



## Operations Research

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Letter to the Editor—A Monte Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems

Martin Pincus,

To cite this article:

Martin Pincus, (1970) Letter to the Editor—A Monte Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems. Operations Research 18(6):1225-1228. <https://doi.org/10.1287/opre.18.6.1225>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

© 1970 INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# A MONTE CARLO METHOD FOR THE APPROXIMATE SOLUTION OF CERTAIN TYPES OF CONSTRAINED OPTIMIZATION PROBLEMS

Martin Pincus

*Polytechnic Institute of Brooklyn, Brooklyn, New York*

(Received October 2, 1969)

This paper considers the problem of minimizing a function  $F(x_1, \dots, x_n)$  over a closed, bounded region  $S$  in  $n$ -dimensional space under the assumption that there exists a unique minimizing point  $(z_1, \dots, z_n) \in S$ . In a previous paper I represented the coordinates of the minimizing point as the limit of a ratio of integrals. The same type of ratio appears, in a different context, in statistical mechanics where a Monte Carlo method has been developed, by METROPOLIS ET AL., for its numerical evaluation. The purpose of this paper is to point out the connection of Metropolis's method with the above type of minimization problem. The idea of the method is to associate with the minimization problem a Markov chain whose sample averages converge with probability one to (approximately) the minimizing point  $(z_1, \dots, z_n)$ . The Markov chain should be easily realizable on a computer. An estimate of the error from sampling over a finite time period is given.

IN REFERENCE 4 I gave a formula for the solution of certain types of constrained minimization problems. More specifically, we have the following:

**THEOREM.** Let  $F(x) = F(x_1, \dots, x_n)$  be a real-valued, continuous function over a closed, bounded domain  $S \subset R^n = n$ -dimensional Euclidean space. Further, assume there is a unique point  $z = (z_1, \dots, z_n) \in S$  at which  $\min_{x \in S} F(x)$  is attained (there are no restrictions on relative minima). Then the coordinates  $z_i$  of the minimizing point are given by

$$z_i = \lim_{\lambda \rightarrow \infty} \left\{ \frac{\int_S x_i \exp[-\lambda F(x_1, \dots, x_n)] dx_1 \cdots dx_n}{\int_S \exp[-\lambda F(x_1, \dots, x_n)] dx_1 \cdots dx_n} \right\} \quad (1)$$

In particular, the hypotheses of the theorem are satisfied when the closed bounded region  $S$  is convex and the objective function  $F$  is strictly convex. The purpose of this note is to outline a method for the approximate evaluation of the coordinates of the minimizing point  $(z_1, z_2, \dots, z_n)$ , i.e., the approximate evaluation of the ratio appearing on the right-hand side of (1).

The first approximation is in taking  $\lambda$  finite. In the proof of the theorem<sup>(4)</sup> it is shown that for given  $\epsilon$  and  $\lambda$  the error

$$\left| z_i - \frac{\int_S x_i \exp[-\lambda F(x)] dx}{\int_S \exp[-\lambda F(x)] dx} \right| \leq \epsilon + A e^{-\lambda \epsilon},$$

where  $A$  and  $\delta$  depend only on  $\epsilon$  (and of course on  $F$  and  $S$ ).  $A$  and  $\delta$  are related to the sharpness of the global minimum. Estimates for  $A$  and  $\delta$  should be obtainable under regularity conditions on  $F$  such as a Lipschitz condition.

For fixed  $\lambda$ , we then have to evaluate the ratio

$$B(\lambda) = \left\{ \int_{\mathcal{S}} x_i \exp[-\lambda F(x)] dx_1 \cdots dx_n \right\} / \left\{ \int_{\mathcal{S}} \exp[-\lambda F(x)] dx_1 \cdots dx_n \right\}.$$

The same type of expression appears in statistical mechanics, where it represents the average of a physical observable. For the purposes of statistical mechanics, METROPOLIS ET AL.<sup>[1]</sup> have developed a Monte Carlo method for approximating  $B(\lambda)$  that has apparently been successfully applied for high-dimensional regions, an outline of it being given in HAMMERSLEY AND HANDSCOMB.<sup>[2]</sup> In the following outline of the method we essentially follow reference 2 with some adaptations for our purposes.

For large  $\lambda$ , the major contribution to the integrals appearing in (1) comes from a small neighborhood of the minimizing point  $z$ . Metropolis's sampling procedure, which is described below, is weighted so that the generated Markov chain spends, in the long run, most of the time visiting states near the minimizing point. Therefore, the procedure should be more efficient than a direct Monte-Carlo estimate of both the numerator and denominator.

The idea of the method is to generate samples with density

$$f(x_1, \dots, x_n) = \begin{cases} \exp[-\lambda F(x)] / \int_{\mathcal{S}} \exp[-\lambda F(u)] du, & (x \in \mathcal{S}) \\ 0, & \text{otherwise,} \end{cases}$$

where  $u = (u_1, \dots, u_n)$  and  $du = du_1 \cdots du_n$ . This is done (at least approximately) as follows.

Partition the region  $\mathcal{S}$  into a finite number  $N$  of mutually disjoint subregions  $S_j$  and replace integrals over  $\mathcal{S}$  by corresponding Riemann sums using the partition  $\{S_j\}$ . Fix a point  $y^j = (y_1^j, \dots, y_n^j) \in S_j$ . Then one constructs an irreducible, ergodic Markov chain  $\{X_k\}$  with state space  $\{y^1, \dots, y^N\}$  and with transition probabilities  $p_{ij}$ ,  $1 \leq i, j \leq N$ , satisfying  $\pi_j = \sum_i \pi_i p_{ij}$ ,  $j = 1, \dots, N$ , where  $\pi_j = \exp[-\lambda F(y^j)] / \sum_{h=1}^N \exp[-\lambda F(y^h)]$ ; that is,  $\{\pi_j\}$  is the invariant distribution for the Markov chain  $\{X_k\}$ . It should be noted that in the last expression for  $\pi_j$  we have assumed, for simplicity, that all subregions  $S_j$  have equal volumes. Then, using the strong law of large numbers for Markov chains,<sup>[1]</sup> we have with probability one

$$\frac{1}{m} \sum_{k=1}^m X_k \xrightarrow{m \rightarrow \infty} \frac{\left\{ \sum_{j=1}^N y_1^j \exp[-\lambda F(y^j)], \dots, \sum_{j=1}^N y_n^j \exp[-\lambda F(y^j)] \right\}}{\sum_{j=1}^N \exp[-\lambda F(y^j)]}$$

$$\approx \frac{\left\{ \int_{\mathcal{I}} x_1 \exp[-\lambda F(x)] dx, \dots, \int_{\mathcal{I}} x_n \exp[-\lambda F(x)] dx \right\}}{\int_{\mathcal{I}} \exp[-\lambda F(x)] dx}$$

$$\approx (z_1, \dots, z_n), \text{ the minimizing point.}$$

In the two ratios above, the denominator divides each component of the vector in the numerator. It should be noted that the sample averages  $(1/m) \sum_{k=1}^{m-1} X_k$  converge with probability one.

So far, it has been shown that, if we can construct a Markov chain with the above properties, we can, by sampling the chain over a 'long enough' time period, obtain an approximation to the minimizing point. In practice, of course, one can sample only over a finite time period.

The order of magnitude of the error introduced by sampling over a finite time period can be obtained for each component  $X_k^i$  of the vector  $X_k$  from the fact<sup>[1,2]</sup> that  $E[(1/m) \sum_{k=1}^{m-1} X_k^i - \mu_i]^2] \leq c/m$ , where  $c$  is an absolute constant and

$$\mu_i = \sum_{j=1}^{j=N} y_j^i \exp[-\lambda F(y^j)] / \sum_{j=1}^{j=N} \exp[-\lambda F(y^j)].$$

From Chebychev's inequality it then follows that

$$P[|(1/m) \sum_{k=1}^{m-1} X_k^i - \mu_i| \geq \epsilon] \leq c/\epsilon^2 m.$$

We now turn to the question of how Metropolis constructs a Markov chain with the required invariant distribution. He starts with a symmetric transition probability matrix  $P^* = (p_{ij}^*) 1 \leq i, j \leq N$ , i.e.,  $p_{ij}^* = p_{ji}^*$ ,  $p_{ij}^* > 0$  and, of course  $\sum_{j=1}^{j=N} p_{ij}^* = 1$ .

The transition matrix of the Markov chain  $\{X_k\}$  we are seeking is  $P = (p_{ij}) 1 \leq i, j \leq N$ , where

$$p_{ij} = \begin{cases} p_{ij}^* \pi_j / \pi_i, & \text{if } \pi_j / \pi_i < 1, & i \neq j, \\ p_{ij}^*, & \text{if } \pi_j / \pi_i \geq 1, & i \neq j, \\ p_{ii}^* + \sum_{j: \pi_j < \pi_i} p_{ij}^* (1 - \pi_j / \pi_i), & i = j. \end{cases}$$

It is shown in reference 2 that a Markov chain with the above transition matrix has the invariant distribution  $\{\pi_i\}$ .

A chain with such a transition matrix can be realized as follows. Given that the chain is in state  $y^i$  at time  $k$ , i.e.,  $\{X_k = y^i\}$ , the state at time  $k+1$  is determined by choosing a new state according to the distribution  $\{p_{ij}^*, j=1, \dots, N\}$ . If the state chosen is  $y^j$ , we then calculate the ratio  $\pi_j / \pi_i$ . If  $\pi_j / \pi_i \geq 1$ , we accept  $y^j$  as the new state at time  $k+1$ ; if  $\pi_j / \pi_i < 1$ , we take  $y^j$  as the state of the Markov chain at time  $k+1$  with probability  $\pi_j / \pi_i$  and take  $y^i$  to be the new state at time  $k+1$  with probability  $1 - \pi_j / \pi_i$ . It is easily shown that this procedure leads to a Markov chain with transition matrix  $P$ .

If one has to evaluate the denominator in the expression for  $\pi_i$ , the whole purpose of the Monte Carlo method would be lost. However, the procedure for determining  $P$  depends only on the ratios  $\pi_j / \pi_i$  and, therefore, the need for evaluat-

ing the denominator in the expressions for the  $\pi_i$  is circumvented. Only the values of  $F$  itself occur in evaluating the ratios  $\pi_j/\pi_i$ .

The author would like to thank the referee for valuable suggestions.

#### REFERENCES

1. K. L. CHUNG, *Markov Chains with Stationary Transition Probabilities*, Second Edition, Springer, Berlin, 1967.
2. J. M. HAMMERSLEY AND D. C. HANDSCOMB, *Monte Carlo Methods*, Methuen's Monographs on Applied Probability and Statistics, London, 1964.
3. N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER, "Equations of State Calculations by Fast Computing Machines," *J. Chem. Physics* **21**, 1087-1092 (1953).
4. M. PINCUS, "A Closed Form Solution of Certain Programming Problems," *Opns. Res.* **16**, 690-694 (1968).

### A COMMENT ON THE MORSE-ELSTON MODEL OF PROBABILISTIC OBSOLESCENCE

Donald H. Kraft

*Purdue University, Lafayette, Indiana*

(Received January 20, 1970)

MORSE AND ELSTON have constructed a probabilistic model of obsolescence and fitted it to the circulation of books in a library. This note presents a modification of their model and fits it to their results to account for time dependence of their original parameter  $\alpha$ . This revised model can be used to describe the phenomena of obsolescence as seen in such a case as the usage over time of books concerned with scientific subjects.

MORSE AND ELSTON<sup>[1]</sup> have introduced a Markovian process to examine obsolescence, especially of library materials, in terms of demand. The applicability of the model in describing the usage of books is discussed more fully in MORSE,<sup>[2]</sup> where empirical data are presented to show how the usage of books changes over time. The primary factor influencing item usage in the Morse and Elston model is the age of the item. The correlation of book usage to age is discussed in great detail by JAIN.<sup>[3,4]</sup> Morse and Elston's model is based on the assumption that

$$Pr\{R(t)=n|R(t-1)=m\}=e^{-(\alpha+bm)}(\alpha+bm)^n/n!, \quad (n=0, 1, 2, \dots) \quad (1)$$

where  $R(t)$  represents the demand for an item  $t$  years old. This implies that

$$\begin{aligned} E[R(t)] &= S(t) = [S(0) - \alpha/1 - b]b^t + \alpha/1 - b \\ &= S(0)b^t + \alpha(1-b^t)/(1-b) = S(0)b^t + \alpha \sum_{i=0}^{t-1} b^i. \end{aligned} \quad (2)$$

Copyright 1970, by INFORMS, all rights reserved. Copyright of Operations Research is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.