ELSEVIER

Operations Research Letters 17 (1995) 77–84

operations research letters

Derivative estimation via stochastic intensities: Event averages in queueing systems[☆]

Michael A. Zazanis

Department of Industrial Engineering and Operations Research, University of Massachusetts, Amherst, MA 01003, USA

Received 1 January 1992; revised 1 September 1994

Abstract

We briefly describe sensitivity analysis methods for simulation via stochastic intensities and give applications on derivative estimation of event averages. We consider a process depending on a parameter and an event-counting process and obtain expressions for the derivatives of the expected number of events occurring while the process is in a given set. The necessary assumptions of these expressions to hold are also examined more specifically for queueing systems. The case of queues with renewal and nonhomogeneous Poisson arrivals is discussed in some detail.

Keywords: Sensitivity analysis; Queues; Stochastic intensities; Event averages

1. Introduction

In this paper a general method for derivative estimation is proposed for a large class of problems that can be cast in a form involving stochastic integrals with respect to a counting process. It uses a compensator identity in conjunction with infinitesimal perturbation analysis (IPA) techniques to provide low variance unbiased estimates at the expense of additional computational requirements. For background on IPA and related techniques the reader is referred to [6]. See also [3] for a recent review of sensitivity analysis techniques. For earlier work on derivative estimation via compensators see [11].

2. Stochastic intensities and derivative estimation for simulations

Let $\{X_t(\theta); t \geqslant 0\}$ be a real valued stochastic process with *left-continuous paths* depending on a parameter θ, and $\{A_t; t \geqslant 0\}$ an embedded point process (e.g. arrivals to, or departures from the system), both defined on a filtered probability space and adapted to $\{\mathscr{F}_t\}_{t \geqslant 0}$. Let λ_t be the $\mathscr{F}_t$-intensity of A_t. Typically, one is

[☆] This research has been supported in part by N.S.F. Grants ECS-8811003 and DDM-8905638.

SSDI 0167-6377(94)00062-X

interested in performance criteria of the form

$$J(\theta) = E\left[\int_0^t f(X_s(\theta))\,\mathrm{d}A_s\right] \tag{1}$$

and the objective is to obtain estimators for $(\mathrm{d}/\mathrm{d}\theta)J(\theta)$. In this paper we confine ourselves to the case $f(x) = \mathbf{1}_B(x)$, the indicator function of a subset B of the state space of the process, and we obtain estimators for

$$\frac{\mathrm{d}}{\mathrm{d}\theta} E\left[\int_0^t \mathbf{1}_B(X_s)\,\mathrm{d}A_s\right].$$

Under additional conditions we will extend these results to obtain steady state estimates for $(\mathrm{d}/\mathrm{d}\theta)P_B(\theta)$, where $P_B(\theta)$ is the steady state probability that an arrival finds the system in B.

The key idea in the method we propose here is the use of the compensator identity

$$E\left[\int_0^t f(X_s)\,\mathrm{d}A_s\right] = E\left[\int_0^t f(X_s)\lambda_s\,\mathrm{d}s\right], \tag{2}$$

which holds provided that $f(X_s)$ is an $\mathscr{F}_t$-predictable process and the expectation exists (e.g. see [2]).

Here and in the sequel we suppress the dependence on θ except when we want to draw attention to it. While IPA methods would use $(\mathrm{d}/\mathrm{d}\theta)\int_0^t f(X_s)\mathrm{d}A_s$ evaluated along a sample path of the process to estimate $(\mathrm{d}/\mathrm{d}\theta)J(\theta)$, we propose to use instead $(\mathrm{d}/\mathrm{d}\theta)\int_0^t f(X_s)\mathrm{d}s$, which in many cases important in applications can also be evaluated along a sample path. The estimators obtained this way have increased computational requirements but are unbiased in many cases where the IPA estimators are biased (see [10]). This is because while it may not be permissible to differentiate with respect to θ inside the expectation on the left-hand side of (2), it is often permissible to do so on the right-hand side as we will see in the examples given in Sections 3 and 4. For a related method (Smoothed Perturbation Analysis) which uses conditioning we refer the reader to [5, 7, 8].

3. Derivative estimates for the expected number of arrivals in a given set

We start with the assumption that the process $\{X_t(\theta); t \geqslant 0\}$ has sample paths which are *left-continuous* with limits from the right w.p.1. Suppose that $\{\mathscr{F}_t\}$ is a filtration which includes the history of the process up to time t, $\sigma - \{X_s(\theta); s \leqslant t\}$. Assume also for simplicity that the counting process $\{A_t; t \geqslant 0\}$ does not depend on θ and denote its $\mathscr{F}_t$-compensator by $\Lambda_t = \int_0^t \lambda_s \mathrm{d}s$. Henceforth we will refer to this process as "arrivals" keeping in mind that it could be any embedded point process that satisfies our assumptions. Without loss of generality we assume that $X_0(\theta) \notin B$. Let $\{u_i(\theta)\}$ and $\{v_i(\theta)\}$, $i \in \mathbb{N}$, be the sequence of entrance and exit times respectively from B which in general will depend on θ. Let $\{U_t(\theta); t \geqslant 0\}$ and $\{V_t(\theta); t \geqslant 0\}$ be the counting processes associated with entrance and exit times from B, respectively. Throughout the paper prime denotes differentiation with respect to θ. To simplify the notation, the dependence on θ is often not made explicit.

We also denote by $\chi_t(\theta) = \mathbf{1}_{(X_t(\theta) \in B)}$ the indicator process corresponding to B. The postulated left-continuity of the sample paths of X guarantees that it is an $\mathscr{F}_t$-predictable process i.e. that, for all t, X_t is measurable w.r.t. the predictable σ-field $\mathscr{P}(\mathscr{F}_t)$ [2, p.8]. While $\chi_t(\theta)$ will *not* have in general left-continuous sample paths, it is also $\mathscr{F}_t$-predictable as long as B is a Borel set (since, for instance, $\{\chi_t(\theta) = 1\} = \{X_t(\theta) \in B\} \in \mathscr{P}(\mathscr{F}_t)$ in view of the fact that X_t is predictable).

Assumption A.1. *The entrance and exit times $u_i(\theta)$, $v_i(\theta)$, are piecewise differentiable functions of $\theta \in [a, b]$ for all $i \in \mathbb{N}$ w.p.1.*

In queueing applications the entrance and exit times will typically be piecewise continuous, differentiable functions of θ, the discontinuities consisting of jumps. It is at these jump points that the derivatives $u_i'(\theta)$, $v_i'(\theta)$ will fail to exist.

Assumption A.2. *With probability* 1 *the sample paths* $\{\chi_s(\theta); 0 \leqslant s \leqslant t\}$ *satisfy the following Lipschitz condition in* θ:

$$\int_0^t |\chi_s(\theta_1) - \chi_s(\theta_2)| \, ds \leqslant K(t,B)|\theta_1 - \theta_2|$$

with $E[K(t,B)] < \infty$.

Assumption A.3. *Assume that*

$$E\left[\sup_{\theta\in[a,b]} \sum_{i=1}^{V_t(\theta)} |v_i'(\theta)|\lambda_{v_i(\theta)}\right] < \infty \quad \textit{and} \quad E\left[\sup_{\theta\in[a,b]} \sum_{i=1}^{U_t(\theta)} |u_i'(\theta)|\lambda_{u_i(\theta)}\right] < \infty.$$

Finally, to simplify the exposition and the expressions obtained we will assume that $P(v_i(\theta) = t) = P(u_i(\theta) = t) = 0$ for all i. At the end of this section (Eq. (12)) we show how the expression for the derivative is modified if an entrance or exit time coincides with the end of the observation interval $[0,t]$ with probability greater than zero.

In view of (2) the reader can easily verify that the expected number of arrivals that finds the system in B is

$$E\int_0^t \chi_s \, dA_s = E\int_0^t \chi_s \lambda_s \, ds = E\sum_{i=1}^{\infty} \int_{u_i \wedge t}^{v_i \wedge t} \lambda_s \, ds. \tag{3}$$

We can now state our main result.

Theorem 1. *Under the above assumptions* $f(\theta) \overset{\text{def}}{=} E\int_0^t \chi_s(\theta)\, dA_s$ *is differentiable with derivative given by*

$$\frac{d}{d\theta} f(\theta) = E\sum_{i=1}^{V_t(\theta)} \lambda_{v_i(\theta)} v_i'(\theta) - E\sum_{i=1}^{U_t(\theta)} \lambda_{u_i(\theta)} u_i'(\theta). \tag{4}$$

Proof. Write the right-hand side of (3) as $\sum_{i=1}^{\infty} g_i(\theta)$, where

$$g_i(\theta) = E[\Lambda_{v_i \wedge t} - \Lambda_{u_i \wedge t}] = E\int_{u_i \wedge t}^{v_i \wedge t} \lambda_s \, ds. \tag{5}$$

(Since λ_s is nonnegative and $v_i \geqslant u_i$ for all ω and i, $\Lambda_{v_i \wedge t} - \Lambda_{u_i \wedge t} = \int_{u_i \wedge t}^{v_i \wedge t} \lambda_s \, ds \geqslant 0$, and Fubini's theorem applies here.)

Let $M \in \mathbb{N}$, define $\lambda_s^M = \lambda_s \wedge M$ for all s, and consider the sequences of functions $\{f_M(\theta)\}_{M\in\mathbb{N}}$ and $\{g_{i,M}(\theta)\}_{i,M\in\mathbb{N}}$, given by

$$f_M(\theta) = E\int_0^t \chi_s(\theta)\lambda_s^M \, ds \tag{6}$$

and

$$g_{i,M}(\theta) = E\int_{u_i \wedge t}^{v_i \wedge t} \lambda_s^M \, ds. \tag{7}$$

To establish (4) we shall first show that

$$\frac{\mathrm{d}}{\mathrm{d}\theta} f_M(\theta) = \sum_{i=1}^{\infty} g'_{i,M}(\theta) = \sum_{i=0}^{\infty} E[\lambda^M_{v_i} v'_i \mathbf{1}_{(v_i \leqslant t)} - \lambda^M_{u_i} u'_i \mathbf{1}_{(u_i \leqslant t)}] \quad \text{for all } M, \tag{8}$$

and then let $M \to \infty$. From A.1 and the definition of λ^M_s we have the inequality

$$\int_0^t |\chi_s(\theta + \delta) - \chi_s(\theta)| \lambda^M_s \, \mathrm{d}s \leqslant MK(t, B)|\delta|. \tag{9}$$

Hence a straightforward application of the dominated convergence theorem allows us to differentiate the right-hand side of (6) inside the expectation to obtain

$$\frac{\mathrm{d}}{\mathrm{d}\theta} f_M(\theta) = E\left[\frac{\mathrm{d}}{\mathrm{d}\theta} \int_0^t \chi_s(\theta) \lambda^M_s \, \mathrm{d}s\right] = \sum_{i=0}^{\infty} E[\lambda^M_{v_i} v'_i \mathbf{1}_{(v_i \leqslant t)} - \lambda^M_{u_i} u'_i \mathbf{1}_{(u_i \leqslant t)}]. \tag{10}$$

Having established (8) we let $M \to \infty$, and show that $\lim_{M \to \infty} (\mathrm{d}/\mathrm{d}\theta) f_M(\theta) = (\mathrm{d}/\mathrm{d}\theta) \lim_{M \to \infty} f_M(\theta) = f'(\theta)$. For this purpose it is sufficient to show (see [1, p. 204]) that

(a) $f'_M(\theta) = \sum_{i=1}^{\infty} g'_{i,M}(\theta)$ converges *uniformly* for $\theta \in [a,b]$ as $M \to \infty$ to $f'(\theta) = \sum_{i=1}^{\infty} g'_i(\theta)$, and

(b) that there exists $\theta_0 \in [a,b]$ for which $f_M(\theta_0) \to f(\theta_0)$.

To obtain the uniform convergence in (a) notice that

$$\left| \sum_{i=1}^{\infty} g'_i(\theta) - \sum_{i=1}^{\infty} g'_{i,M}(\theta) \right| = \left| \sum_{i=1}^{\infty} E[v'_i \mathbf{1}_{(v_i \leqslant t)} (\lambda_{v_i} - \lambda^M_{v_i})] - \sum_{i=1}^{\infty} E[u'_i \mathbf{1}_{(u_i \leqslant t)} (\lambda_{u_i} - \lambda^M_{u_i})] \right|.$$

Applying the triangular inequality one sees that the right-hand side of the above equation is dominated by

$$E\left[\sup_{\theta \in [a,b]} \sum_{i=1}^{\infty} \mathbf{1}_{(v_i \leqslant t)} |v'_i| \lambda_{v_i}\right] + E\left[\sup_{\theta \in [a,b]} \sum_{i=1}^{\infty} \mathbf{1}_{(u_i \leqslant t)} |u'_i| \lambda_{u_i}\right], \tag{11}$$

which is finite because of A.3. An application of the monotone convergence theorem completes the proof of (a). (b) follows immediately (for any $\theta_0 \in [a,b]$) from a straightforward application of the monotone convergence theorem in (6). □

If the simplifying assumption $P(v_i = t) = P(u_i = t) = 0$ does not hold then the counterpart of (4) is

$$\frac{\mathrm{d}}{\mathrm{d}\theta} f(\theta) = E \sum_{i=1}^{\infty} \lambda_{v_i} [v'_i \mathbf{1}_{(v_i < t)} - \mathbf{1}_{(v_i = t)} (v'_i)^-] - E \sum_{i=1}^{\infty} \lambda_{u_i} [u'_i \mathbf{1}_{(u_i < t)} - \mathbf{1}_{(u_i = t)} (u'_i)^-]. \tag{12}$$

4. Application to sensitivity analysis of queues

Consider a queueing system with arrival process $\{A_s;\ s \geqslant 0\}$ and departure process $\{D_s(\theta);\ s \geqslant 0\}$ which depends on a parameter $\theta \in [a,b]$. The number of customers in the system at time t is given by $N_s(\theta) = N_0 + A_s - D_s(\theta)$. We consider the *left-continuous version* of $\{N_s(\theta);\ s \geqslant 0\}$. Let $\{\mathscr{F}_t\}$ be a filtration such that, for each t, $\mathscr{F}_t$ contains the history of the process $\sigma - \{N_s(\theta);\ s \leqslant t\}$, and let λ_t be the $\mathscr{F}_t$-stochastic intensity of the arrival process. Denote by $J(\theta, t)$ the expected number of arrivals that find k or more customers in the system in $[0, t]$ ($k = 1, 2, \ldots$). Without loss of generality, we assume that $N_0 < k$ w.p.1. With the notation introduced in Section 3, $B = \{n \in \mathbb{N}:\ n \geqslant k\}$. $u_i(\theta)$ corresponds to the ith arrival that finds k customers in the system, and $v_i(\theta)$ to the ith departure that leaves $k - 1$ customers behind. In this system, entrances to B occur only at arrival epochs. Let T_j, $j = 1, 2, \ldots$, denote the epoch of the jth arrival and $S_j(\theta)$

the jth departure epoch. To simplify the notation we will often write $N_{S_j}(\theta)$ instead of $N_{S_j(\theta)}(\theta)$. For each θ, the sets $\{v_i(\theta);\ i \in \mathbb{N}\}$ and $\{S_j(\theta);\ j \in \mathbb{N},\ N_{S_j}(\theta) = k\}$ have the same elements since exits from B correspond to departure epochs which leave $k - 1$ customers behind ($N_s(\theta)$ is left-continuous).

The following theorem adapts the results of Section 3 to this framework and provides a set of alternative assumptions that may be easier to check in queueing applications.

Theorem 2. *Suppose that the queueing system described above satisfies Assumptions* A.1–A.3, *or alternatively the following assumptions*:
B.1. *Let $S_i(\theta)$, $i = 1, 2, \ldots$, denote the ith departure time. The derivative with respect to θ, $S_i'(\theta)$, exists w.p.1.*
B.2. $E[\sup_{\theta \in [a,b]} \sum_{i=1}^{D_t(\theta)} |S_i'(\theta)|] < \infty$.
B.3. $E[\sup_{\theta \in [a,b]} \sum_{i=1}^{D_t(\theta)} \lambda_{S_i(\theta)} |S_i'(\theta)|] < \infty$.
Then $J(\theta, t) = E \int_0^t \mathbf{1}_{(N_s(\theta) \geq k)}\, \mathrm{d}A_s$ is differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}\theta} J(\theta, t) = E \sum_{i=1}^{D_t} \lambda_{S_i} \mathbf{1}_{(N_{S_i} = k)} S_i'.$$

Proof. The proof consists in showing that the assumptions of Theorem 2 imply those in Section 3. For each i define $\Phi_i(\theta) = \inf\{j: \sum_{n=1}^{j} \mathbf{1}_{(N_{T_n}(\theta) = k-1)} = i\}$ and $\Psi_i(\theta) = \inf\{j: \sum_{n=1}^{j} \mathbf{1}_{(N_{S_n}(\theta) = k)} = i\}$. Then $u_i(\theta) = T_{\Phi_i(\theta)}$ and $v_i(\theta) = S_{\Psi_i(\theta)}$. Note that $\lim_{\delta \to 0} \Phi_i(\theta + \delta) = \Phi_i(\theta)$ and $\lim_{\delta \to 0} \Psi_i(\theta + \delta) = \Psi_i(\theta)$ w.p.1. Since the arrival process does not depend on θ, $u_i'(\theta) = 0$. Similarly, as a result of B.1, $v_i'(\theta) = S'_{\Psi_i(\theta)}(\theta)$. Hence, in this framework, B.1 implies A.1.

The fact that the arrival process does not depend on θ also allows us to write the following crude inequality:

$$\begin{aligned}\int_0^t |\mathbf{1}_{(N_s(\theta_1) \geq k)} - \mathbf{1}_{(N_s(\theta_2) \geq k)}|\, \mathrm{d}s &\leq \sum_{i=1}^{D_t(\theta_1) \vee D_t(\theta_2)} |S_i(\theta_1) - S_i(\theta_2)| \\ &\leq |\theta_1 - \theta_2| \sup_{\theta \in [a,b]} \sum_{i=1}^{D_t(\theta)} |S_i'(\theta)|.\end{aligned}$$

In the second inequality above we have made use of the mean value theorem $S_i(\theta_1) - S_i(\theta_2) = (\theta_1 - \theta_2) S_i'(\theta_0)$ with $\theta_0 \in (\theta_1, \theta_2)$. Hence B.2 implies A.2 with

$$K(t, k) = \sup_{\theta \in [a,b]} \sum_{i=1}^{D_t(\theta)} |S_i'(\theta)|.$$

Finally,

$$\sum_{i=1}^{V_t(\theta)} |v_i'(\theta)|\, \lambda_{v_i(\theta)} = \sum_{i=1}^{D_t(\theta)} \lambda_{S_i(\theta)} |S_i'(\theta)|\, \mathbf{1}_{(N_{S_i}(\theta) = k)} \leq \sum_{i=1}^{D_t(\theta)} \lambda_{S_i(\theta)} |S_i'(\theta)|$$

(since not all departures correspond to exits from B) and thus B.3 implies A.3. □

The following sections provide illustrations of the above ideas for single server queues with renewal and nonhomogeneous Poisson arrivals.

5. Derivative estimators for customer-stationary probabilities in a GI/G/1 queue

Consider a GI/GI/1 queue with input process (τ_i, σ_i), $i = 1, 2, \ldots$, where τ_i is the interarrival time between the ith and $(i + 1)$th customers, and σ_i is the workload brought by the ith customer to the system. We will assume both $\{\sigma_i\}$ and $\{\tau_i\}$ to the i.i.d., the former with distribution $F(x,\theta)$ depending on a parameter θ, and the latter with distribution G, absolutely continuous with density g.

Without loss of generality, suppose that the first customer arrives to an empty server at time $t = 0$. We will consider the number of customers in the system process $N_t(\theta)$, as θ varies in an interval $[a,b]$, such that $\sup_{\theta\in[a,b]}\int_0^\infty x\,\mathrm{d}F(x,\theta) < E\tau_1$ (to ensure stability) and $\sup_{\theta\in[a,b]}\int_0^\infty x\,\mathrm{d}F(x,\theta) < \infty$. To construct a family of sample paths parametrized by θ on the same probability space let $\sigma(\theta+\delta) = F^{-1}(F(\sigma,\theta),\theta+\delta)$. We will assume that $\sigma_i' = \mathrm{d}\sigma_i/\mathrm{d}\theta$ exists a.s. Then $\{\sigma_i'\}$ is also an i.i.d. sequence. For details we refer the reader to [12]. Notice that $\{N_t(\theta);\ t\geqslant 0\}$ determines $\{N_t(\eta);\ t\geqslant 0\}$ for all $\eta\in[a,b]$. $\{N_t(\theta);\ t\geqslant 0\}$ is defined to be *left-continuous.* We will denote by $\{A_s;\ s\geqslant 0\}$ and $\{D_s;\ s\geqslant 0\}$ the counting processes associated with arrivals and departures from the system. (These are both defined to be right continuous.) As before, $\{\mathscr{F}_t\}$ is a filtration containing the history of the process $\sigma - \{N_s(\theta);\ s\leqslant t\}$. Also define for convenience three additional (left-continuous) processes related to N_s, namely $\{Z_s;\ s\geqslant 0\}$, the time since the last arrival process, $\{\lambda_s;\ s\geqslant 0\}$, the $\mathscr{F}_t$-stochastic intensity process corresponding to the arrivals, and finally $\{Y_s;\ s\geqslant 0\}$, where

$$Y_t = \sum_{i=0}^{D_t}\sigma_i' - \int_0^t Y_s\mathbf{1}_{(N_s=1)}\,\mathrm{d}D_s. \tag{13}$$

Intuitively, if δ is a small perturbation to θ, $Y_t\delta$ is how much ahead (or behind) schedule the server would be as a result of this.

Theorem 3. *In the above framework suppose that*

$$E\sup_{\theta\in[a,b]}\sigma_i'(\theta)^2 < \infty \quad \textit{and} \quad E\left[\sup_{\theta\in[a,b]}\sum_{i=1}^{D_t(\theta)}\frac{g(Z_{S_i(\theta)})}{1-G(Z_{S_i(\theta)})}|S_i'(\theta)|\right] < \infty.$$

Let $\bar{J}(\theta,t) = (1/t)E[\int_0^t \mathbf{1}_{(N_s\geqslant k)}\,\mathrm{d}A_s]$. *Then* $\bar{J}(\theta,t)$ *is differentiable on* $[a,b]$ *and*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\bar{J}(\theta,t) = \frac{1}{t}E\left[\int_0^t \lambda_s Y_s\mathbf{1}_{(N_s=k)}\,\mathrm{d}D_s\right]. \tag{14}$$

Proof. Here, as before, $\{u_i\}$ and $\{v_i\}$ are the sequences of upcrossings from level $k-1$ to k and downcrossings from level k to $k-1$, respectively. Clearly, B.1 is satisfied with $S_i' = Y_{S_i}$, $i = 1,2,\ldots$. We next show that B.2 is satisfied. Indeed from the above remark, (13), and $N_0 = 0$, $|S_i'| \leqslant \sum_{i=1}^{A_t}|\sigma_i'|$ and hence

$$\sup_\theta\sum_{i=1}^{D_t(\theta)}|S_i'(\theta)| \leqslant \left(\sum_{i=1}^{A_t}\sup_\theta|\sigma_i'(\theta)|\right)^2.$$

Furthermore, since A_t is a renewal counting process and $\{\sup_\theta|\sigma_i'(\theta)\}$ is an i.i.d. sequence, independent of A_t, $E(\sum_{i=1}^{A_t}\sup_\theta|\sigma_i'|)^2 \leqslant CE[\sup_\theta(\sigma_1')^2]EA_t^2$ for some numerical constant $C > 0$ (see [9, p. 22]). The observation that $EA_t^r < \infty$ for all $r > 0$ for any renewal process which has interarrival times greater than zero with positive probability and the fact that $E\sup_\theta\sigma_1'^2(\theta) < \infty$ by assumption concludes this argument. B.3 is also satisfied by assumption. A straightforward application of Theorem 2 to $\bar{J}(\theta,t) = (1/t)E\int_0^t\mathbf{1}_{(N_s\geqslant k)}\,\mathrm{d}A_s$ gives

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\bar{J}(\theta,t) = \frac{1}{t}E\left[\sum_{i=1}^{D_t}\lambda_{S_i}S_i'\mathbf{1}_{(N_{S_i}=k)}\right].$$

This, together with the fact that $Y_{S_i} = S_i'$ and $V_t = \int_0^t\mathbf{1}_{(N_s=k)}\,\mathrm{d}D_s$, establishes (14). □

In the next theorem we extend the results to steady state customer averages.

Theorem 4. *Let P_k be the customer stationary probability of k or more customers in the system. Then*

$$\frac{\mathrm{d}}{\mathrm{d}\theta} P_k = \frac{1}{EQ_1} E\left[\int_0^{R_1} \lambda_s Y_s \mathbf{1}_{(N_s=k)} \mathrm{d}D_s\right], \tag{15}$$

where Q_1 and R_1 are the number of customers in the first busy cycle and its length, respectively.

Proof. It is enough to show that

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \lim_{t\to\infty} \bar{J}(\theta,t) = \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}\theta} \bar{J}(\theta,t), \tag{16}$$

the rest following from a standard regenerative argument. The validity of (16) is guaranteed from a standard theorem (e.g. see [1, p. 204]) provided that $\lim_{t\to\infty} \bar{J}(\theta_0,t)$ exists for some $\theta_0 \in [a,b]$ and that $\mathrm{d}/\mathrm{d}\theta\, \bar{J}(\theta,t)$ converges uniformly in $[a,b]$ to *some* limit as $t \to \infty$. The first condition is obviously satisfied because of the regenerative nature of the system while the second is easily seen to hold using a modification of Lorden's inequality for renewal–reward processes [4]. □

Since $S_i = \sum_{j=1}^{i} \sigma_j$ is the ith departure time in the first busy period, (15) can be written in a form suitable for regenerative simulation:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} P_k = \frac{1}{EQ_1} E\left[\sum_{i=1}^{Q_1} \mathbf{1}_{(N_{S_i}=k)} \frac{g(Z_{S_i})}{1-G(Z_{S_i})}\left(\sum_{j=1}^{i} \sigma_j'\right)\right]. \tag{17}$$

Let us note here once again that differentiating inside the expectation to get the right-hand side (14) would not have been possible had we not used the compensator identity (2). Thus the IPA estimate would be biased in that case.

6. Nonhomogeneous Poisson arrivals

We start with a sequence of service times and a deterministic, time varying intensity $\lambda(s,\theta)$ that depends on a parameter $\theta \in \Theta$. Assume that $\lambda(s,\theta)$ is strictly positive on $\mathbb{R}^+ \times \Theta$ and that $(\partial/\partial\theta)\Lambda(s,\theta)$ exists for all $s \in \mathbb{R}^+$. We construct a parametrized workload process $X_s(\theta)$ by letting the nth arrival time $T_n(\theta)$ be

$$T_n(\theta) = \Lambda^{-1}(e_n,\theta),$$

where $\{e_n\}_{n\in\mathbb{N}}$ is a unit rate Poisson process. Then, since $e_n = \Lambda(T_n(\theta),\theta)$, differentiating with respect to θ we obtain

$$\frac{\mathrm{d}T_n}{\mathrm{d}\theta} = -\frac{\mathrm{D}_2\Lambda(T_n,\theta)}{\mathrm{D}_1\Lambda(T_n,\theta)} = -\frac{\mathrm{D}_2\Lambda(T_n,\theta)}{\lambda(T_n,\theta)},$$

where $\mathrm{D}_1\Lambda$ (resp. $\mathrm{D}_2\Lambda$) is the derivative of Λ with respect to its first (resp. second) argument. In that case,

$$\frac{\mathrm{d}}{\mathrm{d}\theta} E\int_0^t \mathbf{1}_B(X_s)\,\mathrm{d}A_s = E\sum_{\{v_i\in(0,t]\}} \lambda_{v_i} v_i' - E\sum_{\{u_i\in(0,t]\}} \lambda_{u_i} u_i' + \sum_{i=1}^{\infty} E[\mathbf{1}_{(v_i\leq t)}[\mathrm{D}_2\Lambda(v_i,\theta)-\mathrm{D}_2\Lambda(u_i,\theta)] + \mathbf{1}_{(u_i<t\leq v_i)}[\mathrm{D}_2\Lambda(t,\theta)-\mathrm{D}_2\Lambda(u_i,\theta)]]. \tag{18}$$

7. Implementation considerations

Notice that (17) can be used to estimate $(\mathrm{d}/\mathrm{d}\theta)P_k$ while observing a single path of the system by simply keeping track of two quantities, namely the age of the busy period Y_t and the age of the arrival process Z_t. When θ is a location (scale) parameter of $F(x, \theta)$, Y_{S_i} becomes the discrete (continuous) age of the busy period at the ith departure epoch, making the implementation of (17) very simple.

Here of course we take advantage of the fact that the stochastic intensity is simply the hazard rate, and the only part of the history of the process necessary to determine λ_t is the age of the arrival process at time t. For most models used in practice, one would be able to compute the stochastic intensity easily. This would be the case for instance for superpositions of renewal processes (in which case one would of course need to know the ages of all the arrival processes involved), for Markov renewal processes (in which case one would need to know the state of the underlying Markov chain and the time since the last arrival), for interrupted renewal processes (such as the output from an upstream server), etc.

References

[1] R.G. Bartle, *The Elements of Real Analysis*, Wiley, New York, 1976.

[2] P. Brémaud, *Point Processes and Queues*, Springer, New York 1981.

[3] P. Brémaud, "Maximal coupling and rare perturbation analysis", *QUESTA* **11**, 307–333, 1992.

[4] P.H. Chen and M. Zazanis, "Steady state IPA estimates for regenerative systems", Technical Report, IEOR Dept., Univ. of Massachusetts, Amherst, 1993.

[5] M.C. Fu and J.Q. Hu, "Smoothed perturbation analysis for general discrete-event systems", *IEEE Trans. Automat. Control* **AC-37**, 1483–1500, 1992.

[6] P. Glasserman, *Gradient Estimation via Perturbation Analysis*, Kluwer Academic Publisher, Cambridge, MA, 1990.

[7] P. Glasserman and W.B. Gong, "Smoothed perturbation analysis for a class of discrete event systems", *IEEE Trans. Automat. Control* **AC-35**, 1218–1230, 1990.

[8] W.B. Gong and Y.C. Ho, "Smoothed perturbation analysis for discrete event systems", *IEEE Trans. Automat. Control* **AC-32**(10) 858–866, 1987.

[9] A. Gut, "*Stopped Random Walks: Limit Theorems and Applications*", Springer, New York, 1988.

[10] P. Heidelberger, X.R. Cao, M.A. Zazanis and R. Suri, "Convergence properties of infinitesimal perturbation analysis", *Management Sci.* **34**(11) 1281–1301, 1988.

[11] M.A. Zazanis, "Compensators and sensitivity analysis of queueing systems", in: *Proc. 26th Allerton Conf. on Communication, Control, and Computing*, pp. 549–554, 1988.

[12] M.A. Zazanis and R. Suri, "Perturbation analysis of the GI/GI/1 queue", *QUESTA* **18**, 199–248, 1994.