Sensitivity Analysis in Monte Carlo Simulation of Stochastic Activity Networks

Michael C. Fu

Robert H. Smith School of Business & Institute for Systems Research
Department of Decision and Information Technologies
University of Maryland
College Park, MD 20742
mfu@rhsmith.umd.edu

Summary. Stochastic activity networks (SANs) such as those arising in Project Evaluation Review Technique (PERT) and Critical Path Method (CPM) are an important classical set of models in operations research. We focus on sensitivity analysis for stochastic activity networks when Monte Carlo simulation is employed. After a brief aside reminiscing on Saul's influence on the author's career and on the simulation community, we review previous research for sensitivity analysis of simulated SANs, give a brief summary overview of the main approaches in stochastic gradient estimation, derive estimators using techniques not previously applied to this setting, and address some new problems not previously considered. We conclude with some thoughts on future research directions.

Key words: Stochastic activity networks; PERT; CPM; project management; Monte Carlo simulation; sensitivity analysis; derivative estimation; perturbation analysis; likelihood ratio/score function method; weak derivatives.

1 Introduction

In the vast toolkit of operations research (OR), two of the most useful methods/models without a doubt are simulation and networks. Numerous surveys of practitioners consistently place these in the top 10 in terms of applicability to real-world problems and solutions. Furthermore, there is a large industry of supporting software for both domains. Thus, almost all degree-granting programs in operations research offer standalone courses covering these two topics. At the University of Maryland's Robert H. Smith School of Business, these two courses at the Ph.D. level have course numbers BMGT831 and BMGT835. They form half of the core of the methodological base that all OR Ph.D. students take, along with BMGT834 Stochastic Models and BMGT830 Linear programming, which Saul Gass taught for much of his career at Maryland, using his own textbook first published in 1958, and translated into numerous other languages, including Polish, Russian, and Spanish. Reflecting back on my own academic career, the first presentation I ever gave at a technical conference was at the 1988 ORSA/TIMS Spring National Meeting in Washington, D.C., in which Saul was the General Chair. When the INFORMS Spring Meeting returned to Washington, D.C. in 1996, Saul was again on the advisory board and delivered the plenary address, and this time, through my connection with him, I served on the Program Committee, in charge of contributed papers.

Saul's contributions to linear programming and his involvement in ORSA and then INFORMS are of course well known, but what may not be as well known are his contributions to the simulation community and simulation research. Saul was heavily involved with the Winter Simulation Conference — the premier annual meeting of stochastic discrete-event simulation researchers, practitioners, and software vendors — during what could be called the formative and critical years of the conference. Saul served as the ORSA representative on the Board of Directors in the early 1980s. During these years, "he contributed much insight into the operation of conferences and added prestige" [21], which clearly helped launch these meetings on the path to success. In addition, Saul has also contributed to simulation research in model evaluation and validation through "(i) the development of a general methodology for model evaluation and validation [15, 18], and (ii) the development of specific validation techniques that used quantitative approaches [16, 17]" [21]. Arjang Assad's article in this volume details numerous additional instances of Saul's involvement with simulation during his early career.

The networks studied in this paper are a well-known class of models in operations research called *stochastic activity networks* (SANs), which include the popular Project Evaluation Review Technique (PERT) and Critical Path Method (CPM). A nice introduction to PERT can be found in the encyclopedia entry written by Arjang Assad and Bruce Golden [2], two long-time colleagues of Saul hired by him in the 1970s while he was chairman of the Management Science & Statistics department at the University of Maryland, and co-authors of other chapters in this volume. Such models are commonly used in resource-constrained project scheduling; see [6] for a review and classification of existing literature and models in this field. In SANs, the most important measures of performance are the completion time and the arc (activity) criticalities. Often just as important as the performance measures themselves are the sensitivities of the performance measures with respect to parameters of the network. These issues are described in the state-of-the-art review [8]. When the networks become complex enough, Monte Carlo simulation is frequently used to estimate performance. However, simulation can be very expensive and time consuming, so finding ways to reduce the computational burden are important. Efficient estimation schemes for estimating arc criticalities and sensitivity curves via simulation are derived in [4, 5].

In this paper, we consider the problem of estimating performance measure sensitivities in the Monte Carlo simulation setting. This problem has been studied by various researchers. In particular, infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method have been used to derive efficient estimators for local sensitivities — see [3] for IPA used with conditional Monte Carlo, and [1] for the LR method used with control variates; whereas [7] use design of experiments methodology and regression to fit performance curves. The use of conditioning in [3] for variance reduction differs from its use in deriving estimators for performance measures to which IPA cannot be applied. Here, we present some alternative estimators: weak derivative (WD) estimators, and smoothed perturbation analysis (SPA) estimators. These can serve as alternatives to IPA and LR estimators. We also consider estimation of performance measure sensitivities that have not been considered in the literature. The rest of the paper is organized as follows. In Section 2, we present the problem setting and briefly review the PA, LR, and WD approaches to derivative estimation. The derivative estimators are presented in Section 3. Some concluding remarks, including extensions and further research, are made in Section 4.

2 Problem Setting

We consider a directed acyclic graph, defined by a set of nodes $\mathcal{N}$ of integers $1, \ldots, |\mathcal{N}|$, and a set of directed arcs $\mathcal{A} \subset \{(i,j) : i,j \in \mathcal{N}; i < |\mathcal{N}|, j > 1\}$, where (i,j) represents an arc from node i to node j, and, without loss of generality, we take node 1 as the source and node $|\mathcal{N}|$ as the sink (destination). For our purposes, we also map the set of directed arcs to the integers $\{1, \ldots, |\mathcal{A}|\}$ by the lexicographically ordering on elements of $\mathcal{A}$. Both representations of the arcs will be used interchangeably, whichever is the most convenient in the context. Let $\mathcal{P}$ denote the set of paths from source to sink. The input random variables are the individual activity times given by X_i, with cumulative distribution function (c.d.f.) $F_i, i = 1, \ldots, |\mathcal{A}|$, and corresponding probability density function (p.d.f.) or probability mass function (p.m.f.) f_i. Assume all of the activity times are *independent*. However, it should be clear that duration of paths in $\mathcal{P}$ will not in general be independent, such as in the following example, where all three of the path durations are dependent, since X_6 must be included in any path.

Example: 5-node network with $\mathcal{A} = \{(1,2),(1,3),(2,3),(2,4),(3,4),(4,5)\}$ mapped as shown in Figure 1; $\mathcal{P} = \{(1,4,6),(1,3,5,6),(2,5,6)\}$.

Let $P^* \in \mathcal{P}$ denote the set of activities on the *optimal* (critical) path corresponding to the project duration (e.g., shortest or longest path, depending on the problem), where P^* itself is a random variable. In this paper, we consider the total project duration, which can be written as

$$Y = \sum_{j \in P^*} X_j. \quad (1)$$

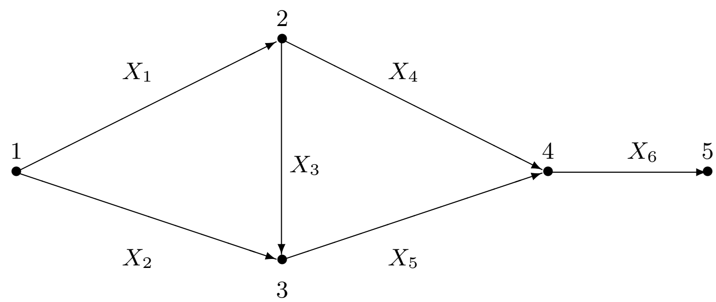

Fig. 1. Stochastic Activity Network.

Another important performance measure is arc criticality, which is the probability that a given activity is on the optimal (or critical) path, i.e., $P(i \in P^*)$ for activity i. As the focus of this paper is sensitivity analysis, we wish to estimate derivatives of performance measures involving Y with respect to a parameter θ. We consider three cases:

1. $dE[Y]/d\theta$, where θ appears in the activity time distributions (i.e., in some p.d.f. f_i);
2. $dP(Y > y)/d\theta$ for some given $y \geq 0$, where again θ appears in the activity time distributions;
3. $dP(Y > \theta)/d\theta$, where θ occurs directly in the tail distribution performance measure (so this is essentially the negative of the density function evaluated at the point θ).

The first case has been addressed previously in [3] using IPA and in [1] using the LR method. We will review these methods briefly, and also present new estimators based on the use of weak derivatives (WD). Neither the second nor the third case has been considered in the literature, but both the LR method and WD approaches can be extended to the second case in a straightforward manner, whereas the IPA estimator would fail for that form of performance measure, requiring the use of smoothed perturbation analysis (SPA) to obtain an unbiased estimator. The third case essentially provides an estimate of the density of Y if taken over all possible values of θ. This form of performance measure presents some additional challenges not seen in previous work.

Before deriving the estimators, we give a brief overview of IPA, SPA, the LR method, and the WD approach. Details can be found in [12]. For illustrative purposes, we will just consider the first case above, where the performance measure is an expectation:

$$J(\theta) = E[Y(\theta)] = E[Y(X_1, \ldots, X_T)]. \qquad (2)$$

Y is the (univariate) output performance measure, $\{X_i\}$ are the input random variables, and T is the number of input random variables. In the SAN setting, $T = |\mathcal{A}|$, and Y is given by (1). Stochastic simulation can be viewed as a means of carrying out the so-called "law of the unconscious statistician" (cf. p.7 in [20]; this term was removed in the 1996 second edition):

$$E[Y(\mathbf{X})] = \int y dF_Y(y) = \int Y(\mathbf{x}) dF_{\mathbf{X}}(\mathbf{x}). \tag{3}$$

Coming into the simulation are input random variables $\{X_i\}$, for which we know the distribution $F_{\mathbf{X}}$; coming out of the simulation is an output random variable Y, for which we would like to know the distribution F_Y; and 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. If we knew the distribution of Y, there would generally be no need for simulation.

For simplicity, we assume for the remainder of the discussion in this section that the parameter θ is scalar, because the vector case can be handled by taking each component individually. In the right-hand side of (3), the parameter appearing directly in the sample performance $Y(\cdot;\theta)$ corresponds to the view of perturbation analysis (PA), whereas its appearance in the distribution $F_{\mathbf{X}}(\cdot;\theta)$ leads to the likelihood ratio (LR) method or weak derivative (WD) approach.

Let f denote the joint density of all of the input random variables. Differentiating (3), and assuming interchangeability of integration and differentiation:

$$\frac{dE[Y(X)]}{d\theta} = \begin{cases} \displaystyle\int_{-\infty}^{\infty} Y(x) \frac{df(x;\theta)}{d\theta} dx, & (4) \\[2ex] \displaystyle\int_0^1 \frac{dY(X(\theta;u))}{d\theta} du, & (5) \end{cases}$$

where x and u and the integrals are all T-dimensional. For notational simplicity, these T-dimensional multiple integrals are written as a single integral throughout, and we also assume that one random number u produces one random variate x (e.g., using the inverse transform method). In (4), the parameter appears in the distribution directly, whereas in (5), the underlying uncertainty is considered the uniform random numbers.

For expositional ease, 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 (5), we use the chain rule to write

$$\begin{aligned} \frac{dE[Y(X)]}{d\theta} &= \int_0^1 \frac{dY(X_1(\theta;u_1), X_2, \ldots)}{d\theta} du \\ &= \int_0^1 \frac{\partial Y}{\partial X_1} \frac{dX_1(\theta;u_1)}{d\theta} du. \end{aligned} \tag{6}$$

In other words, the estimator takes the form

$$\frac{\partial Y(X)}{\partial X_1}\frac{dX_1}{d\theta}, \tag{7}$$

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. This approach is called infinitesimal perturbation analysis (IPA), and the main requirement for it to yield an unbiased estimator is that the sample performance be almost surely continuous, which is not satisfied for certain forms of performance measure (e.g., probabilities) and will be violated for some stochastic systems.

Assume that X_1 has marginal p.d.f. $f_1(\cdot;\theta)$, and that the joint density 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 (4) involving differentiation of a density (measure) can be further manipulated using the product rule of differentiation to yield the following:

$$\frac{dE[Y(X)]}{d\theta} = \int_{-\infty}^{\infty} Y(x)\frac{\partial f_1(x_1;\theta)}{\partial\theta} f_{-1}(x_2,\ldots)dx \tag{8}$$

$$= \int_{-\infty}^{\infty} Y(x)\frac{\partial \ln f_1(x_1;\theta)}{\partial\theta} f(x)dx. \tag{9}$$

In other words, the estimator takes the form

$$Y(X)\frac{\partial \ln f_1(X_1;\theta)}{\partial\theta}. \tag{10}$$

On the other hand, if instead of expressing the right-hand side of (8) as (9), 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:

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

The triple $\left(c(\theta), f_1^{(1)}, f_1^{(2)}\right)$ constitutes a weak derivative (WD) for f_1, which is not unique if it exists. The corresponding WD estimator is of the form

$$c(\theta)\left(Y(X_1^{(2)}, X_2, \ldots) - Y(X_1^{(1)}, X_2, \ldots)\right), \tag{11}$$

where $X_1^{(1)} \sim f_1^{(1)}$ and $X_1^{(2)} \sim f_1^{(2)}$. In other words, the estimator takes the difference of the sample performance at two different values of the input random variable X_1. The term "weak" derivative comes about from the possibility that $\frac{df_1(\cdot;\theta)}{d\theta}$ in (8) may not be proper, and yet 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 (5) 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 allow the interchange as follows:

$$\frac{dE[Y(X)]}{d\theta} = \int_0^1 \frac{dE[Y(X(\theta;u))|Z(u)]}{d\theta} du, \tag{12}$$

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:

$$\frac{\partial E[Y(X)|Z]}{\partial X_1}\frac{dX_1}{d\theta}. \tag{13}$$

Note that taking Z as the entire set leads back to (7), the IPA estimator. The chief difficulty in applying the methodology is determining the appropriate Z such that $E[Y(X)|Z]$ is both smooth, and its derivative can be easily estimated. Further details can be found in [14, 11].

3 Derivations of the Estimators

We begin with $dE[Y]/d\theta$, where θ is some parameter occurring in the distribution of X_1 only, as considered at the end of the last section. Then, the IPA estimator can be obtained by straightforward differentiation of the expression for Y given by (1), noting that θ only affects Y through X_1:

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

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. The LR/SF estimator is given by (10), and the WD estimator is given by (11).

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

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

where f_i is the p.d.f. for X_i. The IPA and LR estimators differ from the respective ones in [3] and [1], in that those both use variance reduction techniques to improve the estimators further.

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.

Examples: Consider two common distributions: the exponential with mean θ and the normal (Gaussian) with mean θ and standard deviation σ. In both cases, let θ_i be the corresponding parameter in X_i. Then we have the following estimators:

(a) exponential distribution, X_i with mean θ_i

$$\text{IPA:} \quad \frac{X_i}{\theta_i}\mathbf{1}\{i \in \mathcal{P}^*\},$$

$$\text{LR:} \quad Y(X)\frac{1}{\theta_i}\left(\frac{X_i}{\theta_i} - 1\right),$$

$$\text{WD:} \quad \frac{1}{\theta_i}\left(Y(X_1,\ldots,X_i^{(2)},\ldots) - Y(X)\right),$$

where $X_i^{(2)}$ has the following Erlang distribution (p.d.f.):

$$\frac{1}{\theta_i^2} x e^{-x/\theta_i}\mathbf{1}\{x > 0\}.$$

(b) normal distribution, X_i with mean θ_i and standard deviation σ_i

$$\text{IPA:} \quad \mathbf{1}\{i \in \mathcal{P}^*\},$$

$$\text{LR:} \quad Y(X)\frac{X_i - \theta_i}{\sigma_i^2},$$

$$\text{WD:} \quad \frac{1}{\sqrt{2\pi}\sigma_i}\left(Y(X_1,\ldots,X_i^{(2)},\ldots) - Y(X_1,\ldots,X_i^{(1)},\ldots)\right),$$

with $X_i^{(1)} = \theta_i - X'$ and $X_i^{(2)} = \theta_i + X'$, where X' has the following Weibull distribution (p.d.f.):

$$8\sigma_i^4 x e^{-(2\sigma_i^2 x)^2}\mathbf{1}\{x > 0\}.$$

If instead of an expectation, we were interested in estimating the tail distribution, e.g., $P(Y > y)$ for some fixed y, the WD and LR/SF estimators would simply replace Y with the indicator function $\mathbf{1}\{Y > y\}$. However, IPA does 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, which we will shortly demonstrate.

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

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

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 handled analogously. 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

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

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

$$\begin{aligned} P_Z(\max_{P \in \mathcal{P}_1} |P| > y) &= P_Z(X_1 + \max_{P \in \mathcal{P}_1} |P|_{-1} > y) \\ &= P_Z(X_1 > y - \max_{P \in \mathcal{P}_1} |P|_{-1}) = \bar{F}_1(y - \max_{P \in \mathcal{P}_1} |P|_{-1}; \theta), \end{aligned}$$

where $\bar{F} \equiv 1 - F$ denotes the complementary c.d.f., we have

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

Differentiating, we get the estimator:

$$\frac{dL_Z}{d\theta} = \frac{\partial \bar{F}_1(y - \max_{P \in \mathcal{P}_1} |P|_{-1}; \theta)}{\partial \theta} \cdot \mathbf{1}\{\max_{P \in \mathcal{P}} |P|_{-1} \leq y\}, \qquad (14)$$

which applies for both continuous and discrete distributions, as the following example illustrates.

Example: For the 5-node example, $\mathcal{P} = \{146, 1356, 256\}$, $\mathcal{P}_1 = \{146, 1356\}$, $|146|_{-1} = X_4 + X_6, |1356|_{-1} = X_3 + X_5 + X_6, |256|_{-1} = X_2 + X_5 + X_6$. If X_1 is exponentially distributed with mean θ, $\partial \bar{F}_1(x;\theta)/\partial\theta = e^{-x/\theta}(x/\theta^2)$, and the estimator is given by

$$\exp\left(-\frac{y - \max(X_3 + X_5, X_4) - X_6}{\theta}\right) \frac{y - \max(X_3 + X_5, X_4) - X_6}{\theta^2} \cdot \mathbf{1}\{\max(X_3 + X_5, X_4, X_2 + X_5) + X_6 \leq y\},$$

whereas if X_1 is Bernoulli — i.e., is equal to x_{low} with probability θ and equal to $x_{high} > x_{low}$ otherwise — then $\partial \bar{F}_1(x;\theta)/\partial\theta = \mathbf{1}\{x_{low} \leq x < x_{high}\}$, and the estimator is given by

$$\mathbf{1}\{x_{low} \leq y - \max(X_3 + X_5, X_4) - X_6 < x_{high}\} \cdot \mathbf{1}\{\max(X_3 + X_5, X_4, X_2 + X_5) + X_6 \leq y\}.$$

Clearly, the estimator (14) was derived without loss of generality, so for the parameter θ being in the distribution of X_i, we have the estimator

$$\frac{\partial \bar{F}_i(y - \max_{P\in\mathcal{P}_i} |P|_{-j};\theta)}{\partial\theta} \cdot \mathbf{1}\{\max_{P\in\mathcal{P}} |P|_{-j} \leq y\}. \quad (15)$$

For the shortest path problem, simply replace "max" with "min" throughout in the estimator (15).

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, i.e.,

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

Note that 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:

$$\tilde{X}_i = X_i - \theta,$$

for arcs i corresponding to directed arcs $(1, j) \in \mathcal{A}$ or $(j, |\mathcal{N}|) \in \mathcal{A}$. In the 5-node example of Figure 1, this would be either {1,2} or {6}. Note that minimal cut sets will not necessarily do the trick. For example, in the 5-node example, {1,5} is a cut set, but both members are contained on the path (1,3,5,6), so subtracting θ from these two arc lengths would lead to possibly erroneous results.

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, making it inapplicable.

Lastly, we apply PA to the problem of estimating $dP(Y > \theta)/d\theta$. We note that this estimation problem only makes sense in the *continuous* distribution case, where it is essentially an estimation of (the negative of) the p.d.f., since in the discrete case, the corresponding derivative is 0; thus, assume henceforth that each X_i has a p.d.f. f_i. Again, this type of performance measure cannot be handled by IPA, so we use SPA. The idea is to condition on a special set of activity times such that both the set itself and its complement have a non-zero probability of having a corresponding activity on the critical path.

Recall the following network definitions. A *cut set* is a set of arcs such that their removal from the network leaves no path from source to sink. A *minimal cut set* is a cut set such that the removal of any arc in the set no longer leaves a cut set. In the 5-node example, the minimal cut sets are {1,2}, {1,5}, {2,3,4}, {4,5}, {6}. By definition, a minimal cut set will have an activity on the critical path. The following observation is key:

Lemma. Let $\mathcal{C}$ be a minimal cut set for the network, and let $Z = \{X_i : i \notin \mathcal{C}\}$. If there exists an $i \notin \mathcal{C}$ such that $P(i \in P^*) > 0$, then $P_Z(Y > \theta)$ is a.s. continuous with respect to θ.

Thus, if one can find a minimal cut set that satisfies the condition in the lemma, one can in principle derive an unbiased derivative estimator for $P(Y > \theta)$ by conditioning on the complement set of activity times and then taking the sample path derivative. Note, however, that finding such a minimal cut set may be a computationally formidable task for large networks. Furthermore, as we shall see, in order to take the sample path derivative in a convenient form, it is desirable that the activities in the cut set actually partition the path space. We illustrate these ideas in an extended example using the 5-node network.

Example: For the 5-node example, we consider all of the minimal cut sets. (i) Using minimal cut set $\mathcal{C} = \{6\}$, we take $Z = \{X_1, X_2, X_3, X_4, X_5\}$, so we have

$$P_Z(Y > \theta) = P_Z(\max_{P \in \mathcal{P}_6} X_6 + |P|_{-6} > \theta) = \bar{F}_6(\theta - \max_{P \in \mathcal{P}_6} |P|_{-6}).$$

Differentiating, the final estimator is given by

$$\frac{dP_Z(Y > \theta)}{d\theta} = -f_6(\theta - \max_{P \in \mathcal{P}_6} |P|_{-6}).$$

Note that the form of the estimator only involves the p.d.f.s of those arcs in the cut set. If X_6 follows an exponential distribution, the specific estimator is given by

$$-\frac{1}{E[X_6]} \exp\left(\frac{\max(X_1 + X_3 + X_5, X_1 + X_4, X_2 + X_5) - \theta}{E[X_6]}\right) \cdot \mathbf{1}\{\max(X_1 + X_3 + X_5, X_1 + X_4, X_2 + X_5) \leq \theta\}.$$

(ii) Using minimal cut set $\mathcal{C} = \{1, 2\}$, we take $Z = \{X_3, X_4, X_5, X_6\}$, so we have

$$\begin{aligned} P_Z(Y > \theta) &= 1 - P_Z(\max_{P \in \mathcal{P}} |P| \leq \theta) \ = \ 1 - P_Z(\max_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} |P| \leq \theta) \\ &= 1 - P_Z(\max_{P \in \mathcal{P}_1} |P| \leq \theta, \max_{P \in \mathcal{P}_2} |P| \leq \theta) \\ &= 1 - P_Z(X_1 + \max_{P \in \mathcal{P}_1} |P|_{-1} \leq \theta, X_2 + \max_{P \in \mathcal{P}_2} |P|_{-2} \leq \theta) \\ &= 1 - F_1(\theta - \max_{P \in \mathcal{P}_1} |P|_{-1}) F_2(\theta - \max_{P \in \mathcal{P}_2} |P|_{-2}), \end{aligned}$$

where we have used the fact that $\mathcal{P}_1$ and $\mathcal{P}_2$ partition the path space, and X_1 and X_2 are independent. Differentiating, the final estimator is given by

$$\begin{aligned} \frac{dP_Z(Y > \theta)}{d\theta} &= -f_1(\theta - \max_{P \in \mathcal{P}_1} |P|_{-1}) F_2(\theta - \max_{P \in \mathcal{P}_2} |P|_{-2}) \\ &\quad - F_1(\theta - \max_{P \in \mathcal{P}_1} |P|_{-1}) f_2(\theta - \max_{P \in \mathcal{P}_2} |P|_{-2}). \end{aligned}$$

Using minimal cut set $\mathcal{C} = \{4, 5\}$ will yield the analogous estimator

$$\frac{dP_Z(Y>\theta)}{d\theta} = -f_4(\theta - \max_{P\in\mathcal{P}_4}|P|_{-4})F_5(\theta - \max_{P\in\mathcal{P}_5}|P|_{-5}) - F_4(\theta - \max_{P\in\mathcal{P}_4}|P|_{-4})f_5(\theta - \max_{P\in\mathcal{P}_5}|P|_{-5}).$$

(iii) Using minimal cut set $\mathcal{C} = \{2,3,4\}$, we take $Z = \{X_1, X_5, X_6\}$, and similar analysis yields

$$P_Z(Y>\theta) = 1 - F_2(\theta - \max_{P\in\mathcal{P}_2}|P|_{-2})F_3(\theta - \max_{P\in\mathcal{P}_3}|P|_{-3})F_4(\theta - \max_{P\in\mathcal{P}_4}|P|_{-4}),$$

again since $\mathcal{P}_2, \mathcal{P}_3$ and $\mathcal{P}_4$ partition the path space, and X_2, X_3, and X_4 are mutually independent. Differentiating, the final estimator is given by

$$\begin{aligned}\frac{dP_Z(Y>\theta)}{d\theta} = &-f_2(\theta - \max_{P\in\mathcal{P}_2}|P|_{-2})F_3(\theta - \max_{P\in\mathcal{P}_3}|P|_{-3})F_4(\theta - \max_{P\in\mathcal{P}_4}|P|_{-4})\\ &- F_2(\theta - \max_{P\in\mathcal{P}_2}|P|_{-2})f_3(\theta - \max_{P\in\mathcal{P}_3}|P|_{-3})F_4(\theta - \max_{P\in\mathcal{P}_4}|P|_{-4})\\ &- F_2(\theta - \max_{P\in\mathcal{P}_2}|P|_{-2})F_3(\theta - \max_{P\in\mathcal{P}_3}|P|_{-3})f_4(\theta - \max_{P\in\mathcal{P}_4}|P|_{-4}).\end{aligned}$$

(iv) Using minimal cut set $\mathcal{C} = \{1,5\}$, we take $Z = \{X_2, X_3, X_4, X_6\}$. Note, however, that $\mathcal{P}_1$ and $\mathcal{P}_5$ do not partition the path space, since path (1,3,5,6) is in both sets. We shall now see how this leads to difficulties:

$$\begin{aligned}P_Z(Y>\theta) &= 1 - P_Z(\max_{P\in\mathcal{P}}|P|\le\theta) \;=\; 1 - P_Z(\max_{P\in\mathcal{P}_1\cup\mathcal{P}_5}|P|\le\theta)\\ &= 1 - P_Z(X_1 + |146|_{-1} \le \theta, X_5 + |256|_{-5} \le \theta,\\ &\qquad X_1 + X_5 + |1356|_{-1,-5} \le \theta),\end{aligned}$$

which cannot be factored nicely as in the previous cases.

In general, if the $\{\mathcal{P}_i\}$ do partition the path space, i.e., $\mathcal{P} = \bigcup_{i\in\mathcal{C}}\mathcal{P}_i$ and $\mathcal{P}_i \cap \mathcal{P}_j = \phi$ for $i \neq j$, then

$$\begin{aligned}P_Z(Y>\theta) &= 1 - P_Z(\max_{P\in\mathcal{P}}|P|\le\theta) \;=\; 1 - P_Z(\max_{P\in\cup_i\mathcal{P}_i}|P|\le\theta)\\ &= 1 - P_Z(\bigcap_{i\in\mathcal{C}}\{\max_{P\in\mathcal{P}_i}|P|\le\theta\})\\ &= 1 - P_Z(\bigcap_{i\in\mathcal{C}}\{X_i + \max_{P\in\mathcal{P}_i}|P|_{-i}\le\theta\})\\ &= 1 - \prod_{i\in\mathcal{C}} P_Z(X_i + \max_{P\in\mathcal{P}_i}|P|_{-i}\le\theta)\\ &= 1 - \prod_{i\in\mathcal{C}} F_i(\theta - \max_{P\in\mathcal{P}_i}|P|_{-i}),\end{aligned}$$

which upon differentiation yields the estimator

$$- \sum_{i \in \mathcal{C}} f_i(\theta - \max_{P \in \mathcal{P}_i} |P|_{-i}) \prod_{j \neq i} F_j(\theta - \max_{P \in \mathcal{P}_j} |P|_{-j}),$$

where the product is equal to 1 if it is empty as in case (i) of the 5-node example just considered.

4 Directions for Future Research

In the general PERT/CPM stochastic activity network framework, we have derived some new unbiased Monte Carlo simulation-based derivative estimators for a setting where there are existing estimators available, and also for new settings. This gives the simulation analyst a wider array of options in carrying out performance evaluation. Depending on the types of networks and performance measures considered, different estimators may be more effective and useful. Clearly, comparisons of variance properties of the various estimators, whether through theoretical analysis and/or empirical/numerical experimental testing, would be beneficial to the potential user of these estimators. Preliminary experiments in [19] indicate that the SPA estimators are quite promising, as they generally show lower variance than the WD and LR estimators. However, in the 5-node Bernoulli example, it turns out that the SPA and one version of the WD estimator coincide. In the cases where they are applicable, the IPA estimators provide the lowest variance, which turns out to be essentially the same as a finite difference estimator with common random numbers, a procedure that can be easily implemented in many settings, but still requires additional simulations for every parameter of interest in the network.

The extension to arc criticalities is an important one. Another interesting extension to consider is the case where the input random variables (individual activity times $\{X_i\}$) are not necessarily independent. Using these derivative estimates in simulation optimization is another fruitful area of research; see, e.g., [9, 10, 12, 13]. Investigating quasi-Monte Carlo estimators and also estimators that are not sample means but quantiles or order statistics are also useful topics for future research.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant DMI-0323220, and by the Air Force Office of Scientific Research under Grant FA95500410210. An earlier version was presented at the International Conference on Automatic Control and Systems Engineering, Cairo, Egypt, in December 2005 [11]. In addition to numerous editorial enhancements, substantive changes in this version include correcting the SPA estimator for $dP(Y > y)/d\theta$ in the 5-node example, adding the discrete (Bernoulli) case for the same example, and greatly expanding the scope of the analysis and adding the 5-node extended example for the $dP(Y > \theta)/d\theta$ SPA estimator.

References

1. V.G. Adlakha and H. Arsham. A simulation technique for estimation in perturbed stochastic activity networks. *Simulation*, 58:258–267, 1992.
2. A.A. Assad and B.L. Golden. PERT. in *Encyclopedia of Statistical Sciences*, Volume 6:691–697, 1992.
3. R.A. Bowman. Stochastic gradient-based time-cost tradeoffs in PERT network using simulation. *Annals of Operations Research*, 53:533–551, 1994.
4. R.A. Bowman. Efficient estimation of arc criticalities in stochastic activity networks. *Management Science*, 41:58–67, 1995.
5. R.A. Bowman. Sensitivity curves for effective project management. *Naval Research Logistics*, 50:481–497, 2003.
6. P. Brucker, A. Drexl, R. Mohring, K. Neumann, and E. Pesch. Resource-constrained project scheduling: Notation, classification, models, and methods. *European Journal of Operational Research*, 112:3–41, 1999.
7. J.G. Cho and B.J. Yum. Functional estimation of activity criticality indices and sensitivity analysis of expected project completion time. *Journal of the Operational Research Society*, 55:850–859, 2004.
8. S.E. Elmaghraby. On criticality and sensitivity in activity networks. *European Journal of Operational Research*, 127:220–238, 2000.
9. M.C. Fu. Optimization via simulation: A review. *Annals of Operations Research*, 53:199–248, 1994.
10. M.C. Fu. Optimization for simulation: Theory vs. practice (Feature Article). *INFORMS Journal on Computing*, 14:192–215, 2002.
11. M.C. Fu. Sensitivity analysis for stochastic activity networks. *Proceedings of the International Conference on Automatic Control and Systems Engineering* (on CD-ROM), 2005.
12. M.C. Fu. Stochastic Gradient Estimation, Chapter 19 in *Handbooks in Operations Research and Management Science: Simulation*, S.G. Henderson and B.L. Nelson, eds., Elsevier, 2006.
13. M.C. Fu, F.W. Glover, and J. April. Simulation optimization: A review, new developments, and applications. *Proceedings of the 2005 Winter Simulation Conference*, 83–95, 2005.
14. M.C. Fu and J.Q. Hu. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Kluwer Academic, Boston, MA, 1997.
15. S.I. Gass. Decision-aiding models: validation, assessment, and related issues for policy analysis. *Operations Research*, 31:601–663, 1983.
16. S.I. Gass. Model accreditation: A rationale and process for determining a numerical rating. *European Journal of Operational Research*, 66:250–258, 1993.
17. S.I. Gass and L. Joel. Concepts of model confidence. *Computers and Operations Research*, 8:341–346, 1987.
18. S.I. Gass and B.W. Thompson. Guidelines for model evaluation: An abridged version of the U. S. general accounting office exposure draft. *Operations Research*, 28:431–479, 1980.
19. C. Gröer and K. Ryals. Sensitivity analysis in simulation of stochastic activity networks: A computational study; project report for BMGT835, Spring 2006. also submitted for presentation and publication, Tenth INFORMS Computing Society Conference on International Impacts of Computing, Optimization, and Decision Technologies.

20. S.M. Ross. *Stochastic Processes*. John Wiley & Sons, New York, NY, 1983.
21. R.G. Sargent. Personal communication, January 2006.