Les cahiers du GERAD ISSN: 0711–2440

On the Convergence Rates of IPA and FDC Derivative Estimators for Finite-Horizon Stochastic Simulations

P. L'Ecuyer, G. Perron

G–90–61

December 1990

On the Convergence Rates of IPA and FDC Derivative Estimators for Finite-Horizon Stochastic Simulations

Pierre L'Ecuyer
Département d'IRO, Université de Montréal
C.P. 6128, Succ. A, Montréal, H3C 3J7, Canada

Gaétan Perron
Département d'informatique, Université Laval
Ste-Foy, Québec, G1K 7P4, Canada

November 1990

Abstract

We show that in most interesting cases where infinitesimal perturbation analysis (IPA) applies for derivative estimation, a finite-difference scheme with common random numbers (FDC) has the same order of convergence, namely $O(n^{-1/2})$, provided that the size of the finite-difference interval converges to zero fast enough. This holds for both one-sided and two-sided FDC. This also holds for different variants of IPA, like for smoothed perturbation analysis (SPA), which is based on conditional expectation. We give some examples and numerical illustrations. Besides their theoretical interest, these results might have practical implications for situations where implementing IPA is more difficult than simply doing finite differences.

Keywords:

Simulation; common random numbers; derivative estimation; perturbation analysis; finite differences

Résumé

Nous montrons que dans presque tous les cas intéressants où l'analyse de perturbation infinitésimale s'applique pour estimer une dérivée, une approche basée sur les différentes finies avec des valeurs aléatoires communes (FDC) possède le même ordre de convergence, soit $O(n^{-1/2})$, pourvu que la taille des intervalles pour les différences finies converge vers zéro suffisamment rapidement. Cela tient autant pour les versions asymétriques que symétriques de FDC. Cela tient aussi pour différentes variantes de l'analyse de perturbation, comme par exemple pour l'analyse de perturbation lissée, basée sur l'espérance conditionnelle. Nous donnons des exemples et des illustrations numériques. Au delà de leur intérêt théorique, ces résultats peuvent avoir des implications pratiques pour les situations où l'implantation de l'analyse de perturbation est significativement plus difficile que l'utilisation de différences finies.

1. Introduction

Infinitesimal perturbation analysis (IPA) is a technique for estimating derivatives by simulation, in the context of discrete-event stochastic dynamic systems. It has received a lot of attention lately (Heidelberger et al. 1988, Ho 1987, Suri 1987, 1989). The main reason why IPA is attractive is that when it applies, it gives an estimator of the derivative of a mathematical expectation with respect to *many* (continuous) parameters with a *single* simulation run. Finite-difference (FD) schemes, in contrast, require *many* simulation runs to do a similar job. It is also widely believed that IPA has a better order of convergence than FD schemes in general. Convergence rates for IPA, FD, and FD with common random numbers (FDC) have been derived, under various sets of assumptions, in Glynn (1989) and Zazanis and Suri (1988), among others.

The aim of this paper is to show that in most interesting cases where IPA applies (i.e. gives an unbiased derivative estimate with finite variance), the finite difference interval in FDC can be taken very small without getting into the problem of a large variance. The variance remains bounded and the bias goes to zero when the size of that interval goes to zero. As a result, if a sample of size n is taken and if the FDC interval size goes to zero fast enough with n (say, as fast as $O(n^{-1/2})$), then the mean square error of the sample mean for FDC is in the order of $1/n$, the same as for IPA. On the other hand, to estimate the derivative with respect to d parameters (a d-dimensional gradient), the computational work required by FDC still increases linearly with d, since at least $d+1$ simulation runs are required for each of the n replications. For IPA, only one simulation run is performed, but the work might also increase linearly with d in some situations because of the number of IPA counters that have to be updated.

In Section 2, we set the mathematical framework and introduce the IPA and FDC derivative estimators. We also show how LR estimators can fit into that framework. In Section 3, we recall (sufficient) unbiasedness and finite variance conditions for IPA. In practice, in most interesting cases where IPA applies, these conditions are satisfied (Glasserman 1988, L'Ecuyer 1990). We show that under these conditions, the mean square error for FDC is in $O(1/n)$. Section 4 gives more precise characterizations of the convergence rates. Numerical examples are given in Section 5. These examples also illustrate how our results apply to smoothed perturbation analysis (SPA) estimators, for which the objective function is replaced by a conditional expectation. [We use the notation $O(f(n))$ to denote the *set* of functions g such that for some constant $c > 0$, $g(n) \leq cf(n)$ for all n.]

2. IPA and FDC derivative estimation

Consider a stochastic system defined over a probability space $(\Omega, \Sigma, P_\theta)$, parametrized by $\theta \in \Theta$, where $\Theta = (a, b)$ is some open interval in $\mathbb{R}$. Let the random variable $h(\theta, \omega)$ represent the "cost' of running the system at parameter level θ for sample point $\omega \in \Omega$. The expected cost at parameter level θ is then

$$\alpha(\theta) = \int_\Omega h(\theta, \omega) dP_\theta(\omega). \tag{1}$$

We want to estimate the derivative $\alpha'(\theta) = d\alpha(\theta)/d\theta$, which is assumed to exist. (Note that for the case where θ is a vector, gradient estimation can be done by considering each component of the gradient individually.)

In typical stochastic simulations, the sample point ω (which represents the "randomness" that drives the system) can be viewed as a sequence of independent uniform variates between 0 and 1. $P_\theta \equiv P$ is then independent of θ and under some additional conditions (see next section), the random variable $h'(\theta, \omega) = dh(\theta, \omega)/d\theta$ can be used as an unbiased estimator of $\alpha'(\theta)$. This is the IPA derivative estimator.

One can get n i.i.d. replicates of $h'(\theta, \omega)$ and use the estimator

$$Y_n^{\mathrm{P}} = \frac{1}{n} \sum_{i=1}^{n} h'(\theta, \omega_i). \tag{2}$$

Here, $\omega_1, \omega_2, \ldots, \omega_n$ are independent sample points generated under the common distribution P.

In practice, for complex simulations, $h'(\theta, \omega)$ is not always easy to compute. There are also many cases where it gives a biased or even totally meaningless derivative estimator. One could then return to more classical finite-difference schemes. (Note that there are other possibilities, e.g. the LR technique. See Glynn 1989, L'Ecuyer 1990, and the references cited there). It is well known that finite-difference schemes with independent streams of random numbers give rise to large variability (Glynn 1989, Meketon 1987, Zazanis and Suri 1988). The convergence rates for forward differences and central differences are $O(n^{-1/4})$ and $O(n^{-1/3})$ respectively. With common random numbers and under a given set of assumptions, Glynn (1989) obtains respective convergence rates of $O(n^{-1/3})$ and $O(n^{-2/5})$. But it turns out that when IPA applies, these assumptions do not hold.

Finite-difference estimators with common random numbers (FDC) can be defined as follows. Let n be the number of replications and $c_n > 0$. Let $\omega_1, \ldots, \omega_n$ be independent sample points generated under P. The *forward* FDC estimator is

$$Y_n^{\mathrm{F}} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(\theta + c_n, \omega_i) - h(\theta, \omega_i)}{c_n}, \tag{3}$$

while the *central* FDC estimator is

$$Y_n^{\mathrm{C}} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(\theta + c_n, \omega_i) - h(\theta - c_n, \omega_i)}{2c_n}. \tag{4}$$

In practical simulation experiments, to compute each term of the sum in (3), one usually performs two different simulation runs, with common random numbers, to obtain $h(\theta,\omega_i)$ and $h(\theta+c_n,\omega_i)$ (and similarly for $h(\theta-c_n,\omega_i)$ and $h(\theta+c_n,\omega_i)$ in (4)). When c_n is very small, two problems might occur. The first one is that the variance of Y_n^{F} or Y_n^{C} could blow out, as is the case with FD *without* common random numbers. In the next section, we show that this is not necessarily so. A second problem has to do with numerical precision. Even when the variance is zero, (3) and (4) become numerically unstable when c_n gets too small. We do not address this (well known) question in this paper.

We could also mention another derivative estimation technique, based on likelihood ratios (LR), which goes as follows (Glynn 1989, L'Ecuyer 1990). Suppose now that P_θ really depends on θ. Rewrite

$$\alpha(\theta) = \int_\Omega H(\theta,\omega)dG(\omega) \tag{5}$$

where $H(\theta,\omega) = h(\theta,\omega)(dP_\theta/dG)(\omega)$, G is a probability measure independent of θ that dominates the P_θ's, and dP_θ/dG denotes the Radon-Nikodym derivative. Now, generate ω according to G and compute $H'(\theta,\omega) = dH(\theta,\omega)/d\theta$ as the LR derivative estimate. This is in fact equivalent to applying IPA *after* replacing h by H and P_θ by G. In other words, LR is equivalent to applying IPA on top of an importance sampling scheme. (Here, ω does not necessarily represent a sequence of $U(0,1)$ variates, but it can be expressed as a function of some $\tilde{\omega}$ which is itself viewed as a sequence of i.i.d. $U(0,1)$ variates, and the mapping from $\tilde{\omega}$ to ω does not depend on θ.) Therefore, our results also apply to LR estimators. In that context, finite-differences are defined using H instead of h.

3. Unbiasedness conditions and convergence rates

We will use the mean square error (MSE) as an efficiency criterion for our estimators. In other words, the *loss function* associated with a derivative estimator Y_n based on n runs is

$$R_n = E[Y_n - \alpha'(\theta_0)]^2 = \text{ Var } (Y_n) + B_n^2 \tag{6}$$

where $B_n = E[Y_n] - \alpha'(\theta_0)$ denotes the bias of Y_n. We say that the *convergence rate* is in $O(f(n))$ if $R_n \in O(f(n)^2)$.

Sufficient conditions for unbiasedness of the IPA estimator are given in Glasserman (1988) and L'Ecuyer (1990). We recall these conditions and extend them to conditions under which convergence rates can be obtained. Suppose that we want to estimate the derivative at $\theta = \theta_0$. Assumption A1 below is parametrized by $\nu > 0$. For example, A1(2) denotes that assumption with $\nu = 2$. A1(2) is not a *necessary* condition for IPA to apply, but according to our experience, for all interesting cases for which IPA applies in practice, A1(2) is satisfied.

ASSUMPTION A1(ν). There is an open neighborhood Υ of θ_0 and a measurable set $\Xi \subseteq \Omega$, such that $P(\Xi) = 1$, and for all $\omega \in \Xi$, $h(\cdot, \omega)$ exists and is continuous everywhere in Υ, and is also differentiable in $D(\omega) \subseteq \Upsilon$, where $\Upsilon \setminus D(\omega)$ is at most a denumerable set. Define

$$\Psi(\omega) = \begin{cases} \sup_{\theta \in D(\omega)} |h'(\theta, \omega)| & \text{if } \omega \in \Xi; \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Assume that there exists a function $\Gamma : \Omega \to \mathbb{R}$, such that $\Psi(\omega) \leq \Gamma(\omega)$ for all $\omega \in \Omega$, and that $\int_\Omega [\Gamma(\omega)]^\nu dP(\omega) = K(\nu) < \infty$. ∎

THEOREM 1. *Under* A1(2), $h'(\theta_0, \omega)$ *is an unbiased estimator of* $\alpha'(\theta_0)$, *with finite variance* $\sigma_{\text{P}}^2 \leq K(2)$. *As a consequence, the mean square error of* Y_n^{P} *is* $R_n^{\text{P}} = \text{ Var } (Y_n^{\text{P}}) = \sigma_{\text{P}}^2/n$ *and the convergence rate is in* $O(n^{-1/2})$.

PROOF. A1(2) implies A1(1) and the unbiasedness then follows from Theorem 1 in L'Ecuyer (1990). Also, Var $[h'(\theta_0, \omega)] \leq E[(h'(\theta_0, \omega))^2] \leq E[\Gamma^2(\omega)] = K(2) < \infty$. ∎

We now show that similar properties hold for FDC. Here, $\sigma_{\text{F}}^2(c_n)$ represents the variance of each term of the sum in (3), R_n^{F} is the mean square error of Y_n^{F}, and similarly for central differences.

THEOREM 2. *Assume* A1(2) *and suppose that* α *is twice continuously differentiable in* Υ. *If* c_n *is small enough such that* $[\theta_0, \theta_0 + c_n] \subset \Upsilon$, *then* $\sigma_{\text{F}}^2(c_n) \leq K(2)$ *and*

$$R_n^{\text{F}} = \sigma_{\text{F}}^2(c_n)/n + [\alpha''(\xi^+)c_n/2]^2 \tag{8}$$

where $\theta_0 \leq \xi^+ \leq \theta_0 + c_n$. *Similarly, if* $[\theta_0 - c_n, \theta_0 + c_n] \subset \Upsilon$, *then* $\sigma_{\text{C}}^2(c_n) \leq K(2)$ *and*

$$R_n^{\text{C}} = \sigma_{\text{C}}^2(c_n)/n + [(\alpha''(\xi^+) - \alpha''(\xi^-))c_n/4]^2 \tag{9}$$

where $\theta_0 - c_n \leq \xi^- \leq \theta_0 \leq \xi^+ \leq \theta_0 + c_n$. *In (8) and (9), the first term is the variance and the second one is the bias. In both cases, as a consequence, the convergence rate is in* $O(n^{-1/2})$ *provided that* c_n *is in* $O(n^{-1/2})$. *Further, if* α *is three times continuously differentiable in a neighborhood of* θ_0, *then the (squared) bias term in (9) can be replaced by* $((\alpha'''(\xi^+) + \alpha'''(\xi^-))c_n^2/12)^2$. *In that case, for the convergence rate to be in* $O(n^{-1/2})$, *it suffices that* $c_n \in O(n^{-1/4})$.

PROOF. A Taylor series development yields

$$E[Y_n^{\mathrm{F}}] = [\alpha(\theta_0 + c_n) - \alpha(\theta_0)]/c_n = \alpha'(\theta_0) + \alpha''(\xi^+)c_n/2$$

for $\theta_0 \leq \xi^+ \leq \theta_0 + c_n$. When c_n is small enough, $\theta_0 + c_n \in \Upsilon$ and from Theorem 8.5.3 in Dieudonné (1969),

$$\left|\frac{h(\theta_0 + c_n, \omega) - h(\theta_0, \omega)}{c_n}\right| \leq \sup_{\theta \in [\theta_0, \theta_0 + c_n] \cap D(\omega)} |h'(\theta, \omega)| \leq \Gamma(\omega).$$

Then, $\sigma_{\mathrm{F}}^2(c_n) \leq K(2)$ and

$$R_n^{\mathrm{F}} = \sigma_{\mathrm{F}}^2(c_n)/n + (E[Y_n^{\mathrm{F}}] - \alpha'(\theta_0))^2 = \sigma_{\mathrm{F}}^2(c_n)/n + (\alpha''(\xi^+)c_n/2)^2.$$

For the central differences, the Taylor series development yields

$$E[Y_n^{\mathrm{C}}] = [\alpha(\theta_0 + c_n) - \alpha(\theta_0 - c_n)]/(2c_n) = \alpha'(\theta_0) + (\alpha''(\xi^+) - \alpha''(\xi^-))c_n/4$$

for $\theta_0 - c_n \leq \xi^- \leq \theta_0 \leq \xi^+ \leq \theta_0 + c_n$. The proof of (9) is then the same as for the forward case. When the third derivative exists, the new bias term can be obtained easily after adding one term to the Taylor series expansion. ∎

4. Convergence in distribution

We now derive stronger convergence results, under additional assumptions. The symbol $\Rightarrow$ denotes convergence in distribution and $N(0,1)$ denotes the standard normal distribution. As seen in the previous section, A1(2) implies that the variance of $(h(\theta_0+c_n,\omega_i)-h(\theta_0,\omega_i))/c_n$, namely $\sigma_{\mathrm{F}}^2(c_n)$, is bounded as $c_n \to 0$ (and similarly for centered differences). For the next result, we also assume that it converges to a constant.

THEOREM 3. *Suppose that* A1$(2+\delta)$ *holds for some* $\delta > 0$, *that* α *is twice continuously differentiable in* Υ, *and that* $\lim_{n\to\infty} \sigma_{\mathrm{F}}^2(c_n) = \sigma_{\mathrm{F}}^2$, *where* $0 < \sigma_{\mathrm{F}} < \infty$. *Then,* $\sigma_{\mathrm{F}}^2 \leq K(2) < \infty$ *and, as* $n \to \infty$:

a) If $n^{1/2}c_n \to \infty$, *then* $n^{1/2}|Y_n^{\mathrm{F}} - \alpha'(\theta_0)| \Rightarrow \infty$;

b) If $n^{1/2}c_n \to 0$, *then* $n^{1/2}(Y_n^{\mathrm{F}} - \alpha'(\theta_0)) \Rightarrow \sigma_{\mathrm{F}}N(0,1)$;

c) If $\alpha''(\theta)$ *is continuous at* $\theta = \theta_0$ *and* $n^{1/2}c_n \to c > 0$, *then* $n^{1/2}(Y_n^{\mathrm{F}} - \alpha'(\theta_0)) \Rightarrow \sigma_{\mathrm{F}}N(0,1) + \alpha''(\theta_0)c/2$.

For centered FDC, with σ_{F}^2 *replaced by* σ_{C}^2, *(b) is also true. Further, if* α *is also three times continuously differentiable in* Υ, *then, as* $n \to \infty$:

d) If $n^{1/4}c_n \to \infty$, *then* $n^{1/4}|Y_n^{\mathrm{F}} - \alpha'(\theta_0)| \Rightarrow \infty$;

e) If $n^{1/4}c_n \to 0$, *then* $n^{1/4}(Y_n^{\mathrm{F}} - \alpha'(\theta_0)) \Rightarrow \sigma_{\mathrm{F}}N(0,1)$;

f) If $\alpha'''(\theta)$ *is continuous at* $\theta = \theta_0$ *and* $n^{1/4}c_n \to c > 0$, *then* $n^{1/4}(Y_n^{\mathrm{F}} - \alpha'(\theta_0)) \Rightarrow \sigma_{\mathrm{F}}N(0,1) + \alpha'''(\theta_0)c^2/6$.

PROOF. Let $\Delta h_{n,i}^{\mathrm{F}} = [h(\theta_0+c_n,\omega_i) - h(\theta_0,\omega_i)]/c_n$ and $\Delta_{n,i} = \Delta h_{n,i}^{\mathrm{F}} - E[\Delta h_{n,i}^{\mathrm{F}}]$. Then

$$\begin{aligned} n^{1/2}(Y_n^{\mathrm{F}} - \alpha'(\theta_0)) &= n^{-1/2}\sum_{i=1}^{n}\Delta_{n,i} + n^{1/2}(E[\Delta h_{n,i}^{\mathrm{F}}] - \alpha'(\theta_0)) \\ &= n^{-1/2}\sum_{i=1}^{n}\Delta_{n,i} + n^{1/2}\alpha''(\xi_n^+)c_n/2 \end{aligned}$$

where ξ_n^+ is as in Section 3. Since $\theta_0 \leq \xi_n^+ \leq \theta_0 + c_n$, $c_n \to 0$, and α'' is continuous, one has $\alpha''(\xi_n^+) \to \alpha''(\theta_0)$. It remains to show that

$$n^{-1/2}\sum_{i=1}^{n}\Delta_{n,i} \Rightarrow \sigma_{\mathrm{F}}N(0,1) \tag{10}$$

as $n \to \infty$ and the results a, b, and c will follow. For that, as in Glynn (1989), we use the Lindeberg-Feller Theorem, which is a central-limit theorem for triangular arrays (see Chung 1974, p.205). We just need to verify Lindeberg's condition and (10) will follow from that

theorem. In our case, the triangular array is $\{n^{-1/2}\Delta_{n,i}/\sigma_{\rm F}(c_n),\ 1 \le i \le n \text{ and } n \ge 1\}$ and Lindeberg's condition is that for each $C > 0$,

$$\lim_{n\to\infty} E[(\Delta_{n,i}/\sigma_{\rm F}(c_n))^2 \mid (\Delta_{n,i}/\sigma_{\rm F}(c_n))^2 > nC] = 0. \tag{11}$$

We have

$$\begin{aligned} E[(\Delta_{n,i}/\sigma_{\rm F}(c_n))^2 \ &| \ (\Delta_{n,i}/\sigma_{\rm F}(c_n))^2 > nC] \\ &\le E[(|\Delta_{n,i}|/\sigma_{\rm F}(c_n))^{2+\delta}/(nC)^{\delta/2} \mid (\Delta_{n,i}/\sigma_{\rm F}(c_n))^2 > nC] \\ &\le E[|\Delta_{n,i}|^{2+\delta}]/[\sigma_{\rm F}(c_n)^{2+\delta}(nC)^{\delta/2}] \\ &\le E[\Gamma^{2+\delta}(\omega)]/[\sigma_{\rm F}(c_n)^{2+\delta}(nC)^{\delta/2}]. \end{aligned}$$

This goes to zero as $n \to \infty$. The proof is similar for the central-difference case. ∎

There exists degenerate cases where A1$(2+\delta)$ holds but where $\sigma_{\rm F}^2(c_n) = 0$. Then, both IPA and FDC have zero variance. This is the case, for example, if $h(\theta,\omega)$ can be decomposed as $h(\theta,\omega) = f(\theta) + g(\omega)$. Another pathological case is when $h(\theta,\omega)$ can be decomposed as $h(\theta,\omega) = f(\theta)g(\omega)$. The variance for IPA is then var $[Y_n^{\rm P}] = n^{-1}(f'(\theta))^2$ var $[g(\omega)]$, which is zero at points θ_0 for which $f'(\theta_0) = 0$. For FDC, the variance then goes to zero as $c_n \to 0$ and $\theta \to \theta_0$ (or $\theta = \theta_0$).

5. Numerical examples

5.1. Average system time for the first t customers in an GI/G/1 queue

Consider a GI/G/1 queue, initially empty, with interarrival and service-time distributions F and G_θ respectively, both with finite expectations and variances. The latter has a density g_θ and depends on a parameter $\theta \in \Theta$. For $1 \leq j \leq t$, A_j denotes the interarrival time between customers $j-1$ and j (A_1 is the arrival time of customer 1 and the system starts at time 0), S_j is the service time for customer j, and X_j is the system time (waiting time in the queue plus service time) for customer j. We have $X_0 = 0$, and $X_{j+1} = S_{j+1} + \max(0,\ X_j - A_{j+1})$ for $j > 0$. Let $h(\theta,\omega)$ be the observed average system time for the first t customers:

$$h(\theta,\omega) = \frac{1}{t}\sum_{j=1}^{t} X_j. \tag{12}$$

Here, $\alpha(\theta)$ is the expected mean system time for the first t customers in the system,

Suppose that each service time S_j is generated by inversion: $S_j = G_\theta^{-1}(U_j)$, where the U_j's are i.i.d. $U(0,1)$ variates. Suppose also that for each $u \in (0,1)$, $\phi(\theta,u) = \partial G_\theta^{-1}(u)/\partial\theta$ exists and that $\sup_{\theta\in\Theta} |\phi(\theta,u)| \leq \Gamma(u)$, where $\Gamma : (0,1) \mapsto \mathbb{R}$ is a measurable function such that $\int_0^1 (\Gamma(u))^3 du < \infty$. The IPA estimator in this case is

$$h'(\theta,\omega) = \frac{1}{t}\sum_{j=1}^{t} X'_j, \tag{13}$$

where $X'_0 = 0$ and $X'_{j+1} = S'_{j+1} + X'_j$ if $X_j > A_{j+1}$, $X'_{j+1} = S'_{j+1}$ otherwise. It is easily seen that for each θ, $h'(\theta,\omega) \leq \sum_{j=1}^{t} \phi(\theta,U_j)$, and that A1(3) holds (see also L'Ecuyer 1990, and L'Ecuyer, Giroux, and Glynn 1990). Therefore, Theorems 1—3 apply.

For a numerical illustration, suppose that G_θ and F are the exponential distributions with means θ and 1, respectively. Let $t = 10$ and $\theta_0 = 0.5$. The true value of the derivative (computed by recursive equations as in L'Ecuyer 1990) is $\alpha'(0.5) = 2.46339$. We performed numerical experiments with IPA and the forward and central versions of FDC, with different subsequences c_n. In each case, we computed the (sample) mean square error R_n for different values of n. The results appear in Table 1. They are in accordance with Theorems 1—3. Common random numbers were used across the lines of the Table. Note in particular the strong similarity between the results of the three methods when c_n is chosen appropriately. For this particular example, the forward and central versions of FDC with $c_n = n^{-1/4}$ also give similar results. One can also see that when c_n goes to zero too fast (like $c_n = n^{-3}$), numerical instability comes in.

5.2. A race between two random variables

This example is inspired from Glasserman and Gong (1990), but our analysis goes further and in different directions. Consider a race between two random variables X_1 and X_2. X_2

	c_n	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$	$n = 10^6$
IPA	—	3.09E-2	3.53E-3	3.54E-4	3.53E-5	3.55E-6
FDC(F)	$n^{-1/4}$	4.60E-2	4.07E-3	4.00E-4	3.77E-5	3.69E-6
"	$n^{-1/2}$	3.58E-2	3.61E-3	3.58E-4	3.54E-5	3.56E-6
"	n^{-1}	3.11E-2	3.53E-3	3.54E-4	3.53E-5	3.55E-6
"	n^{-2}	3.09E-2	3.53E-3	3.54E-4	3.53E-5	3.55E-6
"	n^{-3}	3.09E-2	3.53E-3	3.54E-4	3.44E-5	6.18E-3
FDC(C)	$n^{-1/4}$	2.62E-2	3.05E-3	3.37E-4	3.42E-5	3.50E-6
"	$n^{-1/2}$	3.08E-2	3.41E-3	3.53E-4	3.52E-5	3.55E-6
"	n^{-1}	3.10E-2	3.53E-3	3.54E-4	3.53E-5	3.55E-6
"	n^{-2}	3.09E-2	3.53E-3	3.54E-4	3.53E-5	3.55E-6
"	n^{-3}	3.09E-2	3.53E-3	3.54E-4	3.67E-5	8.13E-3

Table 1: Sample mean square errors for the M/M/1 queue example.

has distribution F, while X_1 has distribution G_θ, which depends on a parameter $\theta \in \Theta$, where Θ is a finite interval. Suppose our reward $h(\theta,\omega)$ is 1 is $X_1 < X_2$ (X_1 wins the race), 0 otherwise. Then, $\alpha(\theta) = P_\theta[X_1 < X_2]$, where P_θ denotes the probability when X_1 has distribution G_θ. As usual, we want to estimate $\alpha'(\theta)$. Of course, the exact value can be computed directly in most cases. But this simple example is a good illustration of what happens in many practical situations where the output is discontinuous with respect to the parameter value. More complicated cases can be treated in a similar way; see, e.g., Glasserman and Gong (1990).

To simplify the analysis, let $G_\theta(0) = F(0) = 0$. Suppose that for each $u \in (0,1)$, $G_\theta^{-1}(u)$ exists, and is non-decreasing and differentiable in θ for $\theta \in \Theta$. Suppose also that $G_\theta(x)$ is twice continuously differentiable with respect to θ, that $dG_\theta(x)/d\theta$ is bounded uniformly: $0 \leq -dG_\theta(x)/d\theta \leq C$ for all $\theta \in \Theta$ and $x \geq 0$, for some constant C, and that F has a density f. Let $X_1(\theta) = G_\theta^{-1}(U_1)$ and $X_2 = F^{-1}(U_2)$, where U_1 and U_2 are i.i.d. $U(0,1)$.

Trying to apply IPA naïvely will not work in this case. If we view ω as representing (U_1, U_2), then $h'(\theta,\omega) = 0$ except when $X_1(\theta) = X_2$, in which case it is undefined. Therefore, $E[Y_n^{\mathrm{P}}] = 0$ and this IPA estimator is clearly biased.

Assume $[\theta + c, \theta - c] \subset \Theta$. One term of the centered FDC estimator is

$$\psi^{\mathrm{C}} = \frac{h(\theta + c,\omega) - h(\theta - c,\omega)}{2c} = \begin{cases} -1/(2c) & \text{if } X_1(\theta - c) < X_2 \leq X_1(\theta + c) \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Using Taylor's development and the uniform bound on the derivative of G_θ with respect to θ, we obtain for some $\xi \in [\theta - c, \theta + c]$,

$$P_\theta[\psi^{\mathrm{C}} = -1/(2c)] \;=\; P_\theta[G_{\theta+c}(X_2) \leq U_1 < G_{\theta-c}(X_2)]$$

$$
\begin{aligned}
&= \int_0^\infty [G_{\theta-c}(x) - G_{\theta+c}(x)]dF(x) \\
&= \int_0^\infty (-2c)\frac{d}{d\theta}G_\theta(x)\,\Big|_{\theta=\xi}\, dF(x) \\
&\leq 2cC.
\end{aligned}
$$

Then,

$$
\text{Var}\,(\psi^{\text{C}}) \leq E[(\psi^{\text{C}})^2] = \frac{P_\theta[\psi^{\text{C}} = -1/(2c)]}{(2c)^2} \leq \frac{C}{2c},
$$

which means that this example satisfies Glynn's assumption (Glynn 1989) that the variance of ψ^{C} is in $O(1/c)$. Therefore, one gets a convergence rate of $n^{-2/5}$, which is obtained with $c_n = n^{-1/3}$.

As suggested in Glasserman and Gong (1990), *smoothed* perturbation analysis can be applied in this case. Basically, this amounts to replacing the "naïve" reward function above by a conditional expectation. Equivalently, this corresponds to viewing ω as representing something different than a sequence of uniform variates. We will examine three ways of applying that to our example, i.e., three different choices for the meaning of ω. (Strictly speaking, when the sample space is "redefined" that way, we assume implicitly that there still exists a "lower-level" probability space on which the probabilities that define $h(\theta,\omega)$ below are well defined. This lower-level probability space is equivalent to the one in which ω can be interpreted as a sequence of i.i.d. $U(0,1)$ variates. We do not discuss that in detail because this is not necessary and would require introducing complicated notation that could just turn the attention away from the fundamental issues.) Let $X_1(\theta)$ and X_2 be as above, and define

$$
I(\theta) = \begin{cases} 1 & \text{if } X_1(\theta) < X_2 \\ 0 & \text{otherwise.} \end{cases}
$$

Suppose we want to estimate the derivative at $\theta = \theta_0$.

Let us redefine $\omega = U_1$ and

$$
h(\theta,\omega) = P[I(\theta) = 1 \mid \omega] = P[X_1(\theta) < X_2 \mid U_1] = \bar{F}(X_1(\theta)),
$$

where $\bar{F} \equiv 1 - F$. Individual terms of the central FDC (4) and IPA estimators are then

$$
\psi^{CS1} = \frac{F(X_1(\theta-c)) - F(X_1(\theta+c))}{2c} \tag{15}
$$

and

$$
\psi^{PS1} = -f(X_1(\theta))X_1'(\theta) \tag{16}
$$

where $X_1'(\theta) = dG_\theta^{-1}(U_1)/d\theta$.

An alternative redefinition could be $\omega = U_2$ and

$$
h(\theta,\omega) = P[I(\theta) = 1 \mid \omega] = P[X_1(\theta) < X_2 \mid U_2] = G_\theta(X_2).
$$

The terms of the central FDC (4) and IPA estimators are then

$$
\psi^{CS2} = \frac{G_{\theta+c}(X_2) - G_{\theta-c}(X_2)}{2c} \tag{17}
$$

and

$$\psi^{PS2} = dG_\theta(X_2)/d\theta. \tag{18}$$

A possible third redefinition is $\omega = (U_1, I(\theta_0))$ and

$$h(\theta,\omega) = P[I(\theta) = 1 \mid \omega] = P[X_1(\theta) < X_2 \mid U_1, I(\theta_0)].$$

Note that $I(\theta)$ is non-increasing in θ. If $I(\theta_0) = 0$ and $\theta \leq \theta_0$, then

$$h(\theta,\omega) = P[X_2 > X_1(\theta) \mid X_2 \leq X_1(\theta_0)] = 1 - \frac{F(X_1(\theta))}{F(X_1(\theta_0))}.$$

If $I(\theta_0) = 1$ and $\theta > \theta_0$, then

$$h(\theta,\omega) = P[X_2 > X_1(\theta) \mid X_2 > X_1(\theta_0)] = \frac{\bar{F}(X_1(\theta))}{\bar{F}(X_1(\theta_0))}.$$

This yields

$$h(\theta,\omega) = \begin{cases} I(\theta_0) + \left(1 - \dfrac{F(X_1(\theta))}{F(X_1(\theta_0))}\right)(1 - I(\theta_0)) & \text{if } \theta \leq \theta_0 \\ \dfrac{\bar{F}(X_1(\theta))}{\bar{F}(X_1(\theta_0))} I(\theta_0) & \text{if } \theta > \theta_0. \end{cases} \tag{19}$$

One term of the central FDC estimator (4) at $\theta = \theta_0$ is then

$$\psi^{CS3} = \frac{h(\theta_0 + c,\omega) - h(\theta_0 - c,\omega)}{2c} = \begin{cases} \dfrac{1}{2c}\left[\dfrac{F(X_1(\theta_0 - c))}{F(X_1(\theta_0))} - 1\right] & \text{if } I(\theta_0) = 0 \\ \dfrac{1}{2c}\left[\dfrac{\bar{F}(X_1(\theta_0 + c))}{\bar{F}(X_1(\theta_0))} - 1\right] & \text{if } I(\theta_0) = 1. \end{cases} \tag{20}$$

An IPA estimator is obtained by derivating (19) for a fixed ω. But here, for a fixed ω, at $\theta = \theta_0$, the left derivative is different from the right derivative. They are respectively:

$$\psi^{PS3L} = \lim_{c \to 0+} \frac{h(\theta_0,\omega) - h(\theta_0 - c,\omega)}{c} = -\frac{f(X_1(\theta_0))X_1'(\theta_0)}{F(X_1(\theta_0))}(1 - I(\theta_0)) \tag{21}$$

and

$$\psi^{PS3R} = \lim_{c \to 0+} \frac{h(\theta_0 + c,\omega) - h(\theta_0,\omega)}{c} = \frac{-f(X_1(\theta_0))X_1'(\theta_0)}{\bar{F}(X_1(\theta_0))} I(\theta_0). \tag{22}$$

Fortunately, both estimators have the same expectation as (16). This can be seen by conditioning on $X_1(\theta_0)$ as follows:

$$\begin{aligned} E[\psi^{PS3L}] &= E[E[\psi^{PS3L} \mid X_1(\theta_0)]] \\ &= E\left[E\left[-\frac{f(X_1(\theta_0))X_1'(\theta_0)}{F(X_1(\theta_0))}(1 - I(\theta_0)) \mid X_1(\theta_0)\right]\right] \\ &= E\left[-\frac{f(X_1(\theta_0))X_1'(\theta_0)}{F(X_1(\theta_0))} E[1 - I(\theta_0) \mid X_1(\theta_0)]\right] \\ &= E[-f(X_1(\theta_0))X_1'(\theta_0)], \end{aligned}$$

and similarly for ψ^{PS3R}. However, (16) should have less variance, because it is a conditional expectation of each of the other two. The estimator suggested in Glasserman and Gong (1990) is equivalent to (22). (Strictly speaking, it is obtained by viewing ω as $\omega = (I(\theta_0), U_1 I(\theta_0) + U_2(1 - I(\theta_0)))$. It remains to verify Assumption A1. In view of our diffentiability assumptions, only the ν-integrability (last assumption) remains to be checked. If it holds for some $\nu > 2$, then Theorem 3 applies. Also, if it holds for (16) for some $\nu \geq 1$, then (16) is unbiased, which implies that (21) and (22) are also unbiased, since they have the same expectation. For (16), a sufficient condition is that

$$E\left[\sup_{\theta\in\Theta}[f(X_1(\theta))X_1'(\theta)]\right]^{\nu} < \infty. \tag{23}$$

For (18), it holds for any $\nu > 0$, since $\psi^{PS2} \leq C$.

As an illustration, suppose that X_1 is exponential with mean θ and X_2 exponential with mean 1. Then $F(x) = 1 - e^{-x}$, $f(x) = e^{-x}$, $X_1(\theta) = -\theta \ln(1 - U_1)$, $X_1'(\theta) = -\ln(1 - U_1) = X_1(\theta)/\theta$. Let $\Theta = [a, b]$, where $0 < a < b$. In this case, since $f(X_1(\theta))X_1'(\theta) \leq X_1'(\theta) = X_1(1)$ and since an exponential of mean one has finite moments, (23) holds with any $\nu > 0$. We performed numerical experiments with these different estimators to estimate the derivative at $\theta_0 = 1.0$. Here, $\alpha(\theta) = 1/(\theta + 1)$ and $\alpha'(\theta) = -1/(\theta + 1)^2$. The experimental procedure was the same as for example 1. The empirical mean square errors appear in Table 2. We see that for FDC without smoothing, with $c_n = n^{-1/3}$, the empirical results agree with the convergence rate of $n^{-2/5}$, while with $c_n = n^{-1}$, things get lost in the noise. After smoothing, IPA and FDC with $c_n = n^{-1}$ have essentially the same behavior, except for the third way of conditioning, where the two IPA estimators are the limits of left and right one-sided differences, respectively.

	c_n	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$	$n = 10^6$	$n = 10^7$
FDC	$n^{-1/3}$	5.273E-3	1.277E-3	2.840E-4	5.918E-5	1.238E-5	2.680E-6
FDC	n^{-1}	6.250E-4	6.250E-5	2.500E-1	6.250E-7	6.250E-8	2.500E-1
FDC-S1	n^{-1}	1.091E-4	1.165E-5	1.174E-6	1.152E-7	1.157E-8	1.158E-9
FDC-S2	n^{-1}	1.157E-4	1.220E-5	1.157E-6	1.162E-7	1.157E-8	1.157E-9
FDC-S3	n^{-1}	4.190E-4	3.593E-5	3.839E-6	3.886E-7	3.816E-8	3.860E-9
IPA-S1	—	1.091E-4	1.165E-5	1.174E-6	1.152E-7	1.157E-8	1.158E-9
IPA-S2	—	1.157E-4	1.220E-5	1.157E-6	1.162E-7	1.157E-8	1.157E-9
IPA-S3L	—	9.093E-4	9.397E-5	9.093E-6	9.129E-7	9.146E-8	9.161E-9
IPA-S3R	—	2.187E-3	1.758E-4	1.855E-5	1.900E-6	1.857E-7	1.878E-8

Table 2: Sample mean square errors for the second example.

5.3. Average number of customers whose system time exceeds a given threshold

We come back to our first example (GI/G/1 queue), with the same assumptions, but with a different cost function. Let $\alpha_j(\theta) = P_\theta(X_j > L)$, for some fixed constant L, and $\alpha(\theta) = (1/t)\sum_{j=1}^{t} \alpha_j(\theta)$. The latter is the expected proportion of customers, among the first t, whose time spent in the system is larger than L.

To construct a gradient estimator for $\alpha(\theta)$, we will examine one customer at a time, i.e. find an estimator for $\alpha'_j(\theta)$. Suppose that ω is viewed as the sequence of all uniform variates that are used to feed the simulation, and that $h_j(\theta,\omega) = I(X_j > L)$, where I denotes the indicator function. Then, as for the previous example, $E[h'_j(\theta,\omega)] = 0 \neq \alpha_j(\theta)$. Now, suppose that only the first $j-1$ interarrival and service times are known. In that case, X_{j-1} is known, and $X_j > L$ if and only if $\max(0, X_{j-1} - A_j) > L - S_j$. This is a race between the two random variables $\max(0, X_{j-1} - A_j)$ and $L - S_j$. Therefore, perhaps we can apply the ideas developped in the previous example. Let us make the same assumptions for G_θ and F as for the previous example. Suppose also that G_θ has a density g_θ that is bounded by a constant K.

Suppose we view ω as the set of uniforms used to generate the first j service times and the first $j-1$ interarrival times, and redefine

$$h_j(\theta,\omega) = P_\theta(X_j > L \mid \omega) = \begin{cases} 1 & \text{if } S_j > L; \\ F(X_{j-1} + S_j - L) & \text{otherwise.} \end{cases} \tag{24}$$

This is an unbiased estimator of $\alpha_j(\theta)$, but unfortunately, it is discontinuous in θ at the value of θ for which $S_j = L$, whenever there is one. Therefore, this particular smoothing is not effective here.

For a second try, suppose we view ω as the set of uniforms used to generate the first $j-1$ service times and the first j interarrival times, and redefine

$$h_j(\theta,\omega) = P_\theta(X_j > L \mid \omega) = 1 - G_\theta(L - \max(0, X_{j-1} - A_j)). \tag{25}$$

The corresponding IPA estimator is then

$$\begin{aligned} h'_j(\theta,\omega) &= -\frac{\partial}{\partial\theta} G_\theta(L - \max(0, X_{j-1} - A_j)) \\ &\quad + g_\theta(L - \max(0, X_{j-1} - A_j)) X'_{j-1} I(X_{j-1} \geq A_j). \end{aligned}$$

This estimator is in fact the same as for the first example of Wardi et al. (1990). Under our assumptions, $h_j(\theta,\omega)$ is continuous and piecewise differentiable for each ω, and

$$h'_j(\theta,\omega) \leq C + K \sum_{k=1}^{j} \Gamma(U_k),$$

so that A1(3) is satisfied and Theorems 1–3 apply. For an estimator of $\alpha'(\theta)$, just take $(1/t)\sum_{j=1}^{t} h'_j(\theta,\omega)$.

To reduce the variance further, one can condition on less, i.e. view ω as containing less information. For example, suppose we view ω as the set of uniforms used to generate the first $j-1$ service times and the first $j-1$ interarrival times. (In the previous example, the equivalent of this would have reduced the variance to zero.) Assume that F has a density, and that $dg_\theta/d\theta$ exists and is bounded by an integrable function over $[0, L]$. Then,

$$\begin{aligned} h_j(\theta,\omega) &= P_\theta(X_j > L \mid \omega) = P_\theta(S_j > L) + P_\theta(S_j \le L,\ A_j < X_{j-1} + S_j - L) \\ &= 1 - G_\theta(L) + \int_0^L F(X_{j-1} - L + s) g_\theta(s) ds, \end{aligned} \tag{26}$$

and Theorems 1–3 apply. The IPA estimator is:

$$h'_j(\theta,\omega) = -\frac{\partial}{\partial\theta} G_\theta(L) + \int_0^L \left(X'_{j-1} f(X_{j-1} - L + s) g_\theta(s) + F(X_{j-1} - L + s) \frac{\partial}{\partial\theta} g_\theta(s) \right) ds. \tag{27}$$

Again, we performed a set of experiments as for the previous examples. The interarrival and service times were exponential exactly as in Example 1. In that case, one has $-\partial G_\theta(s)/\partial\theta = (s/\theta^2)e^{-s/\theta}$, and the expressions (26) and (27) become

$$\begin{aligned} h_j(\theta,\omega) &= e^{-L/\theta} + \int_{\max(0, L - X_{j-1})}^{L} (1 - e^{-(X_{j-1} - L + s)})(1/\theta) e^{-s/\theta} ds \\ &= \begin{cases} \dfrac{1}{\theta+1} \left(e^{-X_{j-1} - L/\theta} + \theta e^{-(L - X_{j-1})/\theta} \right) & \text{if } X_{j-1} < L; \\ 1 + \dfrac{1}{\theta+1} \left(e^{-X_{j-1} - L/\theta} - e^{L - X_{j-1}} \right) & \text{if } X_{j-1} \ge L; \end{cases} \end{aligned}$$

and

$$h'_j(\theta,\omega) = \frac{1}{(\theta+1)^2} \left[\varphi_j(\theta,\omega) + \left((\theta+1)(L/\theta^2 - X'_{j-1}) - 1 \right) e^{-X_{j-1} - L/\theta} \right],$$

where

$$\varphi_j(\theta,\omega) = \begin{cases} \left(1 + (\theta+1)(X'_{j-1} + (L - X_{j-1})/\theta \right) e^{-(L - X_{j-1})/\theta} & \text{if } X_{j-1} < L; \\ \left(1 + (\theta+1) X'_{j-1} \right) e^{L - X_{j-1}} & \text{if } X_{j-1} < L. \end{cases}$$

We took $t = 20$ and $L = 2$. The true value of the derivative (computed by recursive equations) is $\alpha'(0.5) = 0.82484$. The results appear in Table 3. The smoothed estimators (24), (25), and (26) are denoted by S1, S2, and S3, respectively. Again, whenever IPA applies, the results of FDC and IPA are almost identical. Also, straightforward FDC using (24) appears to work reasonably well in this case.

	c_n	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$	$n = 10^6$
FDC	$n^{-1/3}$	3.409E-3	5.899E-4	1.039E-4	1.623E-5	2.786E-6
FDC	n^{-1}	1.542E-2	2.470E-2	1.994E-2	2.811E-2	2.250E-2
FDC-S1	n^{-1}	6.409E-3	2.793E-3	1.241E-3	3.169E-3	1.238E-3
FDC-S2	n^{-1}	5.233E-3	6.404E-4	7.610E-5	7.172E-6	7.326E-7
FDC-S3	n^{-1}	2.842E-3	3.447E-4	4.238E-5	4.012E-6	4.073E-7
IPA-S2	—	5.329E-3	6.369E-4	7.625E-5	7.173E-6	7.326E-7
IPA-S3	—	2.772E-3	3.447E-4	4.238E-5	4.012E-6	4.073E-7

Table 3: Sample mean square errors for Example 3.

6. Conclusion

We have analysed the strong relationship between IPA and FDC for gradient estimation. After reading certain articles, one might be tempted to conclude that IPA estimators have much less variance than their finite-differences counterparts. But we have shown that if common random numbers are used and if the interval sizes for the finite differences are chosen appropriately, this is not really the case.

It is true that IPA estimators require just one simulation, whatever be the number of parameters, while if there are $2d$ parameters, FDC estimators require $2d$ simulations for the central case and $d+1$ simulations for the forward case. IPA can be applied to real-life systems (not just simulations), but not FDC (see Suri 1989). FDC can also have numerical problems when the intervals are too small. Finally, implementing FDC with the proper synchronisation is not always easy in practice, especially for complex systems (see, e.g., Bratley, Fox, and Schrage 1987, and L'Ecuyer, Giroux, and Glynn 1990). On the other hand, IPA is not always trivial to implement either. In some cases, finding out the expression for the IPA estimator and implementing it into a simulation program can be really complicated and/or tedious. FDC might then be a reasonably efficient and less expensive alternative.

Acknowledgements

This work has been supported by NSERC-Canada grant # A5463 and FCAR-Québec grant # EQ2831 to the first author, and a NSERC-Canada scolarship to the second author.

References

Bratley, P., B. L. Fox and L. E. Schrage. 1987. *A Guide to Simulation*, Springer-Verlag, New York, second edition.

Cao, X.-R. 1985. "Convergence of Parameter Sensitivity Estimates in a Stochastic Experiment", *IEEE Trans. on Automatic Control*, **AC-30**, 9, 845–853.

Chung, K.-L. 1974. *A Course in Probability Theory*, Academic Press, New York, second edition.

Glasserman, P. 1988. "Performance Continuity and Differentiability in Monte Carlo Optimization", *Proceedings of the Winter Simulation Conference 1988*, The Society for Computer Simulation, 518–524.

Glasserman, P. and Gong, W.-B. 1990. "Smoothed Perturbation Analysis for a Class of Discrete Event Systems", to appear in *IEEE Trans. on Automatic Control.*

Glynn, P. W. 1989. "Optimization of Stochastic Systems Via Simulation", *Proceedings of the Winter Simulation Conference 1989*, IEEE Press, 90–105.

Glynn, P. W. and W. Whitt. 1989. "The Asymptotic Efficiency of Simulation Estimators", Technical report No. 47, Department of Operations Research, Stanford University.

Heidelberger, P., X.-R. Cao, M. A. Zazanis and R. Suri. 1989. "Convergence Properties of Infinitesimal Perturbation Analysis Estimates", *Management Science*, **34**, 11, 1281–1302.

Ho, Y.-C. 1987. "Performance Evaluation and Perturbation Analysis of Discrete Event Dynamic Systems", *IEEE Transactions of Automatic Control*, **AC-32**, 7, 563–572.

Kushner, H. J. and Clark, D. S. 1978. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, Applied Math. Sciences, vol. 26

L'Ecuyer, P. 1990. "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques", to appear in *Management Science.*

L'Ecuyer, P., N. Giroux and P. W. Glynn. 1990. "Stochastic Optimization by Simulation: Some Experiments With a Simple Steady-State Queue", submitted for publication.

Meketon, M. S. 1987. "Optimization in Simulation: A Survey of Recents Results", *Proceedings of the Winter Simulation Conference 1987*, The Society for Computer Simulation, 58–67.

Suri, R. 1987. "Infinitesimal Perturbation Analysis for General Discrete Event Systems", *J. of the ACM*, **34**, 3, 686–717.

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**, 114–137.

Wardi, Y., Gong, W.-B., Cassandras, C. G., and Kallmes, M. H. 1990. "A New Class of Perturbation Analysis Algorithms for Piecewise-Continuous Sample Performance Functions", Submitted for publication.

Zazanis, M. A. and Suri, R. 1988. "Comparison of Perturbation Analysis with Conventional Sensitivity Estimates for Stochastic Systems", Submitted for publication.