Proceedings of the 2007 INFORMS Simulation Society Research Workshop

STRUCTURED DATABASE MONTE CARLO (SDMC): A NEW EFFICIENT SIMULATION STRATEGY

Gang Zhao & Pirooz Vakili

Center for Information & Systems Engineering
Department of Manufacturing Engineering
Boston University
Boston, MA 02446, U.S.A.

Tarik Borogovac

Center for Information & Systems Engineering
Department of Electrical and Computer Engineering
Boston University
Boston, MA 02446, U.S.A.

1 INTRODUCTION

We introduce a new strategy, called Structured Database Monte Carlo (SDMC), for efficient Monte Carlo (MC) simulation. The term "strategy" is used to emphasize that the approach is not designed to address a specific problem, using a specific method, but rather a new way of designing and implementing MC algorithms. One of the main approaches to achieving efficiency in MC is to reduce the variance of the MC estimator. Methods that follow this approach are generally referred to as variance reduction techniques. The main techniques in this category were introduced many decades ago (see, e.g., Hammersley and Handscomb 1964) and the newer results have generally consisted of adaption of the techniques to specific domains of application. As one author points out "the greatest gains in efficiency from variance reduction techniques result from exploiting specific features of a problem, rather than from generic applications of generic methods." (Glasserman 2004).

1.1 A generic technique for variance reduction

The point of departure of SDMC is an attempt to develop effective variance reduction techniques that are in fact generic methods and can be generically applied. The approach does not require discovering specific features of the problem but rather relies on constructing/imposing a structure that can be used to advantage for variance reduction. More specifically, when considering the basic techniques of variance reduction such as *stratification*, *control variate*, and *importance sampling*, to name a few, SDMC seeks to provide generic methods for defining strata, control variates, and new sampling measures, respectively.

One of our basic assumptions is that the estimation problem at hand depends on a parameter (possibly a vector) θ through the probability measure or the "performance" function whose expected value is to be estimated. SDMC approach gathers and "structures" information at one parameter value, say θ_0, in a "database" and utilizes this information to gain estimation efficiency at neighboring parameters. In other words, when solving additional problem instances at neighboring parameter values, the database of stored information enables more effective use of many variance reduction techniques in a manner that is problem independent.

The idea of carrying over information obtained at one parameter to a neighboring parameter has been applied, for example, for sensitivity estimation. This is done either by a careful study of the impact of a parameter perturbation on a sample performance (evaluating path derivatives) as in Perturbation Analysis (PA) (see, e.g., Ho and Cao 1991 and Glasserman 1991) or by considering how a change in a parameter can be transformed into changes of path weights as in Likelihood Ratio approach (see, e.g., Rubinstein 1989). SDMC is in some sense closer to PA. PA, closely related to the variance reduction technique of common random numbers, pairs paths whose "input" variables are only slightly different (have small perturbations). In the SDMC approach paths at two parameter values are also paired via their reference to the same element of the database. The goal of SDMC, however is to obtain information about the global dependence of the sample performance on sample paths while PA seeks to obtain local information about the dependence of the performance on parameter values.

1.2 Decision making setting

The gathering and structuring of information in a database, required by SDMC, are potentially costly operations. Incurring the cost is only justified if there are significant computational savings or other advantages associated with this method. We specify two sample practical decision making settings where we believe the computational setup cost associated with SDMC may be justified and the method may lead to significant computational savings or significantly better solutions. As-

sume that the parameter value (or vector) θ_0 is used to "construct" the database and let $\mathcal{N}(\theta_0)$ denote a neighborhood of θ_0. Then consider the following settings:

1. The estimation problem needs to be solved for many parameter values in $\mathcal{N}(\theta_0)$.
2. The estimation problem needs to be performed within a limited budget (time or computational resources) at $\theta \in \mathcal{N}(\theta_0)$ where θ is not known in advance and once it becomes known the budget constraint is in force.

Problems of type 1 are common in computational finance where risk valuation and pricing are done repeatedly with only minor changes in the problem setting. Many of these can be viewed as resolving the same problem at different parameters values in the neighborhood of a nominal parameter values.

Problems of type 2 are also common in some computational settings. Consider, for example, the problem of real-time forecasting of the path of a hurricane. The payoff for setting up and structuring a database (or a set of databases) can be a better forecast of the trajectory.

2 PROBLEM DEFINITION AND PRELIMINARIES

We begin with some general comments about the estimation problem via MC. Then, to make things more specific, we narrow down the focus to estimation problems in the context of a specific class of stochastic processes.

In general, estimation via MC simulation corresponds to approximating the expected value of an appropriately defined random variable. Specifically, let $(\Omega, \mathcal{F}, P)$ be a probability space, X a random element of Ω corresponding to the probability measure P (i.e., for all $A \in \mathcal{F}$, $P(X \in A) = P(A)$), and $L(\cdot;\theta) : \Omega \to R$ a parametric family of functions defined on Ω ($\theta \in \Theta$). Let $Y(\theta) = L(X, \theta)$ and define $J(\theta)$ as

$$\begin{aligned} J(\theta) &= E[Y(\theta)] = E[L(X;\theta)] \\ &= \int_\Omega L(X;\theta)dP = \int_\Omega Y(\theta)dP. \end{aligned}$$

The objective is to evaluate $J(\theta)$ efficiently.

Assume that Ω is the path space of a stochastic process, $L(\cdot;\theta)$ is a function that assigns a real number to each path, say the sample payoff, and θ is a parameter of the problem. In simulation there is a chain of transformations that takes place before a path is generated: A set of i.i.d. uniform (0,1) random variables are sampled, these are transformed into a set of more general non-uniform variates from which the path, X, is generated and finally $Y = L(X;\theta)$ is evaluated. Let us explicitly write out these transformations. Let $[0,1]^d$ be the d-dimensional hypercube where the uniform[0,1]'s reside (theoretically, d may be ∞). Let Ω' denote the space of non-uniforms. Then we have

$$[0,1]^d \xrightarrow{T} \Omega' \xrightarrow{K} \Omega \xrightarrow{L(.;\theta)} R.$$

$J(\theta)$ can be written as

$$\begin{aligned} J(\theta) &= \int_{[0,1]^d} L_1(U;\theta)dP_1 \\ &= \int_{R^d} L_2(W;\theta)dP_2 = \int_\Omega L(X;\theta)dP. \end{aligned}$$

where the random elements U and W and measures P_1 and P_2 are appropriately defined.

In some cases it is useful to think of the underlying uniforms as the primitives of the approach, in some, the non-uniform samples can be thought of as the primitives, and yet in others the final paths can be taken as the primitives. As will be discussed below, we select the primitives to be the paths of a vector valued standard Brownian motion driving a class of stochastic processes.

2.1 A class of models & estimation problems

Consider a k-dimensional process $X = \{X(t);\ 0 \le t \le T\}$ satisfying the following class of stochastic differential equations (SDE's):

$$dX(t) = \mu(X(t),t)dt + B(X(t),t)dW(t), \quad X(0) = 0.$$

where $W(t) = (W^1(t), \cdots, W^d(t))'$, $0 \le t \le T$, is a standard Brownian motion in R^d ($'$ is the transpose), and $\mu(\cdot;\cdot)$ and $B(\cdot;\cdot)$ are appropriately defined vector and matrix valued functions. These processes have wide applicability in computational finance and computational physics.

We now take another step towards making things more specific. When Monte Carlo simulation is used, the paths of the stochastic process are sampled at only a finite number of instances within the finite horizon of interest, and the functional is evaluated based on these sampled values.

Let $(t_0, \cdots, t_n)$ be a partition of $[0,T]$, i.e., $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$. Then the driving Brownian motion in R^d, sampled at $t_1 < \cdots < t_{n-1} < t_n = T$ is given by the vector $W = (W_1, \cdots, W_n)$ where $W_i = W(t_i)$. Let $\mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n = \mathcal{F}$ be the filtration defined by $\mathcal{F}_i = \sigma(W_1, \cdots, W_i)$ and let P be the probability

measure on $(\Omega, \mathcal{F})$ consistent with the definition of W above. Let

$$J(\theta) = E[L(Y(\theta))] = E[L((W_1, \cdots, W_n); \theta)].$$

Therefore, the estimation problem can be viewed as the estimation of the expected value of a function that is defined on the sampled values of a standard d-dimensional Brownian motion. Equivalently, it can be viewed as estimating the expected value of a function defined on a finite number of Brownian innovations, i.e, Gaussian vectors.

Note that (1) by adopting the above formulation, we have selected a standard probability space and have pushed out all parameters of interest to components of the L function; (2) a number of sensitivity/derivative estimation problems can be cast in the above framework. Examples are path derivatives, likelihood ratio method based derivatives, and Malliavin derivatives (see, e.g., Chen & Glasserman 2006).

3 SDMC STRATEGY

In order to define the setting for the SDMC approach, we consider yet another "discretization" of the estimation problem, this time by considering a discretized version of the sample space Ω as follows.

Assume a "very large" number of sample paths, i.e., elements of Ω, is generated. Let

$$\Omega_D = \{\omega_1, \omega_2, \cdots, \omega_K\}$$

represent these paths. Let P_D be the uniform measure on Ω_D, i.e., $P_D(\omega_i) = 1/K$ for all $i = 1, \cdots, K$. We assume that K is large enough so that for all practical purposes, the approximation

$$J_D(\theta) = \int_{\Omega_D} L(\theta) dP_D \approx \int_{\Omega} L(\theta) dP = J(\theta)$$

is satisfactory for all $\theta \in \mathcal{N}(\theta_0)$ under consideration. Note that $J_D(\theta)$ is simply the arithmetic average of L over Ω_D, i.e.,

$$J_D(\theta) = \frac{1}{K} \sum_{i=1}^{K} L(\omega_i; \theta)$$

Ω_D can be viewed a very large database of elements of Ω. The precise problem formulation in SDMC, therefore, is the following.

SDMC estimation problem. *Develop efficient MC algorithms to estimate $J_D(\theta)$.*

Note that (1) we are assuming Ω_D is given or constructed. There are interesting and relevant questions one can pose about how to construct an appropriate database Ω_D. We do not address these questions here. (2) We contend that the above formulation is quite broad and encompasses a very large class of estimation problems of interest. (3) The question of interest to us is whether it is possible to structure Ω_D in such a way that benefits solving the SDMC estimation problem.

To structure the database we begin with imposing a linear order on the database using the sample performances at a nominal parameter θ_0, i.e., Let

$$\omega_i \leq \omega_j \Leftrightarrow L(\omega_i; \theta_0) \leq L(\omega_j; \theta_0).$$

Our experimental results suggest that this structure is to some extent maintained when $L(\cdot, \theta)$ is evaluated for some θ is the vicinity of the nominal parameter θ_0 and this is the property we aim to utilize in constructing efficient estimation algorithms. Consider the following example.

Example. This example involves estimating the *action* on paths of a process satisfying a stochastic partial differential equation – specifically a time dependent 1-dimensional Ginzburg-Landau equation. The process is approximated by a 40 point one-dimensional lattice (with periodic boundary conditions) leading to a set of 40 coupled ordinary stochastic differential equations and time discretized using a forward Euler integrator. More specifically, the approximate process $X(t)$ is a 40-dimensional real continuous-time stochastic process on $t = [0, T]$ defined by

$$dX(t) = \mu(X(t), t)dt + \sqrt{2D(X(t), t)}dW(t)$$

Where $W(t)$ is a 40-dimensional standard Brownian motion. We define the k'th element of the forcing function $\mu : \Re^{41} \to \Re^{40}$ to be:

$$\mu_k(X(t), t) = -2\theta X_k(t) + X_k(t) - X_k^{\alpha}(t) + \theta X_{k-1}(t) + \theta X_{k+1}(t)$$

where $k = 1, \ldots, 40$ and we enforce the periodic boundary conditions. Also, we define D (in general a matrix function) to be a scalar constant variance parameter σ^2. α and θ are also scalar constants. For each simulated path, an action, Y, can be calculated as follows:

$$Y(\theta) = \frac{1}{8\sigma\Delta t} \sum_{i=1}^{N} \sum_{k=1}^{40} [X_{k,i} - X_{k,i-1} - \theta(X_{k+1,i} - 2X_{k,i} + X_{k-1,i})\Delta t]^2$$

The objective is to estimate $E[Y(\theta)]$ at different values of θ.

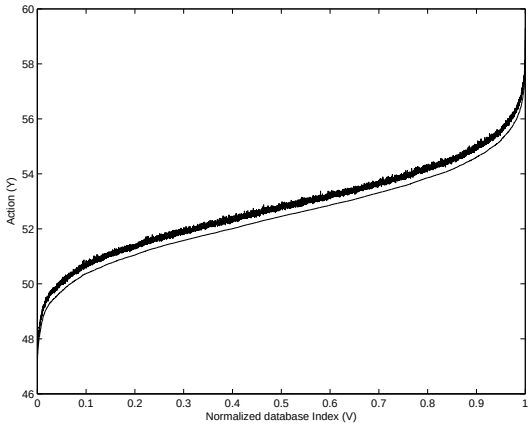

Figure 1: Simulation results of the action of a Ginzburg-Landau process. Lower curve contains sorted results at $\theta = 1$. The top curve contains perturbed results at $\theta = 1.1$

In this example, represented in Figure 1, a database of Brownian paths was generated and ordered based on sample values of $Y(\theta)$ at $\theta_0 = 1$. Figure 1 shows samples of $Y(\theta)$ at $\theta_0 = 1$ and $\theta = 1.1$ (y-axis) as a function of the ordered database. In other words, the x-axis represents the indices of the linearly ordered database starting from 1, representing the smallest to 10000, representing the largest.

Recall that SDMC aims to capture/identify the "structure" of the function L at a nominal parameter value θ_0 and to use this knowledge in designing more effective variance reduction techniques when estimating $J(\theta)$ at "neighboring" values of θ. In the above example, assume the variance reduction approach of *control variate* is under consideration. Then $Y(1)$ is a natural choice of a control for estimating $E[Y(1.1)]$. In fact this choice can lead to order of magnitudes of variance reduction in this case. It is not difficult to come up with specific and effective partitioning of the database for stratification and good choices of sampling measures for importance sampling.

It is instructive to map the database into the interval $[0, 1]$ via $i \to i/K$. Then, the estimation problem is simply a Monte Carlo estimation of the integral of a function L on $[0, 1]$. In general there may be almost no regularity conditions on this function. Structuring the database can be viewed as imposing some regularity condition on the function. This point of view is particularly convenient for analyzing the performance and convergence properties of different strategies in the context of SDMC (for example, in an information-based complexity setting, see, e.g, Traub et al 1988, or a la Kiefer 1957). We have obtained specific results for the stratification technique in the context of SDMC (Zhao & Vakili 2007).

4 PRESENTATION

The presentation will include the following. (1) A general discussion of the SDMC approach and the rationale behind the approach, (2) optimality & convergence properties of some of the algorithms, (3) experimental results, (4) an overall assessment of the SDMC approach and (5) future directions for research.

REFERENCES

Chen, N. and P. Glasserman. 2006. Malliavin Greeks without Malliavin Calculus.

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

Hammersley, J. M. and D. C. Handscomb. 1964. *Monte Carlo Methods*. John Wiley.

Ho, Y. C. 1991. *Perturbation Analysis of Discrete Event Systems*. Kluwer Academic Publishers.

Kiefer, J. 1957. Optimal Sequential Search and Approximation Methods under Minimum Regularity Assumptions. *J. Soc. Indust. Appl. Math.* Vol. 5, No. 3. pp 105-136.

Rubinstein, R. 1989. Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models. *Operations Research*. Vol. 37, pp 72-81.

Traub J. F., H. Wozniakowski, , and G. W. Wasilkowski. 1988. *Information-Based Complexity*. Academic Press.

Zhao, G., and P. Vakili. 2007. Structured Database Monte Carlo & Efficient Simulation via Stratification. Working paper.

5 ACKNOWLEDGEMENT

Research supported in part by the National Science Foundation grants CMMI-0620965 and DGE-0221680.

AUTHOR BIOGRAPHIES

GANG ZHAO is a Ph.D. student in the Department of Manufacturing Engineering at Boston University. His e-mail address is <gzhao@bu.edu>.

TARIK BOROGOVAC is a Ph.D. student in the Department of Electrical and Computer Engineering at Boston University. His e-mail address is <tarikb@bu.edu>.

PIROOZ VAKILI is an Associate Professor in the Department of Manufacturing Engineering at Boston University. His e-mail address is <vakili@bu.edu>.