Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

MONOTONICITY AND STRATIFICATION

Gang Zhao

Division of Systems Engineering
& Center for Information and Systems Engineering
Boston University
15 Saint Mary Street
Brookline, MA 02446, U.S.A.

Pirooz Vakili

Division of Systems Engineering
& Mechanical Engineering Department
Boston University
15 Saint Mary Street
Brookline, MA 02446, U.S.A.

ABSTRACT

In utilizing the technique of stratification, the user needs to first partition/stratify the sample space; the next task is to determine how to allocate samples to strata. How to best perform the second task is well understood and analyzed and there are effective and generic recipes for sample allocation. Performing the first task, on the other hand, is generally left to the user who has limited guidelines at her/his disposal. We review explicit and implicit stratification approaches considered in the literature and discuss their relevance to simulation studies. We then discuss the different ways in which monotonicity plays a role in optimal stratification.

1 INTRODUCTION

To use stratification, the user needs to first partition/stratify the sample space; given such a stratification, the next task is to determine how to allocate samples to strata. The second task is well understood and analyzed where effective and generic recipes have been available for a very long time (for a general discussion see, (Cochran 1977); for applications in the simulation context, see e.g., (Glasserman 2004), (Asmussen and Glynn 2007)). On the other hand, the issue of optimal strata definition has received less attention in the simulation literature. In what follows, we begin with considering three different settings where explicitly or implicitly the issue of strata definition is addressed. They all turn out to be relevant to our discussion.

A. Consider the following problem. Assume that we wish to estimate the average income of wage earners in the US as reported on their tax returns in 2007. Assume this is to be done based on a fixed sample size k. Crude sampling selects k random draws of tax returns and uses the sample average as the estimator. Alternatively, state averages can be estimated for each of the 50 states separately using crude sampling and then assembled into a single estimate. This approach corresponds to the method of *Stratified Sampling* using 50 strata (stratum = state). To reduce the variance of the overall estimator, more samples may be allocated to states that are more populated and/or where the income variability is higher. This is clearly not the only stratification possible. Alternatively, one can consider the stratification of the returns based on income: Assume returns are ordered in *increasing order* of income and then partitioned into 50 strata by selecting 49 interim strata boundaries. This problem has been considered and analyzed as the stratification of a frequency distribution See, e.g., (Cochran 1961) for a comparison of a number of methods for stratifying uni-dimensional frequency distributions; one of the examples in this paper relates to stratification applied to adjusted gross income per tax return for 1951 data. It is worth noting that, as observed in (Cochran 1961), one expects that an effective stratification based on 2007 data (or 1951 data) to remain effective in subsequent years (the economic/social stratification, unfortunately for those in the lower strata, is fairly stable over years!).

B. Consider the problem of evaluating the one dimensional integral

$$\mu = \int_0^1 g(u)du.$$

where g is an *increasing function* on $[0,1]$ (increasing=non-decreasing). Without loss of generality we can assume $g(0) = 0$ and $g(1) = 1$. This problem has been considered and analyzed in the literature on *Information Based Complexity* (for a general discussion see (Traub, Wozniakowski, and Wasilkowski 1988); for a discussion of the above problem see (Kiefer 1957) Section 5, (Sukharev 1987), and (Novak 1992)). Consider the worst case setting where one is to provide deterministic or stochastic error bounds for the estimation problem. Let $\mathcal{G}$ denote the set of all increasing functions g on $[0,1]$ where $g(0) = 0$ and $g(1) = 1$. Assume further that some information, say $I(k)$, in the form of k function evaluations has been gathered. Given this information it is not difficult to obtain integral estimates (denoted by $S(g)$) that minimize the worst case estimation error given $I(k)$ (denoted by $e(S(g),\mu|I(k))$ for both deterministic and stochastic cases). In other words, one needs to solve the following "min-max"problem.

$$\inf_{S(g)}\sup_{g\in\mathscr{G}}\{e(S(g),\mu|I(k))\}.$$

Given the above one can turn to the question of determining how to sample the function (gather information) in order to find the tightest possible error bound. One may consider on the one hand a non-adaptive or an adaptive approach, or on the other a deterministic or a random sampling approach. (Novak 1992) shows that adaptation in the deterministic case does not improve the rate of convergence (the optimal convergence rate is $O(n^{-1})$ in both cases) while it is beneficial in random sampling, improving the convergence rate from $O(n^{-1})$ for the non-adaptive case to $O(n^{-3/2})$ for the adaptive case. He also shows that this rate of convergence is optimal for all random sampling schemes. For our purposes, it is noteworthy that the optimal adaptive algorithm provided in (Novak 1992) is essentially a stratified sampling algorithm.

C. Consider the problem of evaluating an integral on the d-dimensional unit cube $I^d=[0,1]^d$.

$$\mu=\int_{I^d} f(\mathbf{u})d\mathbf{u}.$$

where $\mathbf{u}\in I^d$. As is well known, this problem can be reformulated as evaluating

$$\mu=E[f(\mathbf{U})]$$

where $\mathbf{U}$ is uniformly distributed over I^d and it can be viewed as a general model of a class of estimation via simulation problems where $\mathbf{U}=(U_1,\cdots,U_d)$ is the vector of uninform simulation inputs.

(Cheng and Davenport 1989) provides an insightful discussion of stratification in this setting where the issue of strata selection is explicitly and extensively discussed. Stratification can focus on ways of dissecting the d-dimensional cube (and taking its geometry into account) where the problem becomes more challenging as the dimension increases. Or it can rely on dissecting the range space, $f(I^d)$, a single-dimensional space for all d, and use the pull-back of the stratification of the range to obtain a stratification of the domain I^d. Cheng and Davenport (Cheng and Davenport 1989) note that the second approach represents an ideal case providing the best possible rate of convergence. For practical stratification they propose using one or more *shadow responses* as a way to stratify I^d using values of the shadow responses.

In this paper, we revisit random estimation of μ for example C from the point of stratification and analyze optimal stratification in this setting where the optimality criterion is defined as the optimal rate of convergence as in (Novak 1992). We then briefly consider a parametric version of example C, namely estimating

$$\mu(\theta)=\int_{I^d} f(\mathbf{u};\theta)d\mathbf{u}=E[f(\mathbf{U};\theta)]. \tag{1}$$

To turn the insight obtained from our discussion of optimal stratification into a practical stratification strategy, we consider a "large" sample from I^d, denoted by $DB=\{\mathbf{U}_1,\cdots,\mathbf{U}_N\}$, that we refer to as the *database* and consider the estimation problem

$$\mu(\theta|DB)=E[f(\mathbf{U};\theta)|DB]. \tag{2}$$

Estimation problem (2) can be viewed as an approximation to the original parametric estimation problem (1). The finite sample estimation problem (2) is similar to example A and the optimal stratification applied to problem B has direct implications for this problem.

The rest of the paper is organized as follows. Preliminaries are given in Section 2. Section 3 describes some of the implications of monotonicity for stratification. We give our optimal stratification result in section 4 and briefly describe its connection with the method of Structured Database Monte Carlo (SDMC) in section 5. Some concluding results are given in Section 6.

2 PRELIMINARIES

We begin by specifying an estimation problem and giving a brief description of the stratification method.

2.1 The Estimation Problem

To simplify the discussion and to make the connection with problem C in the introduction more explicit we consider the problem of estimating

$$\mu=\int_{I^d} f(\mathbf{u})d\mathbf{u}=E[f(\mathbf{U})]=E[Y] \tag{3}$$

where $Y=f(U)$ and f is a real-valued function. The discussion to follow applies to more general settings as well.

Let $Y\sim F$, i.e., let F denote the cumulative distribution of (simulation output) Y. Let

$$g(u)=\inf\{y;\ u\le F(y)\}\quad \text{for}\quad u\in(0,1)$$

be the inverse of F. Then we have

$$\mu = \int_0^1 g(u)du = E[g(U)] = E[Y] \tag{4}$$

where $U \sim U(0,1)$. Note that g is a monotone increasing (nondecreasing) function.

Therefore estimation problem (3) can be reformulated as estimating the integral of a monotone function, i.e., a problem of type (4).

We now turn to a brief discussion of the stratification method.

2.2 Stratification

The stratification method involves partitioning the probability space into a finite number, say k, of strata. Then, the original estimation problem turns into that of k estimation subproblems. If the "size" of the strata (their probabilities) is known then one can to assemble the subproblem estimators to construct an estimator for the original problem without introducing additional variance. If the resources (i.e., total number of samples) are appropriately allocated to the estimation subproblems, this approach is guaranteed to reduce the variance (compared to crude MC).

More precisely, let $\{A_1, \cdots A_k\}$ denote a partition of $\Omega = [0,1]^d$. Let $p_i = P(A_i)$, $\mu_i = E[Y_i] = E[Y|\mathbf{U} \in A_i]$ and $\sigma_i^2 = Var[Y_i] = Var[Y|\mathbf{U} \in A_i]$. Let $\hat{\mu}_i$ be an estimator of μ_i for $i = 1, \cdots k$. Then the stratified estimator of μ is

$$\hat{\mu}_{st} = p_1\hat{\mu}_1 + \cdots + p_k\hat{\mu}_k.$$

It is easy to see that the variance of this estimator is

$$Var(\hat{\mu}_{st}) = \sum_{i=1}^{k} p_i Var(\hat{\mu}_i) = E[Var(Y|X \in A_i)] \leq Var(Y).$$

In other words, if the effort to generate a stratified estimator is the same as that of a crude MC estimator then stratification is always beneficial. The magnitude of the benefit depends on the choice of stratification.

Given a fixed partition, it is well known that the optimal allocation of samples is according to quantities q_i

$$q_i = \frac{p_i\sigma_i}{\sum_{j=1}^{k} p_j\sigma_j},$$

i.e., the number of samples out of n allocated to stratum A_i, denoted by n_i, is given by $n_i = \lfloor n * q_i \rfloor$. The minimum variance is given by

$$\sigma^{*2} = (\sum_{i=1}^{k} p_i\sigma_i)^2.$$

Once a partition is selected, optimal sampling within strata requires knowing σ_i's or estimating them. In most cases, these values are not known in advance and need to be estimated via pilot runs.

3 MONOTONICITY

In this section we consider different implications of monotonicity for strata construction.

3.1 Monotone partitioning

The intuition behind stratification is that it eliminates across strata variation. Within strata variation is reduced via sampling. This suggests creating strata in such a way that elements of each stratum lead to "similar" output values and hence to small variance. An implication of this observation is that it is desirable to consider partitions $\{A_1, A_2, \cdots, A_k\}$ that are monotone in the sense that

$$A_1 \leq A_2 \leq \cdots \leq A_k$$

where

$$A_i \leq A_j \Leftrightarrow f(U) \leq f(V) \text{ for all } U \in A_i \text{ \& } V \in A_j.$$

In this case

$$A_1 \leq A_2 \leq \cdots \leq A_k \Leftrightarrow f(A_1) \leq f(A_2) \leq ... \leq f(A_k)$$

where $f(A_i)$ are subsets of the real line and the monotonicity of such subsets is defined naturally as follows: one subset is smaller than another if all its elements are smaller than the elements of the other.

Recall the setting of problem A in the introduction in which the question of optimal boundary selection for optimal partitioning of frequency tables was posed. That question is essentially the same as the problem of optimal selection of a partition of the form $f(A_1) \leq f(A_2) \leq ... \leq f(A_k)$ for the range of values of f, i.e., the "frequency table" of Y. Our discussion above implies that the latter problem is closely related to the optimal stratification of $\Omega = [0,1]^d$.

Let's go one step further. The pull-back of a monotone partition of the range of Y,i.e., $f(\mathbf{U})$, via the monotone function g as defined in (4) is itself a monotone partition of $(0,1)$. Such a partition corresponds to the selection of a finite number of subintervals of $(0,1)$. In other words, monotonicity of g implies a correspondence between partitioning of $(0,1)$ into subintervals and those of range of Y into subintervals. Therefore, optimal partitioning of $\Omega = [0,1]$ can be reformulated as a problem of optimal partitioning of $(0,1)$ given the monotone function g.

This brings us to the question of using stratification to solve problem (4),namely integrating a monotone function on $(0,1)$

3.2 Integration of a Monotone Function

The formulation of problem A in the introduction, namely the problem of optimal stratification of a frequency table, does not take into account the cost of creating the stratification which may require sampling g at many points. To include this cost we consider the formulation of *information based complexity* for the estimation problem (4). We modify problem (4) to be able to call on results available in this setting. Assume f is a bounded function. Therefor Y is a bounded random variable. This allows us to extend the range of g to the closed interval $[0,1]$. Therefore, we consider the following integration problem.

$$\mu = \int_0^1 g(u)du = E[g(U)] = E[Y] \qquad (5)$$

where $U \sim U[0,1]$. Note that g is a monotone increasing (nondecreasing) function on $[0,1]$.

The only known information about g is that it is increasing and no other regularity properties are assumed about g. This is the a priori information. Let $\mathscr{G} = \{g : [0,1] \to R; g \text{ increasing}\}$ be the set of increasing functions on $[0,1]$. Additional information can be obtained by sampling g, i.e., by evaluating g at points in $[0,1]$. Let $x_1, \cdots, x_n$ be n distinct points in $[0,1]$. Let $I(g; x_1, \cdots, x_n) = ((x_1, g(x_1)), \cdots, (x_n, g(x_n)))$ represent the new information about g based on sampling. To simplify notation, we use the above notation for non-adaptive or adaptive and deterministic or stochastic sampling. In other words, x_i may be random variables and the choice of x_i may depend on previous samples. Moreover, to further simplify the notation, we often write $I(g;n)$ or simply I to denote this information.

Consider the stochastic sampling case. In other words, we assume x_i's are stochastically selected. Let

$$N(I) = N(I;n) = \{g' \in \mathscr{G}; I(g';n) = I(g;n)\}.$$

$N(I)$ represents the "uncertainty" associated with the information I and it is the set of functions that are indistinguishable from g given the information I. Let $S : \mathscr{G} \to R$ be the integration operator, i.e., $S(g) = \int_0^1 g(u)du$. Let $c(I) \in R$ denote an estimate of μ based on I (note that $c(I)$ is a random variable). Then for any $g' \in N(I)$, $e(g';c(I)) = |s(g') - c(I)|$ be the magnitude of the error. The *optimal* estimate of μ, denoted by $\phi(I)$, is defined in the following *worst case* sense

$$\phi(I) = \text{argmin}_{c(I)}\{\sup\{e(g';c(I))]; g' \in N(I)\}\}.$$

$e(I;n) = \sup\{e(g';c(I))]; \phi(I)); g' \in N(I)\}$ is the worst case error given I. Let

$$e(n) = \sup\{E[e(I;n)]; g \in \mathscr{G}\}$$

where expectation is with respect to the measure induced by the stochastic (Monte Carlo) sampling algorithm. Note that $e(n)$ depends on the stochastic algorithm used even though this fact is not explicitly indicated in the notation. The value n is a crude yet relevant stand in for the cost of computation.

4 OPTIMAL STRATIFICATION

We are now ready to consider the issue of optimal Monte Carlo algorithms where optimality is interpreted as the asymptotic rate of convergence. We then show consider optimal stratification in this setting.

4.1 Optimal asymptotic rate of convergence

The following results (and their proofs) are provided in (Novak 1992) in the context of the integration of a monotone function g on $[0,1]$, i.e., the problem described in the previous section.

Theorem 1. *For each nonadaptive Monte Carlo method*

$$e(n) \geq \frac{1}{8} n^{-1}.$$

In other words, the optimal rate of convergence of non-adaptive Monte Carlo algorithms cannot be faster than $O(n^{-1})$.

Theorem 2. *For each adaptive Monte Carlo method*

$$e(n) \geq \frac{\sqrt{2}}{32} n^{-3/2}.$$

In other words, the optimal rate of convergence of adaptive Monte Carlo algorithms cannot be faster than $O(n^{-3/2})$. (Novak 1992) provides a specific adaptive algorithm that achieved the $O(n^{-3/2})$ rate of convergence and hence can be viewed as an optimal algorithm in this sense. The algorithm can be considered a stratification algorithm.

We now turn to directly consider optimal stratification algorithms where optimality is defined in terms of rate of convergence. In other words, algorithms with rate of convergence $O(n^{-3/2})$ will be considered optimal.

4.2 Optimal stratification

We first establish some notation. Let $x_0 = 0 < x_1 < \cdots < x_k = 1$ denote the boundaries of a partition of $[0,1]$ into k

subintervals. Let $\delta_i = x_i - x_{i-1}$ and $\delta g_i = g(x_i) - g(x_{i-1})$ for $i = 1, \cdots, k$. Finally, let $\delta g = g(1) - g(0)$.

Let $V \sim U(x_{i-1}, x_i)$ and assume $g(x_{i-1})$ and $g(x_i)$ are given. Then for all h monotone on $[x_{i-1}, x_i]$ where $h(x_{i-1}) = g(x_{i-1})$ and $h(x_i) = g(x_i)$, one can easily show that

Lemma 3.

$$Var(h(V)) \leq \frac{1}{4}(g(x_i) - g(x_{i-1}))^2 = \frac{1}{4}(\delta g_i)^2.$$

Consider the above partition of $[0,1]$. Let, q_i, the proportion of samples allocated to stratum i be given by

$$q_i = \frac{\delta_i \cdot \delta g_i}{\sum_{j=1}^{k} \delta_j \cdot \delta g_j} \tag{6}$$

and let $n_i = nq_i$. Consider a stratified sampling algorithm where n_i samples are randomly allocated to stratum i (in what follows we disregard the minor issue that n_i is not necessarily an integer in which case the integer part of n_i needs to be allocated + a scheme for allocating the remaining samples). Let $\hat{\mu}_{st}$ denote the resulting stratified estimator based on n samples. Then we have

$$Var(\hat{\mu}_{st}) = \frac{1}{n}[\sum_{i=1}^{k} \frac{(\delta_i)^2}{q_i} \sigma_i^2]$$

where σ_i^2 is the variance of a $g(V)$ for $V \sim U(x_{i-1}, x_i)$.

Given $\sigma_i^2 \leq 1/4(\delta g_i)^2$ from above lemma, we have

$$Var(\hat{\mu}_{st}) \leq \frac{1}{4n}(\sum_{i=1}^{k} \delta_i \cdot \delta g_i)^2$$

Now consider the stratification (Strat I) algorithm given by Figure1.

1. Partition $[0,1]$ into k equal length subintervals/strata. Let $x_0, \cdots, x_k$ be the strata boundaries as defined above. Sample the function g at x_i's $i = 1, \cdots, k$.
2. Allocate $n_i = n \cdot q_i$ random samples to stratum i, where q_i is defined by identity 6.
3. Evaluate the stratified sampling estimator.

Figure 1: Strat I Algorithm

Then we have the following result.

Theorem 4. *If Strat I algorithm is used with n strata to estimate* $\mu = \int_0^1 g(u)du$ *then* $e(n) = O(n^{-3/2})$. *In other words, Strat I algorithm is an optimal algorithm in the sense that it has an optimal asymptotic rate of convergence.*

Proof. From our earlier discussion we have

$$\begin{aligned} Var(\hat{\mu}_{stI}) &\leq \frac{1}{4n}(\sum_{i=1}^{n} \delta_i \cdot \delta g_i)^2 \\ &= \frac{1}{4n}(\frac{1}{n})^2(\sum_{i=1}^{n} \delta g_i)^2 = \frac{1}{4n^3}(\delta g)^2. \end{aligned}$$

If we limit ourselves, as in (Novak 1992), to increasing functions such that $\delta g = 1$, the above implies

$$e(n) = \mathrm{O}(n^{-3/2}).$$

We make the following observations.

- Given our criterion for optimality it is sufficient to adapt only at the level of sample allocation. In other words, we note that the stratification is done with no adaption to function g. The only adaptation that uses the information about g is at the level of sample allocation. It is important to emphasize that our result is partially due to the fact that we have a monotone stratification (equivalently g is monotone). This conclusion, i.e., that adaptation at the level of sample allocation is sufficient for achieving optimality, is not valid in general.
- δg_i's, obtained via $n+1$ function evaluations, provide convenient guidelines for optimal sample allocation.
- (Cheng and Davenport 1989) appropriately warn against tailoring stratification to specific outputs of the simulation. They argue that often the objective of a simulation is to estimate several outputs simultaneously and designing stratifications that are best suited for one output may not be appropriate for another. This is a valid argument. We note that for the particular outputs they specify i.e., second moment of Y, $E[Y^2]$, and probabilities of the form $E[I\{Y \leq y\}]$, strat I algorithm remains optimal if sample allocation is appropriately adjusted. These outputs are monotone functions or the original output Y and therefore can be viewed as monotone function on $[0,1]$ in their own right and strat I algorithm can be easily adapted to them. (note that additional sampling may be, and in general will be needed.

The optimality results of this section may provide some guidelines for effective stratification. The key message is that to obtain optimal stratifications one needs to focus on the range space of the random variable of interest rather than its domain ($\Omega = [0,1]^d$ in our case). If such an approach were practical, the problem of dimensionality (of the domain) would be resolved. Stratifying the domain via a pull-back of a stratification of the range space is in general not directly applicable. But, as mentioned above, it may provides clues for effective stratification. (Cheng and Davenport 1989) consider one such approach by utilizing what they call *shadow responses*. In the next section we briefly describe another approach to make use of the insights obtained from the results of this section.

5 STRUCTURED DATABSE MONTE CARLO (SDMC)

In SDMC approach a finite population approximation to the estimation/integration problem (3) is considered by generating N samples uniformly in $\omega = [0,1]^d$ where N is assumed to be large. By structuring, i.e., ordering, ω_i's according to their functional value ($\omega_i \leq \omega_j$ iff $g(\omega_i) \leq g(\omega_j)$ one obtains a finite population approximation to problem (4).

The ordered finite population is called the structured database. Our analysis of stratification for problem (4) now directly applies to, and has practical implications for, the structures database. The stratification of the structured database is also closely related to stratifying a frequency table discussed in problem A of the introduction.

A key question in the SDMC setting (or more generally in the Database Monte Carlo (DBMC) setting) is whether the setup cost of the method justifies the benefits that it can provide. In a way very similar to the stratification of tax returns discussed in the introduction, where the effort in obtaining an effective stratification for one calendar year is expected to accrue estimation benefits in the subsequent years, the SDMC method is intended for repeated use for problem similar to one for which the database is structured. For details, see, e.g., (Zhao, Zhou, and Vakili 2006) and (Zhao, Borogovac, and Vakili 2007).

6 CONCLUSION

We considered the problem of obtaining effective stratification for the variance reduction technique of stratification. We reviewed some work that implicitly or explicitly have addressed this issue. We show that monotonicity in one way or another is intimately connected to designing effective strata. For a particular notion of optimality, namely asymptotic rate of convergence, we provide a generic stratification algorithm. For practice it is worth considering more stringent finite sample optimality criteria and find stratifications that are optimal or near optimal. This problem is a subject of our future research.

ACKNOWLEDGMENTS

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

REFERENCES

Asmussen, S., and P. Glynn. 2007. *Stochastic simulation: Algorithms and analysis*. Springer.

Cheng, R. C., and T. Davenport. 1989. The problem of dimensionality in stratified sampling. *Management Science* 35 (11): 1278–1296.

Cochran, W. G. 1961. Comparison of methods for determining startum boundaries. *Bulletin of The International Statistical Institute* 38:345–358.

Cochran, W. G. 1977. *Sampling techniques*. Wiley.

Glasserman, P. 2004. *Monte carlo methods in financial engineering*. Springer Verlag.

Kiefer, J. 1957. Optimal sequential search and approximation methods under minimum regularity assumptions. *J. Soc. Indust. Appl. Math.* 5 (3): 105–136.

Novak, E. 1992, May. Quadrature formulas for monotone functions. *Proceedings of the American Mathematical Society* 115 (1): 59–68.

Sukharev, A. 1987. The conept of sequential optimality for problems in numerical analysis. *Journal of Complexity* 3:347–357.

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

Zhao, G., T. Borogovac, and P. Vakili. 2007. Efficient estimation of option price and price sensitivities via structured database monte carlo (SDMC). In *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M. H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 984–991.

Zhao, G., Y. Zhou, and P. Vakili. 2006. A new efficient simulation strategy for pricing path – dependent options. In *Proceedings of the 2006 Winter Simulation Conference*, ed. L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 703–710.

AUTHOR BIOGRAPHIES

GANG ZHAO is a Ph.D. student of Systems Engineering at Boston University. His research interests include novel strategies for Monte Carlo simulation, stochastic optimization, American style financial derivative pricing . His email address is <gzhao@bu.edu>.

PIROOZ VAKILI is an Associate Professor in the Division of Systems Engineering and the Department of Mechanical Engineering at Boston University. His research interests include Monte Carlo simulation, optimization, computational finance, and bioinformatics. His email address is <vakili@bu.edu>.