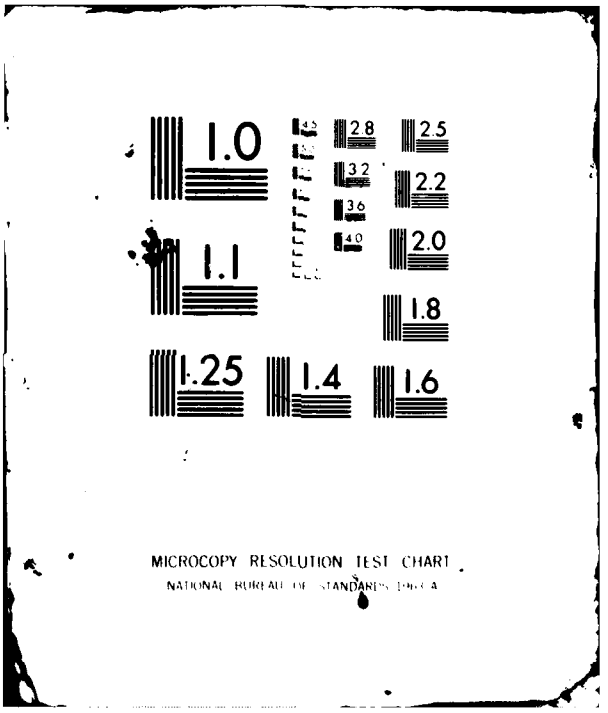
AD-A110 840 STANFORD UNIV CA SYSTEMS OPTIMIZATION LAB F/G 12/1
AN ADAPTIVE IMPORTANCE SAMPLING PROCEDURE.(U)
NOV 81 J H FRIEDMAN, M H WRIGHT N00014-75-C-0267
UNCLASSIFIED SOL-81-23 NL

END
DATE FILMED
3 82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

Systems
Optimization
Laboratory

(12)

AD A110840

DTIC FILE COPY

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

AN ADAPTIVE IMPORTANCE SAMPLING PROCEDURE

by

Jerome H. Friedman†
Margaret H. Wright††

TECHNICAL REPORT SOL 81-23

November 1981

†The research of Jerome H. Friedman was supported by the U.S. Department of Energy under Contract DE-AC03-76-SF00515 at the Stanford Linear Accelerator Center.

††Research and reproduction of this report were partially supported by the Department of Energy Contract AM03-76SF00326, PA# DE-AT03-76ER72018; Office of Naval Research Contract N00014-75-C-0267; National Science Foundation Grants MCS-7926009 and ECS-8012974; and Army Research Office Contract DAA29-79-C-0110, at Stanford University.

An Adaptive Importance Sampling Procedure

Jerome H. Friedman
Stanford Linear Accelerator Center

Margaret H. Wright
Systems Optimization Laboratory
Department of Operations Research

Stanford University
Stanford, California 94305

ABSTRACT

Monte Carlo calculations often require generation of a random sample of n-dimensional points drawn from a specified multivariate probability distribution. We present an importance sampling technique that can often greatly improve the efficiency of an acceptance/rejection generating method. The importance sampling function is defined as piecewise constant on a set of subregions, which are obtained by adaptively partitioning the sampling region so that the variation of density values within each subregion is relatively small. The partitioning strategy is based on multiparameter optimization to estimate the maximum and minimum of the original density function in each subregion.

The work of Jerome H. Friedman was supported by the U.S. Department of Energy under Contract DE-AC03-76-SF[illegible]. The work of Margaret H. Wright was supported by the U.S. Department of Energy Contract DE-AC03-76-SF00326, PA No. DE-AT03-76ER72018; National Science Foundation Grants MCS-7926009 and ECS-8012974; the Office of Naval Research Contract N00014-75-C-0267; and the U.S. Army Research Office Contract DAAG29-79-C-0110.

Keywords: sampling, probability distributions, Monte Carlo procedures, adaptive partitioning.

1. Introduction

In Monte Carlo calculations, it is often necessary to generate a random sample of n-dimensional points drawn from a specified multivariate probability density function $f(x)$. Such sampling can not generally be done directly, except in a few important special cases (see, e.g., Everett and Cashwell, 1972).

When direct sampling is not possible, a method that can always be applied (at least in principle) is the *acceptance/rejection technique*. Let R denote the sampling region of interest (which will be assumed to be hyper-rectangular), and let f_{UB} be an upper bound on the value of f in R. The sample is generated by a sequence of trials, as follows. A random number ρ is uniformly sampled in the interval $[0, f_{UB}]$, and a random point r is sampled with uniform density within the sampling region R. If $f(r) \geq \rho$, the point r is accepted; otherwise, it is rejected. With this technique, the accepted points constitute the desired random sample drawn from the probability density function f.

The efficiency of an acceptance/rejection procedure is defined as the expected number of points accepted per function evaluation, and is given by

$$e = \frac{\bar{f}}{f_{UB}}, \tag{1}$$

where $\bar{f}$ is the mean of f over the sampling region. From (1), it can be seen that the efficiency improves with the quality of the upper bound. However, even at best a large number of function evaluations may be required for every point that is accepted.

An improved efficiency can be achieved if a lower bound f_{LB} for the function f in R is also known. If $\rho \leq f_{LB}$, the corresponding point r can be accepted without evaluating f. In this case, the efficiency is defined as the ratio of the probability of acceptance to the probability of a function evaluation, and is given by

$$e = \frac{\bar{f}}{f_{UB} - f_{LB}}. \tag{2}$$

The optimal efficiency e^* is achieved when the upper and lower bounds are attainable, i.e., when $f_{UB} = f^+$ (the maximum of f in R) and $f_{LB} = f^-$ (the minimum of f in R). Thus,

$$e^* = \frac{\bar{f}}{f^+ - f^-} = \frac{I}{(f^+ - f^-)\,\mathrm{vol}(R)}, \tag{3}$$

where I is the integral of f over R, and $\mathrm{vol}(R)$ is the volume of R (by definition, $I = \bar{f}\,\mathrm{vol}(R)$).

2. An Importance Sampling Technique

The technique of *importance sampling* can sometimes improve the efficiency of sample generation. The idea is to find a second probability density function $g(x)$ — an *importance sampling function* — for which the trial points in R can be generated efficiently. Points with the original probability density function $f(x)$ can then be produced by using the ratio function $f(x)/g(x)$ in an acceptance/rejection procedure (with a suitable choice for the interval from which ρ is drawn). For an appropriate $g(x)$, importance sampling can yield substantial improvements in efficiency compared to simple acceptance/rejection.

In practice it is usually difficult to find an effective importance sampling function $g(x)$ for the particular problem at hand. We present an adaptive algorithm that, given $f(x)$, constructs such a suitable $g(x)$. For this purpose, we initially make two assumptions: (i) that R has been partitioned into a set of M disjoint simply-bounded subregions $\{R_i\}$, $i = 1, \ldots, M$, such that $R = \bigcup_i R_i$; and (ii) that f_i^+ and f_i^- (the maximum and minimum of f in R_i) are known for each R_i. We then define an importance sampling function $\bar{g}(x)$ as the piecewise constant function

$$\bar{g}(x) = \{f_i^+ \mid x \in R_i\}. \tag{4}$$

In order for $\bar{g}$ to be useful as an importance sampling function, it must be possible to generate random sample points with this density function. This is

achieved by randomly choosing a region R_i with probability

$$p_i = \frac{f_i^+ \,\mathrm{vol}(R_i)}{\sum_{j=1}^{M} f_j^+ \,\mathrm{vol}(R_j)}, \tag{5}$$

and then generating a random point r in R_i with uniform density. Applying the accept/reject procedure, a random number ρ is generated in the interval $[0,1]$ with uniform density, and the point r is accepted if $f(r)/f_i^+ \geq \rho$. Note that if $\rho \leq f_i^-/f_i^+$, r is accepted without evaluating f. The accepted points are a random sample from $f(x)$.

In order to calculate the efficiency of an importance sampling proecdure based on $g(x)$, we first consider the probability of acceptance in each region. The probability of choosing the i-th region is, by construction, p_i (5). Given that R_i has been chosen, the probability of accepting a point r uniformly chosen in R_i is $\bar{f}_i/f_i^+$, where $\bar{f}_i$ is the mean of f in R_i. Hence, the overall probability of acceptance (denoted by $Pr(A)$) is the weighted sum of the acceptance probabilities for all the subregions, i.e.

$$\begin{aligned} Pr(A) &= \sum_{i=1}^{M} p_i \frac{\bar{f}_i}{f_i^+} = \sum_{i=1}^{M} \frac{f_i^+ \,\mathrm{vol}(R_i)}{\sum_{j=1}^{M} f_j^+ \,\mathrm{vol}(R_j)} \frac{\bar{f}_i}{f_i^+} \\ &= \frac{I}{\sum_{i=1}^{M} f_i^+ \,\mathrm{vol}(R_i)}, \end{aligned}$$

where I is the integral of f over the entire region R.

The probability of a function evaluation, denoted by $Pr(FE)$, is the weighted sum of probabilities that $f(r)$ exceeds f_i^-, given that r is in R_i, i.e.

$$Pr(FE) = \sum_{i=1}^{M} p_i \frac{f_i^+ - f_i^-}{f_i^+} = \frac{\sum_{i=1}^{M} (f_i^+ - f_i^-)\,\mathrm{vol}(R_i)}{\sum_{i=1}^{M} f_i^+ \,\mathrm{vol}(R_i)}.$$

The efficiency of the importance sampling procedure (denoted by e_{IS}), is then $Pr(A)/Pr(FE)$, or

$$e_{IS} = \frac{I}{\sum_{i=1}^{M} (f_i^+ - f_i^-)\,\mathrm{vol}(R_i)}. \tag{6}$$

If we define the *spread* associated with the i-th region as

$$S_i = (f_i^+ - f_i^-)\,\mathrm{vol}(R_i), \tag{7}$$

then (6) can be written as

$$e_{IS} = \frac{I}{\sum_{i=1}^{M} S_i}. \tag{8}$$

The efficiency e_{IS} of the importance sampling based on $g(x)$ will exceed the efficiency e^* (3) of the original acceptance/rejection procedure if *any* S_i is less than the spread of the entire sampling region R. Furthermore, the efficiency increases as the sum of subregion spreads becomes smaller.

This analysis suggests that an importance sampling procedure based on g could be extremely effective if the sum of spread measures over $\{R_i\}$ is small. In order for this observation to useful, we must show how to contruct a set of subregions $\{R_i\}$ such that $\sum S_i$ is small. In the next section, an adaptive partitioning technique will be described that produces a suitable set of subregions.

3. The Adaptive Refinement Procedure

The adaptive partitioning procedure used to construct the subregions $\{R_i\}$ was originally developed in the context of multidimensional quadrature (Friedman and Wright, 1981a), and is intended to produce a set of subregions with "similar" *spread measures* (i.e., to minimize the sum of spreads). Since this is precisely the quality sought in the case of importance sampling, the same procedure can

be applied to construct $\bar{g}(x)$. This section gives only a brief summary of the adaptive refinement procedure. A detailed description is given in Friedman and Wright (1981a); the associated software is documented in Friedman and Wright (1981b).

Consider a typical hyper-rectangular region R, defined by simple bounds on each coordinate:

$$R = \{x \mid x_i^L \leq x_i \leq x_i^U\}, \tag{9}$$

where x is the vector $(x_1, x_2, \ldots, x_n)^T$. Let $S(R)$ denote the spread measure associated with R:

$$S(R) = (f^+ - f^-)\,\mathrm{vol}(R), \tag{10}$$

where f^+ and f^- denote the maximum and minimum of f in R. The implementation of the adaptive strategy for partitioning R includes three steps:

(1) calculation of the spread measure $S(R)$;

(2) subdivision of R;

(3) processing the new subregions and terminating the partitioning when appropriate.

In order to obtain $S(R)$, the volume of R and the extrema of f in R must be computed. Because of the simply-bounded form (9) of R, the volume is easy to calculate:

$$\mathrm{vol}(R) = \prod_{i=1}^{n} (x_i^U - x_i^L).$$

This simple expression for the volume is the main reason for partitioning R into simply-bounded subregions; even for a region defined by general hyperplanes, calculation of the volume is extremely complicated. Furthermore, efficient uniform sampling within a region of the form (9) is straightforward.

The other term in the expression (10) might seem, at first glance, to cause calculation of the spread to be computationally intractable, since two optimization problems must be solved. However, methods for optimization problems with

simple bounds on the variables are well developed, and thus the sub-problems associated with (10) can be solved quite efficiently if f is a reasonable function. (A non-derivative quasi-Newton method for bound-constrained optimization is used to compute the extrema of f in R; for details, see Friedman and Wright, 1981a.) More importantly, the substantial efficiencies in sampling that result from a well constructed partition ultimately justify the function evaluations expended to compute the extrema.

The second portion of the partitioning algorithm involves dividing R into disjoint simply-bounded *subregions* such that the sum of spread measures is reduced. Let x^+ denote the value of x corresponding to f^+, and x^- denote the value of x corresponding to f^-. An "ideal" partitioning strategy might "split" R into two disjoint subregions: R^+, which contains x^+, and R^-, which contains x^-. These subregions would be separated by the boundary of points x such that $f(x) = \tilde{f}$, where $\tilde{f}$ is chosen so that the spread measures in R^+ and R^- are equal, i.e.

$$(f^+ - \tilde{f})\,\mathrm{vol}(R^+) = (\tilde{f} - f^-)\,\mathrm{vol}(R^-). \tag{11}$$

The strategy just described is impractical — not only because of the difficulties in defining the boundary between R^+ and R^-, but also because the subregions could be of arbitrary shape (which makes calculation of their volumes intractable). Therefore, our partitioning algorithm uses a similar, but simplified motivation, to construct a set of *simply bounded subregions*. In particular, the boundaries of the subregions are defined by "cuts" along each coordinate direction from one of the extrema. The desired cuts are the solution of a system of nonlinear equations similar to (11), and can be computed efficiently using a special secant-like method for solving the associated nonlinear system (see Friedman and Wright, 1981a).

Finally, after R has been partitioned, the daughter subregions are merged into the list of all regions. If the global stopping criteria are satisfied (e.g., the

overall sum of spread measures is sufficiently small) the partitioning terminates. Otherwise, the list of regions is scanned for the one with the largest spread measure, which is then considered for refinement at the next iteration.

4. Efficiency of the Adaptive Importance Sampling Procedure

The procedure described in Section 3 can be applied to define $\bar{g}(x)$ (4) by constructing a set of subregions $\{R_i\}$, along with the corresponding values $\{f_i^+\}$; the resulting $\bar{g}(x)$ can then be used in importance sampling as described in Section 2. This technique will be effective to the extent that the function evaluations needed to construct the partition lead to a sufficiently improved efficiency in the importance sampling.

Let N_p denote the number of function evaluations expended to construct the partition of R. Although e_{IS} (8) will tend to increase monotonically with N_p, the relationship obviously depends on f. It has been found empirically that e_{IS} grows very rapidly with increasing N_p for small values ($N_p \leq 15,000$), and then slows down to a nearly linear dependence for larger values.

Let N_s denote the number of points to be sampled. The efficiency of the suggested adaptive importance sampling procedure is given by

$$E = \frac{N_s}{N_p + N_s/e_{IS}} = \frac{e_{IS}}{1 + (N_p/N_s)e_{IS}}. \tag{12}$$

As we would expect, E is always less than e_{IS}, but will approach e_{IS} as N_s increases.

From (12), we see that the number of function evaluations that should be expended to perform the partitioning in order to obtain maximum efficiency depends on N_s. If only a few points are to be sampled, it is not worthwhile to use many function evaluations in the partitioning phase. On the other hand, a substantial investment in the partitioning will be worthwhile if a great many points are to be sampled.

In order to illustrate the effectiveness of this procedure, we present in Table 1 the results of applying it to several problems. For each example, we give the simple acceptance/rejection efficiency e^* (3), the value of N_p and the resulting number of subregions, and the efficiency e_{IS} (8). The number N_p was taken as a "reasonable" value; as noted above, the best choice of N_p depends on the number of points to be sampled.

For each example, the number of sample points was computed for which the work saved by using the partitioning method is equal to the work necessary to construct the subregions. This "crossover" value N_c is given by

$$N_c = \frac{N_p}{1/e^* - 1/e_{IS}}.$$

Thus, for the first example of Table 1, allowing 250,000 function evaluations for the partitioning would yield a sampling efficiency $e_{IS} \approx 1.0$, with a corresponding crossover value of $N_c = 501$.

As the results indicate, dramatic improvements in sampling efficiency can be achieved by applying the proposed adaptive procedure.

References

Everett, C. J. and Cashwell, E. D. (1972). A Monte Carlo sampler, Report LA-5061-MS, Los Alamos Scientific Laboratory, New Mexico.

Friedman, J. H. and Wright, M. H. (1981a). A Nested Partitioning Procedure for Numerical Multiple Integration, *ACM Trans. Math. Software*, vol. 7, pp. 76–92.

Friedman, J. H. and Wright, M. H. (1981b). DIVONNE4: A Program for Multiple Integration and Adaptive Importance Sampling, CGTM 193, Stanford Linear Accelerator Center.

$$f_n = \frac{1}{2}\left(\frac{10}{\sqrt{\pi}}\right)^n \exp\left(-100\sum_{i=1}^{n}(x_i - \frac{1}{2})^2\right)$$

$$\hat{f}_n = \frac{1}{2}\left(\frac{10}{\sqrt{\pi}}\right)^n \exp\left(-100\sum_{i=1}^{n}(x_i - \frac{1}{3})^2\right) + \exp\left(-100\sum_{i=1}^{n}(x_i - \frac{2}{3})^2\right)$$

$$f_a = \prod_{i=1}^{10} i\, x_i^{i-1}$$

Table 1

Function	e^*	N_p	Regions	e_{IS}	N_c
f_4	.002	50,000	186	0.20	101
f_7	1.1×10^{-5}	152,000	334	0.023	1.7
$\hat{f}_2$	0.008	50,000	182	0.20	476
$\hat{f}_4$	0.0007	50,000	181	0.10	35
f_a	2.8×10^{-7}	250,000	589	0.0025	0.07

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SOL 81-23	2. GOVT ACCESSION NO. AD-A110 849	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN ADAPTIVE IMPORTANCE SAMPLING PROCEDURE		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Jerome H. Friedman and Margaret H. Wright		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0267 DAAG29-81-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research - SOL Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-047-143
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research - Dept. of the Navy 800 N. Quincy Street Arlington, VA 22217		12. REPORT DATE November 1981 13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

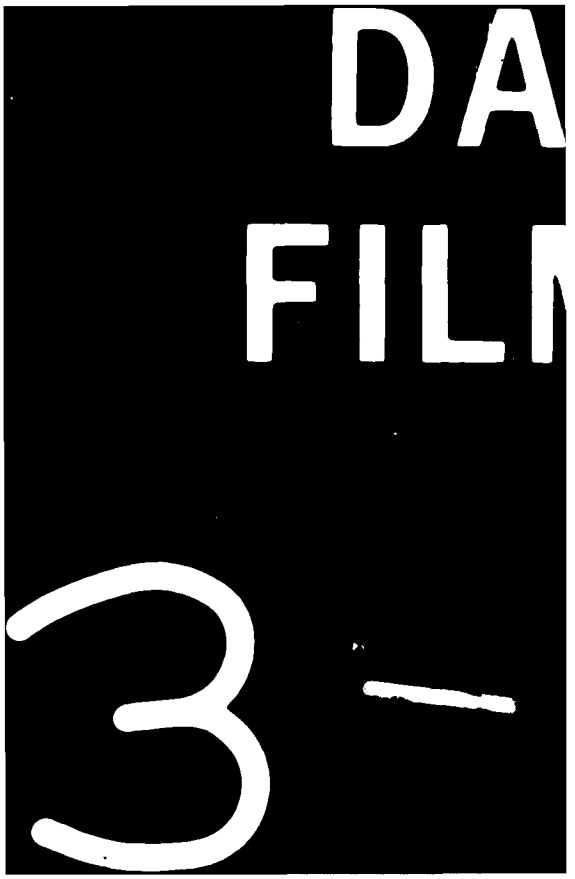
sampling
Monte Carlo procedures
probability distributions,
adaptive partitioning

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Monte Carlo calculations often require generation of a random sample of n-dimensional points drawn from a specified multivariate probability distribution. We present an importance sampling technique that can often greatly improve the efficiency of an acceptance/rejection generating method. The importance sampling function is defined as piecewise constant on a set of subregions, which are obtained by adaptively partitioning the sampling region so that the variation of density values within each subregion is relatively small. The partitioning strategy is based on multiparameter optimization to estimate the maximum and minimum of the original density function in each subregion.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


DA
FIL
3