

DISCRETE STOCHASTIC OPTIMIZATION VIA A MODIFICATION OF THE STOCHASTIC RULER METHOD

Mahmoud H. Alrefaei

Department of Industrial Engineering
University of Wisconsin – Madison
1513 University Avenue
Madison, Wisconsin 53706, U.S.A.

Sigrún Andradóttir

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332, U.S.A.

ABSTRACT

In this paper, we present a modification of the stochastic ruler method for solving discrete stochastic optimization problems. Our method generates a stationary Markov chain sequence taking values in the feasible set of the underlying discrete optimization problem. The number of visits to every state by this Markov chain is used to estimate the optimal solution. Unlike the original stochastic ruler method, our method is guaranteed to converge *almost surely* to a global optimal solution. We present empirical results that illustrate the performance of our method, and we show that these results compare favorably with empirical results obtained using the original stochastic ruler method.

1 INTRODUCTION

Developing efficient methods for solving discrete stochastic optimization problems is an important area of research because such methods could be used to solve important problems in many fields, including manufacturing systems, management science, and logistics. The discrete stochastic optimization problem can be represented as follows:

$$\min_{x \in \mathcal{S}} f(x) = E[h(x, Y_x)], \quad (1)$$

where \mathcal{S} is a discrete set, and it is assumed to be finite throughout this paper, h is a deterministic real-valued function, and Y_x is a random variable that depends on the parameter $x \in \mathcal{S}$. For simplicity define the random variable $H(x) = h(x, Y_x)$ for all $x \in \mathcal{S}$, and let $\mathcal{S}^* = \{x^* \in \mathcal{S} : f(x^*) \leq f(x) \text{ for all } x \in \mathcal{S}\}$ denote the set of global solutions to the optimization problem (1).

Approaches that have been designed to solve the discrete stochastic optimization problem include ranking and selection procedures and multiple comparisons procedures. These procedures can be effi-

cient when the number of alternatives is small. Goldman, Nelson, and Schmeiser (1991) present demonstrations of three methods for discrete stochastic optimization (interactive analysis, ranking and selection, and multiple comparisons) on an airline-reservation-system simulation problem. For more details on ranking and selection and multiple comparisons methods, see Bechhofer, Santner, and Goldsman (1995).

Recently Andradóttir (1995, 1996) has proposed two methods for solving discrete stochastic optimization problems. These two methods involve generating Markov chains. The state that is visited most often by these Markov chains is used as an estimate of the solution. She has shown that these two methods converge to an optimal solution almost surely. Alrefaei and Andradóttir (1995) have proposed another method for discrete stochastic optimization that resembles the simulated annealing algorithm. They use the same criterion that Andradóttir (1995, 1996) used to estimate the solution. They have shown that their method converges to a global optimal solution almost surely. Gelfand and Mitter (1989) proposed to solve the discrete stochastic optimization problem using the simulated annealing algorithm. They show that this algorithm converges to a global optimal solution when the evaluations of the objective function values include noise that is normally distributed with mean zero and a small variance. Lee (1995) generalizes some of the results of Gelfand and Mitter (1989) to the case when the evaluations of the objective function values include noise which is not necessarily normally distributed. Fox and Heine (1995) have also investigated the application of simulated annealing to solve discrete stochastic optimization problems. When the set of feasible solutions is very big, the ordinal optimization approach proposed by Ho, Sreenivas, and Vakili (1992) can be used to narrow the focus on a smaller manageable set that contains a “good enough” (near-optimal) solution with high probability. Then the other methods we have mentioned could

be used to find a solution over this smaller set.

Yan and Mukai (1992) proposed a method for discrete stochastic optimization that is called the *stochastic ruler method*. This method involves comparing the observed objective function values $H(x)$, where $x \in \mathcal{S}$, with a uniform random variable (the stochastic ruler) defined on the range of the observed objective function values. This uniform random variable is used as a scale against which the observations $H(x)$ are measured. This method uses an increasing sequence of observations per iteration and it requires some restrictive assumptions on the neighborhood structure. Yan and Mukai show that their method converges in probability to a global optimal solution. In this paper, we use the idea of the stochastic ruler method and propose a variant of this method that appears to behave better than the original method in practice and also has been shown to converge under more general conditions. Instead of using an increasing sequence of observations per iteration, we draw only a fixed number of observations per iteration, and we use a criterion that is similar to the criterion of Andradóttir (1995, 1996) for estimating the solution. Furthermore, unlike the original stochastic ruler method, our proposed method converges almost surely to a global optimal solution (the original method is only guaranteed to converge in probability). We provide a numerical example that verifies the validity of our method and we compare the performance of our method with that of the original stochastic ruler method.

This paper is organized as follows: In Section 2 we give a brief review of the original stochastic ruler method. In Section 3 we introduce our modified stochastic ruler method. In Section 4 we implement our method to solve a discrete stochastic optimization problem and we compare the results with the performance of the original stochastic ruler method. And finally in Section 5 we give some concluding remarks.

2 THE ORIGINAL STOCHASTIC RULER ALGORITHM

The stochastic ruler method is a random search method that was proposed by Yan and Mukai (1992) to solve the discrete stochastic optimization problem (1). To motivate our new method, in this section we give a brief review of the stochastic ruler method. To proceed we need the following definitions and assumptions that were presented by Yan and Mukai (1992).

Assumption 1 Assume that $E\{H(x)^2\} < \infty, \forall x \in \mathcal{S}$.

Definition 1 For each $x \in \mathcal{S}$, there exists a subset $N(x)$ of $\mathcal{S} \setminus \{x\}$, which is called the set of neighbors of x .

Assumption 2 For any $x, x' \in \mathcal{S}$, x' is reachable from x ; i.e., there exists a finite sequence $\{n_i\}_{i=0}^l$ for some l , such that $x_{n_0} = x, x_{n_1} = x'$, and $x_{n_{i+1}} \in N(x_{n_i}), i = 0, 1, 2, \dots, l-1$.

Definition 2 A function $R : \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ is said to be a transition probability for \mathcal{S} and N if

1. $R(x, x') > 0 \Leftrightarrow x' \in N(x)$, and
2. $\sum_{x' \in \mathcal{S}} R(x, x') = 1, \forall x \in \mathcal{S}$.

Assumption 3 The neighbor system N and the transition probability R are symmetric, i.e.,

1. $x' \in N(x) \Leftrightarrow x \in N(x')$, and
2. $R(x, x') = R(x', x), \forall x, x' \in \mathcal{S}$.

Assumption 4 The parameters $a, b \in \mathfrak{R}$ are selected to cover the range of the observed objective function values $H(x)$, where $x \in \mathcal{S}$.

Assumption 5 A sequence $\{M_k\}$ of positive integers satisfies $M_k \rightarrow \infty$ as $k \rightarrow \infty$.

Instead of solving the original minimization problem (1) directly, Yan and Mukai (1992) solve the following alternative maximization problem

$$\max_{x \in \mathcal{S}} P(x, a, b), \quad (2)$$

where $P(x, a, b) = \text{Prob}\{H(x) \leq \Theta(a, b)\}$, for all $x \in \mathcal{S}$, and $\Theta(a, b)$ is a uniform random variable on the interval (a, b) . The random variable $\Theta(a, b)$ is called the *stochastic ruler* and it is used as a scale against which the observations $H(x)$ are measured. Yan and Mukai (1992) have shown that the solution of the alternative maximization problem (2) is also a solution of the original minimization problem (1) provided that the interval (a, b) is chosen large enough (see Assumption 4). Therefore, they propose the following algorithm to solve the alternative maximization problem (2):

Algorithm 1 (Yan and Mukai (1992))

Parameters: $N, R, \{M_k\}, a, b$.

Step 0: Select a starting point $X_0 \in \mathcal{S}$ and let $k = 0$.

Step 1: Given $X_k = x$, choose a candidate Z_k from $N(x)$ with probability distribution

$$P\{Z_k = z | X_k = x\} = R(x, z), z \in N(x).$$

Step 2: Given $Z_k = z$, draw a sample $h(z)$ from $H(z)$. Then draw a sample θ from $\Theta(a, b)$. If $h(z) > \theta$, then let $X_{k+1} = X_k$ and go to Step 3. Otherwise draw another sample $h(z)$ from $H(z)$ and draw another sample θ from $\Theta(a, b)$. If $h(z) > \theta$, then let $X_{k+1} = X_k$ and go to Step 3. Otherwise continue to draw and compare. If all M_k tests, $h(z) > \theta$, fail, then accept the candidate Z_k and set $X_{k+1} = Z_k = z$.

Step 3: Set $k = k + 1$ and go to Step 1.

Remark 1 Note that Step 2 can be rewritten as: Given $Z_k = z$, set

$$X_{k+1} = \begin{cases} Z_k & \text{with probability } p_{z,k}, \\ X_k & \text{with probability } 1 - p_{z,k}, \end{cases}$$

where

$$\begin{aligned} p_{z,k} &= [P\{H(z) \leq \Theta(a, b)\}]^{M_k} \\ &= [P(z, a, b)]^{M_k}. \end{aligned}$$

Under Assumptions 1 through 5, Yan and Mukai (1992) have shown that the Markov chain $\{X_k\}$ generated by Algorithm 1 converges in probability to a global optimal solution of the discrete stochastic optimization problem (1).

In practice, the sequence $\{M_k\}$ plays a very crucial role in the convergence of Algorithm 1. In particular, if the sequence $\{M_k\}$ is selected to increase very rapidly then the algorithm could converge to a local optimal solution and stay there. On the other hand, if the sequence $\{M_k\}$ is selected to increase very slowly then the algorithm tends to take a longer time to settle down. To ease these difficulties, in the next section we present a new variant of the stochastic ruler algorithm that does not require an increasing sequence of observations to be drawn at every iteration. Furthermore, the new method can handle more general neighborhood structures than the original stochastic ruler method.

3 THE PROPOSED MODIFICATION OF THE STOCHASTIC RULER ALGORITHM

In this section we present a new variant of the stochastic ruler method (Algorithm 1) to solve the discrete stochastic optimization problem (1). This new method- requires only a finite number of observations per iteration instead of an increasing number of observations per iteration. Moreover, in our new method we relax Assumption 3 to include more general neighborhood structures and transition probabilities. To

proceed, we use the following transition probability that also satisfies Definition 2. Let

$$R(x, x') = \frac{R'(x, x')}{D(x)}, \quad (3)$$

where $R' : \mathcal{S} \times \mathcal{S} \rightarrow \mathfrak{R}$ is a function such that $R'(x, x') > 0 \Leftrightarrow x' \in N(x)$, and $D : \mathcal{S} \rightarrow \mathfrak{R}$ is another function such that $D(x) = \sum_{x' \in N(x)} R'(x, x')$, $\forall x \in \mathcal{S}$. Now we use the following assumption instead of Assumption 3:

Assumption 6 Let $R(x, x')$ be defined as in equation (3) and let N satisfy Definition 1. Then we assume that:

1. $x' \in N(x) \Leftrightarrow x \in N(x')$, and
2. $R'(x, x') = R'(x', x)$, $\forall x, x' \in \mathcal{S}$.

As an example that satisfies Assumption 6 consider $R(x, x') = \frac{1}{|N(x)|}$ for all $x' \in N(x)$. Note that this choice of transition probability R only satisfies Assumption 3 if $|N(x)| = |N(x')|$ whenever $x' \in N(x)$.

Our modified stochastic ruler method is given below. Note that $\{X_k\}$ is the Markov chain that is generated by the algorithm, $V_k(x)$ is the number of times the Markov chain $\{X_k\}$ has visited state x up to time k , and $D(x)$ is defined in equation (3). Then the state X_k^* that maximizes $\frac{V_k(X_k)}{D(X_k)}$ is used as an estimate of the solution of the discrete stochastic optimization problem (1).

Algorithm 2

Parameters: N, R, M, a, b .

Step 0: Select a starting point $X_0 \in \mathcal{S}$. Let $V_0(X_0) = 1$, and $V_0(x) = 0$, for all $x \in \mathcal{S}$, $x \neq X_0$. Let $k = 0$ and $X_k^* = X_0$.

Step 1: Given $X_k = x$, choose a candidate Z_k from $N(x)$ with probability distribution

$$P\{Z_k = z | X_k = x\} = R(x, z) = \frac{R'(x, z)}{D(x)},$$

where $z \in N(x)$ and $R'(x, z)$ and $D(x)$ are defined in equation (3).

Step 2: Given $Z_k = z$, draw a sample $h(z)$ from $H(z)$. Then draw a sample θ from $\Theta(a, b)$. If $h(z) > \theta$, then let $X_{k+1} = X_k$ and go to Step 3. Otherwise, draw another sample $h(z)$ from $H(z)$ and draw another sample θ from $\Theta(a, b)$. If $h(z) > \theta$, then let $X_{k+1} = X_k$ and go to Step 3. Otherwise, continue to draw and compare. If all M tests, $h(z) > \theta$, fail, then accept the candidate Z_k and set $X_{k+1} = Z_k = z$.

Step 3: Let $k = k + 1$, $V_k(X_k) = V_{k-1}(X_k) + 1$, and $V_k(x) = V_{k-1}(x)$, for all $x \in \mathcal{S}$, $x \neq X_k$. If $\frac{V_k(X_k)}{D(X_k)} > \frac{V_k(X_{k-1}^*)}{D(X_{k-1}^*)}$, then let $X_k^* = X_k$; otherwise let $X_k^* = X_{k-1}^*$. Go to Step 1.

Remark 2 Note that Step 2 can be rewritten as: Given $Z_k = z$, set

$$X_{k+1} = \begin{cases} Z_k & \text{with probability } p_z, \\ X_k & \text{with probability } 1 - p_z, \end{cases}$$

where $p_z = [P\{H(z) \leq \Theta(a, b)\}]^M = [P(z, a, b)]^M$.

The next theorem has been proved by Alrefaei and Andradóttir (1996). It guarantees that the sequence $\{X_k^*\}$ generated by Algorithm 2 converges almost surely to a global optimal solution of the discrete stochastic optimization problem (1).

Theorem 1 Under Assumptions 1, 2, 4 and 6 the sequence $\{X_k^*\}$ generated by Algorithm 2 converges almost surely to an element of \mathcal{S}^* (in the sense that there exists a set A such that $P(A) = 1$ and for all $\omega \in A$, there exists $K_\omega > 0$ such that $X_k^*(\omega) \in \mathcal{S}^*$ for all $k \geq K_\omega$).

4 NUMERICAL APPLICATION

In this section we implement Algorithm 2 to solve a simple discrete stochastic optimization problem, and compare its performance with that of Algorithm 1 on the same problem. In equation (1), let $\mathcal{S} = \{1, \dots, 10\}$, and let $f(x) = E[Y_x]$ for all $x \in \mathcal{S}$, where Y_x is a uniform random variable on the interval $f(x) \pm 0.5$, for all $x \in \mathcal{S}$, and $f(1), \dots, f(10)$ are 0.3, 0.7, 0.9, 0.5, 1.0, 1.4, 0.7, 0.8, 0.0, and 0.6, respectively. See Figure 1 for the graph of the function. Let $a = -0.5$, $b = 1.9$, and let X_0 be uniformly distributed on the set \mathcal{S} .

We will apply Algorithm 1 to solve this optimization problem with a number of different choices of the parameters M , $\{N(x) : x \in \mathcal{S}\}$, and $\{R(x, x') : x \in \mathcal{S} \text{ and } x' \in N(x)\}$ that satisfy the conditions of Theorem 1 (so that Algorithm 2 is guaranteed to converge). In particular, we use two different values of the constant M : $M = 1$ and $M = 2$. Also, we use three different neighborhood structures $\{N(x) : x \in \mathcal{S}\}$. The first neighborhood structure is given by

$$N(x) = \begin{cases} \{2\} & \text{if } x = 1, \\ \{9\} & \text{if } x = 10, \\ \{x - 1, x + 1\} & \text{otherwise.} \end{cases} \quad (4)$$

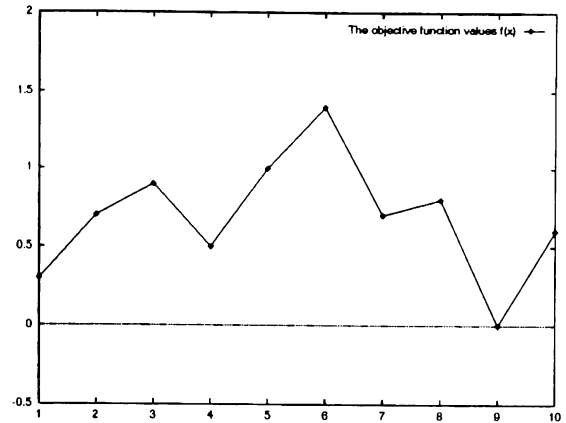


Figure 1: The Objective Function Values $f(x)$, where $f(x) = E[Y_x]$ and Y_x is a Uniform Random Variable on the Interval $f(x) \pm 0.5$.

The second neighborhood structure is given by

$$N(x) = \begin{cases} \{2, 3\} & \text{if } x = 1, \\ \{1, 3, 4\} & \text{if } x = 2, \\ \{7, 8, 10\} & \text{if } x = 9, \\ \{8, 9\} & \text{if } x = 10, \\ \{x \pm 1, x \pm 2\} & \text{otherwise.} \end{cases} \quad (5)$$

The third neighborhood structure is given by

$$N(x) = \mathcal{S} \setminus \{x\}, \quad (6)$$

for all $x \in \mathcal{S}$. In all cases, we let $R(x, x') = \frac{1}{|N(x)|}$, for all $x \in \mathcal{S}$ and $x' \in N(x)$. Note that in the third neighborhood structure (6), we have only one global minimum at $x = 9$. On the other hand, in the first neighborhood structure (4) we have three local minima at $x = 1, 4$, and 7 and one global minimum at $x = 9$, and in the second neighborhood structure (5), we have two local minima at $x = 1$ and $x = 4$ and one global minimum at $x = 9$. Since the first and second neighborhood structures result in more local minima, we expect that Algorithm 2 will converge more slowly in these settings than when the third neighborhood structure is used.

Note that for the first neighborhood structure (see equation (4)) it is impossible to select a transition probability R that satisfies Assumption 3. Therefore, Algorithm 1 is not guaranteed to converge when the first neighborhood structure is used. Also, for the second neighborhood structure (see equation (5)) it is difficult to select a transition probability R that satisfies Assumption 3. Therefore, it is difficult to apply Algorithm 1 using the second neighborhood structure. However, we implement Algorithm 1 using the third neighborhood structure defined in equation (6).

In this case, we let $M_k = \lfloor \log_5(k+10) \rfloor$ for all k ; this choice of the sequence $\{M_k\}$ satisfies the guidelines specified by Yan and Mukai (1992).

Figures 2 and 3 show the results obtained by applying Algorithm 2 to solve this optimization problem with the choices of parameters described above. In particular, the figures show the number of observations of the objective function vs. the number of replications that have converged to the global solution out of a total of 100 replications that were used.

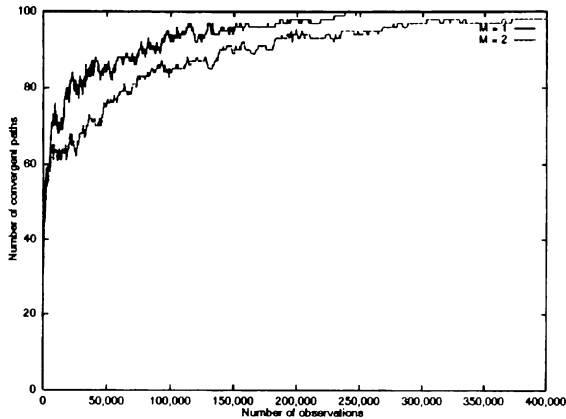


Figure 2: The Performance of Algorithm 2 as a Function of M when the Neighborhood Structure $\{N(x)\}$ given by Equation (4) is used.

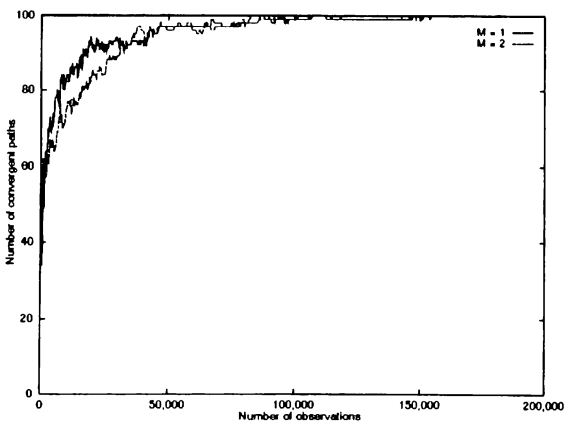


Figure 3: The Performance of Algorithm 2 as a Function of M when the Neighborhood Structure $\{N(x)\}$ given by Equation (5) is used.

Figure 4 includes a comparison between Algorithms 1 and 2 when the third neighborhood structure is used. Note that for Algorithm 2 all 100 replications converge to the true optimizer after 10,000 observations have been drawn. Furthermore, 90 of these

replications converge after only 2,000 observations have been made. On the other hand, the number of convergent replications for Algorithm 1 does not exceed 60 replications after 50,000 observations have been obtained. In fact, in another experiment we completed 100 replications of 10 million iterations of Algorithm 1 (which required about 40 million observations per replication on the average) and we did not get convergence for all 100 replications.

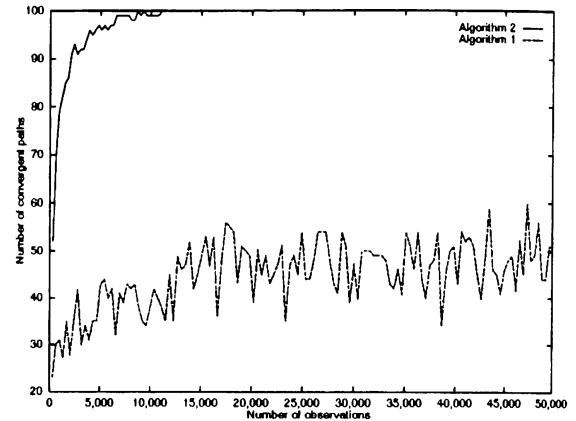


Figure 4: A Comparison Between Algorithms 1 and 2 when the Neighborhood Structure $\{N(x)\}$ given by Equation (6) is used, $M = 1$ in Algorithm 2, and $M_k = \lfloor \log_5(k+10) \rfloor$ for all k in Algorithm 1.

From the results given in Figure 4, we conclude that in this example Algorithm 2 converges much faster than Algorithm 1. Also, as expected, it appears to be better to have larger neighborhoods $N(x)$ where $x \in \mathcal{S}$, and smaller number of observations M in Algorithm 2 (the results for the neighborhood structure (6) are better than those for the neighborhood structures (4) and (5), and the results for $M = 1$ are better than those for $M = 2$).

5 CONCLUSION

In this paper, we have presented a new search method designed to solve the discrete stochastic optimization problem (1). This method resembles the stochastic ruler method of Yan and Mukai (1992). Unlike the original stochastic ruler method, the new method requires only a finite number of observations to be made at every iteration instead of using an increasing sequence of observations per iteration. Furthermore, this new method can handle more general neighborhood structures. It has the feature that it spends most of the time near the optimal solution. Also, unlike the original stochastic ruler method, our method converges *almost surely* to a global optimal solution.

In practice, in our example, it seems to be better to use a small number of observations per iteration to give the method more freedom to move around the state space aggressively and locate the solution very quickly. Also, it is better to use larger neighborhood sets when there are many local minimal solutions. However, we expect that may be better to use smaller neighborhood sets if the objective function is unimodal or has a fair amount of structure. In our example, the new method shows better performance than the original stochastic ruler method. In order to further investigate the efficiency of the proposed method, we are interested in implementing our method to solve more realistic discrete stochastic optimization problems (such as the buffer allocation problem). We are also interested in investigating the application of our method to solve discrete optimization problems in both transient and steady-state simulation.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation under grants DDM-9210679 and DMI-9523111. In addition, the research of the first author was supported in part by the Jordan University of Science and Technology.

REFERENCES

- Alrefaei, M. H., and S. Andradóttir. 1995. A new search algorithm for discrete stochastic optimization. In *Proceedings of the 1995 Winter Simulation Conference*, ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 236-241. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.
- Alrefaei, M. H., and S. Andradóttir. 1996. A modification of the stochastic ruler method. Working paper.
- Andradóttir, S. 1995. A method for discrete stochastic optimization. *Management Science* 41:1946-1961.
- Andradóttir, S. 1996. A global search method for discrete stochastic optimization. To appear in the *SIAM Journal on Optimization*.
- Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: Wiley.
- Fox, B. L., and G. W. Heine. 1995. Probabilistic search with overrides. *The Annals of Applied Probability* 5:1087-1094.

- Gelfand, S. B., and S. K. Mitter. 1989. Simulated annealing with noisy or imprecise energy measurements. *Journal of Optimization Theory and Applications* 62:49-62.
- Goldsman, D., B. L. Nelson, and B. Schmeiser. 1991. Methods for selecting the best system. In *Proceedings of the 1991 Winter Simulation Conference*, ed. B. L. Nelson, W. D. Kelton, and C. M. Clark, 177-186. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.
- Ho, Y. C., R. S. Sreenivas, and P. Vakili. 1992. Ordinal optimization of DEDS. *Journal of Discrete Event Dynamical Systems* 2:61-88.
- Lee, J. 1995. Faster simulated annealing techniques for stochastic optimization problems, with application to queueing network simulation. Ph.D. Thesis, North Carolina State University, Raleigh.
- Yan, D., and H. Mukai. 1992. Stochastic discrete optimization. *SIAM Journal on Control and Optimization* 30:594-612.

AUTHOR BIOGRAPHIES

MAHMOUD HASAN ALREFAEI is pursuing a joint Ph.D. degree in mathematics and industrial engineering at the University of Wisconsin - Madison. He received his B.S. and M.S. degrees in mathematics from Yarmouk University in Jordan. His research interests include discrete stochastic optimization and simulated annealing. He is presently a member of INFORMS and AMS.

SIGRÚN ANDRADÓTTIR is an Associate Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology, on leave from the University of Wisconsin - Madison where she has held a faculty position since 1990. She received a B.S. in Mathematics from the University of Iceland in 1986, an M.S. in Statistics from Stanford University in 1989, and a Ph.D. in Operations Research from Stanford University in 1990. Her research interests include stochastic optimization, simulation, and stochastic processes. She presently serves as Associate Editor for *IIE Transactions*, *Stochastic Models*, and the *ACM Transactions on Modeling and Computer Simulation*. She is a member of INFORMS.