DERIVATIVE ESTIMATES FROM DISCONTINUOUS REALIZATIONS: SMOOTHING TECHNIQUES

Paul Glasserman
Room 3K-324
AT&T Bell Laboratories
Holmdel, NJ 07733

Wei-Bo Gong
Electrical and Computer Engineering
University of Massachusetts
Amherst, MA 01003

ABSTRACT

We develop two methods for estimating derivatives of expectations from simulation of functions whose realizations are discontinuous in the parameter of differentiation. We take as motivating example the estimation of the sensitivity of expected terminal reward for processes on discrete state spaces. Both our methods use conditional expectations to smooth discontinuites. The first smooths the dependence on the differentiation parameter, while the second smooths dependence on the time parameter. The methods are illustrated through examples, including stochastic networks, networks of queues, and Markov processes.

1. INTRODUCTION

This paper considers the problem of estimating derivatives of expectations from realizations that are discontinuous functions of the parameter of differentiation. It is known that — in its simplest form — the method of infinitesimal perturbation analysis (IPA) can only be applied in cases where the realizations are continuous in the parameter (see [4] for a discussion of the role of continuity in IPA). Here, we develop two modifications of IPA that "smooth" discontinuities in order to obtain unbiased derivative estimates.

To motivate the problem of discontinuities, let $Z_t(\theta)$ be a parametric family of processes on a discrete state space S. For f a real-valued function on S, and fixed $t_f > 0$ define the *terminal reward*

$$R(\theta) = f(Z_{t_f}(\theta)). \tag{1}$$

Depending on what Z_t and f represent, $E[R(\theta)]$ could be, for example, a mean queue length at t_f, or the probability that some system is still functioning at t_f. Suppose we want to estimate $dE[R(\theta)]/d\theta$. The IPA estimate is $R'(\theta) = dR(\theta)/d\theta$. But since $Z_{t_f}(\theta)$ takes only discrete values, the same is true of $R(\theta)$; hence $R'(\theta) = 0$ wherever it exists. As a function of θ, R is (at best) piecewise constant. The "sample derivative" $R'(\theta)$ contains no information about the dependence of $E[R(\theta)]$ on θ.

This situation should be contrasted with that of *cumulative reward* given by

$$L(\theta) = \int_0^{t_f} g(Z_t(\theta))dt. \tag{2}$$

Since L takes on a continuum of values, it cannot be *a priori* declared discontinuous the way R can. In fact, a variety of conditions on $Z_t(\theta)$ ensure that $E[L'(\theta)] = E[L(\theta)]'$; some are considered in [4],[5] and [6].

To deal with (1) and similar problems we consider two smoothing techniques. Both use conditional expectations, but in different ways. The first is an application of *smoothed perturbation analysis* (SPA) as developed in Gong and Ho [8] and Glasserman and Gong [7]. Roughly speaking, SPA uses an estimate of the form $E[R(\theta)|\xi]'$ where ξ is determined by $\{Z_t(\theta), t \geq 0\}$. The condition ξ is chosen so that $E[R(\theta+h)|\xi]$ is continuous in h. We apply SPA to (1), after considering simpler problems. Our second approach to smoothing uses Dynkin's formula to convert (1) to (2) and then applies IPA. While the two methods are, in general, distinct, we identify a small class of problems for which both yield the same estimator.

Previous work on calculating $dE[R(\theta)]/d\theta$ (when $Z_t(\theta)$ is Markov) includes the numerical method in Heidelberger and Goyal [9], and the numerical and simulation methods considered in Fox [2]. (See also other references in [2].) These papers partly motivated our work. Our approach based on Dynkin's formula has close ties to the elegant compensator method introduced (earlier, independently) by Zazanis [11] via the single-server queue.

2. SPA: THE BASIC IDEA

Let $\gamma(\theta)$ be a random, piecewise constant function taking only the values $x_1, ..., x_m$. (Think of $\gamma(\theta)$ as the outcome of a simulation conducted at the parameter value θ, in a situation where re-running the simulation using common random numbers with θ slightly perturbed may not change the value of γ.) If $E[\gamma(\theta)]$ is differentiable, then

$$
\begin{aligned}
\frac{dE[\gamma(\theta)]}{d\theta} &= \lim_{h \downarrow 0} E\left[\frac{\gamma(\theta+h) - \gamma(\theta)}{h}\right] \\
&= \lim_{h \downarrow 0} E\left[\frac{E[\gamma(\theta+h) - \gamma(\theta)|\gamma(\theta)]}{h}\right].
\end{aligned} \tag{3}
$$

If the limit and the outer expectation can be interchanged, this becomes

$$
= E\left[\lim_{h \downarrow 0} \frac{E[\gamma(\theta+h) - \gamma(\theta)|\gamma(\theta)]}{h}\right]
$$

$$= \mathbf{E}\left[\lim_{h\downarrow 0}\sum_{j=1}^{m}\frac{1}{h}P(\gamma(\theta+h)=x_j|\gamma(\theta))[x_j-\gamma(\theta)]\right].$$

If we define

$$q(x_i,x_j)=\lim_{h\downarrow 0}\frac{1}{h}P(\gamma(\theta+h)=x_j|\gamma(\theta)=x_i),\qquad (4)$$

then

$$\sum_{j=1}^{m}q(\gamma(\theta),x_j)[x_j-\gamma(\theta)]\qquad (5)$$

is an unbiased estimate of $\mathbf{E}[\gamma(\theta)]'$. Bringing the limit fully inside the expectation in (3) (instead of only partially) would yield the uninteresting estimate $\gamma'(\theta)\equiv 0$.

The estimate in (5) requires calculation of $q(x_i,x_j)$ (whereas $\gamma(\theta)$ is generated from simulation) which can be difficult. Evaluation of the limit in (4) is generally facilitated by conditioning on more information. Suppose, then, that $\xi=\xi(\theta)$ represents additional information available through simulation (including $\gamma(\theta)$ — i.e., $\gamma\in\sigma(\xi)$), and let

$$Q(x_j;\xi)=\lim_{h\downarrow 0}\frac{1}{h}P(\gamma(\theta+h)=x_j|\xi).\qquad (6)$$

Proceeding as in (3)-(5), this yields the estimate

$$\sum_{j=1}^{m}Q(x_j;\xi)[x_j-\gamma(\theta)].\qquad (7)$$

We think of $Q(x_j;\xi)$ as the (conditional) jump rate from $\gamma(\theta)$ to x_j, and $[x_j-\gamma(\theta)]$ as the effect or size of such a jump. Note that $Q(x_j;\xi)$ is, indeed, a conditional intensity for γ viewed as a stochastic process in θ.

Taking this one step further, we can condition different jumps on different information. For $j=1,\ldots,m$ let $\gamma\in\sigma(\xi_j)$. Writing

$$\mathbf{E}[\gamma(\theta+h)-\gamma(\theta)]=\mathbf{E}\left[\sum_{j=1}^{m}P(\gamma(\theta+h)=x_j|\xi_j)[x_j-\gamma(\theta)]\right]$$

leads to the estimate

$$\sum_{j=1}^{m}Q(x_j;\xi_j)[x_j-\gamma(\theta)].\qquad (8)$$

Tailoring ξ_j can make each $Q(x_j;\xi_j)$ easier to calculate than $Q(x_j;\xi)$ with ξ fixed. We will provide explicit expressions for $Q(\cdot;\cdot)$ for a variety of examples.

As a trivial but illustrative example, consider the problem of estimating

$$\frac{d}{d\theta}P(X(\theta)<t)=\partial_\theta F(t,\theta),$$

when the dependence of X on θ is via

$$X(\theta)=X(1)/\theta.\qquad (9)$$

If we set $\gamma(\theta)=1\{X(\theta)<t\}$ then $P(X(\theta)<t)=\mathbf{E}[\gamma(\theta)]$.

We condition on $\gamma(\theta)$. Since X is decreasing in θ, when $\gamma(\theta)=1$ no increase in θ can change the value of γ; hence, when $\gamma(\theta)=1$ the estimate is zero. When $\gamma(\theta)=0$, we need to differentiate the rate at which it jumps to one:

$$\lim_{h\downarrow 0}\frac{1}{h}P(\gamma(\theta+h)=1|\gamma(\theta)=0)$$

$$=\lim_{h\downarrow 0}\frac{1}{h}P(X(\theta+h)<t|X(\theta)>t)$$

$$=\lim_{h\downarrow 0}\frac{1}{h}P(X(\theta)<\frac{\theta+h}{\theta}t|X(\theta)>t)$$

$$=\lim_{h\downarrow 0}\frac{1}{h}\frac{F(\frac{\theta+h}{\theta}t,\theta)-F(t,\theta)}{1-F(t,\theta)}$$

$$=\frac{f(t,\theta)}{1-F(t,\theta)}\frac{t}{\theta},$$

where f is the density of F. The relation (9) implies that $dX/d\theta=-X/\theta$, so this could be expressed as

$$\frac{f(t,\theta)}{1-F(t,\theta)}\left(-\frac{dX}{d\theta}\right)_{X=t}$$

(the second factor is the derivative of X evaluated at $X=t$). This jump rate has a simple interpretation which extends to more general cases: The first factor is the "probability" that X is infinitesimally greater than t, given that it is greater than t; when multiplied by $-dX/d\theta$ (evaluated at $X=t$) this yields the rate at which X and t change order.

To obtain the full estimate, we note that this order change can only occur when $X(\theta)>t$, and when it occurs, γ jumps from 0 to 1, yielding an effect of $[1-0]=+1$. Thus, we get

$$1\{X(\theta)>t\}\frac{f(t,\theta)}{1-F(t,\theta)}\left(-\frac{dX}{d\theta}\right)_{X=t}[+1].\qquad (10)$$

Furthermore, noting that (9) implies $F(x,\theta)=F(x\theta,1)$, differentiation shows that X admits the inversion representation

$$\frac{dX}{d\theta}=-\frac{\partial_\theta F(X,\theta)}{\partial_x F(X,\theta)}.\qquad (11)$$

Substituting (11) into (10) and taking the expectation we get

$$(1-F(t,\theta))\frac{f(t,\theta)}{1-F(t,\theta)}\left(\frac{\partial_\theta F(t,\theta)}{f(t,\theta)}\right)=\partial_\theta F(t,\theta)$$

which is, of course, what we wanted.

Consider, next, a modification of the problem above in which we try to estimate

$$\frac{d}{d\theta}P(X_1(\theta) < X_2(\theta)).$$

Let i_1 and i_2 be, respectively, the indices of the smaller and the larger of the $X_i(\theta)$. We condition on $\xi = (i_1, X_{i_1})$ and consider the rate at which X_1 and X_2 change order:

$$\lim_{h \downarrow 0} \frac{1}{h} P(X_{i_1}(\theta + h) > X_{i_2}(\theta + h) | i_1, X_{i_1}). \qquad (12)$$

Suppose there are functions D_i, $i = 1, 2$ such that

$$\frac{dX_i(\theta)}{d\theta} = D_i(X_i(\theta), \theta) \qquad (13)$$

(more on this condition below). Formally, proceeding as above, we obtain the estimate

$$\frac{f_{i_2}(X_{i_1}(\theta))}{1 - F_{i_2}(X_{i_1}(\theta))} \left[\frac{dX_{i_1}}{d\theta} - D_{i_2}(X_{i_1}) \right]^+ [1\{i_1 = 2\} - 1\{i_1 = 1\}].$$
$$(14)$$

Since $D_{i_2}(X_{i_1})$ has the interpretation

$$D_{i_2}(X_{i_1}) = \left(\frac{dX_{i_2}}{d\theta} \right)_{X_{i_2} = X_{i_1}}$$

the $[\cdot]^+$ term is the rate at which the X_i change order, given that X_{i_2} is just greater than X_{i_1}. The "+" arises from the fact that as $h \downarrow 0$, the conditional probability in (12) is zero unless X_{i_1} is increasing faster than X_{i_2}. Justification of this limit requires some elaboration on the dependence of the X_i on θ. This will be carried out in the next section.

3. A CLASS OF NETWORK PROBLEMS

We now extend the examples of the previous section to some functions of the form

$$\gamma(\theta) = \Gamma(X_1(\theta), ..., X_n(\theta)), \qquad (15)$$

where Γ maps $\mathbf{R}^n$ to a finite set, and the $X_i(\theta)$ are independent. The situation we have in mind is one in which Γ represents a network property and the X_i are stochastic capacities, arc lengths or lifetimes; but this interpretation is not essential. Let $i_1, ..., i_n$ be order indices for the X_i:

$$X_{i_1} \leq X_{i_2} \leq ... \leq X_{i_n}.$$

We restrict attention to Γ determined by a function g via

$$\Gamma(X_1, ..., X_n) = g(i_1, ..., i_n). \qquad (16)$$

As an example, take the X_i to be component lifetimes in a system that functions so long as at least one component in each of a collection of fixed subsets C_j functions. Then if g

is the indicator of the event

$$X_k = \min_j \max_{i \in C_j} X_i,$$

$E[\gamma]$ is the probability that the system ultimately fails due to failure of component k. Other types of Γ can be handled using similar techniques.

We begin by taking up in more detail the calculation of $Q(\cdot; \cdot)$, imposing the following provisional conditions:

(A1'). Each X_i is, with probability one, a differentiable function of θ.

(A2'). Each $X_i(\theta)$ has a density $f_i(\cdot, \theta)$ (with cdf $F_i(\cdot, \theta)$).

For some of our calculations we will need a strengthening of (A1') which generalizes (9). Fix a nominal value of θ and suppose there is a function Φ_i such that

$$\Phi_i(X_i(\theta), h) = X_i(\theta + h). \qquad (17)$$

Think of Φ_i as the general solution to the differential equation specified by (13). Note that $\Phi_i(x, 0) = x$, and

$$\partial_h \Phi_i(x, 0) = \lim_{h \downarrow 0} \frac{\Phi_i(x, h) - \Phi_i(x, 0)}{h}$$
$$= D_i(x) = \left(\frac{dX_i}{d\theta} \right)_{X_i = x} \qquad (18)$$

Also, if Φ_i has an inverse (with respect to its second argument), then (17) shows that it is given by $\Phi_i^{-1}(\cdot, h) = \Phi_i(\cdot, -h)$.

We impose

(A1). For each X_i there is a Φ_i satisfying (17) which is continuously differentiable in both arguments and strictly monotonic in the second argument throughout a neighborhood of θ.

Though somewhat unintuitive, (A1) is fairly general, especially when (A1') is known to hold. (A candidate Φ is provided by the solution to (11).) To show how (A1) is used, we return to (12). We can rewrite it as

$$\lim_{h \downarrow 0} \frac{1}{h} P(\Phi_{i_1}(X_{i_1}(\theta), h) > \Phi_{i_2}(X_{i_2}, h) | i_1, X_{i_1})$$
$$= \lim_{h \downarrow 0} \frac{1}{h} P(X_{i_2} < \Phi_{i_2}(\Phi_{i_1}(X_{i_1}, h), -h) | i_1, X_{i_1})$$
$$= \lim_{h \downarrow 0} \frac{1}{h} \frac{F_{i_2}(\Phi_{i_2}(\Phi_{i_1}(X_{i_1}, h), -h)) - F_{i_2}(X_{i_1})}{1 - F_{i_2}(X_{i_1})},$$

provided

$$\Phi_{i_2}(\Phi_{i_1}(X_{i_1}(\theta), h), -h) > X_{i_1}(\theta), \qquad (19)$$

and zero otherwise. Unless (19) holds for all sufficiently small $h > 0$, the limit is zero. When (19) does hold for all small h, the chain rule yields

$$\frac{f_{i_2}(X_{i_1})}{1 - F_{i_2}(X_{i_1})} \cdot \frac{d}{dh}[\Phi_{i_2}(\Phi_{i_1}(X_{i_1}, h), -h)]_{h=0}.$$

Note that h appears in both arguments of Φ_{i_2}. The derivative can be expressed as

$$\lim_{h \downarrow 0} \frac{1}{h} \Big(\Phi_{i_2}(\Phi_{i_1}(X_{i_1}, h), -h) - \Phi_{i_2}(\Phi_{i_1}(X_{i_1}, 0), -h) +$$

$$\Phi_{i_2}(\Phi_{i_1}(X_{i_1}, 0), -h) - \Phi_{i_2}(\Phi_{i_1}(X_{i_1}, 0), 0) \Big)$$

since the middle terms cancel. If Φ_1 and Φ_2 are C^1, we can take the limit as $h \downarrow 0$ separately for h in the first and second arguments of Φ_{i_2} to get

$$\lim_{h \downarrow 0} \frac{\Phi_{i_2}(\Phi_{i_1}(X_{i_1}, h), 0) - \Phi_{i_2}(\Phi_{i_1}(X_{i_1}, 0), 0)}{h}$$

$$+ \lim_{h \downarrow 0} \frac{\Phi_{i_2}(\Phi_{i_1}(X_{i_1}, 0), -h) - \Phi_{i_2}(\Phi_{i_1}(X_{i_1}, 0), 0)}{h}.$$

Since $\Phi_{i_2}(x, 0) = x = \Phi_{i_1}(x, 0)$, this is

$$\lim_{h \downarrow 0} \frac{\Phi_{i_1}(X_{i_1}, h) - \Phi_{i_1}(X_{i_1}, 0)}{h}$$

$$+ \lim_{h \downarrow 0} \frac{\Phi_{i_2}(X_{i_1}, -h) - \Phi_{i_2}(X_{i_1}, 0)}{h}.$$

Equation (18) now shows that this is

$$\frac{dX_{i_1}}{d\theta} - D_{i_2}(X_{i_1}).$$

Combining this with the case where (19) does not hold for all small h yields (14).

We now apply similar ideas to γ as defined via Γ and g in (15) and (16). Since g could, in general, take up to $n!$ different values, derivation of an SPA estimate for $d\mathrm{E}[\gamma(\theta)]/d\theta$ would seem to require consideration of $n! - 1$ possible changes in γ. Fortunately, most of the changes occur with negligible probability as $h \downarrow 0$. Let $\underline{i} = (i_1, ..., i_n)$ be the sequence of order indices, and let $\underline{i}_j$ be the sequence obtained from $\underline{i}$ by transposing i_j and i_{j+1}. Let $\mathcal{A}(\theta)$ be the set consisting of $\underline{i}(\theta)$ and all $\underline{i}_j(\theta)$, $j = 1, ..., n - 1$. Then (A1') and (A2') imply that

$$P(\underline{i}(\theta + h) \notin \mathcal{A}(\theta)) = o(h); \tag{20}$$

thus, we only need to consider the $n - 1$ possible order changes of consecutive X_{i_j}.

For each of these possible order changes we need to find the conditional jump rate, as in (6). For the jump from $\underline{i}$ to

$\underline{i}_j$ we condition on $\xi_j = (\underline{i}, X_{i_j})$. The general form of the estimate (8) is then given by

$$\sum_{j=1}^{n-1} Q(\underline{i}_j; \xi_j)[g(\underline{i}_j) - g(\underline{i})].$$

(We have abbreviated $Q(g(\underline{i}_j); \xi_j)$ to $Q(\underline{i}_j; \xi_j)$.) As in the example of the previous section, each jump rate takes the form

$$Q(\underline{i}_j; \xi_j) = \hat{f}_j(0) \left[\frac{dX_{i_j}}{d\theta} - D_{i_{j+1}}(X_{i_j}) \right]^+,$$

where $\hat{f}_j$ is the conditional density given ξ_j of the residual lifetime of $X_{i_{j+1}}$. This density is given, up to proportionality, by

$$\hat{f}_j(t) \sim f_{i_{j+1}}(X_{i_j} + t) \prod_{k=j+1}^{n} (1 - F_{i_k}(X_{i_j} + t)).$$

Equality is obtained by dividing the right side by its integral (over $0 \leq t < \infty$) to make it a probability density. Determination of this type of normalization constant is the one non-trivial step in implementing SPA estimates.

We prove unbiasedness of this estimator under two additional conditions, the first a strengthening of (A2'):

(A2). For each i and θ there is a $\lambda_i(\theta)$ such that for all $x \geq 0$, $f_i(x, \theta)/(1 - F_i(x, \theta)) \leq \lambda_i(\theta)$.

(A3). There is a constant $B > 0$ such for every i, with probability one, $|dX_i/d\theta| \leq B(X_i + 1)$.

Integration shows that (A3) implies

$$|X_i(\theta + h) - X_i(\theta)| \leq (e^{Bh} - 1)(X_i(\theta) + 1). \tag{21}$$

Proposition 1. Suppose (A1)-(A3) hold, and suppose that every X_j has a finite expectation. Then, if $\mathrm{E}[\gamma(\theta)]$ is differentiable,

$$\mathrm{E}\left[\sum_{j=1}^{n-1} Q(\underline{i}_j; \xi_j)[g(\underline{i}_j) - g(\underline{i})] \right] = \frac{d\mathrm{E}[\gamma(\theta)]}{d\theta}.$$

Proof. There are two parts to the proof: showing that our expression for $Q(\underline{i}_j; \xi_j)$ is correct, and justifying the interchange of the limit and the expectation in

$$\lim_{h \downarrow 0} \frac{1}{h} \mathrm{E}\left[\sum_{j=1}^{n-1} P(\underline{i}(\theta + h) = \underline{i}_j | \xi_j)[g(\underline{i}_j) - g(\underline{i})] \right] =$$

$$\mathrm{E}\left[\lim_{h \downarrow 0} \frac{1}{h} \sum_{j=1}^{n-1} P(\underline{i}(\theta + h) = \underline{i}_j | \xi_j)[g(\underline{i}_j) - g(\underline{i})] \right].$$

The left side is the derivative of $\mathrm{E}[\gamma(\theta)]$; the right side is the expectation of the estimator. Verification that $Q(\underline{i}; \xi_j)$

is correct is similar to the example carried out above so we omit it. For the interchange, we bound the probability of a jump:

$$\frac{1}{h} P(\underline{i}(\theta + h) = \underline{i}_j | \xi_j)$$

$$\leq \frac{1}{h} P(X_{i_j}(\theta + h) > X_{i_{j+1}}(\theta + h) | \xi_j)$$

$$\leq \frac{1}{h} \left| \frac{F_{i_{j+1}}(\Phi_{i_j}(\Phi_{i_{j+1}}(X_{i_j}, h), -h)) - F_{i_{j+1}}(X_{i_j})}{1 - F_{i_{j+1}}(X_{i_j})} \right|.$$

Applying the mean value theorem and (A2), this is

$$\leq \lambda_{i_{j+1}} \left| \frac{\Phi_{i_j}(\Phi_{i_{j+1}}(X_{i_j}, h), -h)) - X_{i_j}}{h} \right|. \qquad (22)$$

The numerator is bounded by

$$|\Phi_{i_j}(\Phi_{i_{j+1}}(X_{i_j}, h), -h) - \Phi_{i_{j+1}}(X_{i_j}, h)| + |\Phi_{i_{j+1}}(X_{i_j}, h) - X_{i_j}|$$

Using (21) for each term and simplifying, (22) becomes

$$\leq \lambda_{i_{j+1}} \frac{1}{h} (e^{2Bh} - 1)(X_{i_j} + 1).$$

Another application of the mean value theorem yields

$$\leq \lambda_{i_{j+1}} (2Bhe^{2Bh})(X_{i_j} + 1).$$

Since $E[X_{i_j}] < \infty$, the dominated convergence theorem now justifies the interchange.

4. TERMINAL REWARD IN QUEUEING NETWORKS

We return, now, to the problem described in the introduction, in the special case where $Z_t(\theta)$ is the state (vector of queue lengths) in a closed Jackson-like queueing network. Through choice of f in (1), $E[R(\theta)]$ could be, for example, the mean queue length at t_f at a fixed node, or the probability that some node is idle at t_f.

We use the following notation:

τ_i = epoch of ith transition;

Y_i = ith state = $Z_{\tau_i^+}$;

a_i = node from which ith departure occurs;

λ_{aj} = queue joined by jth departure from node a;

X_{aj} = jth service time at a;

f_a, F_a = service time density and cdf at node a.

The service times X_{aj} depend on θ; the sequence, λ, of routing indicators, and the initial state, Y_0, do not.

We first consider a (slightly easier) discrete time version of terminal reward, given by $R_n(\theta) = f(Y_n(\theta))$. We condition

on

$$\xi_n = (Y_0; \tau_1, \ldots, \tau_n; a_1, \ldots, a_{n+1}; \lambda).$$

The node a_{n+1} from which the next departure occurs is included, but the epoch τ_{n+1} of its occurrence is not. Given ξ_n, we can reconstruct $Y_1, \ldots, Y_n$ and $\{Z_t, 0 \leq t \leq \tau_n\}$, but not the entire evolution of Z_t.

Let $a(\theta) = (a_1(\theta), \ldots, a_n(\theta))$. Let $\mathcal{A}_n(\theta)$ be the set consisting of $a(\theta)$, any sequence that can be obtained from $a(\theta)$ by transposing some $a_i(\theta)$ and $a_{i+1}(\theta)$, $i < n$, and also the sequence $a^*(\theta) = (a_1(\theta), \ldots, a_{n-1}(\theta), a_{n+1}(\theta))$. If the service times satisfy (A1') and (A2') then

$$P(a(\theta + h) \notin \mathcal{A}_n(\theta)) = o(h). \qquad (23)$$

In other words, the probability of multiple order changes among the service completions is negligible. Our estimator of the derivative of $E[R_n(\theta)]$ will, therefore, take the form

$$\sum_{a' \in \mathcal{A}_n} Q(a'; \xi_n) E[f(Y_n(\theta + h)) - f(Y_n(\theta)) | \xi_n, a(\theta + h) = a'],$$

where

$$Q(a'; \xi_n) = \lim_{h \downarrow 0} \frac{1}{h} P(a(\theta + h) = a' | \xi_n).$$

Given ξ_n and $a(\theta + h)$, $Y_n(\theta)$ and $Y_n(\theta + h)$ are known. Moreover, Y_n does not depend on the order of $a_1, \ldots, a_n$. Given λ and Y_0, it is enough to know how many departures occurred from each node to determine Y_n; the order in which the departures occur does not change Y_n. If $a(\theta + h) \in \mathcal{A}_n(\theta)$ and $a(\theta + h) \neq a^*$, then $a(\theta + h)$ is just a permutation of $a(\theta)$ and $Y_n(\theta + h) = Y_n(\theta)$. Our estimator therefore reduces to

$$Q(a^*; \xi_n) E[f(Y_n(\theta + h)) - f(Y_n(\theta)) | \xi_n, a(\theta + h) = a^*].$$

If we define Y_n^* to be the nth state when a_n and a_{n+1} change order ($Y_n^* \in \sigma(\xi_n)$), we get simply

$$Q(a^*; \xi_n)[f(Y_n^*) - f(Y_n)].$$

It remains to calculate $Q(a^*; \xi_n)$, which is the rate at which a_n and a_{n+1} change order. The general form of this rate should by now be clear:

$$Q(a^*; \xi_n) = \dot{f}(0) \left[\frac{d\tau_n}{d\theta} - \left(\frac{d\tau_{n+1}}{d\theta} \right)_{\tau_{n+1} = \tau_n} \right]^+ ;$$

$\dot{f}(0)$ is the conditional probability density that a_{n+1} occurs just after a_n, and the $[\cdot]^+$ term is the rate at which they change order, given that they occur simultaneously.

385

Given ξ_n, the residual time (at τ_n) until τ_{n+1} is the residual service time at node a_{n+1} conditioned on being the shortest residual service time at any node. If t_a is the age (at τ_n) of the current service time (if any) at node a, then this density is, for $x \geq 0$,

$$\hat{f}(x) \sim f_{a_{n+1}}(t_{a_{n+1}} + x) \prod_a (1 - F_a(t_a + x)),$$

where the product runs over non-idle nodes a other than a_{n+1}.

The second factor in $Q(a^*; \xi_n)$ is evaluated using IPA. There is a sequence $X_{i_k j_k}$ (determined by ξ_n) such that

$$\tau_n = \sum_k X_{i_k j_k},$$

and

$$\frac{d\tau_n}{d\theta} = \sum_k \frac{dX_{i_k j_k}}{d\theta} = \sum_k D_{i_k}(X_{i_k j_k}).$$

Similarly, if $\tau_{n+1} = \tau_m + X_{a_{n+1} k}$ — i.e., the service time that ends at τ_{n+1} began at τ_m and is the kth service time at a_{n+1} — then

$$\left(\frac{d\tau_{n+1}}{d\theta}\right)_{\tau_{n+1} = \tau_n} = \frac{d\tau_m}{d\theta} + D_{a_{n+1}}(\tau_n - \tau_m).$$

Thus, these terms are easily computed as the simulation evolves.

Theorem 1. Suppose the service times satisfy (A1)-(A3) and have finite expectations. If $\mathbf{E}[R_n(\theta)]$ is differentiable, then

$$\mathbf{E}\left[Q(a^*; \xi_n)[f(Y_n^*) - f(Y_n)]\right] = \frac{d\mathbf{E}[R_n(\theta)]}{d\theta}.$$

This (almost) follows from Theorem 1 of [7], which is a result about generalized semi-Markov processes. A condition less general than (A1) and (A3) is used in [7]; the gap is filled by the argument used to prove Proposition 1.

We now consider the case of $R(\theta) = f(Z_{t_f}(\theta))$. Let N_t be the number of transitions in $[0, t]$. Denote N_{t_f} simply by N and let $a = (a_1, ..., a_N)$. Let $\mathcal{A}$ include a and sequences obtained through a single transposition, as before. But instead of a^*, add to $\mathcal{A}$

$$a_+ = (a_1, ..., a_N, a_{N+1})$$

and

$$a_- = (a_1, ..., a_{N-1}).$$

The sequences a_+ and a_- are, respectively, the result of order changes between τ_{N+1} and t_f, and between τ_N and t_f. If the service times satisfy (A1') and (A2),

$$P(a(\theta + h) \notin \mathcal{A}(\theta)) = o(h).$$

Moreover, of the elements of $\mathcal{A}(\theta)$, only a_+ and a_- result in $Z_{t_f}(\theta + h) \neq Z_{t_f}(\theta)$, so we only need to consider these order changes. Let Z_+ and Z_- denote $Z_{t_f}(\theta + h)$ when $a(\theta + h)$ is $a_+(\theta)$ and $a_-(\theta)$ respectively.

If we condition on $\xi_+ = \xi_N$ and $\xi_- = \xi_{N-1}$, we get

$$Q(a_+; \xi_+) = \hat{f}_+(0) \left[-\left(\frac{d\tau_{N+1}}{d\theta}\right)_{\tau_{N+1} = t_f}\right]^+$$

and

$$Q(a_-; \xi_-) = \hat{f}_-(0) \left[\left(\frac{d\tau_N}{d\theta}\right)_{\tau_N = t_f}\right]^+.$$

To evaluate the densities $\hat{f}_+$ and $\hat{f}_-$, let t_a be the age of the current service time at node a at t_f. Then for all $x \geq 0$,

$$\hat{f}_+(x) \sim f_{a_{N+1}}(t_{a_{N+1}} + x) \prod_a (1 - F_a(t_a + x)),$$

while for $0 \leq x \leq t_f - \tau_N$,

$$\hat{f}_-(x) \sim f_{a_N}(t_{a_N} + x) \prod_a (1 - F_a(t_f - \tau_N - x)).$$

In the first expression, the product runs over non-idle nodes other than a_{N+1}; in the second, it runs over those nodes (if any) that initiate a service time at τ_N.

The following is similar to Theorem 3 of [7]:

Theorem 2. Under the conditions for Theorem 1,

$$\frac{d\mathbf{E}[R(\theta)]}{d\theta} = \mathbf{E}\Big[Q(a_+; \xi_+)[f(Z_+) - f(Z_{t_f})] + Q(a_-; \xi_-)[f(Z_-) - f(Z_{t_f})]\Big].$$

5. REMARKS ON SPA

(5.1). The SPA estimator calculated above for the derivative of expected terminal reward in queueing networks extends to a class of generalized semi-Markov processes (GSMPs). If a_n denotes the nth event to occur, and Y_n is the nth state visited by the GSMP, the main condition we require is that Y_n not depend on the order of $a_1, ..., a_n$. This also happens to be an important condition in applying IPA to cumulative functionals (like (2)) of GSMPs.

(5.2). The function Φ_i required by (A1) never appears in the SPA estimator and need not be known explicitly. All that is needed is D_i — the derivative of Φ_i at $h = 0$. As noted above, a candidate D_i is provided by (11).

(5.3). Evaluation of the densities $f(0)$ that enter into the jump rates $Q(\cdot;\cdot)$ can be difficult. (A notable exception is the case where all X_i are exponential.) In general, conditioning on more information makes calculation of Q easier; but conditioning on too much may fail to smooth jumps. Hence, SPA tentatively prescribes

Condition on the largest $\mathcal{F}$ that makes $\mathbf{E}[\gamma(\theta + h) - \gamma(\theta)|\mathcal{F}]$ continuous in h.

On the other hand, in considering variance, the principle of conditional Monte Carlo (Fox and Glynn [3], Proposition 1) applies: If $\mathcal{F} \subseteq \mathcal{G}$, then for all h,

$$\mathbf{E}\left[\mathbf{E}[\frac{\gamma(\theta + h) - \gamma(\theta)}{h}|\mathcal{F}]^2\right] \le \mathbf{E}\left[\mathbf{E}[\frac{\gamma(\theta + h) - \gamma(\theta)}{h}|\mathcal{G}]^2\right].$$

Modulo an additional interchange of limit and expectation, this implies that an unbiased SPA estimator based on $\mathcal{F}$ has smaller variance than one based on $\mathcal{G}$. Hence, variance reduction suggests

Condition on the smallest $\mathcal{F}$ that makes $\mathbf{E}[\gamma(\theta)|\mathcal{F}]'$ tractable.

In practice, the first of these considerations seems to be the more critical one.

6. TERMINAL REWARD VIA DYNKIN'S FORMULA

We now describe a different method for estimating the sensitivity of expected terminal reward. Let Z_t be a Markov process with infinitesimal generator $\mathbf{Q}$. Dynkin's formula states that if T is a stopping time (for Z_t) with finite expectation,

$$\mathbf{E}[f(Z_T)] = \mathbf{E}[f(Z_0)] + \mathbf{E}\left[\int_0^T \mathbf{Q}f(Z_t)dt\right]. \qquad (24)$$

(See, for example, Karlin and Taylor [10], p.298.) Since $\mathbf{Q}$ can be defined by

$$\mathbf{Q}f(x) = \lim_{h \downarrow 0} \frac{\mathbf{E}[f(Z_h)|Z_0 = x] - f(x)}{h},$$

(24) is reminiscent of the fundamental theorem of calculus. Taking $T = \tau_n$ and $T = t_f$ covers the two cases considered in Section 4.

Suppose, now, that $\mathbf{Q}$ (hence, Z_t) depends on a parameter θ. If Z_0 is fixed, then $\partial_\theta \mathbf{E}[f(Z_0)] = 0$. Thus, in this case, Dynkin's formula transforms a terminal reward problem to a cumulative reward problem with which we can try to use IPA. The problem we end up with does not quite have the

form (2). Instead, we must consider

$$\int_0^{t_f} g(\theta, Z_t(\theta))dt, \qquad (25)$$

where

$$g(\theta, x) = \mathbf{Q}_\theta f(x) = \sum_y \mathbf{q}_\theta(x, y)f(y)$$

depends *explicitly* on θ because $\mathbf{Q}_\theta$ does.

Consider the case of (right-continuous, non-explosive) $Z_t(\theta)$ on a discrete state space. Let $Y_i(\theta)$ be the ith state visited by $Z_t(\theta)$; let $\tau_i(\theta)$ be the epoch of the ith transition of $Z_t(\theta)$; and let $N = N_{t_f}(\theta)$ be the number of transitions in $[0, t_f]$. We may rewrite (25) as

$$\sum_{i=0}^{N-1} g(\theta, Y_i(\theta))[\tau_{i+1} - \tau_i] + (T - \tau_N)g(\theta, Y_N(\theta)).$$

Thus, differentiating (25), we get

$$\int_0^{t_f} \partial_\theta g(\theta, Z_t(\theta))dt + \sum_{i=1}^{N} \frac{d\tau_i}{d\theta}[g(\theta, Y_{i-1}) - g(\theta, Y_i)]. \qquad (26)$$

(In this expression, the integrand is the partial derivative of g with respect to its first argument only; that is, with respect to the explicit dependence on θ.) Sufficient conditions for the IPA estimate (26) to be unbiased are given in [6]. Other conditions on Z_t — based on viewing it as a GSMP with exponential clock times — can be based on the conditions in [5]; cf. Remark 5.1.

This approach to sensitivity analysis of expected terminal reward can be applied even when Z_t is not naturally Markovian; or, to put it another way, when supplementary variables must be appended to make Z_t a Markov process. Burman [1], p.855, exhibits the infinitesimal generator for GSMPs. In [1], reward functions f that depend on clock ages are allowed; but here we restrict attention to functions that depend only on the discrete part of the state. Using s, s' to denote states, α to denote events, $p(s'; s, \alpha)$ for transition probabilities, t_α to denote the age of the α-clock and $\mu_\alpha(t_\alpha)$ for the hazard rate of the α-clock, we get

$$\mathbf{Q}f(s, \{t_\alpha\}) = \sum_\alpha \sum_{s'} p(s'; s, \alpha)\mu_\alpha(t_\alpha)[f(s') - f(s)].$$

(Burman [1] allows speeds.) If only the clock distributions depend on θ, this yields the IPA estimate

$$\int_0^{t_f} \partial_\theta \mathbf{Q}_\theta(Z_t)dt + \sum_{i=1}^{N} \frac{d\tau_i}{d\theta}[\mathbf{Q}_\theta f(Z_{\tau_i^+}) - \mathbf{Q}_\theta f(Z_{\tau_i^-})],$$

where

$$\partial_\theta \mathbf{Q}_\theta f(s, \{t_\alpha\}) =$$

$$\sum_{\alpha, s'} \left(\frac{\partial \mu_\alpha(t_\alpha)}{\partial \theta} + \frac{\partial \mu_\alpha(t_\alpha)}{\partial t_\alpha} \frac{dt_\alpha}{d\theta} \right) p(s'; s, \alpha)[f(s') - f(s)].$$

7. FURTHER REMARKS

(7.1). It should be stressed that, perhaps surprisingly, the domains of SPA estimators for terminal reward, and estimators based on Dynkin's formula plus IPA are roughly the same. The main condition that makes IPA work (see Remark 5.1) also makes SPA feasible. The significance of this condition is that it "localizes" the effect of event order changes, which is important for both methods. The relative efficiency of the two methods is problem-dependent.

(7.2) In SPA, conditioning is used to smooth discontinuities due to changes in θ. In Dynkin's formula, conditioning smooths the dependence on t, via the definition of $\mathbf{Q}$.

(7.3) Let A_t be a point process and λ_t an intensity for A_t. Zazanis [11] considers derivative estimators that *smooth* by replacing an integral against dA_t with one against $\lambda_t dt$, both integrals having the same expectation. Since

$$f(Z_T) - f(Z_0) = \int_0^T df(Z_t),$$

Dynkin's formula (24) can be viewed as replacing an integral against $df(Z_t)$ with one against $\mathbf{Q}f(Z_t)dt = \mathbf{E}[df(Z_t)|Z_t]$, where, again, both have the same expectation.

(7.4) There is a class of problems, which we now describe briefly, for which Dynkin's formula and SPA yield the same estimator. Consider, again, the case of a discrete state Markov process. Call θ a *scale parameter* if whenever $q_\theta(x, y) > 0$, $q'_\theta(x, y)/q_\theta(x, y) = q'_\theta(x)/q_\theta(x)$. This makes $[q_\theta(x, y)/q_\theta(x)]' = 0$, so θ is a parameter of the holding times but not the embedded transition probabilities. Suppose θ is a scale parameter and suppose also that every $q'_\theta(x) \geq 0$ (i.e., the holding times are decreasing in θ). Then the IPA estimator obtained via Dynkin's formula coincides with the SPA estimator for terminal reward based on

$$\xi = (Y_0, \ldots, Y_N; \tau_1, \ldots, \tau_N),$$

where $N \equiv N_{t_f}$ is the number of transitions in $[0, t_f]$. Though we omit the derivation, both estimators become

$$\mathbf{Q}f(Z_{t_f}) \cdot \int_0^{t_f} \frac{q'(Z_t)}{q(Z_t)} dt.$$

(7.5). Our results for expected terminal reward have implications for steady-state derivative estimation. If, say, $Z_t(\theta) \Rightarrow Z_\infty(\theta)$ and

$$\lim_{t \to \infty} \frac{d\mathbf{E}[f(Z_t(\theta))]}{d\theta} = \frac{d\mathbf{E}[f(Z_\infty(\theta))]}{d\theta},$$

then our estimators of $d\mathbf{E}[f(Z_t(\theta))]/d\theta$ are asymptotically unbiased estimators of the derivative of the steady-state mean $\mathbf{E}[f(Z_\infty(\theta))]$.

(7.6). In estimating the derivative of expected terminal reward via likelihood ratios, the fact that R is discontinuous is irrelevant. Even so, Dynkin's formula may have applications here as well. If $L_{t_f}(\theta)$ is the likelihood of $\{Z_t, 0 \leq t \leq t_f\}$ under $\mathbf{Q}_\theta$, then

$$\int_0^{t_f} \mathbf{Q}_\theta f(Z_t) dt \cdot \frac{L'_{t_f}(\theta)}{L_{t_f}(\theta)}$$

may have smaller variance than

$$f(Z_{t_f}) \cdot \frac{L'_{t_f}(\theta)}{L_{t_f}(\theta)}.$$

Fox [2] reports the same order-of-magnitude computational costs for estimating cumulative reward and its derivative as for terminal reward and its derivative, so the transformation is worth considering.

REFERENCES

[1] Burman, D.Y., "Insensitivity in queueing systems", *Advances in Applied Probability*, **13**, pp.846-859, 1981.

[2] Fox, B.L., "Complexity of gradient estimation for transient Markov chains", Technical Report, University of Montreal, 1987.

[3] Fox, B.L. and Glynn, P.W., "Discrete time conversion for simulating finite-horizon Markov processes", Technical Report, University of Colorado, Denver, 1989.

[4] Glasserman, P., "Performance continuity and differentiability in Monte Carlo optimization", *Proceedings of the Winter Simulation Conference*, M. Abrams, P. Haigh and J. Comfort, (eds.), pp.518-524, 1988.

[5] Glasserman, P., "Structural condtions for perturbation analysis derivative estimation: finite-time performance indices", submitted for publication, 1988.

[6] Glasserman, P., "Derivative estimates from simulation of continuous-time Markov chains", submitted for publication, 1989.

[7] Glasserman, P. and Gong, W.B., "Smoothed perturbation analysis for a class of discrete event systems", submitted for publication, 1989.

[8] Gong, W.B. and Ho, Y.C., "Smoothed perturbation analysis of discrete event dynamical systems", *IEEE Transactions on Automatic Control*, **32**, pp.858-866, 1987.

[9] Heidelberger, P. and Goyal, A., "Sensitivity analysis of continuous time Markov chains using uniformization", in *2nd International Workshop on Applied Mathematics and Performance/Reliability Models of Computer/Communication Systems*, University of Rome II, pp.93-104, 1987.

[10] Karlin, S. and Taylor, H.M., *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.

[11] Zazanis, M., "Compensators and Derivative Estimation for Queueing Systems", *Proceedings of the 26th Allerton Conference*, pp.549-555, 1988.

PAUL GLASSERMAN is Member of Technical Staff in the Operations Research Department at AT&T Bell Laboratories. He received a Ph.D. from the Division of Applied Science, Harvard University in 1988, and an A.B. in Mathematics from Princeton University in 1984. His interests include derivative estimation, Monte Carlo optimization and queueing theory.

Room 3K-324
AT&T Bell Laboratories
Crawfords Corner Rd.
Holmdel, NJ 07733

WEI-BO GONG received the M.S. degree in engineering from the China University of Sciences and Technology in 1981. He worked as an Instructor at the same university for two years. In 1987 he received the Ph.D. degree from Harvard University. He is on the faculty of the Department of Electrical and Computer Engineering, University of Massachusetts at Amherst. His research interests include analysis and optimization of discrete event systems, such as computer-communication networks and flexible manufacturing systems.

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