TECHNICAL RESEARCH REPORT

Stochastic Gradient Estimation

by Michael Fu

TR 2005-93

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2005		**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Stochastic Gradient Estimaiton	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Office of Scientific Research,875 North Randolph Street,Arlington,VA,22203-1768**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**31**	

Stochastic Gradient Estimation[1]

Michael C. Fu
University of Maryland, College Park, mfu@rhsmith.umd.edu

Abstract

We consider the problem of efficiently estimating gradients from stochastic simulation. Although the primary motivation is their use in simulation optimization, the resulting estimators can also be useful in other ways, e.g., sensitivity analysis. The main approaches described are finite differences (including simultaneous perturbations), perturbation analysis, the likelihood ratio/score function method, and the use of weak derivatives.

1 Introduction

For optimization problems with *continuous-valued* controllable parameters, the availability of gradients is clearly helpful in obtaining improved solutions based on an iterative scheme, and can play a critical role in making a particular problem tractable. This is true in stochastic optimization, as well, especially for such problems based on an underlying simulation model. However, in this stochastic setting, since the outputs are themselves random, finding or deriving *stochastic* gradient estimators can itself be a challenging problem.

We write the general simulation optimization problem as follows:

$$\min_{\theta \in \Theta} J(\theta), \tag{1}$$

where $\theta \in \Theta$ is the controllable parameter and $\Theta \subset \mathcal{R}^d$ is the feasible region, i.e., θ is a d-dimensional vector.

We describe three examples that will be for illustrative purposes. The first example is a stochastic activity network. The input random variables are the individual activity times $X_1, X_2, \ldots, X_d$, and the objective function is the total project duration. The parameters are the mean activity times, i.e., $\theta = (\theta_1, \ldots, \theta_d)$, where θ_i is the mean of the ith activity, and the dimension of the parameter vector is equal to the number of activities. The optimization problem is to minimize the expected project duration as a function of the mean activity times, where a cost is attached to each choice of mean activity time.

The second example is the familiar first-come, first-served (FCFS), single-server queue, with the input random variables being the interarrival and service times, and the output performance measure being time spent in the system. The objective function includes a cost on the service rate, and the optimization problem is

$$\min_{\theta \in (0, 1/\lambda)} E[T(\theta)] + c/\theta, \tag{2}$$

where c is the service rate cost, $T(\theta)$ is the average system time, θ is the mean service time (hence θ is scalar), and λ is the arrival rate. Usually the problem is considered in steady state, and often specializing to the $M/M/1$ queue, because this leads to an analytically tractable solution that can be compared with that obtained using the simulation optimization algorithm.

The third example is an (s, S) inventory control system. Here, the input random variables are related to demand (amount and possibly timing), with the objective function being a total cost function associated with inventory levels and order amounts, and the optimization problem being to minimize expected total cost by the selection of the inventory control parameters s and S, i.e., $\theta = (s, S)$ is a two-dimensional vector.

Although the main application of gradient estimation emphasized in this paper is simulation-based optimization, derivative estimation has other important applications in simulation, most notably *sensitivity analysis*. This can be useful in many different contexts, e.g., factor screening to decide which factors are the most critical, and hedging of financial instruments and portfolios. The rest of this paper is organized as follows. A brief overview of gradient-based simulation optimization is provided. Then the main approaches for stochastic gradient estimation are developed in some detail, including examples, some discussion of desirable theoretical properties and computational requirements. Specific applications in different areas are then

[1]This is a pre-print version of Chapter 19 in *Handbooks in Operations Research and Management Science: Simulation*, S.G. Henderson and B.L. Nelson, eds., Elsevier.

described. Other than a few exceptions — such as acknowledging a specific key result — references to the literature will be provided in (deferred to) Sections 8 and 9 on applications and probing further.

We conclude this introduction by explicitly stating an implicit assumption that pervades stochastic gradient estimation research (as well as much of the research reported in this handbook).

Key Implicit Assumption: Each estimate of J at a given θ is expensive to generate.

This is not a very rigorous statement, as it is not even mathematical, but the gist of its implication is that because estimating J requires a nontrivial amount of effort, it behooves us to make use of its output more efficiently. The costliness can be due to a number of reasons:

(a) It is expensive to generate input random variables $\{X_i\}$ used to produce an estimate of J.

(b) A lot of input random variables need to be generated, either because each estimate involves a large number of input random variables, or because a lot of estimates (simulation replications) need to be generated to achieve a desired level of precision.

(c) It is a nontrivial task to go from the input random variables $\{X_i\}$ to the estimate of J.

In most stochastic discrete-event simulation models of practical interest, both (b) and (c) are true. If none of these conditions hold, then the simulation user should probably just use "brute force" finite difference techniques, which are described in Section 3.

2 Gradient-Based Simulation Optimization

The two main approaches for conducting simulation-based optimization can be roughly characterized as follows:

- Carry out all of the simulations first — generally a very large number of replications — and then store things appropriately (random number seeds, the random numbers themselves, or realizations of random variates or possibly sample paths), converting the stochastic problem into a deterministic problem that is based on a large enough set of samples to well approximate the desired problem.

- Carry out a relatively small set of simulations and iteratively improve upon the current solution (or set of solutions in a population-based approach) until a sufficiently good solution is reached or found.

Previously, the first approach might have been hindered for large problems by constraints in computer memory, but that has become less of an issue over the past decade, making it a more attractive option, due to its conceptual simplicity and ability to use the arsenal of deterministic nonlinear optimization algorithms available. Since our main concern is not deterministic optimization, we will not delve further along this line; see the last section for some references to the development of the theory and application of this approach, which has a number of different names: sample average approximation, sample path optimization, and stochastic counterpart. We note that the gradient estimates discussed here can also be incorporated into, and often play a critical role in, these algorithms. Often this involves "freezing" the random numbers, to be able to use them to generate a value of the performance measure at any value of the parameter, and this value is treated as a deterministic quantity rather than a sample estimate.

The second approach uses the stochastic analog of gradient-based optimization, which is converted to a zero-finding problem (for the gradient of the objective function) and then addressed using stochastic approximation, which we now describe.

2.1 Stochastic Approximation

A natural adaptation of "steepest descent" in deterministic nonlinear optimization is stochastic approximation (SA), an iterative update scheme on the parameter that takes the following general form for finding a zero of the objective function gradient:

$$\theta_{n+1} := \Pi_\Theta \left(\theta_n - a_n \widehat{\nabla} J(\theta_n) \right), \tag{3}$$

where the "hat" notation denotes an estimate of the gradient $\nabla J(\theta_n)$, $\{a_n\}$ denotes the "gain" (step-size multiplier) sequence, and Π_Θ denotes a projection back into the feasible region Θ when the update (3) takes θ out of Θ. Whereas in second-order Newton-Raphson schemes, a key enhancer is to use the inverse Hessian to estimate the optimal step size, this is much less of a concern in SA, especially in the early stages of the algorithm. In fact, to guarantee almost sure (a.s.) convergence, the gain sequence must vanish in the limit, but not too quickly. The usual condition is that

$$\sum_n a_n = \infty, \quad \sum_n a_n^2 < \infty.$$

However, in practice, one often decreases the step size to some value at which it is kept constant. Theoretically, this leads to weak convergence (in distribution), at best, which might be unsatisfactory. Note that the sequence need not be deterministic either, in which case the conditions above have to be modified accordingly. There are also Central Limit Theorem results for the asymptotic behavior of θ_n, but these are beyond the scope of this work.

When $\widehat{\nabla} J(\theta_n)$ is an unbiased estimator of $\nabla J(\theta_n)$, the SA algorithm is generally referred to as being of the Robbins-Monro type, whereas if $\widehat{\nabla} J(\theta_n)$ is only asymptotically unbiased, e.g., using a finite difference estimate with the difference going to zero at an appropriate rate, then the algorithm is referred to being of the Kiefer-Wolfowitz type. The Robbins-Monro SA algorithm generally has a canonical asymptotic convergence rate of $n^{-1/2}$, in contrast to $n^{-1/3}$ for the Kiefer-Wolfowitz SA algorithm, which takes the following form in the scalar parameter case:

$$\theta_{n+1} := \Pi_\Theta \left(\theta_n + a_n \frac{\widehat{J}(\theta_n + c_n) - \widehat{J}(\theta_n - c_n)}{2c_n} \right), \tag{4}$$

where $\widehat{J}$ denotes an estimate of J and $\{c_n\}$ denotes a difference sequence that must also decrease to zero at an appropriate rate satisfying

$$\sum_n a_n c_n < \infty, \quad \sum_n a_n^2 / c_n^2 < \infty.$$

Thus, in addition to having a slower canonical asymptotic convergence rate, a Kiefer-Wolfowitz SA algorithm involves the additional selection of an appropriate difference sequence. In certain special cases involving common random numbers, however, the best $n^{-1/2}$ rate can also be achieved in practice. A common form for the two sequences is $a_n = an^{-1}$ for some positive a (harmonic series) and $c_n = cn^{-1/6}$ for positive c. If

$$\frac{\widehat{J}(\theta_n + c_n) - \widehat{J}(\theta_n)}{c_n} \tag{5}$$

is used instead for the gradient estimator in the Kiefer-Wolfowitz SA algorithm (i.e., one-sided forward difference gradient estimator), then $c_n = cn^{-1/4}$ is commonly used.

Perhaps one of the key drawbacks of using an SA algorithm, especially for the new or inexperienced user, is the sensitivity of the early "transient" convergence rate to the choices of these sequences. For example, if the common sequences just mentioned are used, then the behavior will depend on the choices of a and c. If a is too small, then the algorithm will "crawl" towards the optimum, even at the $1/\sqrt{n}$ asymptotic rate. On the other hand, if a is chosen too large, then extreme oscillations may occur, resulting in an "unstable" progression. As far as we know, all of the theoretical results relating to convergence rate of these types of algorithms are *asymptotic* results, which may be of little practical use in some applications where all of the action occurs in a relatively short time.

Another drawback is that in general the SA algorithm converges only to a local optimum, although this can also be strengthened by the appropriate introduction of "noise" into the algorithm. Finally, in some settings, it may not be obvious how to project onto the feasible region Θ, which might be specified indirectly (e.g., in a mathematical programming formulation) and possibly involve "noisy" constraints that also have to be estimated along with the objective function. These are practical issues that still have not been adequately addressed for simulation optimization.

3 Indirect Gradient Estimation

We divide the approaches to stochastic gradient estimation into two main categories — indirect and direct — which we now define. An indirect gradient estimator usually has two characteristics: (i) it only estimates an *approximation* of the true gradient value, e.g., via a secant approximation in the scalar case; and (ii) it uses only function evaluations (performance measure output samples) from the original (unmodified) system of interest. A direct gradient estimator tries to estimate the true gradient using some additional analysis of the underlying stochastics of the model. More specifically, we will refer to the indirect gradient estimation approach as one in which the simulation output is treated as coming out of a given black box, by which we mean it satisfies two assumptions: (i) no knowledge of the underlying mechanics of the simulation model is used in deriving the estimators, such as knowing the input probability distributions; and (ii) no changes are made in the execution of the simulation model itself, such as changing the input distribution for importance sampling. Note that this entails satisfying *both* assumptions; many of the direct gradient estimation techniques can be *implemented* without changing anything in the underlying simulation, but they may require some knowledge of the simulation model, such as the input distributions or some of the system dynamics. In the case of stochastic simulation, as opposed to online estimation based on an actual system, it could be argued that to carry out the simulation most of these mechanics need to be known, i.e., one cannot carry out a stochastic simulation without specifying the input distributions. Here, we simply use the two assumptions to distinguish between the two categories of approaches and not to debate whether an estimator is "model" dependent or not. In terms of stochastic approximation algorithms, indirect and direct gradient estimators generally correspond to Kiefer-Wolfowitz and Robbins-Monro algorithms, respectively.

We describe two indirect gradient estimators: finite differences and simultaneous perturbations. Following our definition, these approaches require no knowledge of the workings of the simulation model, which is treated as a black box.

3.1 Finite Differences

The "brute force" or "naïve" method for estimating a gradient at θ is simply to use finite differences, i.e., perturbing the value of each component of θ separately while holding the other components at the nominal value. If the value of the perturbation is too small, the resulting difference estimator could be extremely noisy, because the output is stochastic; hence there is a trade-off between bias and variance in making this selection, and unless all components of the parameter vector are suitably "standardized" a priori, this choice must be done for each component separately, which could be a burdensome task for high-dimensional problems. If the goal is sensitivity analysis rather than optimization, then one would generally err on the side of selecting a relatively larger value for the perturbation.

The ith component of the one-sided forward difference gradient estimator is given by

$$\frac{\widehat{J}(\theta + c_i e_i) - \widehat{J}(\theta)}{c_i}, \tag{6}$$

where c is the vector of differences (c_i the perturbation in the ith direction) and e_i denotes the unit vector in the ith direction.

The ith component of the two-sided symmetric (or central) difference gradient estimator is given by

$$\frac{\widehat{J}(\theta + c_i e_i) - \widehat{J}(\theta - c_i e_i)}{2c_i}, \tag{7}$$

which corresponds to the estimator used in the original Kiefer-Wolfowitz SA algorithm given by (4).

Again, it should be noted that in stochastic simulation, using common random numbers can reduce the variance of the gradient estimators substantially, although in practice synchronization is clearly an issue, since merely using the same random number seeds is typically not effective. The symmetric difference estimator given by (7) is more accurate, but it requires $2d$ objective function estimates (simulation replications) per gradient estimate, as opposed to $d+1$ function estimates (simulation replications) for the one-sided estimator given by (6). For example, in the stochastic activity network, the symmetric difference estimator would involve performing simulations by varying the mean of each activity i individually by $\pm c_i$ while holding the other activity means at their nominal values, i.e., at parameter settings $\{\theta_i \pm c_i, \theta_j, j \neq i\}$, for $i = 1, \ldots, d$.

3.2 Simultaneous Perturbations

This approach has the advantage that the number of simulation replications needed to form an estimator of the gradient is independent of the dimension of the parameter vector.

The ith component of the simultaneous perturbations (SP) gradient estimator is given by

$$\frac{\widehat{J}(\theta + c\Delta) - \widehat{J}(\theta - c\Delta)}{2c_i\Delta_i}, \tag{8}$$

where $\Delta = (\Delta_1, \ldots, \Delta_d)$ is a d-dimensional vector of perturbations, which are generally assumed i.i.d. as a function of iteration and independent across components. In this case, c is again the set of differences for each component, but it is a diagonal matrix with the differences $\{c_i\}$ on the diagonal. The key difference between this estimator and a finite difference estimator is that the *numerator* of (8) — corresponding to a difference in the function estimates — is the *same* for all components (i.e., independent of i), whereas the numerator in the symmetric difference estimator given by (7) involves a different pair of function estimates for each component (i.e., is a function of i). Thus, the full gradient estimator requires only *two* function estimates, regardless of the size of the parameter dimension d. However, since d random numbers must be generated to produce the perturbation sequence Δ at each iteration, if generating the function estimates $\widehat{J}$ is relatively "cheap," then this procedure is likely to be inferior to the previous finite difference approaches. However, our key implicit assumption is that function estimates are expensive, which is generally true in the context of stochastic discrete-event simulation. Furthermore, this procedure may also be of benefit in deterministic situations where J is expensive to generate, e.g., requires a computationally intensive finite element analysis program.

The key requirement on the perturbation sequence to guarantee a.s. convergence when used in a simultaneous perturbation stochastic approximation (SPSA) algorithm is that each term have mean zero and finite inverse second moments. Thus, the normal (Gaussian) distribution is prohibited, and the most commonly used distribution is the symmetric Bernoulli, whereby the perturbation takes the positive and negative (equal in magnitude, e.g., ± 1) value w.p. 0.5. Intuitively, convergence comes about from the averaging property of the random directions selected at each iteration, i.e., in the long-run, each component will converge to the correct gradient even if at any particular iteration the estimator may appear odd. This estimator was designed for optimization via stochastic approximation, and is of limited use in sensitivity analysis, although perhaps averaging over a large number of replications might make it more applicable.

A very similar gradient estimator for use in SA algorithms is the random directions gradient estimator, whose ith component is given by

$$\frac{\left(\widehat{J}(\theta + c\Delta) - \widehat{J}(\theta - c\Delta)\right)\Delta_i}{2c_i}. \tag{9}$$

Instead of dividing by the perturbation component, the difference term multiplies the component. Thus, normal distributions can be used for the perturbation sequence, and convergence requirements translate the moment condition to a bound on the second moment, as well as zero mean. Of course, a correspondence to the SP estimator can be made by simply taking the componentwise inverse, but in practice the performance of the two resulting SA algorithms differs substantially.

A more recent development in the application of SPSA is to use deterministic sequences for the perturbation sequences $\{\Delta\}$. The idea is analogous to the use of quasi-Monte Carlo, whereby the gradient averaging is the critical factor. There are also relevant connections to literature in the design of experiments methodology. However, more theoretical work explaining their effectiveness and more numerical examples establishing their advantage over stochastic sequences is needed.

4 Direct Gradient Estimation

Although indirect gradient estimation offers greater generality, direct gradient estimation has the following advantages:

- It usually provides an unbiased estimator, which leads to faster convergence rates when implemented in a simulation optimization algorithm, e.g., stochastic approximation.

- It eliminates the need to determine appropriate values for the finite difference perturbations — c in the estimators given by (4), (5), (6), (7), (8), (9), — which influences the accuracy of the estimator. Smaller values lead to lower bias but generally at the cost of increased variance, to the point of possibly giving the wrong sign for small enough values.
- The resulting estimators are usually more computationally efficient. This is almost universally true when compared to high-dimensional brute force finite differences, but not necessarily the case when compared with SP estimators. When used in stochastic approximation, this can result in faster convergence rates, as discussed earlier.

Potential challenges in applying direct gradient estimation include the following:

- They require more "off-line" work, which might be as simple as computing some derivatives of density functions, but could involve quite a bit of problem-specific analysis requiring sophisticated applied probability tools.
- The implementation usually requires some coding inside the simulation model, and sometimes involves some changes in the way the simulation is actually carried out.

For expositional purposes, we will assume that the objective function is an expectation, specifically

$$J(\theta) = E[Y(\theta)] = E[Y(X_1, \ldots, X_T)], \tag{10}$$

where Y is the (univariate) output performance measure, $\{X_i\}$ are the input random variables, and T is a fixed finite number. This covers many types of performance measures, including distributional performance measures such as the tail distribution, handled using indicator functions. However, the "dual" performance measure involving quantiles (e.g., median), and also measures such as the mode, cannot be absorbed into this framework. The cases where T is random (a stopping time) or as T goes to infinity (steady state), can also be handled, but require some additional technicalities.

In the right-most expression for the performance measure given in (10), the dependence on θ has not been displayed, because where θ appears influences which direct gradient estimation technique is most applicable. In particular, we distinguish between two main dependencies:

- sample (pathwise);
- measure (distributional).

It is important to note that for many performance measures of interest, Y can be expressed such that either dependency can be used. We will see examples of this in the discussion that follows.

To derive direct gradient estimators, we write the expectation using what is sometimes called the law of the unconscious statistician:

$$E[Y(\mathbf{X})] = \int y dF_Y(y) = \int Y(\mathbf{x}) dF_{\mathbf{X}}(\mathbf{x}). \tag{11}$$

In fact, one of our chief views of stochastic simulation is a way of carrying out this relationship, i.e., coming into the simulation are input random variables, for which we know the distribution; coming out of the simulation are output random variables, for which we would like to know the distributions. However, what we have is a way to generate samples of the output random variables as a function of the input random variables via the simulation model. Of course, if we really knew the distribution of the sample performance output r.v. Y, then there would probably be little reason to have to simulate the system.

For simplicity in discussion, we will assume henceforth that the parameter θ is scalar, because the vector case can be handled by taking each component individually. In view of the right-hand side of (11), we revisit the question as to the location of the parameter in a stochastic setting. Putting it in the sample performance $Y(\cdot;\theta)$ corresponds to the view of perturbation analysis (PA), whereas if it is absorbed in the distribution $F(\cdot;\theta)$, then the approach follows that of the likelihood ratio (LR) method (also known as the score function (SF) method) or weak derivatives (WD) (also known as measure-valued differentiation). In the general setting where the parameter is a vector, it is possible that some of the components would be most naturally located in the sample performance, while others would be easily retained in the distributions, giving rise to a mixed approach. For example, in the (s, S) inventory control case with random order arrival

times, it might be most effective to use PA for the control parameters, WD for demand parameters, and LR/SF for order parameters.

Naming the parameter "Spot" for good measure, we pose a question and offer several observations:

(i) Where is Spot?
This may determine which approach is most appropriate. LR/SF and WD estimators consider distributional parameters, so for them the parameter should appear in the input distribution(s).

(ii) Spot can move!
For the same system, the problem may be formulated such that Spot appears in either spot. We will demonstrate this in the examples. In the first two examples, it is simply a matter of interpretation of the underlying stochastics (probability measures), whereas in the (s, S) inventory system, a change of variables is necessary.

(iii) Spot can make repeat appearances.
For example, in the single-server queue example, where the parameter appears in the service time distribution, the parameter must be considered every time a service time is generated.

(iv) Spot can be in two places at once!
It is possible that the parameter appears in both the distribution and in a sample pathwise manner. Also, the parameter could appear in more than one distribution (as opposed to being in the same distribution multiple times, as in the previous item).

Let f denote the joint p.d.f. of all of the input random variables. Differentiating (11), and assuming an interchange of integration and differentiation is permissible, we write two cases:

$$\frac{dE[Y(X)]}{d\theta} = \begin{cases} \displaystyle\int_{-\infty}^{\infty} Y(x) \frac{df(x;\theta)}{d\theta} dx & (12) \\ \displaystyle\int_0^1 \frac{dY(X(\theta;u))}{d\theta} du, & (13) \end{cases}$$

where x and u and the integrals are all T-dimensional. For notational simplicity, throughout these T-dimensional multiple integrals are written as a single integral, and we also assume one random number u produces one random variate x. In (12), the parameter appears in the distribution directly, whereas in (13), the underlying uncertainty is considered the uniform random numbers. These correspond to what we called the distributional (measure) and pathwise (sample) dependencies, respectively.

For expositional ease in introducing the approaches, we begin by assuming that the parameter only appears in X_1, which is generated *independently* of the other input random variables. So for the case of (13), we use the chain rule to write

$$\begin{aligned} \frac{dE[Y(X)]}{d\theta} &= \int_0^1 \frac{dY(X_1(\theta;u_1), X_2, \ldots))}{d\theta} du \\ &= \int_0^1 \frac{\partial Y}{\partial X_1} \frac{dX_1(\theta;u_1)}{d\theta} du. \end{aligned} \tag{14}$$

In other words, the estimator takes the form

$$\frac{\partial Y(X)}{\partial X_1} \frac{dX_1}{d\theta}, \tag{15}$$

where the parameter appears in the transformation from random number to random variate, and the derivative is expressed as the product of a sample path derivative and derivative of a random variable. The issue of what constitutes the latter will be taken up shortly, but this approach is called infinitesimal perturbation analysis (IPA). For the $M/M/1$ queue, the sample path derivative could be derived using Lindley's equation, relating the time in system of a customer to the service times (and interarrival times, which are not a function of the parameter).

Assume that X_1 has marginal p.d.f. $f_1(\cdot;\theta)$ and that the joint p.d.f. for the remaining input random variables $(X_2,\ldots)$ is given by f_{-1}, which has no (functional) dependence on θ. Then the assumed independence gives $f = f_1 f_{-1}$, and the expression (12) involving differentiation of a density (measure) can be further manipulated using the product rule of differentiation to yield the following:

$$\frac{dE[Y(X)]}{d\theta} = \int_{-\infty}^{\infty} Y(x)\frac{\partial f_1(x_1;\theta)}{\partial\theta} f_{-1}(x_2,\ldots)dx \tag{16}$$

$$= \int_{-\infty}^{\infty} Y(x)\frac{\partial \ln f_1(x_1;\theta)}{\partial\theta} f(x)dx. \tag{17}$$

In other words, the estimator takes the form

$$Y(X)\frac{\partial \ln f_1(X_1;\theta)}{\partial\theta}. \tag{18}$$

Since the term $\frac{\partial \ln f_1(\cdot;\theta)}{\partial\theta}$ is the well-known (efficient) score function in statistics, this approach has been called the score function (SF) method. The other name given to this approach — the likelihood ratio (LR) method — comes from the closely related likelihood ratio function given by

$$\frac{f_1(\cdot;\theta)}{f_1(\cdot;\theta_0)},$$

which when differentiated with respect to θ gives

$$\frac{\partial f_1(\cdot;\theta)/\partial\theta}{f_1(\cdot;\theta_0)},$$

which is equal to the score function upon setting $\theta_0 = \theta$.

On the other hand, if instead of expressing the right-hand side of (16) as (17), the density derivative is written as

$$\frac{\partial f_1(x_1;\theta)}{\partial\theta} = c(\theta)\left(f_1^{(2)}(x_1;\theta) - f_1^{(1)}(x_1;\theta)\right),$$

it leads to the following relationship:

$$\begin{aligned}\frac{dE[Y(X)]}{d\theta} &= \int_{-\infty}^{\infty} Y(x)\frac{\partial f(x;\theta)}{\partial\theta}dx,\\ &= c(\theta)\left(\int_{-\infty}^{\infty} Y(x)f_1^{(2)}(x_1;\theta)f_{-1}(x_2,\ldots)dx - \int_{-\infty}^{\infty} Y(x)f_1^{(1)}(x_1;\theta)f_{-1}(x_2,\ldots)dx\right).\end{aligned}$$

The triple $\left(c(\theta), f_1^{(1)}, f_1^{(2)}\right)$ constitutes a weak derivative (WD) for f_1, which is in general not unique, as we shall shortly see. The corresponding WD estimator is of the form

$$c(\theta)\left(Y(X_1^{(2)}, X_2,\ldots) - Y(X_1^{(1)}, X_2,\ldots)\right), \tag{19}$$

where $X_1^{(1)} \sim f_1^{(1)}$ and $X_1^{(2)} \sim f_1^{(2)}$. The term "weak" derivative comes about from the possibility that $\frac{df_1(\cdot;\theta)}{d\theta}$ in (16) may not be proper, but its *integral* may be well-defined, e.g., it might involve delta-functions (impulses), corresponding to mass functions of discrete distributions.

If in the expression (13), the interchange of expectation and differentiation does not hold (e.g., if Y is an indicator function), then as long as there is more than one input random variable, appropriate conditioning will often allow the interchange as follows:

$$\frac{dE[Y(X)]}{d\theta} = \int_0^1 \frac{dE[Y(X(\theta;u))|Z(u)]}{d\theta}du, \tag{20}$$

where $Z \subset \{X_1,\ldots,X_T\}$. This conditioning is known as smoothed perturbation analysis (SPA), because it is intended to "smooth" out a discontinuous function. It leads to an estimator of the following form:

$$\frac{\partial E[Y(X)|Z]}{\partial X_1}\frac{dX_1}{d\theta}, \tag{21}$$

Note that taking Z as the entire set leads back to (15).

Remark For SPA, the conditioning in (20) was done with respect to a subset of the input random variables only. Further conditioning can be done on events in the system, which leads to an estimator of the following general form:

$$\frac{dY}{d\theta} + E_Z[\delta Y|\mathcal{B}]\frac{dP_Z(\mathcal{B})}{d\theta}, \tag{22}$$

where the subscript indicates a corresponding conditional expectation/probability, $\mathcal{B}$ is an appropriately selected event, and δY represents a change in the performance measure under the conditioned (usually called "critical") event. In this case, if the probability rate $\frac{dP_Z(\mathcal{B})}{d\theta}$ is 0, the estimator (22) also reduces to IPA. On the other hand, if the IPA term $\frac{dY}{d\theta}$ is zero, the estimator may coincide with the WD estimator in certain cases, with correspondences between $c(\theta)$ and the probability rate, and between the difference term in (19) and the conditional expectation in (22).

4.1 Desirable Properties

Direct gradient estimators are estimators like any other estimators; they just happen to be estimating derivatives. Thus, like other estimators, we would like them to be unbiased.

Definition. The stochastic gradient estimator $\widehat{\nabla} J(\theta)$ is *unbiased* if $E[\widehat{\nabla} J(\theta)] = \nabla_\theta J(\theta)$.

We would also like the estimators to have low variance. In other words, we want them to be both correct and precise, as opposed to being wrong and noisy. Strong consistency is another desirable property. For finite horizon performance measures, we simply invoke the strong law of large numbers. For infinite horizon or steady-state performance measures, the usual ergodicity considerations come into play, and just as in steady-state output analysis methodology, this can involve a lot of theoretical tools from applied probability.

Basically, unbiasedness requires the exchange of the operations of differentiation (limit) and integration (expectation), as was assumed in deriving (12) and (13). Mathematically, this boils down to whether or not the dominated convergence theorem can be applied (see Section 6). In the case of PA, the bounding involves properties of the performance measure, whereas in LR/SF and WD, the bounding involves the distribution functions. In either case, the primary difficulty is in being able to implement the derivative/gradient, just as building a simulation model is a non-trivial task in implementation, although conceptually there may be little difficulty.

The applicability of IPA may depend on how the input processes are constructed/generated (see the next section). In applying SPA, there is the choice of conditioning quantities (cf. (20)/(21)), which affects how easily the resulting conditional expectation can be estimated from sample paths. There are also a multitude of choices of WD triples for a given input distribution, and this determines both the amount of additional simulation required and the variance of the resulting WD output gradient estimator. For LR/SF estimators, the variance of the estimator could also be a problem if care is not taken in implementation, e.g., the naïve estimator will lead to a linear increase in variance with respect to the simulation horizon.

4.2 Derivatives of Random Variables

PA estimators — e.g., those shown in (15), (21), (22) — require the notion of derivatives of random variables. The mathematical problem for defining such derivatives consists of constructing a family of random variables parameterized by θ on a *common* probability space, with the point of departure being a set of parameterized distribution functions $\{F(\cdot;\theta)\}$. We wish to construct $X(\theta) \sim F(\cdot;\theta)$ s.t. $\forall\theta \in \Theta$, $X(\theta)$ is differentiable w.p.1. The sample derivative is then defined in the intuitive manner as

$$\frac{dX(\theta,\omega)}{d\theta} = \lim_{\Delta\theta\to 0} \frac{X(\theta+\Delta\theta,\omega) - X(\theta,\omega)}{\Delta\theta},$$

where ω denotes a sample point in the underlying probability space. If the distribution of X is known, we have

$$\frac{dX(\theta)}{d\theta} = -\frac{\partial F(X;\theta)/\partial\theta}{\partial F(X;\theta)/\partial X}, \tag{23}$$

where we use the (slightly abusive) notation $\dfrac{\partial F(X;\theta)}{\partial X} = \left.\dfrac{\partial F(x;\theta)}{\partial x}\right|_{x=X}$.

Definition. For a distribution function $F(x;\theta)$, θ is said to be a *location* parameter if $F(x+\theta;\theta)$ does not depend on θ; θ is said to be a *scale* parameter if $F(x\theta;\theta)$ does not depend on θ; and θ is said to be a *generalized scale* parameter if $F(\bar{\theta}+x\theta;\theta)$ does not depend on θ, for some fixed $\bar{\theta}$ (usually a location parameter) not dependent on θ.

In these special cases, one can use the following sample derivatives for the three respective cases (location, scale, generalized scale):

$$\frac{dX}{d\theta} = 1, \quad \frac{dX}{d\theta} = \frac{X}{\theta}, \quad \frac{dX}{d\theta} = \frac{X-\bar{\theta}}{\theta}.$$

The most well-known example is the normal distribution, with the mean being a location parameter and the standard deviation a generalized scale parameter. Similarly, the two parameters in the Cauchy, Gumbel (extreme value), and logistic distributions also correspond to location and generalized scale parameters. Other examples include the mean of the exponential being a scale parameter; and the mean of the uniform distribution being a location parameter, with the half-width being a generalized scale parameter. In the special case $U(0,\theta)$, the single parameter is an ordinary scale parameter. Also, for $N(\theta,(\theta\sigma)^2)$, θ is an ordinary scale parameter.

Lastly, note that for a given distribution, there may be multiple ways to generate a random variate, which leads to different derivatives, some of which may be unbiased and some of which may not. This is called the role of representations, which we illustrate with a simple example.

Example 1 Let

$$X \sim \begin{cases} U(1,2) & \text{w.p. } \theta, \quad \theta \in (0,1), \\ U(0,1) & \text{w.p. } 1-\theta, \end{cases}$$

a mixture of two uniform distributions, with $dE[X]/d\theta = 1$. A straightforward construction/representation using two random numbers is

$$X = \mathbf{1}\{U_1 \le \theta\}(1+U_2) + \mathbf{1}\{U_1 > \theta\}U_2, \tag{24}$$

where $U_1,\ U_2 \sim U(0,1)$ are *independent* and $\mathbf{1}\{\cdot\}$ denotes the indicator function. However, since

$$\frac{dX}{d\theta} = 0 \ \text{ w.p.1},$$

this clearly leads to a biased estimator. Note that viewed as a function of θ, X jumps from U_2 to $1+U_2$ at $\theta = U_1$. However, an unbiased estimator can be obtained by using the following construction in which the "coin flipping" and uniform generation are correlated:

$$X = \mathbf{1}\{U \le \theta\}\left(1 + \frac{\theta - U}{\theta}\right) + \mathbf{1}\{U > \theta\}\left(\frac{1-U}{1-\theta}\right), \ \text{ where } U \sim U(0,1),$$

$$\Longrightarrow \frac{dX}{d\theta} = \mathbf{1}\{U \le \theta\}\frac{U}{\theta^2} + \mathbf{1}\{U > \theta\}\frac{1-U}{(1-\theta)^2},$$

which is unbiased (has expectation equal to $dE[X]/d\theta = 1$). This construction is based on the property that the distributions of the random variable $(\theta - U)/\theta$ under the condition $\{U < \theta\}$ and the random variable $(1-U)/(1-\theta)$ under the condition $\{U \ge \theta\}$ are both $U(0,1)$. From a simulation perspective, this representation has the additional advantage of requiring only a single random number to generate X instead of two as in the previous construction. In terms of the derivative, the crucial property is that X is continuous across $\theta = U$. One can easily construct other single random number representations that do not have this desirable characteristic, e.g.,

$$X = \mathbf{1}\{U \le \theta\}(1 + \frac{U}{\theta}) + \mathbf{1}\{U > \theta\}\frac{U-\theta}{1-\theta}, \ \text{ where } U \sim U(0,1),$$

$$\Longrightarrow \frac{dX}{d\theta} = \mathbf{1}\{U \le \theta\}\frac{-U}{\theta^2} + \mathbf{1}\{U > \theta\}\left[-\frac{1-U}{(1-\theta)^2}\right],$$

which is biased (has expectation -1), the intuitive reason being the discontinuity of X at $U = \theta$.

For the first representation, where the "natural" construction leads to a biased estimator, we demonstrate the use of conditioning (SPA) to rectify the situation. First, we derive an alternative estimator by conditioning on U_2, i.e., let

$$X = E[X_1|U_2] = (1 + U_2)\theta + U_2(1 - \theta) = \theta + U_2,$$

leading to the obviously unbiased estimator $dX/d\theta = 1$.

Alternatively, consider the event

$$\mathcal{B}(\Delta\theta) = \{U_1 \in (\theta, \theta + \Delta\theta]\}, \quad \Delta\theta > 0.$$

Intuitively, this event corresponds to those samples such that a perturbation $\Delta\theta > 0$ causes a "jump" in the sample performance from U_2 (when $U_1 > \theta$) to $1 + U_2$ (when $U_1 \leq \theta + \Delta\theta$). The complement of this event corresponds to the "smooth" case of IPA. Thus, we write for the representation defined by (24):

$$\begin{aligned}
\frac{dE[X]}{d\theta} &= \lim_{\Delta\theta\to 0} \frac{E[X(\theta + \Delta\theta) - X(\theta)]}{\Delta\theta} \\
&= \lim_{\Delta\theta\to 0} \frac{E[(X(\theta + \Delta\theta) - X(\theta))\mathbf{1}(\mathcal{B}^c(\Delta\theta))]}{\Delta\theta} \\
&\quad + \lim_{\Delta\theta\to 0} \frac{E[(X(\theta + \Delta\theta) - X(\theta))\mathbf{1}(\mathcal{B}(\Delta\theta))]}{\Delta\theta} \\
&= E\left[\frac{dX}{d\theta}\right] + \lim_{\Delta\theta\to 0} E[X(\theta + \Delta\theta) - X(\theta)|\mathcal{B}(\Delta\theta)] \lim_{\Delta\theta\to 0} \frac{P(\mathcal{B}(\Delta\theta))}{\Delta\theta} \qquad (25) \\
&= \lim_{\Delta\theta\to 0} (E[(1 + U_2) - U_2]) \lim_{\Delta\theta\to 0} \frac{\Delta\theta}{\Delta\theta} \;=\; 1,
\end{aligned}$$

where $\mathcal{B}^c$ denotes the complement of $\mathcal{B}$. In this case, because we can evaluate the difference $(1 + U_2) - U_2$ analytically, the final "estimator" is a deterministic quantity equal to the value to be estimated. In more complicated systems, the challenge is usually in estimating this difference of the sample performance under two different conditions on the sample path, given by the limiting conditional expectation in (25).

4.3 Derivatives of Measures

As we have seen already, both the LR/SF and WD estimators rely on differentiation of the underlying measure, so the parameters of interest should appear in the underlying (input) distributions. If this is not the case, then one approach is to try to "push" the parameter out of the performance measure into the distribution, so that the usual differentiation can be carried out. This is achieved by a change of variables, which is problem dependent.

Recall that we introduced the idea of a weak derivative by expressing the derivative of a density as the difference of two measures, i.e.,

$$\frac{\partial f(x;\theta)}{\partial\theta} = c(\theta)\left(f^{(2)}(x;\theta) - f^{(1)}(x;\theta)\right).$$

This idea can be generalized without the need for a differentiable density, as long as the integral exists with respect to a certain set of (integrable) "test" functions, say $\mathcal{L}$, e.g., the set of continuous bounded functions.

Definition. The triple $(c(\theta), F^{(1)}, F^{(2)})$ is called a *weak derivative* for distribution (c.d.f.) F if for all functions $g \in \mathcal{L}$ (not a function of θ),

$$\frac{d}{d\theta}\int g(x)dF(x;\theta) = c(\theta)\left(\int g(x)dF^{(2)}(x;\theta) - \int g(x)dF^{(1)}(x;\theta)\right).$$

Remarks. As mentioned earlier, the derivative is "weak" in the sense that the density derivative may not be defined in the usual sense, but in terms of generalized functions integrable with respect to the functions in $\mathcal{L}$, as in the "definition" of a delta function in terms of its integral, i.e., they could include discrete mass

functions, as well. The concept of a weak derivative need not be restricted to probability measures, but any finite signed measures. Lastly, note that a WD gradient estimate may require as many as $2d$ additional simulations for the vector case (a pair for each component), unlike LR/SF and IPA estimators, which will always require just a single simulation.

One choice for the weak derivative (density) that is readily available is

$$\frac{df}{d\theta} = c\,(f_2 - f_1)\,, \tag{26}$$

where

$$f_1 = \frac{1}{c}\left(\frac{df}{d\theta}\right)^-, \quad f_2 = \frac{1}{c}\left(\frac{df}{d\theta}\right)^+, \tag{27}$$

$x^+ \equiv \max(x,0), x^- \equiv \max(-x,0)$, and $c = \int \left(\frac{df}{d\theta}\right)^+ dx = \int f_2 dx = \int \left(\frac{df}{d\theta}\right)^- dx$, using the fact that

$$\int f(x)dx = 1 \quad \Longrightarrow \quad \int \frac{df}{d\theta}dx = 0.$$

The representation given by (26) and (27) is the Hahn-Jordan decomposition, which will always exist for probability measures, and results in a decomposition involving two measures with complementary support.
Remarks. The representation is clearly not unique. In fact, for any non-negative integrable function h, we have

$$\frac{df}{d\theta} = c\,([f_1 + h] - [f_2 + h]) = c'\left([f_1 + h]/(1 + \int h) - [f_2 + h]/(1 + \int h)\right),$$

where $c' = c(1 + \int h)$. Thus, one way to obtain the estimator using the original simulation is to choose a representation in which both measures have the same support as the original measure. Then importance sampling can be applied, so that the original simulation can be used to generate the estimator without the need for simulating the system under alternative input distributions. Perhaps the most direct way to achieve this is to add the original measure itself to both f_1 and f_2 and renormalize appropriately, i.e., choose $h = f$ above:

$$\frac{df}{d\theta} = 2c\,([f_1 + f]/2 - [f_2 + f]/2)\,.$$

4.4 Input Distribution Examples

We now demonstrate some of these concepts on a single random variable. Section 5 will consider the examples introduced at the beginning.

Example 2 Let $X \sim \exp(\theta)$, an exponential random variable with mean θ, with p.d.f. given by

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}\mathbf{1}\{x > 0\}.$$

The usual construction of the r.v. is as follows:

$$X(\theta;u) = -\theta \ln u,$$

where u represents a random number. Differentiating, we get

$$\begin{aligned}
\frac{df(x;\theta)}{d\theta} &= \left[\frac{x}{\theta^2}\frac{1}{\theta}e^{-x/\theta} - \frac{1}{\theta^2}e^{-x/\theta}\right]\mathbf{1}\{x > 0\} \\
&= f(x;\theta)\left[\frac{x}{\theta^2} - \frac{1}{\theta}\right] \\
&= \frac{1}{\theta}\left[\frac{x}{\theta^2}e^{-x/\theta}\mathbf{1}\{x > 0\} - f(x;\theta)\right] \\
&= \frac{1}{\theta e}\left[\frac{e}{\theta}\left(\frac{x}{\theta} - 1\right)e^{-x/\theta}\mathbf{1}\{x > \theta\} - \frac{e}{\theta}\left(1 - \frac{x}{\theta}\right)e^{-x/\theta}\mathbf{1}\{0 < x \le \theta\}\right], \\
\frac{dX(\theta;u)}{d\theta} &= -\ln u \;=\; \frac{X(\theta;u)}{\theta}.
\end{aligned}$$

In the third and fourth lines above, the density derivative (which is itself *not* a density) has been expressed as the difference of two densities multiplied by a constant. This demonstrates that the weak derivative representation is not unique, with the last decomposition being the Hahn-Jordan decomposition, noting that $x = \theta$ is the point at which $df(x;\theta)/d\theta$ changes sign. The following correspond to the LR/SF, WD (a) & (b), and IPA estimators, respectively:

$$Y(X)\frac{1}{\theta}\left(\frac{X_1}{\theta} - 1\right),$$

$$\frac{1}{\theta}\left[Y(X_{1a}^{(2)},\ldots) - Y(X_{1a}^{(1)},\ldots)\right], \quad \frac{1}{\theta e}\left[Y(X_{1b}^{(2)},\ldots) - Y(X_{1b}^{(1)},\ldots)\right],$$

$$\frac{dY}{dX_1}\frac{X_1}{\theta},$$

where $X_{1a}^{(1)}$ and $X_{1a}^{(2)}$ are random variables distributed as $exp(\theta)$ and $Erl(2,\theta)$, respectively, and $X_{1b}^{(1)}$ and $X_{1b}^{(2)}$ are random variables with densities $\frac{e}{\theta}\left(1 - \frac{x}{\theta}\right)e^{-x/\theta}, 0 < x \leq \theta$, and $\frac{e}{\theta}\left(\frac{x}{\theta} - 1\right)e^{-x/\theta}, x > \theta$, respectively.

The following is a simple example that demonstrates that the WD estimator is more broadly applicable than the LR/SF estimator.

Example 3 Let $X \sim U(0,\theta)$. Then we have the following:

$$\begin{aligned}
f(x;\theta) &= \frac{1}{\theta}\mathbf{1}\{0 < x < \theta\}, \\
X(\theta;u) &= u\theta, \\
\frac{df(x;\theta)}{d\theta} &= \frac{1}{\theta}\left[\delta(\theta - x) - \frac{1}{\theta}\mathbf{1}\{0 < x < \theta\}\right] \\
&= \frac{1}{\theta}\left[\delta(\theta - x) - f(x;\theta)\right], \\
\frac{dX(\theta;u)}{d\theta} &= u \quad = \quad \frac{X(\theta;u)}{\theta},
\end{aligned} \tag{28}$$

where we define the Dirac δ-function as the "derivative" of a step function by

$$\mathbf{1}\{x \geq \theta\} = \int_{-\infty}^{x} \delta(y - \theta)dy. \tag{29}$$

On the right-hand side of Equation (28), we have the difference of densities for a mass at θ and the original $U(0,\theta)$ distribution, respectively, i.e., the weak derivative representation $(1/\theta, \theta, F)$, where θ indicates a deterministic distribution with mass at θ. So, for example, the estimator in (19) would be given by

$$\frac{1}{\theta}\left(Y(\theta, X_2,\ldots) - Y(X_1, X_2,\ldots)\right).$$

This is a case where the LR/SF estimator is ill-defined, due to the δ-function. Another example is the following.

Example 4 Let $X \sim Par(\alpha,\theta)$, which represents the Pareto distribution with shape parameter α and scale parameter θ, and p.d.f. given by

$$f(x) = \theta^{\alpha}\alpha x^{-(\alpha+1)}\mathbf{1}\{x \geq \theta\}.$$

Once again the LR/SF estimator does not exist, due to the appearance of the parameter in the indicator function that controls the support of the distribution, whereas WD estimators can be derived (see Table 1 at the end of the section).

However, the very general exponential family of distributions leads to a nice form for the LR/SF estimator.

Example 5 Let θ denote the vector of parameters in a p.d.f. that can be written in the following form:

$$f(x;\theta) = k(\theta)\exp\left(\sum_i v_i(\theta)t_i(x)\right)h(x),$$

where the functions h and $\{t_i\}$ are independent of θ, and the functions k and $\{v_i\}$ do not involve the argument. Then it is straightforward to derive

$$\frac{\partial \ln f(x;\theta)}{\partial \theta} = \frac{\nabla k(\theta)}{k(\theta)} + \sum_i \nabla v_i(\theta) t_i(x).$$

Examples include the normal, gamma, Weibull, and exponential, for the continuous case, and the binomial, Poisson, and geometric for the discrete case.

Discrete distributions present separate challenges for the two approaches. Basically, when the parameter appears in the support *probabilities*, then LR/SF can be easily applied, whereas IPA is in general not applicable. The reverse is true, however, if the parameter appears instead in the support *values*. Here are two examples to demonstrate this, where we work directly with probability mass functions $p(x;\theta) = P(X = x)$, instead of densities with δ-functions. Let $Ber(p; a, b)$ denote a Bernoulli distribution that takes value a with probability p and b with probability $1 - p$.

Example 6 Let $X \sim Ber(\theta; a, b)$, which has mass function

$$p(x;\theta) = \theta \cdot \mathbf{1}\{x = a\} + (1-\theta) \cdot \mathbf{1}\{x = b\},$$

so

$$\frac{\partial p}{\partial \theta} = \mathbf{1}\{x = a\} - \mathbf{1}\{x = b\},$$

which can be viewed as the difference of two (deterministic) masses at a and b (with $c(\theta) = 1$), and is the Hahn-Jordan decomposition in this case. For the LR/SF estimator, we have

$$\frac{\partial \ln p}{\partial \theta} = \frac{\mathbf{1}\{x = a\} - \mathbf{1}\{x = b\}}{\theta \mathbf{1}\{x = a\} + (1-\theta)\mathbf{1}\{x = b\}} = \frac{1}{\theta}\mathbf{1}\{x = a\} - \frac{1}{1-\theta}\mathbf{1}\{x = b\}.$$

In this case, there is no way to construct X such that it will be differentiable w.p.1. For example, the natural construction/representation

$$X = a\mathbf{1}\{U \le \theta\} + b\mathbf{1}\{U > \theta\}$$

yields $dX/d\theta = 0$ w.p.1, so IPA is not applicable.

Example 7 Let $X \sim Ber(p; \theta; 0)$, which has mass function

$$p(x;\theta) = p \cdot \mathbf{1}\{x = \theta\} + (1-p) \cdot \mathbf{1}\{x = 0\},$$

which is not differentiable with respect to θ, so LR/SF and WD estimators cannot be derived.

The natural random variable construction

$$X = \theta \cdot \mathbf{1}\{U \le p\}$$

leads to the unbiased

$$\frac{dX}{d\theta} = \mathbf{1}\{U \le p\} = \mathbf{1}\{X = \theta\}$$

(which in this example also happens to equal X/θ). The IPA estimator $dX/d\theta = \mathbf{1}\{X = \theta\}$ holds even if any number of additional values are added to the underlying support, if all of them do not involve θ. If θ enters into them, then the estimator can be easily modified to reflect that.

It is straightforward to verify the corresponding derivatives of some common distributions displayed in Table 1, where $geo(p)$ indicates a geometric distribution with mass function

$$(1-p)^{n-1}p, \quad x = 1, 2, \ldots;$$

$bin(n, p)$ indicates a binomial distribution with mass function

$$\left(\begin{array}{c} n \\ p \end{array} \right) p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots;$$

$negbin(n,p)$ indicates a negative binomial distribution with mass function

$$\left(\begin{array}{c} x-1 \\ n-1 \end{array} \right) (1-p)^{x-n} p^n, \quad x = n, n+1, \ldots;$$

$Poi(\lambda)$ indicates a Poisson distribution (mean/variance λ) with mass function

$$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, \ldots;$$

$Erl(n,\beta)$ indicates an Erlang distribution with p.d.f.

$$\frac{\beta^{-n} x^{n-1} e^{-x/\beta}}{(n-1)!} \mathbf{1}\{x > 0\};$$

$Mxw(\mu, \sigma^2)$ indicates a double-sided Maxwell distribution with p.d.f.

$$\frac{(x-\mu)^2}{\sqrt{2\pi}\sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma}};$$

$Wei(\alpha,\beta)$ indicates a Weibull distribution with shape parameter α and scale parameter β, and p.d.f.

$$\alpha\beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \mathbf{1}\{x > 0\};$$

$gam(\alpha,\beta)$ indicates a gamma distribution with shape parameter α and scale parameter β, and p.d.f.

$$\frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} \mathbf{1}\{x > 0\},$$

where

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt = (\alpha-1)\Gamma(\alpha-1);$$

recalling that $Erl(n,\beta) = gam(n,\beta)$ for positive integer n; $Wei(1,\beta) = gam(1,\beta) = exp(\beta)$; and the special distribution $F^*(\alpha,\theta)$ has p.d.f. (does this fall into any of the well-known families?)

$$\alpha\beta^{-2\alpha} x^{2\alpha-1} e^{-(x/\beta)^\alpha}.$$

Note that our definition of the second parameter β in the gamma and Weibull distributions (also corresponding to the only parameter in the exponential distribution) is the inverse of what is usually found in the literature (Law and Kelton 2000 being a notable exception), but leads to it being a scale parameter under our definition.

5 Examples

5.1 Stochastic Activity Network

Let X_i have p.d.f. f_i, and assume all of the activity times are *independent*. Let P^* denote the set of activities on the optimal path corresponding to the project duration (e.g., shortest or longest path, depending on the problem), so we can write

$$Y = \sum_{j \in P^*} X_j,$$

where P^* itself is a random variable. We wish to estimate $dE[Y]/d\theta$. Let θ be a parameter in the distribution of X_1, i.e., in f_1 only. Then, the IPA estimator is given by

$$\frac{dY}{d\theta} = \frac{dX_1}{d\theta} \mathbf{1}\{1 \in P^*\}.$$

input dist $X \sim F$	WD $(c, F^{(2)}, F^{(1)})$	IPA $\frac{dX}{d\theta}$	LR/SF $\frac{d \ln f(X)}{d\theta}$
$Ber(\theta; a, b)$	$(1, a, b)$	NA	$\frac{1}{\theta}\mathbf{1}\{X = a\}$ $-\frac{1}{1-\theta}\mathbf{1}\{X = b\}$
$Ber(p; \theta, b)$	NA	$\mathbf{1}\{X = \theta\}$	NA
$geo(\theta)$	$(\frac{1}{\theta}, geo(\theta), negbin(2, \theta))$	NA	$\frac{1}{\theta} + \frac{1-X}{1-\theta}$
$bin(n, \theta)$	$(n, 1 + bin(n-1, \theta), bin(n-1, \theta))$	NA	$\frac{X}{\theta} - \frac{n-X}{1-\theta}$
$Poi(\theta)$	$(1, 1 + Poi(\theta), Poi(\theta))$	NA	$\frac{X}{\theta} - 1$
$N(\theta, \sigma^2)$	$\left(\frac{1}{\sqrt{2\pi}\sigma}, \theta + Wei(2, \frac{1}{2\sigma^2}), \theta - Wei(2, \frac{1}{2\sigma^2})\right)$	1	$\frac{X-\theta}{\sigma^2}$
$N(\mu, \theta^2)$	$(\frac{1}{\theta}, Mxw(\mu, \theta^2), N(\mu, \theta^2))$	$\frac{X-\mu}{\theta}$	$\frac{1}{\theta}\left[\left(\frac{x-\mu}{\theta}\right)^2 - 1\right]$
$U(0, \theta)$	$(\frac{1}{\theta}, \theta, U(0, \theta))$	$\frac{X}{\theta}$	NA
$U(\theta - \gamma, \theta + \gamma)$	$(\frac{1}{2\gamma}, \theta + \gamma, \theta - \gamma)$	1	NA
$U(\mu - \theta, \mu + \theta)$	$(\frac{1}{\gamma}, Ber(0.5; \mu - \theta, \mu + \theta), U(\mu - \theta, \mu + \theta))$	$\frac{X-\mu}{\sigma}$	NA
$exp(\theta)$	$(\frac{1}{\theta}, Erl(2, \theta), exp(\theta))$	$\frac{X}{\theta}$	$\frac{1}{\theta}\left(\frac{X}{\theta} - 1\right)$
$Wei(\alpha, \theta)$	$(\frac{\alpha}{\theta}, F^*(\alpha, \theta), Wei(\alpha, \theta))$	$\frac{X}{\theta}$	$\frac{1}{\theta}\left[\left(\frac{X}{\theta}\right)^\alpha - \alpha\right]$
$gam(\alpha, \theta)$	$(\frac{\alpha}{\theta}, gam(\alpha + 1, \theta), gam(\alpha, \theta))$	$\frac{X}{\theta}$	$\frac{1}{\theta}\left(\frac{X}{\theta} - \alpha\right)$
$Par(\alpha, \theta)$	$(\frac{\alpha}{\theta}, Par(\alpha, \theta), \theta)$	$\frac{X}{\theta}$	NA

Table 1: Derivatives for some common/simple input distributions ("NA" = not applicable).

The LR/SF estimator is given by

$$Y \frac{\partial \ln f_1(X_1; \theta)}{\partial \theta}.$$

The WD estimator is of the form

$$c(\theta)\left(Y(X_1^{(2)}, X_2, \ldots, X_T) - Y(X_1^{(1)}, X_2, \ldots, X_T)\right)$$

where $X_1^{(j)} \sim F_1^{(j)}, j = 1, 2$, and $(c(\theta), F_1^{(2)}, F_1^{(1)})$ is a weak derivative for F_1.

If we allow the parameter to possibly appear in all of the distributions, then the IPA estimator is found by applying the chain rule:

$$\frac{dY}{d\theta} = \sum_{i \in P^*} \frac{dX_i}{d\theta};$$

whereas the LR/SF and WD estimators are derived by applying the product rule of differentiation to the underlying input distribution, applying the independence that allows the joint distribution to be expressed as a product of marginals. In particular, the LR/SF estimator is given by

$$Y(X)\left(\sum_{i=1}^{d} \frac{\partial \ln f_i(X_i; \theta)}{\partial \theta}\right).$$

The WD estimator is of the form

$$\sum_{i=1}^{T} c_i(\theta)\left(Y(X_1, \ldots, X_i^{(2)}, \ldots, X_T) - Y(X_1, \ldots, X_i^{(1)}, \ldots, X_T)\right),$$

where $X_i^{(j)} \sim F_i^{(j)}, j = 1, 2; i = 1, \ldots, T$, and $(c_i(\theta), F_i^{(2)}, F_i^{(1)})$ is a weak derivative for F_i.

If instead we were interested in estimating $P(Y > y)$ for some fixed y, the WD and LR/SF estimators would simply replace Y with the indicator function $\mathbf{1}\{Y > y\}$. However, IPA will not apply, since the indicator function is inherently discontinuous, so an extension of IPA such as SPA is required. On the other hand, if the performance measure were $P(Y > \theta)$, then since the parameter does not appear in the distribution of the input random variables, WD and LR/SF estimators cannot be derived without first carrying out an appropriate change of variables.

To derive a PA estimator for the derivative of $P(Y > y)$, we first define the following:

$$\begin{aligned} \mathcal{P}_j &= \{P \in \mathcal{P} \mid j \in P\} = \text{ set of paths containing arc } j, \\ |P|_{-j} &= \text{ length of path } P \text{ with } X_j = 0. \end{aligned}$$

The idea will be to condition on all activity times except a set that includes activity times dependent on the parameter. In order to proceed, we need to specify the form of Y. We will take it to be the longest path. Other forms, such as shortest path, are analogously handled. Again, assuming that θ occurs in the density of X_1 and taking the conditioning quantities to be everything except X_1, i.e., $Z = \{X_2, \ldots, X_T\}$, we have

$$L_Z(\theta) = P_Z(Y > y) \equiv E[\mathbf{1}\{Y > y\}|X_2, \ldots, X_T] = \begin{cases} 1 & \text{if} \max_{P_i \in \mathcal{P}} |P_i|_{-1} > y; \\ P_Z\left(\max_{P_i \in \mathcal{P}_1} |P_i| > y\right) & \text{otherwise;} \end{cases}$$

where P_Z denotes the conditional (on Z) probability. Since

$$P_Z(\max_{P_i \in \mathcal{P}_1} |P_i| > y) = P_Z(X_1 + \max_{P_i \in \mathcal{P}_1} |P_i|_{-1} > y) = P_Z(X_1 > y - \max_{P_i \in \mathcal{P}_1} |P_i|_{-1}) = \bar{F}_1(y - \max_{P_i \in \mathcal{P}_1} |P_i|_{-1}),$$

$$L_Z(\theta) = \bar{F}_1(y - \max_{P_i \in \mathcal{P}_1} |P_i|_{-1}) \cdot \mathbf{1}\{\max_{P_i \in \mathcal{P}} |P_i|_{-1} \leq y\} + \mathbf{1}\{\max_{P_i \in \mathcal{P}} |P_i|_{-1} > y\}.$$

Differentiating,

$$\frac{dL_Z}{d\theta} = \frac{\partial \bar{F}_1}{\partial \theta}(y - \max_{P_i \in \mathcal{P}_1} |P_i|_{-1}) \cdot \mathbf{1}\{\max_{P_i \in \mathcal{P}} |P_i|_{-1} \leq y\}.$$

For the shortest path problem, simply replace "max" with "min" throughout.

To derive an LR/SF or WD derivative estimator for the performance measure $P(Y > \theta)$, there are two main ways to do the change of variables: subtraction or division.

$$P(Y - \theta > 0), P(Y/\theta > 1).$$

To achieve this requires translating the operation on the output performance measure back to a *change of variables* on the input random variables, so this clearly requires some additional knowledge of the system under consideration. In this particular case, it turns out that two properties make it amenable to a change of variables: (i) additive performance measure; (ii) clear characterization of paths that need to be considered. The simplest change of variables is to take

$$\tilde{X}_i = X_i/\theta \ \forall i \in \mathcal{A},$$

so that θ now appears as a scale parameter in *each* distribution f_i. If $\tilde{Y}$ represents the performance measure after the change of variables, then we have

$$P(Y > \theta) = P(\tilde{Y} > 1),$$

and this can be handled as previously discussed.

Another change of variables that will work in this case is to subtract the quantity θ from an appropriate set of arc lengths. In particular, the easiest sets are the nodes leading out of the source or the nodes leading into the sink. Note that minimal cut sets will not necessarily do the trick. Again, if $\tilde{Y}$ represents the performance measure after the change of variables, then we have

$$P(Y > \theta) = P(\tilde{Y} > 0).$$

Now the parameter θ appears in the distribution, specifically as a location parameter, but only in a relatively small *subset* of the $\{f_i\}$. Since this transformation results in the parameter appearing in fewer number of input random variables, it may be preferred, because for both the LR/SF and WD estimators, the amount of effort is proportional to the number of times the parameter appears in the distributions. The extra work for a large network can be particularly burdensome for the WD estimator. However, for the LR estimator, this type of location parameter is problematic, since it changes the support of the input random variable.

Application of PA to the problem of estimating $dP(Y > \theta)/d\theta$ also requires SPA, the idea being to condition on all activity times except a special (minimal) set that includes at least one activity that must be on the optimal (critical) path. Taking any minimal cut set, one can derive an unbiased derivative estimator for $P(Y > \theta)$ by conditioning on the complement set and then taking the sample path derivative. The details are contained in Fu (2005).

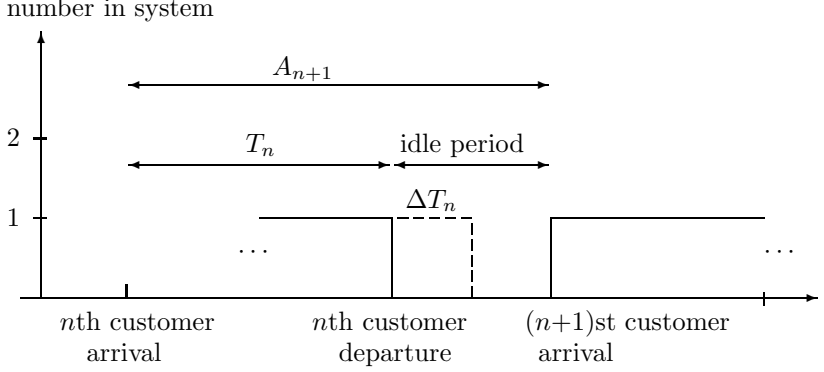

Figure 1: Quantities used in deriving FCFS single-server queue SPA estimator.

5.2 Single-Server Queue

We illustrate all of the gradient estimators for the queueing example. Let A_n be the interarrival time between the $(n-1)$st and nth customer (i.i.d. with PDF f_1 and CDF F_1), X_n the service time of the nth customer (i.i.d. with PDF f_2 and CDF F_2), and T_n the system time (in queue plus in service) of the nth customer. We consider the case where θ is a parameter in the service time distribution, and the sample performance of interest is the average system time over the first N customers $\overline{T}_N = \frac{1}{N}\sum_{n=1}^{N} T_n$. Assume that the system starts empty, so that the times of the first N service completions are completely determined by the first N interarrival times and first N service times.

The system time of a customer for a FCFS single-server queue satisfies the well-known recursive Lindley equation:

$$T_{n+1} = X_{n+1} + (T_n - A_{n+1})^+. \tag{30}$$

The IPA estimator is obtained by differentiating (30):

$$\frac{dT_{n+1}}{d\theta} = \frac{dX_{n+1}}{d\theta} + \frac{dT_n}{d\theta}\mathbf{1}\{T_n \geq A_{n+1}\}. \tag{31}$$

Using the above recursion, the IPA estimator for the derivative of average system time is given by

$$\frac{d\overline{T}_N}{d\theta} = \frac{1}{N}\sum_{n=1}^{N}\frac{dT_n}{d\theta} = \frac{1}{N}\sum_{m=1}^{M}\sum_{i=n_{m-1}+1}^{n_m}\sum_{j=n_{m-1}+1}^{i}\frac{dX_j}{d\theta}, \tag{32}$$

where M is the number of busy periods observed and n_m is the index of the last customer served in the mth busy period ($n_0 = 0$ and $n_M = N$ for M complete busy periods). Implementation of the estimator involves keeping track of two running quantities, one for (31) and another for the summation in (32); thus, the additional computational overhead is minimal, and *no alteration of the underlying simulation is required.*

The implicit assumption used in deriving an IPA estimator is that small changes in the parameter will result in small changes in the sample performance. Thus, in the above, this means that the boundary condition in (31) is unchanged by differentiation. In general, the interchange (12) will typically hold if the sample performance is continuous with respect to the parameter. For the Lindley equation, although T_{n+1} in (30) has a "kink" at $T_n = A_{n+1}$, it is still continuous at that point. This intuitively explains why IPA works. Unfortunately, the "kink" means that the derivative given by (31) has a discontinuity at $T_n = A_{n+1}$, so that IPA will fail for the second derivative.

An unbiased SPA second derivative estimator can be derived by conditioning on all *previous* interarrival and service times at each departure, which determines the system time, say T_n, with the corresponding next interarrival time, A_{n+1}, unconditioned. We provide a brief informal derivation based on sample path intuition (refer to Figure 1). For the right-hand estimator, in which we assume $\Delta T_n > 0$ (technically it should refer to $\Delta\theta$), the only "critical" events are those departures that terminate a busy period, with the possibility that two busy periods coalesce (idle period disappears) due to a perturbation. The corresponding

probability rate (conditional on T_n) is then calculated as follows:

$$\lim_{\Delta\theta\to 0}\frac{P(T_n+\Delta T_n\geq A_{n+1}|T_n<A_{n+1})}{\Delta T_n}=\frac{f_1(T_n)}{1-F_1(T_n)}\frac{dT_n}{d\theta},$$

and the corresponding effect would be that the ΔT_n perturbation would be propagated to the next busy period. The complete SPA estimator is given by

$$\left(\frac{d^2\overline{T}_N}{d\theta^2}\right)_{SPA}=\frac{1}{N}\sum_{m=1}^{M}\sum_{i=n_{m-1}+1}^{n_m}\sum_{j=n_{m-1}+1}^{i}\frac{d^2X_j}{d\theta^2}+\frac{1}{M}\sum_{m=1}^{M}\frac{f_1(T_{n_m})}{1-F_1(T_{n_m})}\left(\frac{dT_{n_m}}{d\theta}\right)^2,$$

where $d^2X/d\theta^2$ is well-defined when $F_2(X;\theta)$ is twice differentiable, $\frac{d^2X}{d\theta^2}=0$ for location, scale, and generalized scale parameters.

To derive the LR/SF estimator, since all the interarrival and service times are independently generated, the joint p.d.f. f on X will simply be the product of the p.d.f.s of the interarrival and service time distributions given by

$$f(\theta,A_1,...,A_N,X_1,...,X_N)=\prod_{i=1}^{N}f_1(A_i)\prod_{i=1}^{N}f_2(X_i;\theta). \tag{33}$$

Thus, the straightforward estimator would be given by

$$\left(\frac{d\bar{T}_N}{d\theta}\right)_{LR}=\frac{1}{N}\sum_{i=1}^{N}T_i\sum_{j=1}^{N}\frac{\partial\ln f_2(X_j;\theta)}{\partial\theta}, \tag{34}$$

where expressions for some common input distributions can be found in Table 1. However, for large N, the variance of the estimator, which increases linearly with N, will render it practically useless, so some sort of truncation is necessary, and in this particular example, the regenerative structure provides such a mechanism. Using regenerative theory, the mean steady-state system time can be written as a ratio of expectations:

$$E[T]=\frac{E[Q]}{E[\eta]},$$

where η is the number of customers served in a busy period and Q is the sum of the system times of customers served in a busy period. Differentiation yields

$$\frac{dE[T]}{d\theta}=\frac{dE[Q]/d\theta}{E[\eta]}-\frac{dE[\eta]/d\theta}{E[\eta]}E[T].$$

Now, use the natural LR/SF estimators for each of the terms separately to obtain the following regenerative estimator over M busy periods:

$$\left(\frac{d\bar{T}_N}{d\theta}\right)_{LR}=\frac{1}{N}\sum_{m=1}^{M}\left\{\sum_{i=n_{m-1}+1}^{n_m}T_i\sum_{i=n_{m-1}+1}^{n_m}\frac{\partial\ln f_2(X_i;\theta)}{\partial\theta}\right\}-\frac{1}{N}\sum_{m=1}^{M}\left\{(n_m-n_{m-1})\sum_{i=n_{m-1}+1}^{n_m}\frac{\partial\ln f_2(X_i;\theta)}{\partial\theta}\right\}\frac{1}{N}\sum_{j=1}^{N}T_j.$$

The advantage of these estimators is that the summations are bounded by the length of the busy periods, so as long as the busy periods are not too lengthy, the variance of the estimators should be acceptable. Furthermore, the second derivative estimator is relatively easy to derive, as well:

$$\left(\frac{d^2\bar{T}_N}{d\theta^2}\right)_{LR}=\frac{1}{N}\sum_{m=1}^{M}\left\{\sum_{i=1}^{n_m}T_i\sum_{i=n_{m-1}+1}^{n_m}\left[\frac{\partial^2\ln f_2(X_i;\theta)}{\partial\theta^2}+\left(\frac{\partial\ln f_2(X_i;\theta)}{\partial\theta}\right)^2\right]\right\}-\frac{1}{N}\sum_{m=1}^{M}\left\{(n_m-n_{m-1})\sum_{i=n_{m-1}+1}^{n_m}\left[\frac{\partial^2\ln f_2(X_i;\theta)}{\partial\theta^2}+\left(\frac{\partial\ln f_2(X_i;\theta)}{\partial\theta}\right)^2\right]\right\}\frac{1}{N}\sum_{j=1}^{N}T_j.$$

The WD estimator is also relatively straightforward, just incorporating the product rule of differentiation as before:

$$\left(\frac{d\bar{T}_N}{d\theta}\right)_{WD} = c(\theta)\sum_{i=1}^{N}\left[\bar{T}_N(A_1,\ldots,A_N,\ldots,X_i^{(2)},\ldots) - \bar{T}_N(A_1,\ldots,A_N,\ldots,X_i^{(1)},\ldots)\right],$$

where $X_i^{(j)} \sim F_2^{(j)}, j = 1,2; i = 1,\ldots,N$ for $(c(\theta), F_2^{(1)}, F_2^{(2)})$ a weak derivative of F_2. A second derivative estimator would take exactly the same form, the only difference being that $(c(\theta), F_2^{(1)}, F_2^{(2)})$ should be a weak second derivative of F_2. Note that in general, implementation of the estimator requires $2T$ separate sample paths and resulting sample performance estimates, because the parameter appears in T input random variables. Although the variance of the estimator does not increase with T, implementation may not be practical for large T. However, in many cases, the expression can be simplified, making the computation more acceptable. The variance properties of the estimators depend heavily on the particular weak derivative(s) used.

What happens when T is random and when $T \longrightarrow \infty$ is also of theoretical interest. In these settings, the estimators are extended in the natural way, but proving theoretical properties becomes a more challenging task.

5.3 (s, S) Inventory System

We now consider the single-item periodic review (s, S) inventory system, in which once every period the inventory level is reviewed and, if necessary, orders are placed to replenish depleted inventory. An (s, S) ordering policy specifies that an order be placed when the level of inventory on hand plus that on order (called *inventory position*) falls below the level s, and that the amount of the order be the difference between S and the present inventory position, i.e., order amounts are placed "up to S." For expositional ease, we describe only the gradient estimate for average inventory level with respect to the policy parameters s and $q = S - s$, which do not enter through probability distributions as in the previous examples.

We consider the model where all excess demand is backlogged and eventually filled, and replenishment orders are immediately received (zero lead time), so that inventory level and inventory position coincide. We assume that during the period, demand is satisfied, and then the order replenishment decision is made at the end of the period. Let D_n be the demand in period n, which is assumed i.i.d. with respective density and distribution functions given by f and F, and let V_n be the inventory level in period n *after* demand satisfaction at the end of the period, but just *prior* to the order replenishment decision. This quantity satisfies a recursive equation somewhat analogous to the Lindley equation:

$$V_{n+1} = \begin{cases} V_n - D_{n+1} & \text{if } V_n \geq s, \\ S - D_{n+1} & \text{if } V_n < s. \end{cases} \tag{35}$$

The average inventory level over N periods is given by

$$\overline{V}_N = \frac{1}{N}\sum_{n=1}^{N} V_n.$$

From a sample path point of view, the key discrete event in the system is the ordering decision each period. A change in s, with q held fixed, has no effect on these decisions, so infinitesimal perturbations in s result in infinitesimal changes in the inventory level and in the sample performance function $\overline{V}_N$. In particular, for a perturbation of size Δs (of any size, not necessarily infinitesimal), $V_n(s + \Delta s) = V_n(s) + \Delta s$, and thus $\partial \overline{V}_N / \partial s = 1$ is an unbiased estimator for $\partial E[\overline{V}_N]/\partial s$. Intuitively, the shape of the sample path is unaltered by changes in s if q is held constant; the entire sample path is merely shifted by the size of the change (assuming starting inventory of $V_0 = S = s + q$; else, the statement only holds starting from the first order point). The IPA estimator can also be obtained by simply differentiating the recursive relationship (35), noting that D_n does not depend on s or q:

$$\frac{dV_{n+1}}{d\theta} = \begin{cases} \dfrac{dV_n}{d\theta} & \text{if } V_n \geq s, \\ 1 & \text{if } V_n < s, \end{cases}$$

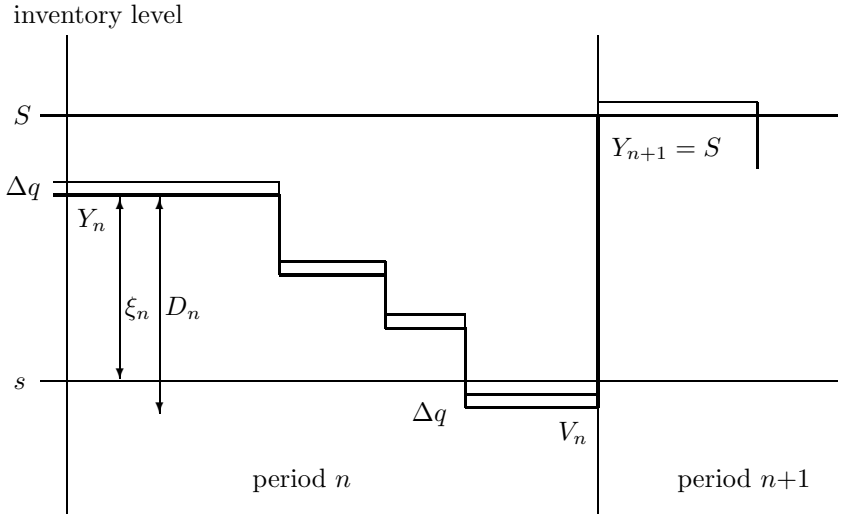

Figure 2: Quantities used in (s, S) inventory system.

for either θ=s or θ=q. Under the assumption that V_0=S=$s + q$, the expression reduces to 1 for all n when $\theta = s$.

On the other hand, a change in q with s held fixed may cause a change in the set of ordering decisions, resulting in a drastic change in the sample path and hence in the performance measure $\overline{V}_N$. Thus, SPA is required to derive an unbiased gradient estimator for $\theta = q$. We provide a brief informal derivation for a right-hand SPA estimator, i.e., $\Delta q > 0$, in which a period where a replenishment order was originally placed could become one where an order is not placed, for sufficiently large Δq (refer to Figure 2). To calculate the probability rate for such an event requires knowing the quantity above the reorder point s prior to demand being realized in the period, given by $\xi_n = Y_n - s$ in Figure 2, where Y_n is the inventory position/level prior to demand being realized. Conditioning on all demands except the one in the period of interest (in which an order is placed), D_n, the corresponding probability rate is then given by

$$\lim_{\Delta q \to 0} \frac{P(\xi_n + \Delta q \geq D_n | \xi_n < D_n)}{\Delta q} = \frac{f(\xi_n)}{1 - F(\xi_n)}.$$

The resulting conditional expected effect on the performance measure requires some further analysis, and is omitted here, and we simply give the final SPA estimator for $\partial E[\overline{V}_N]/\partial q$, which can be easily and efficiently estimated from the original sample path:

$$1 + \frac{1}{N} \sum_{n \leq N: V_n < s} \frac{f(\xi_n)}{1 - F(\xi_n)} \left[s - E[D] - \overline{V}_N \right],$$

where $E[D]$ is mean demand.

The LR/SF and WD methods require a change of variables in order to move the parameters into the distribution, which requires some non-trivial analysis; see Pflug and Rubinstein (2002) for details.

6 Basic Theoretical Tools

The key result used in the theoretical proofs of unbiasedness is the (Lebesgue) dominated convergence theorem that allows for the exchange of limit and expectation operators required in Equations (12) or (13):

Theorem 1 (Dominated Convergence Theorem) If $\lim_{n\to\infty} g_n = g$ a.s. and $|g_n| \leq M \ \forall n$ a.s. with $E[M] < \infty$, then $\lim_{n\to\infty} E[g_n] = E[g]$.

Take $\Delta\theta \to 0$ instead of $n \to \infty$, and g is the gradient estimator, so $g_{\Delta\theta}$ is the limiting sequence that defines the sample (path) gradient. Verifying that an actual bound exists is often a non-trivial task in applications, especially in the case of perturbation analysis.

Looking at (12), we translate these conditions to

$$g_{\Delta\theta} = \frac{Y(\theta + \Delta\theta) - Y(\theta)}{\Delta\theta},$$

$$g_{\Delta\theta} = Y(x)\frac{f(x;\theta + \Delta\theta) - f(x;\theta)}{\Delta\theta},$$

for IPA and LR/SF, respectively.

For IPA, the following generalization of the mean value theorem is most useful:

Theorem 2 (Generalized Mean Value Theorem) Let Y be a continuous function that is differentiable on a compact set $\tilde{\Theta} = \Theta\backslash\tilde{D}$, where $\tilde{D}$ is a set of countably many points. Then, $\forall\theta, \theta + \Delta\theta \in \Theta$,

$$\left|\frac{Y(\theta + \Delta\theta) - Y(\theta)}{\Delta\theta}\right| \leq \sup_{\theta\in\tilde{\Theta}} \left|\frac{dY}{d\theta}\right|.$$

If $Y(\theta)$ can be shown to be continuous and piecewise differentiable on Θ w.p.1, then it just remains to show

$$E\left[\sup_{\theta\in\tilde{\Theta}} \left|\frac{dY}{d\theta}\right|\right] < \infty,$$

to satisfy the conditions required for unbiasedness via the dominated convergence theorem. Basically, in order for the chain rule to be applicable, the sample performance function needs to be continuous with respect to the underlying random variable(s). This translates into requirements on the form of the performance measure and on the dynamics of the underlying stochastic system.

For the LR/SF method, the bound is applied to the (joint) density (or mass) function. Note that the bound is for $f(x;\theta)$ with respect to the parameter θ and not its usual argument. For WD, the required interchange is guaranteed by the definition of the weak derivative, but the sample performance must be in the set of "test" functions $\mathcal{L}$ in the definition, which again generally requires the dominated convergence theorem (or uniform integrability, which is usually difficult to check directly).

The previous examples can be used to show in very simple cases where difficulties arise.

Consider the p.d.f.

$$f(x;\theta) = \frac{1}{\theta}\mathbf{1}\{0 < x < \theta\},$$

where the LR/SF method does not apply. In this case, f viewed as a function of θ for fixed x has a discontinuity at $\theta = x$.

Consider the function

$$P(Y > y) = E[\mathbf{1}\{Y > y\}],$$

in which IPA will not work. In this case, the performance measure is an indicator function, which is discontinuous in its argument.

Thus, in both simple examples, the dominated convergence theorem is not applicable as the required quantity cannot be bounded. However, it is only a sufficient condition, not necessary, so in some (very) special cases, unbiasedness may hold even without the boundedness (continuity).

7 Simple Guidelines for the Simulation Practitioner

We summarize the most important considerations in applying the three direct gradient procedures (PA, LR/SF, WD):

- IPA: cannot handle "bad performance measures, e.g., indicator functions, or non-smooth systems; one way of checking the latter is the commuting condition, which checks event sequences in the system (cf. Glasserman 1991); smoothness may depend on system representation (see discussion on role of representations);
- SPA: choice of what to condition on, and how to compute (estimate) resulting conditional expectation; may require many additional simulations;
- LR/SF and WD: may encounter difficulties in handling parameters not in distribution (so-called "structural" parameters); may need to try a change of variables;
- LR/SF: if the parameter appears in an input distribution that is reused frequently (parameter makes too many repeat appearances), e.g., i.i.d. service times in a queueing system, then may need to find a way to truncate the estimator to mitigate the linear increase in variance;
- WD: choice of which (non-unique) WD representation to use; also high-dimensional vectors may require many simulations;
- For discrete distributions, IPA works if the parameter occurs in the support *values*, whereas LR/SF and WD work if the parameter occurs in the support *probabilities*;
- Higher derivative estimates are usually easiest to derive using LR/SF, but even then the variance of the resulting estimators may be problematic.

The characteristics/choices in applying SPA are strikingly similar to the use of conditional Monte Carlo for variance reduction. The choice of what to condition on in applying SPA also has analogs to the WD representation choice. The simplest procedures to use are the LR/SF and IPA estimators. The WD estimator may be easier to apply than SPA, because one has a set of "standard" choices for a large class of distributions. However, there is no guarantee that these choices are necessarily good for a particular application.

It should be clear from this brief summary that the direct gradient estimation techniques may require some effort on the part of the simulation user, whereas the indirect techniques are straightforward to apply. To put it another way, for direct gradient estimation sometimes it takes some art to do the science.

8 Applications

In discussing applications, the focus is on the direct gradient estimators, since application of the indirect gradient estimates is generally more straightforward and domain independent. The most dominant application of direct gradient estimation in the research literature is clearly queueing systems; however, inventory management and financial engineering have employed many of these results in real-world practice.

For derivatives with respect to distributional parameters, IPA can be applied in any single-class Jackson-like network. Here, "Jackson-like" (also called a "generalized Jackson network") means that the network retains all of the usual characteristics of a Jackson network except that service times and interarrival times (in an open network) do not have to be exponentially distributed. This is also true for the LR/SF method and the WD method. In the latter case, there are often choices of the WD used. However, if the parameter is the routing probabilities, then IPA cannot be applied directly. Also, if the network is extended to multiple classes of customers, then IPA is often not applicable.

It is relatively straightforward to apply the LR/SF method to queueing systems, the only caveat being the potential variance problem mentioned in the previous section. It is also relatively straightforward to derive weak derivative estimators, but there are generally many choices of distributions, and often one or more additional simulations using different input distributions will be necessary.

Inventory systems is a domain in which direct gradient estimation has been successfully applied in real-world applications. Because the parameters of interest are usually those controlling replenishment decisions

as in the (s, S) inventory control example of Section 5, PA is most relevant. For so-called base-stock policies, often IPA suffices. Such IPA estimators are being used to optimize inventory management in the worldwide supply chain of Caterpillar, and the success of the approach is reported in a *Fortune* magazine article, "New Victories in the Supply Chain Revolution" (by Philip Siekman, October 30, 2000); technical details of the approach can be found in Kapuscinski and Tayur (1999). For more complicated ordering policies such as an (s, S) policy that comes about with the inclusion of a fixed order cost, extensions of IPA are required; see Fu and Healy (1992, 1997), Fu (1994b), Bashyam and Fu (1994), Fu and Hu (1994), Zhang and Fu (2005), Pflug and Rubinstein (2002) for an LR/SF estimator using the push-out method with conditioning, and especially Chapter 7 of Fu and Hu (1997).

Because on Wall Street and elsewhere in the global financial world, Monte Carlo simulation is routinely used for pricing and hedging derivatives, it is now commonly included in any finance text that addresses numerical solution methods. Simulation can easily handle the pricing of high-dimensional derivatives, such as path-dependent claims or systems with a large number of underlying assets and/or uncertainties (e.g., stochastic volatility and interest rate models). In hedging, gradients are the most critical ingredients in determining what positions need to be taken in any portfolio. In equities, they are usually referred to as "Greeks," because they are represented by Greek letters. For example, the most common one is the Delta, which is the sensitivity of a derivative price with respect to the underlying security price, e.g., the price of a call option with respect to the underlying stock price on which the contract is written. Delta hedging and other types of hedging are described in most elementary finance textbooks on derivatives such as Hull (2000), and the text by Clewlow and Strickland (1999) includes simple IPA estimators for calculating Greeks using Monte Carlo simulation (though they do not use the term IPA nor stochastic gradient estimation). Fu and Hu (1995) and Broadie and Glasserman (1996) were the first to develop stochastic gradients in these settings; see also Glasserman (2004). Heidergott (2001b) considered weak derivatives in a similar setting as Fu and Hu (1995). In fixed income securities, the analogous quantities go by the name of duration and convexity. In Chen and Fu (2002), IPA is applied to the pricing and hedging of mortgage-backed securities, where both first and second derivatives are required. For just five parameters, if symmetric differences were used, this would require 236 (1 + 10 + 225) simulations for each set of performance measures and gradient estimates. In actual implementation, there was a resulting dramatic 97.2% reduction in computation time. Another important finance application is the pricing of American-style derivatives: contingent claims in which early exercise is permitted. One of the earliest proposed approaches to this problem was to parameterize the early exercise boundary, thus casting the optimal stopping problem as a stochastic optimization problem with respect to the parameters. The earliest example of applying PA and SA to such an option pricing problem is given in Fu and Hu (1995). Earlier editions of Hull (2000) and other finance texts claimed that simulation could be used only to price European options, so this is one of many approaches that dispelled that belief. See Glasserman (2004) for other approaches and references on this topic.

Other applications include stochastic activity networks (see Rubinstein and Shapiro 1993 for the LR/SF method, Bowman 1994 for IPA, and Fu 2005 for SPA and WD); preventive maintenance (Fu, Hu, and Shi 1993; Heidergott 1999, 2001a; L'Ecuyer, Martin, and Vazquez-Abad 1999); statistical process control (Fu and Hu 1999); traffic light signal control (Howell and Fu 2003; also mentioned in Rubinstein and Shapiro 1993, p.3, as an example).

9 Probing Further

Other approaches not treated here include frequency domain experimentation (Schruben and Cogliano 1981; Schruben 1986; Jacobson, Buss, and Schruben 1991; Jacobson 1994); and Malliavin calculus, which has been used primarily in financial applications (e.g., Fournié et al. 1999, Benhamou 2002, and references therein).

For gradient-based simulation optimization, further discussion can be found in the papers of Fu (2002), Andradóttir (1998), Fu (1994a), and Jacobson and Schruben (1989); see also the books by Rubinstein and Shapiro (1993), Pflug (1996), Fu and Hu (1997), and Spall (2003), as well as Fu (2001b). Both of the well-known simulation textbooks by Law and Kelton (2000) and Banks et al. (2000) also devote sections to the topic, but the latter does not discuss gradient estimation, per se. Applications to the single-server queue example considered here include the theoretical convergence results of Fu (1990), Chong and Ramadge (1992), and L'Ecuyer and Glynn (1994), and the in-depth numerical comparisons of L'Ecuyer, Girouz, and Glynn (1994) and Andradóttir (1998). Andradóttir (1996) considers the more general setting of using the LR/SF

estimators in SA algorithms, and Tang, L'Ecuyer, and Chen (1999) analyze the asymptotic efficiency of an averaging version of SA using perturbation analysis estimators. Work on sample path optimization (called the stochastic counterpart method in Rubinstein and Shapiro 1993) includes Plambeck et al. (1996), Robinson (1996), Gürkan, Özge, and Robinson (1999), Homem-de-Mello, Shapiro, and Spearman (1999); Dussault et al. (1997) combines the approach with stochastic approximation. See also the chapter by Shapiro (2003) on using Monte Carlo methods for stochastic programming. The original papers on stochastic approximation are Robbins and Monro (1951) and Kiefer and Wolfowitz (1952). For a more recent general book on stochastic approximation, see Kushner and Yin (1997); L'Ecuyer and Yin (1998) discusses convergence rates as a function of computational budget. Spall (2000) provides both indirect and direct gradient-based SA methods for obtaining near-optimal or optimal convergence rates via stochastic analogues to the deterministic Newton-Raphson algorithm, using Hessian matrix estimation. SPSA was introduced by Spall (1992), although the random directions method was proposed in Kushner and Clark (1978); see `http://www.jhuapl.edu/SPSA/` for an extensive annotated bibliography. Application to the settings of this handbook is described in Fu and Hill (1997), and using deterministic sequences for SPSA is considered in Bhatnagar et al. (2003) and Xiong, Wang, and Fu (2002). Andradóttir (1998) contains further useful discussion on the application of SA algorithms in simulation optimization, especially regarding the issues of choosing the gain sequence and modifying the algorithm in cases where the objective function is not well-behaved. Most stochastic approximation algorithms involve known constraints; Bashyam and Fu (1998) consider the case of a noisy constraint that must also be estimated, in particular a service level constraint for an (s, S) inventory system. Studies on choosing the finite difference used in the Kiefer-Wolfowitz version of SA include Zazanis and Suri (1993), and L'Ecuyer and Perron (1994).

For perturbation analysis, the books by Glasserman (1991), Ho and Cao (1991), and Cao (1994) treat IPA extensively (in particular, see Glasserman 1991 for a complete treatment of the commuting condition; see also Cao 1985 and Heidelberger et al. 1988) and SPA to some extent, whereas the book by Fu and Hu (1997) is a comprehensive treatment of SPA, which was introduced by Gong and Ho (1987) and Suri and Zazanis (1988); see also Fu and Hu (1992), Fu (2001a), and Fu and Hu (1991, 1993), where higher derivatives for multi-server queues are treated. Many of the examples here are adopted from Fu and Hu (1997). More recent work connecting IPA with Markov decision processes using the idea of potentials, which also relates to the LR/SF method, include Cao (2000, 2003ab). The field of perturbation analysis for gradient estimation includes numerous other extensions and variations on IPA not discussed here, including rare perturbation analysis (Brémaud and Vázquez-Abad 1992); structural IPA (Dai and Ho 1995); discontinuous perturbation analysis (Shi 1996); and augmented IPA (Gaivoronski, Shi, and Sreenivas 1992). Seminal papers applying perturbation analysis for estimating the effects of *finite* perturbations in the parameter include Ho and Li (1988), Cassandras and Strickland (1989), and Vakili (1991).

For the likelihood ratio or score function method, a good reference is the book by Rubinstein and Shapiro (1993), which also includes some discussion of IPA (Chapter 5 on the "push in" method), as well as both first and second derivative estimators for most of the entries in Table 1 (Section 2.2); see also Reiman and Weiss (1989), Glynn (1990), and Andradóttir (1996). The "push out" method for handling structural parameters was introduced in Rubinstein (1992); see also Section 2.5.4 in Rubinstein and Shapiro (1993). Using conditional expectation to reduce variance in the LR/SF method was considered in McLeish and Rollins (1992); see also Section 3.4 in Rubinstein and Shapiro (1993). A "unified" view of IPA and LR/SF is presented in L'Ecuyer (1990), which allows the parameter to appear in both the sample performance and the distribution (see also L'Ecuyer 1995 for further discussion and some technical corrections).

The weak derivative method was introduced by Pflug (1989, 1990, 1996). More recent work attempting to unify the approach with others such as rare perturbation analysis and smooth perturbation analysis, includes Heidergott and Vazquez-Abad (2000, 2001), in the setting of general Markov processes, which may not provide as convenient a framework for the discrete-event simulation setting as generalized semi-Markov processes (GSMPs). Derivations for many of the entries in Table 1 can be found in Heidergott, Pflug, and Vazquez-Abad (2003).

10 Future Research Directions

Stochastic gradient estimation research has matured over the last decade to the point that much of the analysis has become theoretical rather than algorithmic. This line of research addresses many issues that

also arise in traditional steady-state output analysis – only now for stochastic gradient estimators – and can involve advanced probabilistic tools. In contrast, two possible directions for future research described briefly here are more motivated by simulation practice and have much algorithmic room left in them to grow.

How is gradient estimation best employed in simulation optimization? One recent development is the use of fluid models in this setting. A discrete-event simulation model for which application of the direct derivative estimation may be difficult is approximated by a stochastic fluid model, which is used to derive IPA estimators that are then implemented on either the (simulated) stochastic fluid model or sometimes on the original stochastic discrete-event simulation model. In the former case, where the IPA estimates are generally unbiased, the optimal settings often provide a good approximation for the optimum in the original model. In the latter case, the new IPA estimators do not yield unbiased estimators when implemented in the discrete-event model (so are generally not useful in sensitivity analysis, per se), but they often provide a good approximation of the zero gradient location when used in optimization; however, the theoretical basis for this good fortune is still not fully understood. Some papers in this area include Wardi et al. (2002), Cassandras et al. (2002), Sun et al. (2004), and Panayiotou, Howell, and Fu (2005).

The LR/SF method is closely related to importance sampling. Investigating this connection more thoroughly for variance reduction purposes, and doing so in a more general stochastic gradient setting, would be of benefit to simulationists. Some work along this line in the setting of stochastic programming is contained in Shapiro and Homem-de-Mello (1998).

Acknowledgments

This work was supported in part by the National Science Foundation under Grants DMI 9988867 and DMI 0323220, and by the Air Force Office of Scientific Research under Grants F496200110161 and FA95500410210. The author thanks Shane Henderson and Barry Nelson for their detailed comments and suggestions that have led to an improved exposition.

References

Andradóttir, S., 1996. Optimization of the transient and steady-state behavior of discrete event systems, *Management Science* 42, 717-737.

Andradóttir, S., 1998. Simulation optimization, Chapter 9 in J. Banks, ed., *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John Wiley & Sons, New York.

Banks, J., J.S. Carson, B.L. Nelson, and D.M. Nicol, 2000. *Discrete Event Systems Simulation*, 3rd edition, Prentice Hall, Englewood Cliffs, NJ.

Bashyam, S. and M.C. Fu, 1994. Application of perturbation analysis to a class of periodic review (s, S) inventory systems, *Naval Research Logistics* 41, 47-80.

Bashyam, S. and M.C. Fu, 1998. Optimization of (s, S) inventory systems with random lead times and a service level constraint, *Management Science* 44, S243-S256.

Benhamou, E., 2002. Smart Monte Carlo: various tricks using Malliavin calculus, *Quantitative Finance* 2, 329-336.

Bhatnagar, S., M.C. Fu, S.I. Marcus, and I.J. Wang, 2003. Two-timescale simultaneous perturbation stochastic approximation using deterministic perturbation sequences, *ACM Transactions on Modeling and Computer Simulation*, 180-209.

Bowman, R.A., 1994. Stochastic gradient-based time-cost tradeoffs in PERT network using simulation, *Annals of Operations Research* 53, 533-551, 1994.

Brémaud, P. and F.J. Vázquez-Abad, 1992. On the pathwise computation of derivatives with respect to the rate of a point process: The phantom RPA method, *Queueing Systems: Theory and Applications* 10, 249-270.

Broadie, M. and P. Glasserman, 1996. Estimating security price derivatives using simulation, *Management Science* 42, 269-285.

Cao, X.R., 1994. *Realization Probabilities: The Dynamics of Queuing Systems*, Springer-Verlag, New York.

Cao, X.R., 1985. Convergence of parameter sensitivity estimates in a stochastic experiment, *IEEE Transactions on Automatic Control* 30, 834-843.

Cao, X.R., 2000. A unified approach to Markov decision problems and performance sensitivity analysis, *Automatica* 36, 771-774.

Cao, X.R., 2003a. From perturbation analysis to Markov decision processes and reinforcement learning, *Discrete Event Dynamic Systems: Theory and Applications* 13, 9-39.

Cao, X.R., 2003b. Semi-Markov decision problems and performance sensitivity analysis, *IEEE Transactions on Automatic Control* 48, 758-769.

Cassandras, C.G. and S.G. Strickland, 1989. On-line sensitivity analysis of Markov chains, *IEEE Transactions on Automatic Control* 34, 76-86.

Cassandras, C.G., Y. Wardi, B. Melamed, G. Sun, and C.G. Panayiotou, 2002. Perturbation analysis for on-line control and optimization of stochastic fluid models, *IEEE Transactions on Automatic Control* 47, 1234-1248.

Chen, J. and M.C. Fu, 2002. Hedging beyond duration and convexity, *Proceedings of the 2002 Winter Simulation Conference*, 1593-1599.

Chong, E.K.P. and P.J. Ramadge, 1992. Convergence of recursive optimization algorithms using infinitesimal perturbation analysis, *Discrete Event Dynamic Systems: Theory and Applications* 1, 339-372.

Clewlow, L.J. and C.R. Strickland, 1999. *Implementing Derivative Models*, Wiley.

Dai, L. and Y.C. Ho, 1995. Structural infinitesimal perturbation analysis for derivative estimation in discrete event dynamic systems, *IEEE Transactions on Automatic Control* 40, 1154-1166.

Dussault, J.P., D. Labrecque, P. L'Ecuyer, and R. Y. Rubinstein, 1997. Combining the stochastic counterpart and stochastic approximation methods, *Discrete Event Dynamic Systems: Theory and Applications* 7, 5-28.

Fournié, E., J.M. Lasry, J. Lebuchoux, P.L. Lions, and N. Touzi, 1999. Applications of Malliavin calculus to Monte Carlo methods in finance, *Finance and Stochastics* 3, 391-412.

Fu, M.C., 1990. Convergence of a stochastic approximation algorithm for the GI/G/1 queue using infinitesimal perturbation analysis, *Journal of Optimization Theory and Applications* 65, 149-160.

Fu, M.C., 1994a. Optimization via simulation: A review. *Annals of Operations Research* 53, 199-248.

Fu, M.C., 1994b. Sample path derivatives for (s, S) inventory systems, *Operations Research* 42, 351-364.

Fu, M.C., 2001a. Perturbation analysis, *Encyclopedia of Operations Research and Management Science*, 2nd edition, S. Gass and C. Harris, eds., Kluwer Academic Publishers, 608-611.

Fu, M.C., 2001b. Simulation optimization, *Encyclopedia of Operations Research and Management Science*, 2nd edition, S. Gass and C. Harris, eds., Kluwer Academic Publishers, 756-759.

Fu, M.C., 2002. Optimization for simulation: Theory vs. practice (Feature Article), *INFORMS Journal on Computing* 14, 192-215.

Fu, M.C., 2005. Sensitivity analysis for stochastic activity networks, working paper; also submitted to the 2005 International Conference on Automatic Control and System Engineering.

Fu, M.C. and K. Healy, 1992. Simulation optimization of (s, S) inventory systems, *Proceedings of the Winter Simulation Conference*, 506-514.

Fu, M.C. and K.J. Healy, 1997. Techniques for simulation optimization: An experimental study on an (s, S) inventory system. *IIE Transactions* 29, 191-199.

Fu, M.C. and S.D. Hill. 1997. Optimization of discrete event systems via simultaneous perturbation stochastic approximation, *IIE Transactions* 29, 233-243.

Fu, M.C. and J.Q. Hu, 1991. On choosing the characterization for smoothed perturbation analysis, *IEEE Transactions on Automatic Control* 36, 1331-1336.

Fu, M.C. and J.Q. Hu, 1992. Extensions and generalizations of smoothed perturbation analysis in a gener-

alized semi-Markov process framework, *IEEE Transactions on Automatic Control* 37, 1483-1500.

Fu, M.C. and J.Q, Hu, 1993. Second derivative sample path estimators for the GI/G/m Queue, *Management Science* 39, 359-383.

Fu, M.C. and J.Q. Hu, 1994. (s, S) inventory systems with random lead times: Harris recurrence and its implications in sensitivity analysis, *Probability in the Engineering and Informational Sciences* 8, 355-376.

Fu, M.C. and J.Q. Hu, 1995. Sensitivity analysis for Monte Carlo simulation of option pricing. *Probability in the Engineering and Information Sciences* 9, 417-446.

Fu, M.C. and J.Q. Hu, 1997. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*, Kluwer Academic, Boston, MA.

Fu, M.C. and J.Q. Hu, 1999. Efficient design and sensitivity analysis of control charts using Monte Carlo simulation, *Management Science* 45, 395-413.

Fu, M.C., J.Q. Hu, and L. Shi, 1993. An application of perturbation analysis to replacement problems in maintenance, *Proceedings of the 1993 Winter Simulation Conference*, 329-337.

Gaivoronski, A., L. Shi, and R.S. Sreenivas, 1992. Augmented infinitesimal perturbation analysis: an alternate explanation, *Discrete Event Dynamic Systems: Theory and Applications* 2, 121-138.

Glasserman, P., 1991. *Gradient Estimation via Perturbation Analysis*, Kluwer Academic, Boston, MA.

Glasserman, P., 2004. *Monte Carlo Methods in Financial Engineering*, Springer, New York.

Glynn, P.W., 1990. Likelihood ratio gradient estimation for stochastic systems, *Communicatios of the ACM* 33, 75-84.

Gong, W.B. and Y.C. Ho, 1987. Smoothed perturbation analysis of discrete-event dynamic systems, *IEEE Transactions on Automatic Control* 32, 858-867.

Gürkan, G., A.Y. Özge, and S.M. Robinson, 1999. Sample-path solution of stochastic variational inequalities. *Mathematical Programming* 84, 313-333.

Heidelberger, P., X.R. Cao, M. Zazanis, and R. Suri, 1988. Convergence properties of infinitesimal perturbation analysis estimates, *Management Science* 34, 1281-1302.

Heidergott, B., 1999. Optimization of a single-component maintenance system: A smoothed perturbation analysis approach, *European Journal of Operational Research*, 181-190.

Heidergott, B., 2001a. A weak derivative approach to optimization of threshold parameters in a multi-component maintenance system. *Journal of Applied Probability* 38, 386-406.

Heidergott, B., 2001b. Option pricing via Monte Carlo simulation: A weak derivative approach. *Probability in Engineering and Informational Sciences* 15, 335-349.

Heidergott, B., G. Pflug, and F. Vázquez-Abad, 2003. Measure-valued differentiation for stochastic systems: From simple distributions to Markov chains, manuscript; available at `http://staff.feweb.vu.nl/bheidergott/`.

Heidergott, B. and F. Vázquez-Abad, 2000. Measure-valued differentiation for stochastic processes: the finite horizon case, EURANDOM Report 2000-033.

Heidergott, B. and F. Vázquez-Abad, 2001. Measure-valued differentiation for stochastic processes: the random horizon case, GERAD Report G-2001-18.

Ho, Y.C. and X.R. Cao, 1991. *Discrete Event Dynamics Systems and Perturbation Analysis*, Kluwer Academic, Boston, MA.

Ho, Y.C. and S. Li, 1988. Extensions of infinitesimal perturbation analysis, *IEEE Transactions on Automatic Control* 33, 827-838.

Homem-de-Mello, T., A. Shapiro, and M.L. Spearman, 1999. Finding optimal material release times using simulation based optimization, *Management Science* 45, 86-102.

Howell, W.C. and M.C. Fu, 2003. Application of perturbation analysis to traffic light signal timing, *Proceedings of the 42nd IEEE Conference on Decision and Control*, 4837-4840.

Hull, J.C., 2000. *Options, Futures, and Other Derivative Securities*, 4th edition, Prentice Hall, Englewood Cliffs, NJ.

Jacobson, S.H. and L.W. Schruben, 1989. A review of techniques for simulation optimization. *Operations Research Letters* 8, 1-9.

Jacobson, S.H., 1994. Convergence results for harmonic gradient estimators, *ORSA Journal on Computing* 6, 381-397.

Jacobson, S.H., A. Buss, and L.W. Schruben, 1991. Driving frequency selection for frequency domain simulation experiments, *Operations Research* 39, 917-924.

Kapuscinski, R. and S.R. Tayur, 1999. Optimal policies and simulation based optimization for capacitated production inventory systems, Chapter 2 in S.R. Tayur, R. Ganeshan, M.J. Magazine, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic, Boston, MA.

Kiefer, J. and J. Wolfowitz, 1952. Stochastic estimation of the maximum of a regression function, *Annals of Mathematical Statistics* 23, 462-466.

Kushner, H.J. and D.C. Clark, 1978. *Stochastic Approximation Methods for Constrained and Unconstrained Systems,* Springer-Verlag, New York.

Kushner, H.J., and G.G. Yin, 1997. *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York.

Law, A.M. and W.D. Kelton, 2000. *Simulation Modeling and Analysis*, 3rd edition, McGraw-Hill, New York.

L'Ecuyer, P., 1990. A unified view of the IPA, SF, and LR gradient estimation techniques, *Management Science* 36, 1364-1383.

L'Ecuyer, P., 1995. On the interchange of derivative and expectation for likelihood ratio derivative estimators, *Management Science* 41, 738-748.

L'Ecuyer, P. and P. W. Glynn, 1994. Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady-state, *Management Science* 40, 1562-1578.

L'Ecuyer, P., N. Giroux, and P. W. Glynn, 1994. Stochastic optimization by simulation: Numerical experiments with a simple queue in steady-state, *Management Science*, **40** 1245-1261.

L'Ecuyer, P., B. Martin, and F. Vázquez-Abad, 1999. Functional estimation for a multicomponent age-replacement model, *American Journal of Mathematical and Management Sciences* 19, 135-156.

L'Ecuyer, P. and G. Perron, 1994. On the convergence rates of IPA and FDC derivative estimators, *Operations Research* 42, 643-656.

L'Ecuyer, P. and G. Yin, 1998. Budget-dependent convergence rate of stochastic approximation, *SIAM Journal on Optimization* 8, 217-247.

McLeish, D.L. and S. Rollins, 1992. Conditioning for variance reduction in estimating the sensitivity of simulations, *Annals of Operations Research* 39, 157-173.

Panayiotou, C.G., W.C. Howell, and M.C. Fu, 2005. Online traffic light control through gradient estimation using stochastic fluid models, *Proceedings of the IFAC 16th Triennial World Congress*.

Pflug, G.C., 1989. Sampling derivatives of probabilities, *Computing* 42, 315-328.

Pflug, G.C., 1990. On-line optimization of simulated Markovian processes, *Mathematics of Operations Research* 15, 381-395.

Pflug, G.C., 1996. *Optimization of Stochastic Models*, Kluwer Academic, Boston, MA.

Pflug, G.C. and Rubinstein R.Y., 2002. Inventory processes: Quasi-regenerative property, performance evaluation and sensitivity estimation via simulation, *Stochastic Models* 18, 469-496.

Plambeck, E.L, B.-R. Fu, S.M. Robinson, and R. Suri, 1996. Sample-path optimization of convex stochastic performance functions. *Mathematical Programming* 75, 137-176.

Robinson, S.M., 1996. Analysis of sample-path optimization. *Mathematics of Operations Research* 21,

513-528.

Reiman, M.I. and A. Weiss, 1989. Sensitivity analysis for simulations via likelihood ratios, *Operations Research* 37, 830-844.

Robbins, H. and S. Monro, 1951. A stochastic approximation method, *Annals of Mathematical Statistics* 22, 400-407.

Rubinstein, R.Y., 1989. Sensitivity analysis of computer simulation models via the score efficient, *Operations Research* 37, 72-81.

Rubinstein, R.Y., 1992. Sensitivity analysis of discrete event systems by the 'push out' method, *Annals of Operations Research* 39, 229-251.

Rubinstein, R.Y. and A. Shapiro, 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, John Wiley & Sons, New York.

Schruben, L.W., 1986. Simulation optimization using frequency domain methods, *Proceedings of the Winter Simulation Conference*, 366-369.

Schruben, L.W. and V. J. Cogliano, 1981. Simulation sensitivity analysis: A frequency domain approach, *Proceedings of the Winter Simulation Conference*, 455-459.

Shapiro, A., 2003. Monte Carlo Sampling Methods, Chapter 6 in A. Ruszczynski and A. Shapiro, eds., *Stochastic Programming, Handbook in Operations Research and Management Science*, Elsevier.

Shapiro, A. and T. Homem-de-Mello, 1998. A simulation-based approach to two-stage stochastic programming with recourse, *Mathematical Programming* 81, 301-325.

Shi, L.Y., 1996. Discontinuous Perturbation Analysis of Discrete Event Dynamic Systems, *IEEE Transactions on Automatic Control* 41, 1676-1681.

Spall, J.C., 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control* 37, 332-341.

Spall, J.C., 2000. Adaptive stochastic approximation by the simultaneous perturbation method, *IEEE Transactions on Automatic Control* 45, 1839-1853.

Spall, J.C., 2003. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control,* Wiley, Hoboken, NJ.

Sun, G., Cassandras, C.G., Wardi, Y., Panayiotou, C.G., and Riley, G.F., 2004. Perturbation analysis and optimization of stochastic flow networks, *IEEE Transactions on Automatic Control* 49, 2143-2159.

Suri, R. and M.A. Zazanis, 1988. Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue. *Management Science* 34, 39-64.

Tang, Q.-Y., P. L'Ecuyer, and H.-F. Chen, 1999. Asymptotic efficiency of perturbation analysis-based stochastic approximation with averaging, *SIAM Journal on Control and Optimization* 37, 1822-1847.

Vakili, P., 1991. Using a standard clock technique for efficient simulation, *Operations Research Letters* 10, 445-452.

Wardi, Y., B. Melamed, C.G. Cassandras, and C.G. Panayiotou, 2002. IPA gradient estimators in single-node stochastic fluid models, *Journal of Optimization Theory and Applications* 115, 369-406.

Xiong, X., I.J. Wang, and M.C. Fu, 2002. Randomized-direction stochastic approximation algorithms using deterministic perturbation sequences, *Proceedings of the 2002 Winter Simulation Conference*, 285-291.

Zazanis, M.A. and Suri, R., 1993. Convergence rates of finite-difference sensitivity estimates for stochastic systems, *Operations Research* 41, 694-703.

Zhang, H. and M.C. Fu, 2005. Sample path derivatives for (s, S) inventory systems with price determination, *The Next Wave in Computing, Optimization, and Decision Technologies*, Bruce L. Golden, S. Raghavan, Edward A. Wasil, editors, Kluwer Academic Publishers, 229-246.