance derivatives. As seen in section 4, even though the derivation of the second-order derivative estimator is fairly elaborate, the actual algorithm for implementing it on line is relatively simple. We have also established the unbiasedness of our estimators.

As mentioned in the introduction, a major motivation for this work is the possibility of using the first and second derivatives of performance metrics of complex DEDS in order to construct a global response surface. Recent developments exploiting Padé approximation techniques (Gong, Nananukul, and Yan, submitted) have made this possibility very real. Our results in section 7 indicate that the entire throughput response surface of a serial closed queueing network can be constructed with remarkable accuracy using only the first and second derivative estimates we have developed. Moreover, we believe that the basic approach presented here may be extended to more complex network topologies, which is the subject of ongoing research.

Acknowledgment

The authors would like to thank Prof. Wei-Bo Gong for several suggestions and helpful comments, especially for the work presented in section 6.

Appendix

We present here the proofs of Lemmas 1 through 4 as well as a number of auxiliary results (Lemmas 5 through 8) that will be needed in the course of these proofs. While so far we

described the sample paths of the system solely in terms of departure times, in some of the Lemmas that follow, it will be useful to use both arrival times at, and departure times from, the nodes of the network. We will denote by $A_{ik}(\theta)$ the arrival time at node k of the customer who becomes C_{ik} in the nominal sample path, and by $A_{ik}(\theta + \Delta\theta)$ the same arrival time in the perturbed path. Thus $A_{ik}(\theta) = D_{i\hat{k}}$ and $A_{ik}(\theta + \Delta\theta) = D_{i^*,k^*}$. Finally, we will denote by $\eta_{ik}(\theta)$ the total number of events up to time $D_{ik}(\theta)$ (i.e., $\eta_{ik}(\theta) = \sum_{j=1}^{\infty} \sum_{l=1}^m \mathbf{1}(D_{jl}(\theta) \leq D_{ik}(\theta))$). We start with a result due to Shanthikumar and Yao (1989).

LEMMA 6 *i) The departure times are nondecreasing in θ , i.e. for all i, k , and $\theta' \geq \theta$,*

$$D_{ik}(\theta') \geq D_{ik}(\theta).$$

ii) Furthermore, the departure times $D_{ik}(\theta)$ are absolutely continuous functions of θ w.p. 1.

iii) The arrival times $A_{ik}(\theta)$ are nondecreasing in θ , i.e. for all i, k , and $\theta' \geq \theta$,

$$A_{ik}(\theta') \geq A_{ik}(\theta).$$

iv) $A_{ik}(\theta)$ is an absolutely continuous function of θ w.p. 1. for all i, k . Finally,

v) $\eta_{ik}(\theta)$ is decreasing in θ .

Proof: See Shanthikumar and Yao (1989) which proves the monotonicity results i), iii), v). The absolute continuity of $D_{ik}(\theta)$ and $A_{ik}(\theta)$ also follows easily from their arguments. ■

We next proceed to obtain two bounds that will be used repeatedly in the sequel. Note that, w.p.1, $\frac{d}{d\theta} D_{ik}(\theta)$ exists for almost all $\theta \in \Theta$ and, as a result of **A.1** is nonnegative. Furthermore, by assumptions **A.1**, **A.2**, for every i, k , and θ , $dS_{ik}/d\theta \leq c_1 S_{ik} + c_2$. Thus

$$0 \leq \frac{dD_{ik}}{d\theta} \leq c_1 D_{ik}(\theta) + c_2 \eta_{ik}(\theta). \quad (53)$$

Also, in view of the absolute continuity of D_{ik} we have

$$D_{ik}(\theta + \Delta\theta) - D_{ik}(\theta) = \int_{\theta}^{\theta + \Delta\theta} \frac{d}{du} D_{ik}(u) du. \quad (54)$$

In light of the fact that $\eta_{ik}(\theta) \geq \eta_{ik}(u)$ for $u \geq \theta$ (Lemma 6), (54), (53) imply in turn that $D_{ik}(\theta + \Delta\theta) - D_{ik}(\theta) \leq \int_{\theta}^{\theta + \Delta\theta} [c_1 D_{ik}(u) + c_2 \eta_{ik}(\theta)] du$, which, upon integration, gives the bound

$$\begin{aligned} 0 \leq D_{ik}(\theta + \Delta\theta) - D_{ik}(\theta) &\leq (e^{c_1 \Delta\theta} - 1)(D_{ik}(\theta) + \frac{c_2}{c_1} \eta_{ik}(\theta)) \\ &\leq \frac{c_1 \Delta\theta}{1 - c_1 \Delta\theta} (D_{ik}(\theta) + \frac{c_2}{c_1} \eta_{ik}(\theta)), \end{aligned} \quad (55)$$

the last inequality holding provided that $\Delta\theta \leq 1/c_1$, as a result of the elementary inequality $e^y - 1 \leq \frac{y}{1-y}$, $y < 1$.

Proof of Lemma 1: This proof is along the lines of the proof of Lemma 3 in Glasserman and Gong (1991) with some modifications. We remind the reader that, because all service time distributions have hazard rates bounded above by γ , (Assumption **A.4**) the point process that consists of all departures in the network has stochastic intensity which is bounded above by $m\gamma$. Since by the assumptions of the lemma at least one event intervenes between D_{ik} and D_{jl} in the nominal sample path, given the history of the process up to time D_{ik} , Λ_{ik} , $D_{jl} - D_{ik}$ is stochastically larger than the sum of two independent, exponential random variables, each with rate $m\gamma$, i.e.

$$P(D_{jl}(\theta) - D_{ik}(\theta) \leq x | \Lambda_{ik}) \leq 1 - (1 + m\gamma x)e^{-m\gamma x} \leq (m\gamma x)^2,$$

where, we have used the elementary inequality $e^{-m\gamma x} \geq 1 - m\gamma x$. In view of (54) and the fact that $D_{jl}(\theta + \Delta\theta) \geq D_{jl}(\theta)$ we have the inequality

$$\begin{aligned} P(D_{ik}(\theta + \Delta\theta) > D_{jl}(\theta + \Delta\theta) | \Lambda_{ik}) &\leq P(D_{ik}(\theta + \Delta\theta) > D_{jl}(\theta) | \Lambda_{ik}) \\ &= P(D_{jl}(\theta) - D_{ik}(\theta) < \Delta D_{ik} | \Lambda_{ik}) \\ &\leq (m\gamma \Delta D_{ik})^2 \\ &\leq m\gamma^2 (e^{c_1 \Delta\theta} - 1)^2 \left(D_{ik}(\theta) + \frac{c_2}{c_1} \eta_{ik}(\theta) \right)^2 \\ &\leq (m\gamma)^2 \frac{c_1^2 \Delta\theta^2}{(1 - c_1 \Delta\theta)^2} \\ &\quad \times \left(D_{ik}(\theta) + \frac{c_2}{c_1} \eta_{ik}(\theta) \right)^2, \end{aligned} \quad (56)$$

the last inequality holding provided that $\Delta\theta \leq 1/c_1$, as a result of the elementary inequality $e^y - 1 \leq \frac{y}{1-y}$, $y < 1$. Let

$$V_{ik} := m\gamma^2 \frac{c_1^2}{(1 - c_1 \Delta\theta)^2} \left(D_{ik}(\theta) + \frac{c_2}{c_1} \eta_{ik}(\theta) \right)^2.$$

The proof of the lemma will be complete, provided we show that $EV_{ik} < \infty$. Anticipating the requirements of the proof of Lemma 2 we will show that

$$E \left[\left(D_{ik}(\theta) + \frac{c_2}{c_1} \eta_{ik}(\theta) \right)^3 \right] < \infty.$$

We do this next:

It is easy to see that the number of service completions by time t in our network is stochastically smaller than the superposition of m independent renewal processes with interevent time distributions F_1, \dots, F_m . As a result we conclude that $E\eta_{ik}^r < \infty$ for all $r > 0$ from the corresponding result for renewal processes, provided of course that $F_i(0) < 1$ (e.g. see Çinlar 1975), a condition which is satisfied automatically under our assumptions. To provide a bound for the time to complete i services at node k , consider the same network *with a single customer* starting his service at node l at time 0 and denote by d_{ik} the time he will complete i services at node k . Suppose that the starting node l is such

that, in the original network, the initial number of customers, $n_l \geq 1$. Based on the results of Shanthikumar and Yao (1989)

$$d_{ik} \geq_{\text{st}} D_{ik}.$$

Examining the underlying Markov–Renewal process (Çinlar 1975) we see that **A.5** implied that $Ed_{ik}^3 < \infty$ and, as a result of the above display, that $ED_{ik}^3 < \infty$. Finally, from the “ c_r inequality” (e.g. see Loève 1977) we have

$$E \left(D_{ik} + \frac{c_2}{c_1} \eta_{ik} \right)^3 \leq 4 \left(ED_{ik}^3 + \left(\frac{c_2}{c_1} \right)^3 E\eta_{ik}^3 \right) < \infty.$$

This establishes the proof of Lemma 1 since $EV_{ik} < \infty$. ■

Proof of Lemma 2: Since the proofs of the three parts are similar, we focus only on part (1). Recall that

$$W_{jl}^p = (D_{j^*l^*}^p - D_{j-1,l}^p)^+ = (A_{jl}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta))^+ \quad (57)$$

or, equivalently,

$$W_{jl}^p = (A_{jl} - D_{j-1,l} + \Delta A_{jl} - \Delta D_{j-1,l})^+. \quad (58)$$

We will show that

$$\frac{1}{\Delta\theta^2} E[(A_{jl} - D_{j-1,l} + \Delta A_{jl} - \Delta D_{j-1,l})^+ | \Lambda_{j\hat{l}}] \leq K_{jl}^W(\Delta\theta), \quad (59)$$

for some random variable $K_{jl}^W(\Delta\theta) \in \Lambda_{j\hat{l}}$ which is integrable for all $\Delta\theta$ and $\lim_{\Delta\theta \rightarrow 0} K_{jl}^W(\Delta\theta) = 0$ w.p. 1. There are three cases to consider for which $(j, l) \in Q \setminus Q^*$:

- (i) $D_{j-1,l} = D_{j\hat{l}}$;
- (ii) $D_{j-1,l} = \widehat{D}_{j+1,\hat{l}}$; and
- (iii) $e_{j-1,l}, e_{j\hat{l}}$ are not adjacent events.

It will be sufficient to show that in these three cases $EK_{jl}^W(\Delta\theta) < \infty$ and $\lim_{\Delta\theta \rightarrow 0} K_{jl}^W(\Delta\theta) = 0$. We do this next.

(i) $(j, l) \in Q$ and $D_{j-1,l} = D_{j\hat{l}}$ (i.e. the customer $C_{j-1,l}$ departing at $D_{j-1,l}$ from node l is immediately routed back to node l and becomes C_{jl}). In this case, $A_{jl}(\theta) = D_{j\hat{l}} = D_{j-1,l}(\theta)$ and hence from (58)

$$W_{jl}^p = (\Delta A_{jl} - \Delta D_{j-1,l})^+.$$

Note that, since the routing decisions do not change in the perturbed path, either $A_{jl}(\theta + \Delta\theta) = D_{j-1,l}(\theta + \Delta\theta)$ or, if there is a change in the order of events, $A_{jl}(\theta + \Delta\theta) < D_{j-1,l}(\theta + \Delta\theta)$. Write

$$\begin{aligned} (\Delta A_{jl} - \Delta D_{j-1,l})^+ &= (\Delta A_{jl} - \Delta D_{j-1,l})^+ \mathbf{1}(D_{j-1,l}(\theta + \Delta\theta) = A_{jl}(\theta + \Delta\theta)) \\ &\quad + (\Delta A_{jl} - \Delta D_{j-1,l})^+ \mathbf{1}(D_{j-1,l}(\theta + \Delta\theta) > A_{jl}(\theta + \Delta\theta)) \end{aligned} \quad (60)$$

and note that $\Delta A_{jl} - \Delta D_{j-1,l} = 0$ when $D_{j-1,l}(\theta) = A_{jl}(\theta)$ (which is part of $\Lambda_{j-1,l}$ in this case) and $D_{j-1,l}(\theta + \Delta\theta) = A_{jl}(\theta + \Delta\theta)$. Hence, the first term on the rhs of the above equation vanishes. On the other hand, when $D_{j-1,l}(\theta + \Delta\theta) > A_{jl}(\theta + \Delta\theta)$, we must necessarily have $\Delta D_{j-1,l} > \Delta A_{j-1,l}$ (since $D_{j-1,l}(\theta) = A_{jl}(\theta)$) implying that $(\Delta A_{jl} - \Delta D_{j-1,l})^+ = 0$ and hence that the second term also vanishes. Therefore $W_{jl}^p = 0$ w.p.1 for all $\Delta\theta$ and the Lemma is trivially true in this case.

(ii) $(j, l) \in Q$ and $D_{j-1,l} = D_{j+1,\hat{l}}$ (i.e. the customer departing at $D_{j-1,l}$ from node l is immediately routed back to node l and becomes the arrival of $C_{j+1,l}$ while the j th customer is in service at l .) Equivalently, $D_{j-1,l} = A_{j+1,l}$. Again, the routing indicator $U_{j-1,l}$ dictates that customer $C_{j-1,l}$ will return to node l immediately (regardless of the value of the parameter θ). Suppose that, in the perturbed path, $C_{j-1,l}$ leaves behind him q customers. Then $D_{j-1,l}(\theta + \Delta\theta) = A_{j+q,l}(\theta + \Delta\theta)$. Since $A_{j+q,l}(\theta + \Delta\theta) \geq A_{jl}(\theta + \Delta\theta)$ we see from (57) that $W_{jl}^p = 0$ w.p. 1 for all $\Delta\theta$ and again the Lemma is trivially true.

(iii) $(j, l) \in Q$ and $e_{j-1,l}, e_{j\hat{l}}$ are not adjacent events. We start with the observation that $D_{j-1,l} \geq D_{j\hat{l}}$ since we are assuming $(j, l) \in Q$. In other words, in the nominal sample path C_{jl} has to wait. Furthermore, since the two events in question are assumed to be nonadjacent, $(D_{j\hat{l}}, D_{j-1,l})$ contains at least one event. In view of (57) we see that a necessary condition for $W_{jl}^p > 0$ is

$$\max\{D_{r\hat{l}}(\theta + \Delta\theta); r = 1, 2, \dots, j\} > D_{j-1,l}(\theta + \Delta\theta). \quad (61)$$

To see this, observe that if (61) does not hold, then $A_{jl}(\theta + \Delta\theta) < D_{j-1,l}(\theta + \Delta\theta)$ since the routing indicators remain the same. Let σ be the value of r corresponding to the maximum in (61) and note that when (61) holds, then $A_{jl}(\theta + \Delta\theta) \leq D_{\sigma\hat{l}}(\theta + \Delta\theta)$. (We clarify that σ is an index whose value is determined “at the perturbed path,” i.e., with parameter value $\theta + \Delta\theta$. Later on, when we refer to the event $D_{\sigma\hat{l}}(\theta)$, σ maintains the value determined from (61).) Based on these remarks we see that

$$\begin{aligned} W_{jl}^p &= (A_{jl}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta))^+ \\ &= (D_{\sigma\hat{l}}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta))^+. \end{aligned} \quad (62)$$

Write $D_{\sigma\hat{l}}(\theta + \Delta\theta) = D_{\sigma\hat{l}}(\theta) + \Delta D_{\sigma\hat{l}}$ and observe that

$$D_{\sigma\hat{l}}(\theta) \leq D_{j\hat{l}}(\theta) < D_{j-1,l}(\theta) \leq D_{j-1,l}(\theta + \Delta\theta),$$

the second inequality following by the assumption that $(j, l) \in Q$. From the above inequality and (62) we obtain

$$W_{jl}^p = (\Delta D_{\sigma\hat{l}} + D_{\sigma\hat{l}}(\theta) - D_{j-1,l}(\theta + \Delta\theta))$$

$$\begin{aligned}
& \times \mathbf{1}(D_{\hat{\sigma}i}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta) > 0) \\
& \leq \Delta D_{\hat{\sigma}i} \mathbf{1}(D_{\hat{\sigma}i}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta) > 0) \\
& \leq \frac{c_1 \Delta\theta}{(1 - c_1 \Delta\theta)} \left(D_{ji}(\theta) + \frac{c_2}{c_1} \eta_{ji}(\theta) \right) \mathbf{1}(D_{\hat{\sigma}i}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta) > 0),
\end{aligned}$$

for $\Delta\theta \leq 1/c_1$. In the last inequality we have used the fact that

$$\Delta D_{\hat{\sigma}i} \leq \frac{c_1 \Delta\theta}{(1 - c_1 \Delta\theta)} \left(D_{\hat{\sigma}i}(\theta) + \frac{c_2}{c_1} \eta_{\hat{\sigma}i}(\theta) \right) \leq \frac{c_1 \Delta\theta}{(1 - c_1 \Delta\theta)} \left(D_{ji}(\theta) + \frac{c_2}{c_1} \eta_{ji}(\theta) \right), \quad (63)$$

the first inequality above following immediately from (54) while the second from Lemma 6 and the fact that $D_{\hat{\sigma}i}(\theta) \leq D_{ji}(\theta)$. Thus, for sufficiently small $\Delta\theta$,

$$\begin{aligned}
\frac{1}{\Delta\theta^2} E[W_{ji}^p | \Lambda_{ji}] & \leq \frac{1}{\Delta\theta^2} E \left[\frac{c_1 \Delta\theta}{(1 - c_1 \Delta\theta)} \left(D_{ji}(\theta) + \frac{c_2}{c_1} \eta_{ji}(\theta) \right) \right. \\
& \quad \left. \times \mathbf{1}(D_{\hat{\sigma}i}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta) > 0) | \Lambda_{ji} \right] \\
& \leq \frac{1}{\Delta\theta^2} \frac{c_1 \Delta\theta}{(1 - c_1 \Delta\theta)} \left(D_{ji}(\theta) + \frac{c_2}{c_1} \eta_{ji}(\theta) \right) \\
& \quad \times P \left(D_{\hat{\sigma}i}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta) > 0 | \Lambda_{ji} \right).
\end{aligned}$$

Recall however that, since by assumption $(D_{\hat{\sigma}i}, D_{j-1,l})$ contains at least one event, so does $(D_{\hat{\sigma}i}, D_{j-1,l})$. Using an argument identical to that given in the proof of Lemma 1 and (63) we obtain the bound

$$P \left(D_{\hat{\sigma}i}(\theta + \Delta\theta) - D_{j-1,l}(\theta + \Delta\theta) > 0 | \Lambda_{ji} \right) \leq \left(\frac{\Delta\theta m \gamma c_1}{1 - c_1 \Delta\theta} \left(D_{\hat{\sigma}i}(\theta) + \frac{c_2}{c_1} \eta_{\hat{\sigma}i}(\theta) \right) \right)^2$$

Hence

$$\frac{1}{\Delta\theta^2} E[W_{ji}^p | \Lambda_{ji}] \leq \Delta\theta m^2 \gamma^2 \left(\frac{c_1}{1 - c_1 \Delta\theta} \left(D_{\hat{\sigma}i}(\theta) + \frac{c_2}{c_1} \eta_{\hat{\sigma}i}(\theta) \right) \right)^3 \stackrel{\text{def}}{=} K_{ji}^W(\Delta\theta).$$

To conclude the proof recall that in the last paragraph of Lemma 1 we showed that the expectation of the above quantity is finite. \blacksquare

Proof of Lemma 3: We examine only the idle period case in detail, as the other two cases are similar. We begin with the observation that necessarily $\hat{l} \in \mathcal{M}(D_{j-1,l})$, i.e., the node which generates the arrival of j th customer at node l must be busy at time $D_{j-1,l}$. Hence, the idle period of length I_{jl} is the residual service time of the customer in \hat{l} th node, at time $D_{j-1,l}$, defined as $S_{\hat{l}}^r(D_{j-1,l})$. Equivalently, we write

$$I_{jl} = S_{\hat{l}}(D_{j-1,l}) - S_{\hat{l}}^a(D_{j-1,l}).$$

Because of the independence assumptions regarding the service processes at the nodes of the network, the relevant part of $\hat{z}_{j-1,l}$ is the set of active nodes, $\mathcal{M}(D_{j-1,l})$, the ages of

the service processes at the active nodes given by $\{S_q^a(D_{j-1,l}); q \in \mathcal{M}(D_{j-1,l})\}$, the identity of the next event, and the residual service times *at the end of the idle period* I_{jl} , $\{S_q^r(D_{j\hat{l}}); q \in \mathcal{M}(D_{j-1,l})\}$.

Under the assumption of the lemma, $e_{j\hat{l}}$ is the next event, i.e., no events occur during the idle period I_{jl} . This translates into the condition

$$S_i^l(D_{j-1,l}) - S_i^a(D_{j-1,l}) = S_q(D_{j-1,l}) - S_q^a(D_{j-1,l}) - S_q^r(D_{j\hat{l}}),$$

$$\text{for all } q \in \mathcal{M}(D_{j-1,l}) \setminus \{\hat{l}\},$$

where $S_q(D_{j-1,l}) - S_q^a(D_{j-1,l})$ is the residual service time at node q at the *beginning* of the idle period, and $S_q^r(D_{j\hat{l}})$ is the residual service time at node q at the *end* of the idle period.

Denoting by $g^l(\cdot|z_{j-1,l})$ the conditional density of I_{jl} , we then have

$$g^l(x|z_{j-1,l}) dx \propto P \left(\begin{aligned} &S_i^l(D_{j-1,l}) - S_i^a(D_{j-1,l}) \in dx, \\ &S_q(D_{j-1,l}) - S_q^a(D_{j-1,l}) - S_q^r(D_{j\hat{l}}) \in dx, \\ &q \in \mathcal{M}(D_{j-1,l}) \setminus \{\hat{l}\} \mid \\ &S_i^a(D_{j-1,l}), S_q^a(D_{j-1,l}), S_q^r(D_{j\hat{l}}), q \in \mathcal{M}(D_{j-1,l}) \setminus \{\hat{l}\} \end{aligned} \right).$$

It is then straight-forward to obtain the following expression:

$$g^l(x|z_{j-1,l}) = \frac{f_i(S_i^a(D_{j-1,l}) + x) \left[\prod_{\substack{q \in \mathcal{M}(D_{j-1,l}) \\ q \neq \hat{l}}} f_q(S_q^a(D_{j-1,l}) + S_q^r(D_{j\hat{l}}) + x) \right]}{\int_0^\infty f_i(S_i^a(D_{j-1,l}) + u) \left[\prod_{\substack{q \in \mathcal{M}(D_{j-1,l}) \\ q \neq \hat{l}}} f_q(S_q(D_{j-1,l}) + S_q^r(D_{j\hat{l}}) + u) \right] du}.$$
(64)

Taking into account that

$$S_q(D_{j-1,l}) = S_q^a(D_{j-1,l}) + I_{jl} + S_q^r(D_{j\hat{l}}), \quad q \in \mathcal{M}(D_{j-1,l}),$$

(64) can be rewritten as

$$g^l(x|z_{j-1,l}) = \frac{f_i(S_i^a(D_{j-1,l}) + x) \left[\prod_{\substack{q \in \mathcal{M}(D_{j-1,l}) \\ q \neq \hat{l}}} f_q(S_q(D_{j-1,l}) - I_{jl} + x) \right]}{\int_0^\infty f_i(S_i^a(D_{j-1,l}) + u) \left[\prod_{\substack{q \in \mathcal{M}(D_{j-1,l}) \\ q \neq \hat{l}}} f_q(S_q(D_{j-1,l}) - I_{jl} + u) \right] du}.$$
(65)

■

The following auxiliary result is needed in the proof of Lemma 4.

LEMMA 7 *For the sets Q^* , R^* , R' defined in (20)*

(i) *if $(j, l) \in Q^*$, then*

$$E[W_{jl}^p | z_{j\hat{l}}] = E[W_{jl}^n | z_{j\hat{l}}] + o(\Delta\theta^2)$$
(66)

(ii) if $(j, l) \in R^*$, then

$$E[I_{jl}^p | z_{j-1,l}] = E[I_{jl}^n | z_{j-1,l}] + o(\Delta\theta^2) \quad (67)$$

(iii) if $(j, l) \in R'$, then

$$E[X_{jl}^p | z_{j\hat{l}}] = -E[X_{jl}^n | z_{j\hat{l}}] + o(\Delta\theta^2) \quad (68)$$

where $W_{jl}^n, I_{jl}^n, X_{jl}^n$ are defined as follows:

$$\begin{aligned} W_{jl}^n &= (D_{j\hat{l}}^p - D_{j-1,l}^p)^+ = (\Delta D_{j\hat{l}} - \Delta D_{j-1,l} - W_{jl})^+ \\ I_{jl}^n &= (D_{j-1,l}^p - D_{j\hat{l}}^p)^+ = (\Delta D_{j-1,l} - \Delta D_{j\hat{l}} - I_{jl})^+ \\ X_{jl}^n &= (D_{j\hat{l}}^p - D_{j+1,\hat{l}}^p)^+ = (\Delta D_{j\hat{l}} - \Delta D_{j+1,\hat{l}} - X_{jl})^+ \end{aligned} \quad (69)$$

Proof: Part (i): From (58), the fact that $W_{jl} = D_{j-1,l} - A_{jl}$, and the above definitions one can check that

$$\begin{aligned} \frac{1}{\Delta\theta^2} |W_{jl}^p - W_{jl}^n| &= \frac{1}{\Delta\theta^2} \left[(\Delta A_{jl} - \Delta D_{j-1,l} - W_{jl})^+ \right. \\ &\quad \left. - (\Delta D_{j\hat{l}} - \Delta D_{j-1,l} - W_{jl})^+ \right] \\ &\leq \frac{1}{\Delta\theta^2} \left| \Delta D_{j\hat{l}} - \Delta A_{jl} \right| \left(\mathbf{1}(W_{jl} < \Delta D_{j\hat{l}} - \Delta D_{j-1,l}) \right. \\ &\quad \left. + \mathbf{1}(W_{jl} < \Delta A_{jl} - \Delta D_{j-1,l}) \right) \\ &\leq \frac{1}{\Delta\theta^2} \left| \Delta D_{j\hat{l}} - \Delta A_{jl} \right| \left(\mathbf{1}(W_{jl} < \Delta D_{j\hat{l}}) + \mathbf{1}(W_{jl} < \Delta A_{jl}) \right), \end{aligned}$$

the last inequality following from the fact that $\Delta D_{j-1,l}$ is nonnegative. Note that $\Delta D_{j\hat{l}}, \Delta A_{jl} \in z_{j\hat{l}}$, and hence, the conditional expectation of the rhs of the above inequality given $z_{j\hat{l}}$ can be written as

$$\left| \frac{\Delta D_{j\hat{l}}}{\Delta\theta} - \frac{\Delta A_{jl}}{\Delta\theta} \right| \left(\frac{1}{\Delta\theta} P(\Delta D_{j\hat{l}} > W_{jl} | z_{j\hat{l}}) + \frac{1}{\Delta\theta} P(\Delta A_{jl} > W_{jl} | z_{j\hat{l}}) \right) \quad (70)$$

Observe that, as $\Delta\theta \rightarrow 0$, both $\frac{\Delta D_{j\hat{l}}}{\Delta\theta}$ and $\frac{\Delta A_{jl}}{\Delta\theta}$ converge w.p. 1 to $\frac{\partial D_{j\hat{l}}}{\partial\theta}$ and thus their difference vanishes. On the other hand

$$\begin{aligned} \frac{1}{\Delta\theta} P(\Delta D_{j\hat{l}} > W_{jl} | z_{j\hat{l}}) &= \frac{1}{\Delta\theta} \int_0^{\Delta D_{j\hat{l}}} g_{jl}^W(x | z_{j\hat{l}}) dx \\ &= \int_0^{\frac{\Delta D_{j\hat{l}}}{\Delta\theta}} g_{jl}^W(y \Delta\theta | z_{j\hat{l}}) dy \longrightarrow_{\Delta\theta \rightarrow 0} \frac{\partial D_{j\hat{l}}}{\partial\theta} g_{jl}^W(0 | z_{j\hat{l}}) \text{ w.p.1.} \end{aligned}$$

A similar calculation can be carried out for the second term inside the parenthesis of (70).

From the above we conclude that

$$\frac{1}{\Delta\theta^2}E[W_{jl}^n - W_{jl}^p|z_{j\hat{i}}] \rightarrow 0.$$

Part (ii): The proof is similar to part (i).

Part (iii): We remind the reader (8) that $X_{jl}^p = D_{j^*l^*}^p - D_{j\hat{i}}^p$. First, observe that $D_{j^*l^*}^p = \min\{D_{j+r,l}^p; r = 0, 1, 2, \dots\}$. To be specific, let ρ denote the value of r for which the minimum achieved so that $D_{j^*l^*}^p = D_{j+\rho,l}^p$. If $\rho = 0$ then both $X_{jl}^p = 0$ and $X_{jl}^n = 0$, the second equality following from definition (69) since $\rho = 0$ implies $D_{j+1,\hat{i}}^p > D_{j\hat{i}}^p$. If $\rho = 1$ then $X_{jl}^p = D_{j+1,\hat{i}}^p - D_{j\hat{i}}^p = -X_{jl}^n$. Therefore

$$X_{jl}^p + X_{jl}^n = (D_{j+\rho,\hat{i}}^p - D_{j+1,\hat{i}}^p)\mathbf{1}(\rho \geq 2). \quad (71)$$

From the definition of ρ we also have

$$D_{j+\rho,\hat{i}}^p \leq D_{j\hat{i}}^p, \quad (72)$$

while from the monotonicity of the departure process (Lemma 6) $D_{j+1,\hat{i}}^p \geq D_{j+1,\hat{i}}^n \geq D_{j\hat{i}}^n$. From these inequalities and (71) we obtain

$$X_{jl}^p + X_{jl}^n \leq (D_{j\hat{i}}^p - D_{j\hat{i}}^n)\mathbf{1}(\rho \geq 2) = \Delta D_{j\hat{i}}\mathbf{1}(\rho \geq 2).$$

Hence, since $\Delta D_{j\hat{i}} \in z_{j\hat{i}}$,

$$\frac{1}{\Delta\theta^2}E[X_{jl}^p + X_{jl}^n|z_{j\hat{i}}] \leq \Delta D_{j\hat{i}} \frac{1}{\Delta\theta^2}P(\rho \geq 2|z_{j\hat{i}}). \quad (73)$$

Note that from the definition of ρ and (72), $\{\rho \geq 2\} = \{D_{j+\rho,\hat{i}}^p(\theta + \Delta\theta) \leq D_{j\hat{i}}^p(\theta + \Delta\theta)\}$ and there is at least one event in the interval $(D_{j\hat{i}}^p(\theta), D_{j+\rho,\hat{i}}^p(\theta))$ in the nominal sample path}. An argument similar to that used in the proof of Lemma 1 shows that

$$P(\rho \geq 2|z_{j\hat{i}}) \leq (m\gamma)^2 \frac{c_1^2 \Delta\theta^2}{(1 - c_1 \Delta\theta)^2} \left(D_{j\hat{i}}^p(\theta) + \frac{c_2}{c_1} \eta_{j\hat{i}}(\theta) \right)^2$$

From the above inequality, the bound obtained for $\Delta D_{j\hat{i}}$ from (54), and (73) we obtain

$$\frac{1}{\Delta\theta^2}E[X_{jl}^p + X_{jl}^n|z_{j\hat{i}}] \leq \Delta\theta m^2 \gamma^2 \left(\frac{c_1}{1 - c_1 \Delta\theta} \left(D_{j\hat{i}}^p(\theta) + \frac{c_2}{c_1} \eta_{j\hat{i}}(\theta) \right) \right)^3$$

which concludes the proof. ■

The following lemma, also needed in the proof of Lemma 4, is taken from Bao, Cassandras, and Zazanis (submitted) and presented here for convenience.

LEMMA 8 Let $\text{supp}F_1(\cdot; \theta)$ denote the support of $F_1(\cdot; \theta)$ and $\Phi_{\Delta\theta}: \text{supp}F_1(\cdot; \theta) \rightarrow \mathbb{R}^+$ the function $F_1^{-1}(F_1(\cdot; \theta); \theta + \Delta\theta)$. Keeping θ fixed, consider the family $\{\Phi_{\Delta\theta}; \Delta\theta \geq 0\}$. Then, for any $S \in \text{supp}F_1(\cdot; \theta)$,

$$\lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta} [\Phi_{\Delta\theta}(S + y\Delta\theta) - S] = \frac{\partial S}{\partial \theta} + y.$$

Proof: Observe first that Φ_0 is the identity map on the support of $F_1(\cdot; \theta)$ and

$$\lim_{\Delta\theta \rightarrow 0} \Phi_{\Delta\theta}(x) = x \quad \text{for } x \in \text{supp } F_1. \quad (74)$$

Since $\Phi_{\Delta\theta}(S + y\Delta\theta) - S = \Phi_{\Delta\theta}(S + y\Delta\theta) - \Phi_{\Delta\theta}(S) + \Phi_{\Delta\theta}(S) - S$, and

$$\frac{\Phi_{\Delta\theta}(S) - S}{\Delta\theta} \rightarrow \frac{\partial S}{\partial \theta} \quad \text{for all } S \in \text{supp } F_1(\cdot; \theta),$$

it is enough to show that

$$\frac{\Phi_{\Delta\theta}(S + y\Delta\theta) - \Phi_{\Delta\theta}(S)}{\Delta\theta} \rightarrow y \quad \text{for all } S \in \text{supp } F_1(\cdot; \theta). \quad (75)$$

Suppose that for some x_0 , $f(x_0; \theta) > 0$. Note that, in view of the continuity of $f(x; \theta)$, there exists ϵ such that when $|x - x_0| < \epsilon$, $\Delta\theta < \epsilon$, $f(x; \theta + \Delta\theta) > 0$. Also, (74) implies that there exists $\delta > 0$ such that for $\Delta\theta < \delta$, $|\Phi_{\Delta\theta}(x_0) - x_0| < \epsilon$. From this follows that, for $\Delta\theta < \delta$, $f(\Phi_{\Delta\theta}(x_0); \theta + \Delta\theta) > 0$, and hence that, for $x_0 \in \text{supp } F_1(\cdot; \theta)$,

$$\frac{\partial}{\partial x} \Phi_{\Delta\theta}(x) = \frac{f(x; \theta)}{f(\Phi_{\Delta\theta}(x); \theta + \Delta\theta)} \quad \text{for } \Delta\theta < \delta.$$

Therefore, from the mean value theorem,

$$\frac{1}{\Delta\theta} [\Phi_{\Delta\theta}(S + y\Delta\theta) - \Phi_{\Delta\theta}(S)] = y \frac{f(\zeta; \theta)}{f(\Phi_{\Delta\theta}(\zeta); \theta + \Delta\theta)}$$

where $\zeta \in [S, S + y\Delta\theta]$.

Letting $\Delta\theta \rightarrow 0$, and invoking the continuity of f and (74), establishes (75). \blacksquare

We can now provide the proof of Lemma 4:

Proof of Lemma 4: To avoid repetition we only examine the behavior of $E[I_{jl}^p | z_{j-1,l}]$ and establish (i). The proofs of (ii) and (iii) are similar. In view of Lemma 1 it is enough to examine the behavior of

$$I_{jl}^n = [\Delta D_{j-1,l} - \Delta D_{j\hat{l}} - I_{jl}]^+. \quad (76)$$

Setting

$$G_{jl}(\Delta\theta) := \sum_{(q,r) \in Q_{jl}} W_{qr}^p + \sum_{(q,r) \in R_{jl}} I_{qr}^p + X_{qr}^p,$$

we can easily see that $\frac{1}{\Delta\theta}G_{j-1,l}(\Delta\theta) \rightarrow 0$ as $\Delta\theta \rightarrow 0$. In particular, using (17) we have

$$\Delta D_{j-1,l} = \sum_{(q,1) \in P_{j-1,l}} \Delta S_{q1} + G_{j-1,l}(\Delta\theta), \quad (77)$$

and

$$\Delta D_{j\hat{l}} = \sum_{(q,1) \in P_{j\hat{l}}} \Delta S_{q1} + \Delta S_{j\hat{l}} + G_{j\hat{l}}(\Delta\theta), \quad (78)$$

with both $G_{j-1,l}(\Delta\theta)$ and $G_{j\hat{l}}(\Delta\theta) \in z_{j-1,l}$ when $(j,l) \in R^*$. We will also need the quantity

$$\overline{\Delta D_{j\hat{l}}} = \Delta D_{j\hat{l}} - \Delta S_{j\hat{l}} \quad (79)$$

which, unlike $\Delta D_{j\hat{l}}$, belongs to $z_{j-1,l}$ when $(j,l) \in R^*$. We now distinguish two cases:

Case 1: $\hat{l} \neq 1$. In this case, since we have assumed that the parameter θ only affects the service distribution of node 1, we have $\Delta S_{j\hat{l}} = 0$. Thus

$$E[I_{j\hat{l}}^n | z_{j-1,l}] = \int_0^\infty [\Delta D_{j-1,l} - \Delta D_{j\hat{l}} - x]^+ g_{j\hat{l}}^I(x | z_{j-1,l}) dx \quad (80)$$

$$= \Delta\theta^2 \int_0^\infty \left[\frac{\Delta D_{j-1,l}}{\Delta\theta} - \frac{\Delta D_{j\hat{l}}}{\Delta\theta} - y \right]^+ g_{j\hat{l}}^I(y\Delta\theta | z_{j-1,l}) dy, \quad (81)$$

the second equation following from the change of variables $x = y\Delta\theta$. A straightforward application of the Dominated Convergence Theorem gives then

$$\lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[I_{j\hat{l}}^n | z_{j-1,l}] = \int_0^\infty \left[\lim_{\Delta\theta \rightarrow 0} \frac{\Delta D_{j-1,l}}{\Delta\theta} - \frac{\Delta D_{j\hat{l}}}{\Delta\theta} - y \right]^+ g_{j\hat{l}}^I(0 | z_{j-1,l}) dy. \quad (82)$$

Finally, observe that since in this case $\frac{\Delta S_{j\hat{l}}}{\Delta\theta} = 0$, (26) reduces to

$$Y_{j\hat{l}}^I = \sum_{(q,r) \in P_{j-1,l}} \frac{\partial S_{qr}}{\partial\theta} - \sum_{\substack{(q,r) \in P_{j\hat{l}} \\ (q,r) \neq (j\hat{l})}} \frac{\partial S_{qr}}{\partial\theta}. \quad (83)$$

From (82), (83), the fact that for any real a $\int_0^\infty [a - y]^+ dy = \frac{1}{2}(a^+)^2$, and Lemma 1 we obtain

$$\lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[I_{j\hat{l}}^p | z_{j-1,l}] = \lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[I_{j\hat{l}}^n | z_{j-1,l}] = \frac{1}{2} g_{j\hat{l}}^I(0 | z_{j-1,l}) (Y_{j\hat{l}}^I)^2. \quad (84)$$

Case 2: $\hat{l} = 1$. Using the same notation as before, let us set in addition:

$$\Delta S_{j\hat{l}} = \begin{cases} F_1^{-1}(F_1(S_{j1}; \theta); \theta + \Delta\theta) & \text{if } l = 2, \\ 0 & \text{otherwise.} \end{cases} \quad (85)$$

Then, using the definition in Lemma 8 and observing that $S_{j\hat{l}} = S_{j\hat{l}}^a + I_{j\hat{l}}$ (since $D_{j-1,l}$ immediately precedes $D_{j\hat{l}}$ when $(j, l) \in R^*$), we have

$$\Delta S_{j\hat{l}} = \Phi_{\Delta\theta}(S_{j\hat{l}}^a + I_{j\hat{l}}) - S_{j\hat{l}}^a - I_{j\hat{l}}.$$

Thus,

$$\begin{aligned} E[I_{j\hat{l}}^n | z_{j-1,l}] &= E[\Delta D_{j-1,l} - \overline{\Delta D_{j\hat{l}}} - \Delta S_{j\hat{l}} - I_{j\hat{l}} | z_{j-1,l}] \\ &= E \int_0^\infty [\Delta D_{j-1,l} - \overline{\Delta D_{j\hat{l}}} - \Phi_{\Delta\theta}(S_{j\hat{l}}^a + x) \\ &\quad + S_{j\hat{l}}^a - 2x]^+ g_{j\hat{l}}^I(x | z_{j-1,l}) dx \\ &= \Delta\theta^2 \int_0^\infty \left[\Delta_{j-1,l} - \overline{\Delta D_{j\hat{l}}} \right. \\ &\quad \left. - \frac{\Phi_{\Delta\theta}(S_{j\hat{l}}^a + y\Delta\theta) - S_{j\hat{l}}^a}{\Delta\theta} - 2y \right]^+ g_{j\hat{l}}^I(y\Delta\theta | z_{j-1,l}) dy. \end{aligned} \quad (86)$$

However, Assumption A.2 and the triangle inequality lead to the bound

$$\begin{aligned} |\Phi_{\Delta\theta}(S_{j\hat{l}}^a + y\Delta\theta) - S_{j\hat{l}}^a| &\leq |\Phi_{\Delta\theta}(S_{j\hat{l}}^a + y\Delta\theta) - S_{j\hat{l}}^a - y\Delta\theta| + y\Delta\theta \\ &\leq c_1\Delta\theta + c_2S_{j\hat{l}}^a + (c_2 + 1)y\Delta\theta. \end{aligned} \quad (87)$$

Furthermore, from Lemma 8,

$$\lim_{\Delta\theta \rightarrow 0} \frac{\Phi_{\Delta\theta}(S_{j\hat{l}}^a + y\Delta\theta) - S_{j\hat{l}}^a}{\Delta\theta} = \frac{\partial S_{j\hat{l}}^a}{\partial\theta} + y. \quad (88)$$

The above equation together with (79) and (26) shows that

$$\lim_{\Delta\theta \rightarrow 0} \left[\frac{\Delta D_{j-1,l}}{\Delta\theta} - \frac{\overline{\Delta D_{j\hat{l}}}}{\Delta\theta} + \frac{\Phi_{\Delta\theta}(S_{j\hat{l}}^a + y\Delta\theta) - S_{j\hat{l}}^a}{\Delta\theta} \right] = Y_{j\hat{l}}^I + y \quad \text{w.p.1.} \quad (89)$$

We can now divide (86) by $\Delta\theta^2$ and take the limit as $\Delta\theta \rightarrow 0$. A Dominated Convergence argument together with a computation similar to the one used to obtain (84) leads to

$$\begin{aligned} \lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[I_{j\hat{l}}^n | z_{j-1,l}] &= \lim_{\Delta\theta \rightarrow 0} \frac{1}{\Delta\theta^2} E[I_{j\hat{l}}^n | z_{j-1,l}] \\ &= \int_0^\infty [Y_{j\hat{l}}^I - y]^+ g_{j\hat{l}}^I(0 | z_{j-1,l}) dy \\ &= \frac{1}{2} g_{j\hat{l}}^I(0 | z_{j-1,l}) ([Y_{j\hat{l}}^I]^+)^2. \end{aligned} \quad \blacksquare$$

Proof of Lemma 5: Indeed, from Assumption **A.3** and Taylor's theorem, $\Delta S_{j1} = \frac{\partial}{\partial \theta} S_{j1}(\theta) \Delta \theta + \frac{1}{2} \frac{\partial^2}{\partial \theta^2} S_{j1}(\theta + \beta_j \Delta \theta) \Delta \theta^2$, where $\beta_j \in (0, 1)$. Observe that

$$\begin{aligned} & \lim_{\Delta \theta \rightarrow 0} \frac{1}{\Delta \theta^2} E \left[\sum_{(j,1) \in P_{ik}} \Delta S_{j1} - \frac{\partial}{\partial \theta} S_{j1}(\theta) \Delta \theta - \frac{1}{2} \frac{\partial^2}{\partial \theta^2} S_{j1}(\theta) \Delta \theta^2 \right] \\ &= \lim_{\Delta \theta \rightarrow 0} E \left[\sum_{(j,1) \in P_{ik}} \frac{\partial^2}{\partial \theta^2} S_{j1}(\theta + \beta_j \Delta \theta) - \frac{\partial^2}{\partial \theta^2} S_{j1}(\theta) \right]. \end{aligned} \quad (90)$$

Let ν be the time of the first departure from node 1 after D_{ik} . Clearly ν is a stopping time (with respect to the history of the process) with finite expectation. The quantity inside the expectation on the rhs of the above equation is dominated by

$$\sum_{j=1}^{\nu} 2 \sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \theta^2} S_{j1}(\theta) \right|$$

which has finite expectation (as a result of Assumption **A.3** and Wald's lemma). Therefore we can pass the limit inside the expectation on the rhs of (89) and take advantage of the continuity of $\frac{\partial^2}{\partial \theta^2} S_{j1}(\theta)$ to show that the limit vanishes. ■

Notes

1. This work was supported in part by the National Science Foundation under Grants EID-9212122 and ECS-9311776, and by a Grant from United Technologies/Otis Elevator Co.

References

- Baker, G. A. 1975. *Essentials of Padé Approximants*. Academic Press.
- Bao, G., Cassandras, C. G., and Zazanis, M. 1996. First and second derivative estimations for cyclic closed queueing networks. *IEEE Trans. on Automatic Control* 41(8): 1116–1136.
- Bertsekas, D. P. 1982. Dynamic behavior of shortest path routing algorithms for communication networks. *IEEE Trans. on Automatic Control* AC-27(1): 60–74.
- Bertsekas, D. P., Gafni, E. M., and Gallager, R.G. 1984. Second derivative algorithms for minimum delay distributed routing in networks. *IEEE Trans. on Communications* COM-32(8): 911–919.
- Brémaud, P., and Vázquez-Abad, F. J. 1992. On the pathwise computation of derivatives with respect to the rate of a point process: The phantom RPA method. *Queueing Systems* 10: 249–270.
- Cao, X. R. 1987. Realization probability in closed Jackson queueing networks and its application. *Adv. Appl. Prob.* 19: 708–738.
- Cao, X. R. 1990. The basic concepts of perturbation analysis of closed queueing networks with general service time distributions. *Proc. of 29th IEEE Conf. Decision and Control*, pp. 2833–2838.
- Cassandras, C. G., Abidi, M., and Towsley, D. 1990. Distributed routing with on-line marginal delay estimation. *IEEE Trans. on Communications* COM-38(3): 348–359.
- Çınlar, E. 1975. *Introduction to Stochastic Processes*. Prentice-Hall.
- Chong, E. K. P., and Ramadge, P. J. 1992. Convergence of recursive optimization algorithms using IPA derivative estimates, to appear in *J. of Discrete Event Dynamic Systems*.
- Fu, M. C., and Hu, J. Q. 1992. Extensions and generalizations of smoothed perturbation analysis in a generalized semi-Markov process framework. *IEEE Trans. on Automatic Control* AC-37(10): 1483–1450.

- Fu, M. C., and Hu, J. Q. 1997a. Second derivative sample path estimators for the GI/G/m Queue, submitted to *Management Science*.
- Fu, M. C., and Hu, J. Q. 1997b. Addendum to "Extensions and generalizations of smoothed perturbation analysis in a generalized semi-Markov process framework." *IEEE Trans. on Automatic Control* preprint, submitted to *IEEE Trans. on Automatic Control*.
- Fu, M. C., and Hu, J. Q. 1997c. On unbounded hazard rates for smoothed perturbation analysis, to appear in *J. Appl. Prob.*
- Glasserman, P., and Gong, W. B. 1991. Smoothed perturbation analysis for a class of discrete-event systems. *IEEE Transactions on Automatic Control* AC-35(11): 1218–1230.
- Glasserman, P. 1991. *Gradient Estimation Via Perturbation Analysis*. Kluwer Academic Publishers.
- Glynn, P. 1987. Likelihood ratio gradient estimation: An overview. *Proceedings of the 1987 Winter Simulation Conference*, pp. 336–375.
- Gong W. B., and Ho, Y. C. 1987. Smoothed perturbation analysis of discrete-event dynamic systems. *IEEE Transactions on Automatic Control* AC-32(10): 858–866.
- Gong, W., Nananukul, S., and Yan, A. Padé approximation for stochastic discrete event systems, submitted for publication.
- Ho, Y. C., and Cao, X. R. 1983. Perturbation analysis and optimization of queueing networks. *J. of Optim. Theory and Applications* 40: 559–582.
- Ho, Y. C., Cao, X. R., and Cassandras, C. G. 1983. Infinitesimal and finite perturbation analysis for queueing networks. *Automatica* 19: 439–445.
- Ho, Y. C., and Cao, X. R. 1991. *Perturbation Analysis of Discrete Event Dynamic Systems*. Kluwer Academic Publishers.
- Loève, M. 1977. *Probability Theory I*, 4th ed. Springer-Verlag.
- Reiman, M., and Weiss, A. 1989. Sensitivity analysis for simulations via likelihood ratios. *Operations Research* 37: 830–844.
- Rubinstein, R. Y. 1986. *Monte Carlo Optimization, Simulation, and Sensitivity Analysis of Queueing Networks*. John Wiley.
- Shanthikumar, J. G., and Yao, D. D. 1989. Stochastic monotonicity in general queueing networks. *J. Appl. Prob.* 26: 413–417.
- Suri, R. 1989. Perturbation analysis: The state of the art and research issues explained via the GI/G/1 queue. *Proceedings of the IEEE* 77(1): 114–137.
- Suri, R., and Zazanis, M. A. 1988. Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue. *Mgmt. Sci.* 34(1): 39–64.
- Zazanis, M. A., and Suri, R. 1994. Perturbation analysis of the GI/GI/1 queue. *Queueing Systems* 18: 199–248.
- Zazanis, M. A. 1995. Derivative estimation via compensators: Event averages in queueing systems, to appear in *Letters of Operations Research*.