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,

Published in Operations Research on December 01, 1970 as DOI: 10.1287/opre.18.6.1225.
This article has not been copyedited or formatted. The final version may differ from this version.

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 $F(x_1, \cdots, x_n)$ over a closed, bounded region S in n-dimensional space under the assumption that there exists a unique minimizing point $(z_1, \cdots, z_n) \epsilon S$. 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 $(z_1, \cdots, z_n)$. 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 $F(x) = F(x_1, \cdots, x_n)$ 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 $z = (z_1, \cdots, z_n) \epsilon S$ at which $\min_{x \epsilon S} F(x)$ is attained (there are no restrictions on relative minima). Then the coordinates z_i of the minimizing point are given by*

$$z_i = \lim_{\lambda \to \infty} \left\{ \int_S x_i \exp[-\lambda F(x_1, \cdots, x_n)]\, dx_1 \cdots dx_n \right\} \Big/ \left\{ \int_S \exp[-\lambda F(x_1, \cdots, x_n)]\, dx_1 \cdots dx_n \right\}. \tag{1}$$

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 $(z_1, z_2, \cdots, z_n)$, i.e., the approximate evaluation of the ratio appearing on the right-hand side of (1).

The first approximation is in taking λ finite. In the proof of the theorem[4] it is shown that for given ϵ and λ the error

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

Published in Operations Research on December 01, 1970 as DOI: 10.1287/opre.18.6.1225.
This article has not been copyedited or formatted. The final version may differ from this version.

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

For fixed λ, we then have to evaluate the ratio

$$B(\lambda)=\left\{\int_S x_1\exp[-\lambda F(x)]\,dx_1\cdots dx_n\right\}\Big/\left\{\int_S \exp[-\lambda F(x)]\,dx_1\cdots dx_n\right\}.$$

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.[3] 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.[2] In the following outline of the method we essentially follow reference 2 with some adaptations for our purposes.

For large λ, 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

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

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

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

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

Published in Operations Research on December 01, 1970 as DOI: 10.1287/opre.18.6.1225.
This article has not been copyedited or formatted. The final version may differ from this version.

$$\approx \frac{\left\{\int_a x_1 \exp[-\lambda F(x)]\,dx, \cdots, \int_a x_n \exp[-\lambda F(x)]\,dx\right\}}{\int_a \exp[-\lambda F(x)]\,dx}$$

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

In the two ratios above, the denominator divides each component of the vector in the numerator. It should be noted that the sample averages $(1/m)\sum_{k=1}^{k=m} X_k$ 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[1,2] that $E[((1/m)\sum_{k=1}^{k=m} X_k{}^i - \mu_i)^2] \leqq c/m$, where c is an absolute constant and

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

From Chebychev's inequality it then follows that

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

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 $P^* = (p_{ij}^*) 1 \leqq i, j \leqq N$, 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 $P = (p_{ij}) 1 \leqq i, j \leqq N$, where

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

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 $\{p_{ij}^*, j = 1, \cdots, N\}$. If the state chosen is y^j, we then calculate the ratio π_j/π_i. If $\pi_j/\pi_i \geqq 1$, 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 π_j/π_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 π_i, the whole purpose of the Monte Carlo method would be lost. However, the procedure for determining P depends only on the ratios π_j/π_i and, therefore, the need for evaluat-

Published in Operations Research on December 01, 1970 as DOI: 10.1287/opre.18.6.1225.
This article has not been copyedited or formatted. The final version may differ from this version.

ing the denominator in the expressions for the π_i is circumvented. Only the values of F itself occur in evaluating the ratios π_j/π_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 α. 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[1] 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,[2] 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.[3,4] Morse and Elston's model is based on the assumption that

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

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

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

Published in Operations Research on December 01, 1970 as DOI: 10.1287/opre.18.6.1225.
This article has not been copyedited or formatted. The final version may differ from this version.