ORDINAL OPTIMIZATION: A NONPARAMETRIC FRAMEWORK

Peter W. Glynn

Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

Sandeep Juneja

School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, 400 005, INDIA

ABSTRACT

Simulation-based ordinal optimization has frequently relied on large deviations analysis as a theoretical device for arguing that it is computationally easier to identify the best system out of d alternatives than to estimate the actual performance of a given design. In this paper, we argue that practical implementation of these large deviations-based methods need to estimate the underlying large deviations rate functions of the competing designs from the samples generated. Because such rate functions are difficult to estimate accurately (due to the heavy tails that naturally arise in this setting), the probability of mis-estimation will generally dominate the underlying large deviations probability, making it difficult to build reliable algorithms that are supported theoretically through large deviations analysis. However, when we justify ordinal optimization algorithms on the basis of guaranteed finite sample bounds (as can be done when the associated random variables are bounded), we show that satisfactory and practically implementable algorithms can be designed.

1 INTRODUCTION

Suppose that a simulationist is faced with the problem of determining the 'best' of d alternative designs for a system, on the basis of Monte Carlo simulation of each of the designs. This describes the basic 'ordinal optimization' problem, in that our goal is only to identify the best design and not to actually estimate the performance. More precisely, we assume that the simulationist is comparing each of the d different designs on the basis of an associated (random) performance measure $X(i), i \leq d$, and that the goal is to identify the parameter i^* (assumed unique) for which

$$\mu(i^*) = \max_{1 \leq j \leq d} \mu(j),$$

where $\mu(j) \triangleq EX(j), 1 \leq j \leq d$. We further assume that the simulationist has the ability to generate independent and identically distributed (iid) realizations of each of the d random variables (rv's) $X(j), 1 \leq j \leq d$.

A great deal is known about the construction of appropriate algorithms when the X(i)'s are Gaussian rv's (i.e., normally distributed); see, for example, Goldsman and Nelson (2001), Kim and Nelson (2003), Nelson et. al. (2001), Branke, Chick and Schmidt (2005). Our goal, in this paper, is to describe a framework for the study of such ordinal optimization problems when the underlying rv's are non-Gaussian, as is typical in the simulation context.

We discuss two different types of algorithms for ordinal optimization in the nonparametric setting, the first based on large-sample asymptotics as the theoretical justification and the second based on the use of additional problem structure to construct guaranteed bounds that hold for any sample size. We introduce this differentiation between algorithms that come with guarantees and those based on large-sample approximations in Section 2, in the simpler and more easily understood context of confidence interval construction. In Section 3, we turn to the ordinal optimization problem, and point out that there are two natural asymptotic regimes that arise in this setting and that can be used as theoretical justification for guiding the construction of ordinal optimization algorithms. One is based on asymptotics that arise when the 'indifference margin' is small (and can be analyzed through central limit approximations), whereas the other arises in settings where one is insisting that the probability of incorrectly eliminating the best system be small (leading to consideration of large deviation asymptotics). Our principal interest here is in the second context. We argue that any reasonable algorithm will need to estimate the final probability of incorrect selection from the sample, thereby leading to a requirement that the large deviations rate functions for the underlying random variables be estimated from the data. However, the empirical quantities that arise in estimating rate functions tend to be heavy tailed, making it problematic to estimate the probability of incorrect selection without introducing errors of a (much) larger order of magnitude than that associated with the probability of incorrect selection. As a consequence, ordinal optimization algorithms that are justified through large deviations asymptotics face fundamental difficulties in their implementation. On the other hand, we argue in Section 3 that when additional problem structure is imposed (e.g. assuming the underlying random variables are bounded), it is easy to derive probability guarantees that hold for all sample sizes, thereby providing ordinal optimization algorithms that meet the probability of incorrect selection requirement.

2 THE BASIC FRAMEWORK FOR SIMULATION-BASED PERFORMANCE EVALUATION

We review here the basic approach used within the simulation community to study simulation-based algorithms for purposes of performance evaluation. This discussion will help set the stage for our ordinal optimization framework.

Suppose that we wish to use Monte Carlo simulation to compute EX to a given precision ε. Because the estimator for EX contains sampling-based randomness, this uncertainty is usually communicated via a confidence interval for EX. In particular, given $\varepsilon > 0$ and $\delta \in (0,1)$, our goal is to then produce a confidence interval $[L, R]$ for which $|L - R| \le 2\varepsilon$ and

$$P(EX \in [L, R]) \ge 1 - \delta. \tag{1}$$

A commonly described (two-stage) algorithm for producing such confidence intervals is discussed next.

Algorithm **A1**

1. Find z so that $P(N(0,1) > z) = \delta/2$, where $N(0,1)$ is a Normal rv with mean zero and unit variance.
2. Choose m large (say, $m = \lfloor 1/\varepsilon \rfloor$).
3. Generate m iid copies $X_1', \ldots, X_m'$ of the rv X and compute

$$s_m^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i' - \frac{1}{m} \sum_{i=1}^{m} X_j')^2$$

4. Set $n = \lfloor (z^2/\varepsilon^2) s_m^2 \rfloor$.
5. Generate n iid copies $X_1, X_2, \ldots, X_n$ of the rv X, and compute

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

6. Return $L = \bar{X}_n - \varepsilon$, and $R = \bar{X}_n + \varepsilon$.

The theoretical support for the above algorithm is provided by the following result; see (Nakayama 1994) for related literature.

Proposition 1 Suppose that $0 < varX \triangleq \sigma^2 < \infty$. Then, $R - L = 2\varepsilon$, and

$$\lim_{\varepsilon \downarrow 0} P(EX \in [L, R]) = 1 - \delta.$$

Note that this theoretical support is only asymptotic. This is not surprising as much of stochastic analysis relies on asymptotic simplifications to make the analysis tractable. However, from an algorithmic viewpoint, this is somewhat unsatisfactory, as one has no guarantee that for a given instance of X, and specified values of ε and δ, that $[L, R]$ does indeed satisfy (1). Nevertheless, the simulation (and statistics) communities have generally accepted Proposition 1 as adequate theoretical support for methods like Algorithm **A1**.

If one desires an algorithm for which the guarantee (1) can be made, one needs to assume more about the distribution of X. Such guarantees are easy to derive when X is a bounded rv for which the simulationist has knowledge of bounds a and b for which $P(X \in [a, b]) = 1$, provided that one suitably modifies Algorithm **A1**.

Algorithm **A2**

1. Choose $n = \lceil \frac{(b-a)^2}{4\varepsilon^2 \delta} \rceil$.
2. Generate n iid copies $X_1, X_2, \dots, X_n$ of the rv X, and compute

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

3. Return $L = \bar{X}_n - \varepsilon$ and $R = \bar{X}_n + \varepsilon$.

The support for **A2** is derived from Chebyshev's inequality. Observe that

$$P(|\bar{X}_n - EX| > \varepsilon) \le \frac{var\,X}{n\varepsilon^2}.$$

In order that $\frac{var\,X}{n\varepsilon^2} \le \delta$, we must require that

$$n \ge \frac{var\,X}{\delta\varepsilon^2}.$$

But, as is well known, $var\,X \le (b-a)^2/4$ for a rv supported on $[a, b]$, since the maximum variance rv has mass $1/2$ on both a and b. This discussion can be summarized via our next proposition.

Proposition 2 Suppose $P(X \in [a, b]) = 1$. If $[L, R]$ is constructed through Algorithm **A2**, then $R - L = 2\varepsilon$ and

$$P(EX \in [L, R]) \ge 1 - \delta.$$

An improved version of **A2** can also be supplied.

Algorithm **A3**

1. Choose $n = \lceil \frac{(b-a)^2 \log(2/\delta)}{2\varepsilon^2} \rceil$.
2. Generate n iid copies $X_1, X_2, \dots, X_n$ of rv X, and compute

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

3. Return $L = \bar{X}_n - \varepsilon$ and $R = \bar{X}_n + \varepsilon$.

The theoretical support for **A3** rests upon Hoeffding's inequality for bounded rv's:

$$P(|\bar{X}_n - EX| > \varepsilon) \leq 2\exp(-\frac{2n\varepsilon^2}{(b-a)^2}).$$

This inequality immediately leads to the following guarantee for Algorithm **A3**.

Proposition 3 Suppose $P(X \in [a,b]) = 1$. If $[L,R]$ is constructed through Algorithm **A3**, then $R - L = 2\varepsilon$ and

$$P(EX \in [L,R]) \geq 1 - \delta.$$

Algorithms **A2** and **A3**, despite their coverage guarantees, are rarely used in practice because they impose a significant additional cost on the simulationist in terms of extra sampling. In particular, when ε is small, the value of n dictated by **A1** satisfies

$$n \sim var\, X \frac{z^2}{\varepsilon^2}$$

as $\varepsilon \downarrow 0$ ($(a_\varepsilon : \varepsilon > 0)$ is said to be asymptotic to $(b_\varepsilon : \varepsilon > 0)$, denoted by $a_\varepsilon \sim b_\varepsilon$ as $\varepsilon \downarrow 0$, if $\lim_{\varepsilon \downarrow 0} \frac{a_\varepsilon}{b_\varepsilon} = 1$).
Because

$$P(N(0,1) > x) \leq \exp(-x^2/2)/2$$

for $x > 0$ (see, e.g., Feller 1971), evidently the z value satisfying $P(N(0,1) > z) = \delta/2$ is such that $z \leq \sqrt{2\log(1/\delta)}$. Consequently, for ε small, the sample size n of **A1** grows no faster than $var\, X \frac{2}{\varepsilon^2}\log(1/\delta)$. Of course, when X lies in $[a,b]$, $var\, X \leq (b-a)^2/4$; the resulting bound $(b-a)^2 \cdot (\frac{1}{2\varepsilon^2})\log(1/\delta)$ is smaller than the sample sizes determined by either **A2** or **A3**. The key point is that the typical choice made by simulationists is to dispense with the coverage guarantee in order to obtain a smaller sample size (and thereby faster computation times). Of course even without the coverage guarantee, one still has asymptotic coverage as described in Proposition 1.

3 AN ORDINAL OPTIMIZATION FRAMEWORK

We now turn to the ordinal optimization setting that is our principal interest in this paper. We shall adopt an indifference zone perspective, so that goal is to screen the d alternative designs so as to eliminate from further competition those that are clearly not optimal (by an amount ε).

Specifically, given sample sizes $n_1, n_2, \ldots, n_d$, we independently simulate n_i iid copies $X_1(i), X_2(i), \ldots, X_{n_i}(i)$ of the rv $X(i)$, and form the sample mean

$$\bar{X}_{n_i}(i) = \frac{1}{n_i}\sum_{j=1}^{n_i} X_j(i)$$

for $1 \leq i \leq d$. For $\varepsilon > 0$, we then set

$$\hat{\mathscr{P}} = \{i : \bar{X}_{n_i}(i) \leq \max_{1 \leq k \leq d} \bar{X}_{n_k}(k) - \varepsilon\};$$

$\hat{\mathscr{P}}$ is the (estimated) set of eliminated alternatives. Given $\varepsilon > 0$ and $\delta > 0$, our goal is to develop an algorithm for which

$$P(i^* \in \hat{\mathscr{P}}) \leq \delta; \tag{2}$$

of course, $P(i^* \in \hat{\mathscr{P}})$ is the probability that we incorrectly eliminate the true maximizer i^* from further consideration as a result of our screening process.

In parallel with our development of Section 2, we consider first an asymptotic regime in which we can potentially apply asymptotic analysis to offer theoretical support for our screening algorithm. One possibility is to consider the behavior of the algorithm when $\varepsilon \downarrow 0$. When the 'indifference margin' ε becomes small, we need to accurately estimate the magnitude of the difference between $\mu(i^*)$ and $\mu(j)$ in order to determine j's membership in the set of eliminated alternatives. This puts us in a setting in which the algorithmic design parameters will be determined by central limit theorem (i.e., mean/variance) considerations. However, as mentioned in the Introduction, the ordinal optimization philosophy is one in which one wishes to take advantage of the fact that one is interested only in identifying i^* and not in estimating its actual performance. As a consequence, much of the stochastic simulation literature on ordinal optimization has utilized large deviations theory rather than central limit arguments to conduct asymptotic analysis; see e.g., Ho, Sreenivas and Vakili (1992), Dai (1996), Glynn and Juneja (2004), Chen et. al. (2000). In such settings, the large deviations rate function (as opposed to just mean and variances) now plays a key role both in theoretical analysis and in algorithmic design.

The asymptotic setting in which such large deviations becomes relevant is one in which we send the probability of incorrect selection to zero, so that this is of principal interest in settings in which we desire a very high probability that our screening algorithm has not inadvertently eliminated the best design from further consideration. Suppose, in particular, that the rv's $X(1),\ldots,X(d)$ are light-tailed, so that $E\exp(\theta X(i)) < \infty$ for θ in a neighborhood of the origin, and that $n_i \sim np_i$ as $n \to \infty$, where $p_1 + \cdots + p_d = 1$ and $0 < p_i < 1$, $1 \le i \le d$. In this setting, it is easy to establish (see, Glynn and Juneja 2004 for a closely related analysis) under suitable additional hypothesis that

$$\frac{1}{n}\log P(i^* \in \hat{\mathscr{P}}) \to -I \tag{3}$$

as $n \to \infty$, where I can be computed easily from the large deviations rate functions asssociated with each of the rvs $X(1),\ldots,X(d)$. Hence, by setting

$$n = \lceil \frac{1}{I}\log(1/\delta)\rceil, \tag{4}$$

we can offer theoretical support for (2). Specifically, with this choice of n,

$$\log P(i^* \in \hat{\mathscr{P}}) \sim \log(\delta) \tag{5}$$

as $\delta \downarrow 0$; this limit theorem is, in some sense, an analog to that of Proposition 1.

However, there is an important distinction to be drawn here. Note that **A1** includes steps 1-4; these steps are used to compute a value of n that will meet (at least approximately) the coverage requirement (1). The analog, in the current setting, would be a corresponding set of steps at which I would be estimated (via $\hat{I}$, say) through a set of trial runs, imposing an additional computational burden of at most order $\log(1/\delta)$ or smaller (for, otherwise, the trial runs would dominate the entire calculation from a computer time perspective). Observe that on the event $\{\hat{I} > I + \gamma\},(\gamma > 0)$, one would compute too small a value of n from the sample version of (4), this suggests that (5) would fail for the second stage runs, conditional on this event. In order that the resulting algorithm's overall false selection probability satisfy (5), this would therefore require that
$$\log P(\hat{I} > I + \gamma) = O(\log(\delta)) \tag{6}$$
as $\delta \downarrow 0$, where function $g(x)$ is said to be $O(f(x))$ if there exists a constant $K > 0$ and an $x_0 > 0$ such that $|g(x)| \le K|f(x)|$ for all $0 < x < x_0$.

The difficulty is that the natural sample estimator for $\hat{I}$ involves the empirical moment generating function of the rv's $X(1),X(2),\ldots,X(d)$. Consider, for example, the setting in which $X(i)$ has an (asymptotically) exponential right tail, so that
$$P(X(i) > y) \sim c_i e^{-\lambda_i y}$$

as $y \rightarrow \infty$. In this case,

$$
\begin{aligned}
P(e^{\theta X(i)} > y) &= P(X(i) > \frac{1}{\theta} \log y) \\
&\sim c_y \exp(-\frac{\lambda_i}{\theta} \log y) \\
&= c_i y^{-\lambda_i/\theta}
\end{aligned}
$$

as $y \rightarrow \infty$. It follows that when the $X(i)'s$ have asymptotically exponential right tails, $e^{\theta X(i)}$ is heavy-tailed for $\theta > 0$. If $0 < \theta < \lambda_i$, $E \exp(\theta X(i)) < \infty$ and for $y > E e^{\theta X(i)}$,

$$
P(\frac{1}{n} \sum_{j=1}^{n} e^{\theta X_j(i)} > y) \sim nP(e^{\theta X(i)} > ny)
$$

as $n \rightarrow \infty$; see, e.g., Borovkov and Borovkov (2008). As a consequence, the empirical estimator of the moment generating function $E e^{\theta X(i)}$ based on a sample of size n makes an error of (absolute) magnitude γ' with a probability of order $n^{1-\lambda_i/\theta}$. Hence, if n is of order $\log(1/\delta)$, an error of size γ' is incurred with a probability of order $\log(1/\delta)^{1-\lambda_i/\theta}$. Since we expect the event $\{\hat{I} > I + \gamma\}$ to occur when one or more of the moment generating functions $E e^{\theta X(i)}, (1 \leq i \leq d)$ is mis-estimated by an amount γ', this suggests that $P(\hat{I} > I + \gamma)$ is of order $\log(1/\delta)^{-\beta}$ (for some $\beta > 0$). This, in turn, suggests that $\log P(i^* \in \hat{\mathscr{P}})$ is (at least) of order $-\beta \log \log(1/\delta)$. This contradicts (5), which requires that $\log P(i^* \in \hat{\mathscr{P}}) \sim -\log(1/\delta)$ as $\delta \downarrow 0$.

The above heuristic argument strongly suggests that the need to estimate the rate constant I leads to an algorithm that is in violation of (5). The basic problem is that that because the empirical moment generating function is much more heavy tailed than is $X(i)$ itself, one makes errors in estimating I (via $\hat{I}$) with a probability of much larger order then the theoretical probability δ of incorrect selection as dictated through (2).

One might hope to circumvent this difficulty through a posterior error estimate. Specifically, suppose that we decide a priori to put $n_i = \lceil \log(1/\delta) \rceil$ for $1 \leq i \leq d$, run the corresponding simulations for each of the d designs, and compare $\hat{\mathscr{P}}$. The relation (3) then applies to the $\hat{\mathscr{P}}$ so constructed. While this is an interesting fact theoretically, the simulationist (practically speaking) needs an estimate for the probability $P(i^* \in \hat{\mathscr{P}})$ of incorrect selection in order that the algorithm be useful. In this setting, the obvious a posteriori estimate for $P(i^* \in \hat{\mathscr{P}})$ is $\exp(-n\hat{I})$, where $\hat{I}$ is an estimator for I constructed from the $d \lceil \log(1/\delta) \rceil$ simulations conducted. Of course, for the same reasons as discussed above, we expect that the estimator $\hat{I}$ for I will differ from I by at least γ with a probability far exceeding the desired probability δ of incorrect selection.

The above discussion suggests that supporting an ordinal optimization algorithm on the basis of an asymptotic analysis in which $\delta \downarrow 0$ is problematic. An alternative is to consider more restricted problem contexts in which guarantees can be made (rather than through a small δ approximation).

Specifically, suppose that $X(i) \in [a,b]$ with probability one for $1 \leq i \leq d$, and $n_i = \lceil n/d \rceil$ for $1 \leq i \leq d$. Then, $X(i) - X(j)$ has a distribution with support in $[a-b, b-a]$, so that

$$
\begin{aligned}
P(i^* \in \hat{\mathscr{P}}) \quad &= \quad P(\bar{X}_{n_{i^*}}(i^*) \leq \max_{1 \leq k \leq d} \bar{X}_{n_k}(k) - \varepsilon) & (7) \\
&\leq \quad \sum_{k \neq i^*} P(\bar{X}_{n_{i^*}}(i^*) \leq \bar{X}_{n_k}(k) - \varepsilon) & (8) \\
&= \quad \sum_{k \neq i^*} P(\bar{X}_{n_{i^*}}(i^*) - \bar{X}_{n_k}(k) - (\mu(i^*) - \mu(k)) \leq -\varepsilon - (\mu(i^*) - \mu(k))) & (9) \\
&\leq \quad \sum_{k \neq i^*} P(\bar{X}_{n_{i^*}}(i^*) - \bar{X}_{n_k}(k) - (\mu(i^*) - \mu(k)) \leq -\varepsilon) & (10) \\
&\leq \quad \sum_{k \neq i^*} P(|\bar{X}_{n_{i^*}}(i^*) - \bar{X}_{n_k}(k) - (\mu(i^*) - \mu(k))| \geq \varepsilon) & (11) \\
&\leq \quad 2(d-1)\exp(-\frac{n\varepsilon^2}{2(b-a)^2}), & (12)
\end{aligned}
$$

where we used Hoeffding's inequality at the last inequality. This calculation, suggests the following algorithm

Algorithm **B1**

1. Set $n = \lceil \log(2(d-1)/\delta)2(b-a)^2/\varepsilon^2 \rceil$.
2. For $1 \leq i \leq d$, independently generate $\lceil n/d \rceil$ iid copies $X_1(i), X_2(i), \ldots, X_{\lceil n/d \rceil}(i)$ of $X(i)$.
3. Form

$$
\bar{X}(i) = \frac{1}{\lceil n/d \rceil} \sum_{i=1}^{\lceil n/d \rceil} X_j(i)
$$

   for $1 \leq i \leq d$.
4. Return $\hat{\mathscr{P}} = \{j : \bar{X}(j) \leq \max_{1 \leq k \leq d} \bar{X}(k) - \varepsilon\}$.

Given the calculation leading to (12), the following result is immediate.

Proposition 4 If $P(X(i) \in [a,b]) = 1$ for $1 \leq i \leq d$, then the set $\hat{\mathscr{P}}$ constructed through the algorithm **B1** satisfies

$$
P(i^* \in \hat{\mathscr{P}}) \leq \delta.
$$

A key element in Proposition 4 is that the total number n of samples taken is of order $\log(1/\delta)$ as $\delta \downarrow 0$ so that the sample size increases slowly with $1/\delta$.

Of course, a similar growth in n (of order $\log(1/\delta)$) would ensue if the sample size n were instead determined at Step 1 of Algorithm **B1** on the basis of normal approximations to the distribution of the $\bar{X}(i)'s$. (This is essentially the approximation used within 'selection of best system' algorithms.) The problem is that because the central limit approximation fails for large deviations probabilities of the type that arise when $\delta \downarrow 0$, the resulting algorithm would fail to satisfy (2) or (5).

We conclude that Algorithm **B1** is a suitable 'ordinal optimization' analog to Algorithm **A3**.

REFERENCES

Borovkov, A. A., and K. A. Borovkov. 2008. *Asymptotic Analysis of Random Walks : Heavy-Tailed Distributions*: Cambridge University Press.

Branke J, S. E. Chick, C Schmidt, 2007, "Selecting a Selection Procedure", *Management Science*, 53 (12): 1916-1932.

Chen, C.H., J. Lin, E. Yucesan and S. E. Chick. 2000. "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization". *Journal of Discrete Event Dynamic Systems: Theory and Applications*, 10: 251-270.

Dai, L. 1996. "Convergence Properties of Ordinal Comparison in the Simulation of Discrete Event Dynamic Systems". *Journal of Optimization Theory and Applications,* 91 (2): 363-388.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications,* Vol. II, New York: Wiley.

Glynn, P. W. and S. Juneja. 2004. "A Large Deviations Perspective on Ordinal Optimization". In *Proceedings of the 2004 Winter Simulation Conference*. Edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 577-585. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Goldsman, D., and B. L. Nelson. 2001. "Statistical Selection of Best System". In *Proceedings of the 2001 Winter Simulation Conference*. Edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 139-146. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Ho, Y. C., R. S. Sreenivas, and P. Vakili. 1992. "Oridnal Optimization of DEDS". *Discrete Event Dynamic Systems: Theory and Applications*, **2**: 61-88.

Kim, S. H., and B. L. Nelson. 2003. "Selecting the Best System: Theory and Methods". In *Proceedings of the 2003 Winter Simulation Conference.* Edited by S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, 101-112. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Nakayama, M. 1994. "Two-stage stopping procedures based on standardized time series". *Management Science* 40: 1189-1206.

Nelson, B. L., and F. J. Matejcik. 1995. "Using Common Random Numbers for Indifference-Zone Selection and Multiple Comparisons in Simulation". *Management Science* **41**: 1935-1945.

Nelson, B. L., J. Swann, D. Goldsman and W. Song. 2001. "Simple Procedures for Selecting the Best Simulated System when the Number of Alternatives is Large". *Operations Research* **49**: 950-963.

AUTHOR BIOGRAPHIES

PETER W. GLYNN is a Professor and Chair, Management Science and Engineering at Stanford University. He received his Ph.D in Operations Research from Stanford University in 1982. He then joined the faculty of the University of Wisconsin at Madison, where he held a joint appointment between the Industrial Engineering Department and Mathematics Research Center, and courtesy appointments in Computer Science and Mathematics. In 1987, he returned to Stanford, where he joined the Department of Operations Research. He is now the Thomas Ford Professor of Engineering in the Department of Management Science and Engineering, and also holds a courtesy appointment in the Department of Electrical Engineering. From 1999 to 2005, he served as Deputy Chair of the Department of Management Science and Engineering, and was Director of Stanford's Institute for Computational and Mathematical Engineering from 2006 until 2010. He is a Fellow of INFORMS and a Fellow of the Institute of Mathematical Statistics, has been co-winner of Best Publication Awards from the INFORMS Simulation Society in 1993 and 2008, was a co-winner of the Best (Biannual) Publication Award from the INFORMS Applied Probability Society in 2009, and was the co-winner of the John von Neumann Theory Prize from INFORMS in 2010. His research interests lie in computational probability, queueing theory, statistical inference for stochastic processes, and stochastic modeling. His email address is glynn@stanford.edu and his web page is http://www.stanford.edu/~glynn/.

SANDEEP JUNEJA is a Professor in School of Technology and Computer Science at the Tata Institute of Fundamental Research. He received his Ph.D in Operations Research from Stanford University in 1994. His research interests include applied probability, computational finance, simulation and rare-event analysis. He serves on the editorial board on Mathematics of Operations Research. His email address is juneja@tifr.res.in and his web page is http://www.tcs.tifr.res.in/~sandeepj.