New Monte-Carlo Procedures for Numerical Integration

by

John R. Birge
Robert L. Smith

Technical Report 82-8

Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

May, 1982

New Monte-Carlo Procedures for Numerical Integration

by

John R. Birge
Department of Industrial and
Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

Robert L. Smith
Department of Industrial and
Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

Abstract

A method for generating uniformly distributed points within arbitrary bounded convex regions is used to evaluate integrals over such regions. The procedure involves approximating the region by a mesh of boxes and of estimating the content medians in each direction. It is most efficient for regions that are nearly rectangular and may be quite useful in evaluating integrals over polytopes.

New Monte-Carlo Procedures for Numerical Integration

by

J. R. Birge and R. L. Smith

1. Introduction

The numerical evaluation of integrals over multi-dimensional regions has been a problem of interest since Clerk-Maxwell's [2] formulas for rectangles in 1877. Approximation procedures have also been successfully applied to standard lower dimensional regions, such as cubes, simplices and spheres. (See Stroud [7]). In higher dimensional problems, especially when the region can be represented as the Cartesian product of lower dimensional standard regions, Monte-Carlo methods become generally more efficient. Difficulties arise, however, when the functional domain is more complicated because of reasons discussed below.

The problem we consider is to evaluate numerically:

$$(1) \qquad S = \int_D f(x)dx,$$

where f is a continuous function from R^n into R, $x \in R^n$, D is a contented region in R^n and S is the value of the integral.

Deterministic methods for evaluating S are based on a so-called product rule (Stroud [7]). In these methods, m points, $x_1, x_2, \ldots, x_m$, in D are chosen and weights, $w_1, w_2, \ldots, w_m$, are assigned to these points. The result is an approximation of S,

$$(2) \qquad S^m = C(D) \sum_{i=1}^{m} w_i \, f(x_i)$$

where $\sum_{i=1}^{m} w_i = 1$ and C(D) is the n-dimensional content of D. The weights and points in (2) are generally chosen so as to ensure that $S^m = S$ for polynomials of degree up to some integer k. The weights are found by solving a system of equations which represents relationships among the coefficients in polynomials obtained as integrals over D of all polynomial functions of degree less than or equal to k. Using symmetrical properties of D, lower dimensional results can be applied to higher dimensional regions (Hammer ans Stroud [4]).

As dimensionality increases, however, these approximation techniques require prohibitively many function evaluations and prove inadequate for complicated regions. When n is large or when D is not an elementary region, Monte-Carlo procedures are generally considered the only practical alternative. The simplest Monte-Carlo estimate (Hammersly and Handscomb [5]) is:

$$S = \frac{C(D) \sum_{i=1}^{k} f(X_i)}{k} \qquad (3)$$

where $X_1, X_2, \ldots, X_k$ are points uniformly distributed throughout D. Two major difficulties may, however, occur in evaluating (3) because of the complexity of D:

a) $X_1, X_2, \ldots, X_k$ may not be efficiently generatable in D, and

b) C(D) may be difficult to estimate.

The classical techniques for solving (b) are i) to use the "hit-or-miss" procedure of generating points in a simpler region of known content that encloses D and of rejecting points that land outside of D and to estimate C(D) by this fraction, or ii) to decompose D into simpler regions of known content, such as simplices. The rejection procedure may prove inadequate in higher dimensions because the content of D may be quite small in comparison to the content of an enclosing region such as a sphere. Decomposing D may also be extremely difficult if D is irregular. For example, even decomposing a polytope into simplices requires the evaluation of all extreme points. In the discussion below, a method is given which avoids the difficulties inherent in these two approaches.

2. A Markov Chain Approach for Generating Uniform Sample Points, $X_1, X_2,\ldots,X_k$.

Boneh and Golan [1] and Smith [6] have developed procedures for generating uniformly distributed points over bounded regions. Their procedures are based on a symmetric mixing algorithm that we will adapt to our example.

We first need some basic definitions and assumptions about the region D. An n-dimensional rectangular submesh M is defined as:

$$(4) \qquad M \equiv \bigcup_{j=0}^{N} B_j$$

where $N < \infty$, $B^j = \{x: \{x: \ell_i^j \delta \leq x_i \leq \ell_i^j + 1)\delta,\ i = 1,\ldots,n\}$ is a box of M,δ is a small positive number, and ℓ_i^j are distinct integers for all i and j. A slab or level of M, L_i^k, is defined as:

$$(5) \qquad L_i^k = M \cap \{x: k\,\delta \leq x_i \leq (k + 1)\delta\},$$

where k is an integer.

M is <u>connected</u> if the graph, $G = (N, A)$ is connected where the nodes in N correspond to the centers of each box B^j and the arcs in A connect all nodes corresponding to adjacent (i.e., facet sharing) boxes in M.

M is strongly connected if M is connected <u>and</u> L_i^k is connected for all i and k.

We assume that D is a strongly connected n-dimensional rectangular submesh and note that any bounded convex region may be approximated in content arbitrarily well by such a rectangular mesh. The algorithm which follows for generating uniformly distributed points in D is a discrete version of a symmetric mixing algorithm.

<u>Algorithm</u>:

1) Let x^0 be the center of some box in D, set $i = 0$.

2) Choose a random direction D uniformly distributed over $\{e_1, e_2, \ldots, e_n\}$. Let $D = e_J$, let $Y(\lambda) = (X_1^i, \ldots, X_J^i + \lambda\delta, X_{J+1}^i, \ldots, X_n^i)$ and let $Y(\underline{\lambda}) = \min_{\lambda \varepsilon Z}\{Y(\lambda) \mid Y(\lambda) \varepsilon D\}$ and $Y(\bar{\lambda}) = \max_{\lambda \varepsilon Z}\{Y(\lambda) \mid Y(\lambda) \varepsilon D\}$, where Z is the set of all integers.
   Choose λ^* uniformly distributed on $\{\underline{\lambda}, \underline{\lambda} + 1, \ldots, -1, 0, 1, \ldots, \bar{\lambda} - 1, \bar{\lambda}\}$. Let $X^{i+1} = (X_1^i, \ldots, X_J^i + \lambda^* \delta, \ldots, X_n^i)$. Set $i = i + 1$.

3) If $i = K$, an upper bound on the number of iterations, STOP. Else, return to Step 2.

As will be shown, this algorithm asymptotically approaches the uniform distribution over M. We have discretized D in terms of M in order to find a confidence bound on the algorithm's rate of convergence to uniformity. We first need the following lemma on box-to-box transitions in D.

Lemma 1: There exists an integer $\tau < \infty$ such that

$$\varepsilon = \min_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}} P(X^{\tau} \varepsilon B^{j} \mid X^{0} \varepsilon B^{i}) > 0. \tag{6}$$

Proof: We wish to show that τ is an upper bound on the maximum number of transitions necessary in passing from any box to any other box. Let $h_i = \bar{\ell}_i - \underline{\ell}_i$ where $\bar{\ell}_i = \min\{\ell_i : \ell\ \delta \geq x_i,$ for all $x\ \varepsilon\ M\}$ and $\underline{\ell}_i = \max\{\ell_i : \ell\ \delta \leq x_i$ for all $x\ \varepsilon\ M\}$. h_i then is the maximum deviation in boxes of M in direction i and $h_i < \infty$ since N, the number of boxes, is finite. h_i is then a side length of the smallest enclosing rectangle about D.

We claim that $\tau \leq \sum_{i=1}^{n} h_i$. To see this, assume that we start at $X^0\ \varepsilon\ B^0$ and we will construct a path leading to $X^j\ \varepsilon\ B^j$ such that coordinate values are monotone increasing. On the first transition, we move without loss of generality to B^1 such that $\ell_i^1\ \delta = x_i \leq (\ell_i^1 + 1)\delta$ for $x\ \varepsilon\ B^1$ and require that $(\ell_i^0 + 1) \leq \ell_i \leq \ell_i + 1 \leq \ell_i^j$ for all i.

Since D is connected and the slab L_i^1 which includes B^1 is connected, the half-mesh $D^1 = D \cap \{x: x_i \geq \ell_i^1\ \delta\}$ is connected. Therefore, we can continue the algorithm on D^1. At the next transition, we can again divide D^1 and continue. We never return to a level previously crossed by the path. Hence, the maximum total number of possible transitions is equal to the sum of the number of slabs in each direction, $\tau \leq \nu \equiv \sum_{i=1}^{n} h_i < \infty$, where ν is defined as the sum of the side lengths of the enclosing rectangle. ■

The worst case possibility for τ occurs when the region is a <u>string</u> of boxes each connected to only one other box in the mesh. For this case, the path from the lowest box, $\underline{B} = \{ x \mid \underline{\ell}_i \delta \leq x_i \leq (\underline{\ell}_i + 1)\delta,\ i = 1,\ldots,n\}$, to the highest box, $\bar{B} = \{x \mid (\bar{\ell}_i - 1)\delta \leq x_i \leq \bar{\ell}_i \delta,\ i = 1,\ldots,n\}$, requires a direction change at each step. A string can be more generally viewed as a collection of $\nu - n + 1$ strongly connected rectangular regions. In fact, any mesh M can be decomposed into $\bar{\nu}$ rectangular regions, where

$$M = \bigcup_{j=1}^{\bar{\nu}} R^j, \tag{7}$$

and where R^j has sides of lengths, $s_1^j, s_2^j, \ldots, s_n^j$.

Now, we can reach any point in R^j in n steps, so, for general M, another upper bound on the number of required transitions in $\tau \leq n\,\bar{\nu}$. This bound may not be as convenient as ν to measure, but the decomposition can also give us a minimum probability for passing from any box to any other. For any rectangle, this minimum probability for R^j is

$$\eta^j = \prod_{i=1}^{n} \frac{1}{s_i^j}\,. \tag{8}$$

Next let us assume that at every step of the algorithm we cannot remain in the same box. By this assumption, we in general restrict the number of possible directions we may follow. Let $\bar{n}^j$ be the maximum number of admissible directions in exiting from R^j to any other rectangle $R^{j'}$.

We can now obtain the minimum transition probability ε in M. It is the product of the minimum transition probabilities within each rectangle and the minimum probabilities for exiting from one rectangle to another. That is,

$$\varepsilon \geq \prod_{j=1}^{\bar{\nu}} \frac{1}{\left[\prod_{\substack{i=1 \\ i \neq i_{min}}}^{n} \max\{s_i^j - 1, 1\}(s^j_{i_{min}})\right](\bar{n}^j)}, \tag{9}$$

where $s^j_{i_{min}} = \min_i \{s_i^j\}$. The number of boxes in M is $N = \sum_{j=1}^{\bar{\nu}} \left(\prod_{i=1}^{n} s_i^j\right)$,

and we define $\gamma = N\varepsilon$, where

$$\gamma \geq \sum_{j=1}^{\bar{\nu}} \prod_{k \neq j}^{\bar{\nu}} \left(\frac{1}{\bar{n}^k \left(\prod_{\substack{i=1 \\ i \neq i_{min}}}^{n} \max\{s_i^k - 1, 1\} s^k_{i_{min}}\right)} \right) \left(\frac{s^j_{i_{min}}}{s^j_{i_{min}} + 1} \right). \tag{10}$$

Lemma 2: For a mesh, M, of N boxes, and a minimum transition probability within M of ε, $\gamma = N\varepsilon \geq \dfrac{\nu - n + 1}{2^{(\nu - n + 1)}}$.

Proof: We will proceed by induction to show that the string has the lowest possible γ. First, we observe that the construction of the string forces a single box at every δ increment in each direction. Starting from the first box, this leads to $\nu - n$ additional boxes. Hence, for the string, $N = \nu - n + 1$.

For $\nu - n + 1$ even, the string can be decomposed into $\bar{\nu} = \dfrac{\nu - n + 1}{2}$ rectangles with all sides of length 1 except one side of length 2. The maximum number of possible directions in any of these rectangles is 2, hence

$$\gamma = N\varepsilon \geq (\nu - n + 1)\left(\prod_{j=1}^{\frac{\nu-n+1}{2}} \frac{1}{2^2}\right)$$

$$= \frac{\nu - n + 1}{2^{(\nu - n + 1)}}. \tag{11}$$

For $\nu - n + 1$ odd, there are $\frac{\nu - n}{2}$ rectangles of 2 boxes and a single box. In the single box, only one direction is possible, so

$$\gamma \geq N \varepsilon \geq (\nu - n + 1) \left(\frac{1}{1 \cdot 2 \cdot 1}\right) \left(\prod_{j=1}^{\frac{\nu-n}{2}} \frac{1}{2^2}\right)$$

$$= \frac{\nu - n + 1}{2^{(\nu-n) + 1}}. \qquad (12)$$

Every mesh of given dimensions $h_1, h_2, \ldots, h_n$ summing to ν must have at least $\nu - n + 1$ boxes since that is the minimum number of boxes that can be traversed in following the algorithm from $\underline{B}$ to $\bar{B}$. We wish to show that the mesh has the lowest γ over all meshes of $\nu - n + 1$ boxes. To do this, we proceed by displacing boxes from the mesh while preserving the connectedness properties. The number of possible directions in the new mesh will not increase because the path from $\underline{B}$ to $\bar{B}$ must cross over every other box and, therefore, only two directions are admissible from any box. This observation shows that we may still consider the mesh as the union of one-and two-box rectangles. Hence, γ is still greater than or equal to $\frac{\nu - n + 1}{2^{(\nu-n+1)}}$.

Starting from a mesh of $\nu - n + 1$ boxes, any mesh enclosed by a rectangle of side lengths $h_1, h_2, \ldots, h_n$ can be constructed by adding boxes. We will proceed by induction on the number of boxes, N_b, in the mesh. Let ε^{N_b} and $\gamma^{N_b} = N_b \varepsilon^{N_b}$ be the minimum values ε and γ for a mesh of N_b boxes. We know $\gamma^{N_b} \geq \frac{\nu - n + 1}{2^{(\nu-n+1)}}$ for $N_b = \nu - n + 1$, and we assume that

$$\gamma^{N_b} \geq \frac{\nu - n + 1}{2^{(\nu-n+1)}} \quad \text{for all } N_b \leq k > \nu - n + 1.$$

Assume that the k + 1st box is added to rectangle R^1 of some mesh of N_b boxes. If the k + 1st box is not adjacent to any box in a minimum transition probability path, then γ^{N_b} remains the same. If it is adjacent to a path then we consider that either the number of possible directions from R^1 increases from $\bar{n}^1$ to $\bar{n}^1 + 1$ or the length of some side i of R^1 has increased to $s_i^1 + 1$. The new box must also be connected to some other rectangle along a minimum probability path in order to maintain connectedness. Therefore, in proceeding along a path between two boxes most distant from each other, we can either follow one of the previous paths or proceed through the new box. The minimum probability is then either

$$\varepsilon^{k+1} \geq \left(\frac{\bar{n}^1}{\bar{n}^1 + 1}\right)\varepsilon^k + \left(\frac{1}{\bar{n}+1}\right)\varepsilon^k = \varepsilon^k, \text{ or}$$

$$\varepsilon^{k+1} \geq \left(\frac{s_i^1}{s_i^1 + 1}\right)\varepsilon^k + \left(\frac{1}{s_i^1 + 1}\right)\varepsilon^k = \varepsilon^k, \text{ or}$$

$$\varepsilon^{k+1} \geq \left(\frac{s_{min}^1 + 1}{s_{min}^1 + 2}\right)\varepsilon^k + \left(\frac{1}{s_{min}^1 + 2}\right)\varepsilon^k \qquad (13)$$

$$= \varepsilon^k.$$

Since, the number of boxes, k + 1 > k, we have $\gamma^{k+1} > \gamma^k$. Hence, the induction is proved. ■

Given Lemmas 1 and 2, we have the following theorem.

Theorem: For X_0, $X_1, \ldots,$ generated by the symmetric mixing algorithm over a strongly connected region $D = \bigcup_{i=1}^{N} B_i$ and where the uniform distribution on D is defined as $\mu(B_i) = \frac{\delta^n}{C(D)}$,

$$|P(x^m \in B_i \mid x^0 \in B_0) - \mu(B_i)| < (1-\gamma)^{[(\frac{m}{\tau}) - 1]} \qquad (14)$$

where $\tau \leq \nu - n + 1$ and $\gamma \geq \frac{(\nu - n + 1)}{2^{(\nu-n+1)}}$.

Proof: Follows from Lemma 1, Lemma 2, and Doob [3], p. 173. ■

The bound for γ found above does not lead to quick convergence to uniformity. The bound in (14) should, however, be taken as the worst case. It is a bound on deviation from uniformity within any box. Uniformity may also be achieved with some confidence on a certain proportion of the boxes.

For more computational utility, we may update N, ε and τ as we proceed with the algorithm. We begin by assuming that $\tau = \nu-n+1 = \sum_{i=1}^{n} h_i - n + 1$,

$$\varepsilon = \frac{1}{2^{(\nu-n+1)}} \text{ and } N = \nu - n + 1.$$

ε is updated by considering the maximum possible movement found by the algorithm in each direction, λ_i^{max}. When a new maximum movement is found, we wish to find the largest rectangle enclosing this line segment. Assume that we have found B' and B" such that the ith coordinate of the center of B', x_i^1, is equal to $\delta\lambda_i^{max} + x_i''$, the ith coordinate of the center of B". Given B' and B", we find the boundaries of D in every direction from B' and B". We take the smallest common intervals in each direction from B' and B" to find the enclosing rectangle, which is in D because of the connected ness property.

We repeat this search procedure whenever a new maximum coordinate diameter is found for D. We also restrict the rectangles to be disjoint of any formed previously. In this way, we construct n rectangles with side lengths $\lambda_i^1, \lambda_i^2, \ldots, \lambda_i^n$ for each direction i. We can also determine the minimum proportion of boxes within each rectangle from which we can enter another rectangle and the minimum probability of entering another rectangle. We call this value ρ_i, where $\rho_i \geq \dfrac{1}{\max_j\{\lambda_i^j\}} \cdot \dfrac{1}{n}$.

The minimum number of transitions is now the sum of the constructed rectangles around the maximum deviation in each direction and the worst case string for the remaining portion of the enclosing rectangle. We, therefore, have

$$\tau \leq n^2 + \sum_{i=1}^{n} (n_i - \lambda_i^{max}) + 1. \tag{15}$$

The minimum transition probability is the product of transition probabilities within the constructed rectangles and within the remaining string. Hence,

$$\varepsilon \geq \prod_{i=1}^{n} \left(\rho_i \cdot \frac{1}{\prod_{j=1}^{m} \lambda_i^j}\right) \cdot \frac{1}{2^{\nu} - \sum_{i=1}^{n} \lambda_i^{max} + 1} \cdot \tag{16}$$

Upper and lower bounds on the number of boxes N may be obtained by keeping a list of different boundary points obtained during the algorithm. For example, if $X^k = (X_1^k, X_2^k, \ldots, X_n^k)$ is the next chosen iteration point, j is the direction chosen at this iteration, and there do not exist boundary points, ω^ℓ and $\omega^{\ell+1}$, such that $\omega_i^\ell = \omega_i^{\ell+1} = X_i^k$ for all $i \neq j$, then there exist $\bar{\lambda} + \underline{\lambda}$ boxes between the boundary points, $X^k - \underline{\lambda}\,\delta e_j$ and $X^k + \bar{\lambda}\,\delta e_j$. We check the list of boundary points to find the number β of these boxes that has previously been found. We update a lower bound N_L on N by $N_L = N_L + \bar{\lambda} + \underline{\lambda} - \beta$. We can also update an upper bound N^U on N by considering boxes outside of D. Again, we can use the other boundary points to determine the number ξ of box centers x^1 on the line $X^k + \alpha e_j$ and not in D that have already been identified. N^V is then updated by $N^U = N^U - h_j + \bar{\ell}_i + \bar{\ell}_j + \xi$.

The convergence to uniformity accelerates quickly as $\bar{\nu}$, the minimum number of rectangles required to partition D, decreases. We consider the class of meshes with equally dimensioned rectangular subregions. The worst case is the string composed of unit boxes. In this case, there are $\nu - n + 1$ rectangles of a single box. The best case is when there is one rectangle with sides of length $\frac{\nu}{n}$. Here,

$$1 = \frac{\nu}{\left(\frac{\nu}{n}\right)} - n + 1.$$

The other elements of this class contain rectangles of sides with length $s\frac{\nu}{n}$ where $\frac{n}{\nu} \leq s \leq 1$. For this class, we can proceed as in the string and reduce the number of rectangles to $(\frac{n}{s} - n + 1)/2$ rectangles with sides of length $s(\frac{\nu}{n})$ except for one side of length $2s(\frac{\nu}{n})$. (We omit the possible cases of non-integer numbers of rectangles for ease of exposition.)

For each class, the minimum number of transitions is

$$\tau \geq \frac{n}{s} - n + n = \frac{n}{s}, \tag{17}$$

since one transition is required to locate in the region of the rectangle from which the next rectangle is admissible, one transition is required to enter the next rectangle, and n transitions are required to obtain any box in the destination rectangle. The probability of making the first of these transitions is $(\frac{1}{n})\ (\frac{1}{2})$ and the probability of the second transition is also $\frac{1}{2n}$. Hence

$$\varepsilon \geq \left\{ \prod_{j=1}^{(\frac{n}{s}-n+1)/2} \frac{1}{4n^2} \right\} \cdot \left(\frac{1}{(\frac{s\nu}{n})^n} \right)$$
$$= \left(\frac{1}{4n^2} \right)^{(\frac{n}{s} - n+1)/2} \left(\frac{1}{(\frac{s\nu}{n})^n} \right), \tag{18}$$

where $\frac{1}{(\frac{s\nu}{n})^n}$ is the minimum probability of an n-step transition in the destination rectangle.

We also have $N = (\frac{n}{s} - n + 1)\ (\frac{s\nu}{n})^n$, so

$$\gamma \geq (\frac{n}{s} - n + 1)(\frac{s\nu}{n})^n\ (\frac{1}{4n^2})^{(\frac{n}{s} - n + 1)/2}\ \frac{1}{(\frac{s\nu}{n})^n} \cdot$$

$$= (\frac{n}{s} - n + 1)\ (\frac{1}{4n^2})^{(\frac{n}{s} - n + 1)/2} . \tag{19}$$

We can improve further on these bounds by considering a larger τ in the evaluation of ε. We define

$$\bar{\tau} = \left\{(\frac{n}{s} - n + 1)/2\right\} 2n + n = n^2(\frac{1-s}{s}) + 3n, \tag{20}$$

where we allow for ^{2}n steps within each rectangle. Now, the conditional distribution, given that we do no exit a rectangle, is uniform in each rectangle. In a transition from rectangle R^1 to R^{11}, we allow n steps for uniformity in R^1 and another n steps for uniformity in R^{111}, the rectangle formed by half of R^1 and half of R^{11}. We condition first on not exiting R^1

and second on not exiting R^{111}. A minimum probability for entering R^{111} is the product of the probability of staying in R^1, the probability of entering R^{111}, the probability of staying in R^{111}, and the probability of entering R^{11}, $(1 - \frac{1}{2n})^n \cdot \frac{1}{2} \cdot (1 - \frac{1}{2n})^n \cdot \frac{1}{2} = (\frac{2n-1}{n})^{2n} (\frac{1}{4})$. Hence

$$\varepsilon \geq \{ \prod_{j=1}^{(n/s-n+1)/2} (\frac{2n-1}{2n})^{2n} (\frac{1}{4})\} (\frac{1}{(\frac{s\nu}{n})^n}) , \qquad (21)$$

and

$$\gamma \geq (\frac{n}{s} - n+1) (\frac{2n-1}{2n})^{n(\frac{n}{s}-n+1)} (\frac{1}{4})^{(\frac{n}{s} - n+1)/2} . \qquad (22)$$

In the string case, this bound is not tight since we assumed that only one direction was admissible in each box. The bound in (22) does also not strictly apply to the case where s = 1 since only one rectangle is present in that case. These examples for worst case, $s = \frac{n}{\nu}$, and best case, s = 1, do provide a range for γ. For $s = \frac{n}{\nu}$,

$$\gamma \geq (\nu - n + 1) (\frac{1}{4})^{(\nu-n+1)/2} (\frac{2n-1}{2n})^{n(\nu-n+1)},$$

and for s = 1,

$$\gamma \geq \frac{1}{2} (\frac{2n-1}{2n})^n ,$$

For $s = \frac{n}{2\nu}$,

$$\gamma \geq (\frac{\nu}{2} - n + 1) (\frac{1}{4})^{(\frac{\nu}{2} - n+1)/2} (\frac{2n-1}{2n})^{n(\frac{\nu}{2} -n+1)},$$

and for $s = \frac{1}{2}$,

$$\gamma \geq (n + 1) (\frac{1}{4})^{(n+1)/2} (\frac{2n-1}{2n})^{n(n+1)} .$$

We can see how the exponential decrease in γ as s approaches 1 leads to quicker convergence to uniformity on an example. We let $n = 10$ and $\nu = 100$. We wish to find m such that

$$\frac{| P(x^m \varepsilon B_i \mid x^0 \varepsilon B^0) - \mu(B_i) |}{\mu(B_i)} \leq \alpha \tag{23}$$

From (14), we want m such that

$$(1 - \gamma) \left(\frac{m}{\tau} - 1 \right) \leq \alpha\, \mu(B_i) = \frac{\alpha}{N} , \tag{24}$$

or

$$m > \tau \left(\frac{- \ell n\, \alpha/N}{\ell n (1 - \gamma)} - \ell n\ (1-\gamma) \right) \tag{25}$$

If we begin with the string case where $s = \frac{n}{\nu} = \frac{1}{10}$, then $\tau = 91$, and $N = 91$ and from (11), $\gamma \geq 91/2^{91}$. Here, $m > 2 \times 10^{28}$. For $s = 2/3$, $\bar{\tau} = 80$, $N = 10^9$, $\gamma \geq 4.3 \times 10^{-3}$, and we find $m > 4.4 \times 10^5$. We note that these cases are taken from a class that essentially represents the worst case for regions composed of $\frac{n}{s} - n + 1$ rectangles since each rectangle is connected to at most two others. We would expect that general regions have more interconnections.

3. The Use of $X_1, X_2, \ldots, X_k$ to Estimate S.

The next step of the procedure is to investigate each coordinate direction to estimate the content median in that direction, that is, the point m_i such that $C[D \cap \{x: x_i \geq m_i\}] = C[D \cap \{x: x_i \leq m_i\}]$. We can define a distribution on L_i by

$$g(L_i^k) = \frac{C(L_i^k)}{C(D)} , \qquad k = \underline{\ell}_i, \underline{\ell}_{i+1}, \ldots, \bar{\ell}_i. \tag{26}$$

Following the symmetric mixing algorithm can be viewed as taken a random sample, $y_i^1, y_i^2, \ldots, y_i^N$ from $L_i = \{L_i^{\underline{\ell}_i}, L_i^{\underline{\ell}+1}, \ldots, L_i^{\bar{\ell}_i}\}$ where $y_i^j = L_i^j$ if $x^j \varepsilon L_i^j$. We order $\{y_i^j\}$ so that $y_i^1 \leq y_i^2 \leq \cdots \leq y_i^N$, and define $z_i^j = P\{L_i \leq y_i^j\}$. We also let $m_{L_i} = L_i$ such that $m_i \varepsilon L_i$.

$$P\{y_i^j \le m_{L_i}\} = P\{z_i^j \le \frac{1}{2}\} \tag{27}$$

$$= \sum_{q=j}^{k} \binom{k}{q} \left(\frac{1}{2}\right)^q \left(\frac{1}{2}\right)^{k-q} .$$

Hence, we can find a confidence interval on m_i by

$$P\{y_i^{j-1} \le m_i \le y_i^j\} = 2 \sum_{q=j}^{k} \binom{k}{q} 2^{-k} . \tag{28}$$

For $y_i^{j-1} = y_i^j = L_i^k$ we estimate m_i by $\hat{x}_i = \dfrac{\ell_i^k\delta + (\ell_i^k + 1)\delta}{2}$. If $L_i^{j-1} = y_i^{j-1} < y_i^j = L_i^j$, we let $\hat{x}_i = \dfrac{\ell_i^{j-1}\delta + (\ell_i^j + 1)\delta}{2}$. We do this for each i and obtain $\hat{x} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$ which belongs to some box, $\hat{B}$, an unbiased estimator of the box which contains the median $m = (m_1, m_2, \ldots, m_n)$.

Using the planes $x_i = \hat{x}_i$, we can cut D into 2^n pieces, $\zeta_0, \zeta_2, \ldots, \zeta_{2^n-1}$, of mean content $C(D)/2^n$. Each ζ_j is defined

$$\zeta_j = D \cap H_1^{j_1} \cap H_2^{j_2} \cap \cdots \cap H_n^{j_n} , \tag{29}$$

where $j = j_1 + 2j_2 + \cdots + 2^{n-1}j_n$ and $H_i^{j_i} = \{x: x_i \le x_i^n\}$ if $j_i = 0$ and $H_i^{j_i} = \{x: x_i \ge \hat{x}_i\}$ if $j_i = 1$. These pieces can be viewed as an n-dimensional simplex σ^n with a single irregular face. The vertices of this simplex are the apex, $\hat{x}$, and $x_i = \bar{x}_i = \max\{x_i: x \in D \text{ and } x_j = \hat{x}_j \text{ for all } j \neq i\}$ or $x_i = \underline{x}_i = \min\{x_i: x \in D \text{ and } x_j = \hat{x}_j \text{ for all } j = i\}$. Letting $y_1, y_2, \ldots, y_n$ be the non-apex vertices for a given piece, the irregular face is the surface corresponding to an n-1 dimensional simplex σ^{n-1} determined by $y_1, y_2, \ldots, y_n$. The content of a single piece can be estimated by first choosing M uniformly distributed i.i.d. points in σ^{n-1}. For each of these points, z_j, an orthonormal vector v_j to σ^{n-1} is calculated and we let:

$$h(z_j) = \max_{\lambda}\{z_j + \lambda v_j \mid z_j + \lambda v_j \in D\}, \text{ and let} \tag{30}$$

$$g(z_j) = \lambda \text{ such that } h(x_j) = x_j + \lambda v_j. \tag{31}$$

The content of a piece ζ can then be estimated as

$$C(\zeta) \cong C(\sigma^n) + \frac{1}{C(\sigma^{n-1})} \sum_{j=1}^{M} \frac{g(z_j)}{M}, \tag{32}$$

as in the Monte-Carlo procedure for estimating content and where

$$C(\sigma^n) = \frac{1}{n!} \det \begin{pmatrix} y_1 \\ y_2 \\ y_n \end{pmatrix} .$$

We estimate $C(\zeta)$ for a sample of k of the 2^n pieces. For k large, we can apply the Central Limit Theorem to obtain an estimate, $\hat{C}(\zeta)$, of the average content of a piece. An unbiased estimate of the content of D is then:

$$\hat{C}(D) = 2^n \hat{C}(\zeta). \tag{33}$$

By letting $s^2 = \sum_{j=1}^{M} (C(\zeta_j) - \hat{C}(\zeta))/M-1$ be an estimate of the variance, we obtain a confidence interval on $\mu = C(D)/2^n$ as

$$P\{\mu - t_{\alpha/2} \frac{s}{\sqrt{M}} \leq \hat{C}(\tau) \leq \mu + t_{\alpha/2} \frac{s}{\sqrt{M}}\} = 1 - \alpha, \tag{34}$$

where $t_{\alpha/2} = \{y \mid P\{T \leq y\} = \alpha/2\}$ for T having the t-distribution with M degrees of freedom. We note that if D has a great deal of symmetry then the sample variance should be small and the error in estimating $\hat{C}(\zeta)$ will be reduced.

4. Conclusion

We have shown that the integral of a function over any bounded convex region in R^n can be estimated by efficient Monte-Carlo procedures. Our procedure involves approximating the region by a rectangular mesh of arbitrary box size and by generating uniformly distributed points within this region. Uniformity is approached by the symmetric mixing algorithm at a faster rate as the fractional content γ of D to an enclosing rectangle increases. This property is important for D, a polytope, as in $D = \{x: Ax = b\}$ where A is an m x n matrix and b is an m-vector. In most applications, A is very sparse and γ is large relative to arbitrary regions.

Given the uniform distribution of m points in D, the procedure used non-parametric estimation to find the content median in each dimension. These medians were then used to divide D into pieces whose content we could more easily approximate. The estimates of these piece contents were then used to estimate C(D).

Our method is an extension of the "sample-mean" [5] or crude Monte-Carlo procedure which has lower variance than the hit-or-miss method. Other variance reduction techniques such as importance sampling and stratified sampling require information about the region which we have not assumed. They may, however, be used to refine the procedure above in specific cases.

REFERENCES

[1] Boneh, A. and A. Golan, "Constraint redundancy and feasible point generator," presented at EURO III, Amsterdam, 1979 also in Karwar, M.H., J. Telgen and S. Zionts (eds) Redundancy in Mathematical Programming, Springer-Verlag, New York-Heidelberg, Berlin (1981).

[2] Clerk-Maxwell, J., "On approximate multiple integration between limits of summation," Proceedings, Cambridge Philosophical Society, Vol. 3 (1877), pp. 39-47.

[3] Doob, J.L., Stochastic Processes, John Wiley and Sons, New York, 1953.

[4] Hammer, P.C. and A.H. Stroud, "Numerical evaluation of integrals I," Mathematical Tables and Aids for Computation, Vol. 11 (1957), pp. 59-67.

[5] Hammersly, J.M. and D.C. Handscomb, Monte Carlo Methods, John Wiley and Sons, New York, 1964.

[6] Smith, R.L., "Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions," Technical Report No. 81-1, The University of Michigan, Department of Industrial and Operations Engineering, 1981.

[7] Stroud, A.H., Approximate Calculation of Multiple Integrals, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.