Stochastic Quadrature Formulas

By Seymour Haber

Abstract. A class of formulas for the numerical evaluation of multiple integrals is described, which combines features of the Monte-Carlo and the classical methods. For certain classes of functions—defined by smoothness conditions—these formulas provide the fastest possible rate of convergence to the integral. Asymptotic error estimates are derived, and a method is described for obtaining good a posteriori error bounds when using these formulas. Equal-coefficients formulas of this class, of degrees up to 3, are constructed.

1. **Introduction.** In the simplest "Monte-Carlo" scheme for numerically approximating the integral

$$I = \int_{G_s} f(\mathbf{x}) d\mathbf{x}$$

($\mathbf{x} = (x^1, x^2, \cdots, x^s)$ is a vector in s-dimensional Euclidean space; G_s is the closed unit cube $0 \leqq x^i \leqq 1$, $i = 1, 2, \cdots, s$), one chooses N points $\mathbf{x}_1, \cdots, \mathbf{x}_N$ at random in G_s and takes the quantity

$$J_0 = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i)$$

as an estimate of I. The error of the estimate is stated probabilistically; J_0 is regarded as a random variable, defined in terms of the random variables $\mathbf{x}_1, \cdots, \mathbf{x}_N$ which are taken to be independent (or at least pairwise independent), and uniformly distributed on G_s. Then the mean value $m(J_0)$ is I, and the standard deviation $\sigma(J_0)$ is $C_0 N^{-1/2}$ (for $f \in L^2(G_s)$), where C_0 is a constant depending on f; it is usual to consider 3σ (or even 2σ) as a reliable upper bound on $|I - J_0|$.

In [1] and [2] I proposed two modifications of this scheme: In the first, for $N = K^s$, K an integer, G_s is partitioned into N congruent subcubes; ordering these in some manner, a single point $\mathbf{x}_i$ is chosen at random from the ith subcube, and

$$J_1 = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i)$$

is taken as the estimate of I. In the second, the same partition of G_s is used, and, for each i, $\mathbf{x}_i$ is chosen as before and the point symmetrically opposite it in the ith subcube is denoted $\mathbf{x}_i'$; and the integral is estimated by

$$J_2 = \frac{1}{2N} \sum_{i=1}^{N} (f(\mathbf{x}_i) + f(\mathbf{x}_i')) .$$

These estimates have the following property: Let $D_s{}^n$ denote the set of functions f defined on G_s and such that

---

$$\frac{\partial^{n_1+\cdots+n_s}}{(\partial x^1)^{n_1} \cdots (\partial x^s)^{n_s}} f(x^1, \cdots, x^s)$$

is continuous on G_s whenever $n_1 + n_2 + \cdots + n_s \leqq n$. Then for $f \in D_s^1$

$$\sigma(J_1) = (C_1(f) + o(1))N^{-(1/s+1/2)}$$

and for $f \in D_s^2$

$$\sigma(J_2) = (C_2(f) + o(1))N^{-(2/s+1/2)}.$$

The surprising point about these estimates is that if one evaluates I by any classical quadrature method—Newton-Cotes, Gaussian, etc.—the best that could be said in general for $f \in D_s^1$ is that the error is $O(N^{-1/s})$ and for $f \in D_s^2$ is that the error is $O(N^{-2/s})$. Thus for the function classes involved, the probabilistic estimates J_1 and J_2 simply add the $\frac{1}{2}$ which is characteristic of Monte-Carlo to the best exponents of convergence obtainable by classical methods.

This is in fact part of a general situation. N. S. Bahvalov [3], in a study of lower bounds on quadrature errors, showed that for the class D_s^n the error of any deterministic method is $\Omega(N^{-n/s})$*; for methods with probabilistic features the lower bound he found was $\sigma = \Omega(N^{-(n/s+1/2)})$. For the set of periodic functions in D_s^n he constructed a probabilistic method for which in fact $\sigma = O(N^{-(n/s+1/2)})$.

In this paper I shall describe a large class of formulas which combine the Monte Carlo approach with classical considerations to give errors of the order of $N^{-(n/s+1/2)}$ for the class D_s^n; and I will construct, in the case $s = 2$, formulas of this class having some especially desirable properties. The basic idea is the modification of the Monte-Carlo sampling and averaging to produce formulas which give exact answers for certain classes of polynomials. This lets us obtain the additional accuracy which arises from the possibility of approximating the functions to be integrated by polynomials—and it is this possibility which underlies the classical approach to numerical quadrature. Most of these results were announced without proof in [4].

For convenience I shall deal with integration, not over G_s, but over the larger cube $-1 \leqq x^i \leqq 1$, $i = 1, 2, \cdots, s$ which I shall call H_s.

2. Stochastic Quadrature Formulas.

Definition. A "stochastic quadrature formula (s.q.f.) of degree n (for H_s)" is a sequence of 1-dimensional random variables $A_1, \cdots, A_k$ and s-dimensional random variables $\mathbf{X}_1, \cdots, \mathbf{X}_k$ such that:

(1) $\sum_{i=1}^k A_i P(\mathbf{X}_i) \equiv \int_{H_s} P$ for every polynomial P (in s variables) of degree n or lower; and there is some polynomial of degree $n + 1$ for which this equality does not hold.

(2) For any $f \in L^2(G_s)$, $m(\sum_{i=1}^k A_i f(\mathbf{X}_i)) = \int_{H_s} f$ (where "$m(\cdot)$" denotes the mean value of a random variable).

For example, $\mathbf{X}_1$ uniformly distributed over H_s, $\mathbf{X}_2 = -\mathbf{X}_1$, and $A_1 \equiv A_2 \equiv 2^{s-1}$ define an s.q.f. of degree 1.

I shall write "$Q(f)$" for $\sum_{i=1}^k A_i(f(\mathbf{X}_i))$, and speak of "the quadrature formula Q", or the "k-point quadrature formula Q." In the usual way one may apply an

---

* Hardy's notation: $f = \Omega(g)$ iff $g = O(f)$.

s.q.f. Q to any region A obtainable from G_s by an affine transformation, without changing its degree. The adapted formula will be denoted "Q_A." I shall denote by "Q_M" the formula resulting from partitioning H_s into M congruent subcubes and applying Q independently to each. The letter N will be reserved for the number of function evaluations used in a quadrature formula; if Q is a k-point s.q.f., then for Q_M, $N = kM$.

THEOREM 1. *There exist s.q.f.'s of any degree.*

Proof. Let n be any integer $\geqq 0$. Let $\phi_1, \phi_2, \cdots, \phi_L$, $L = (n + s)!/n!s!$, be the distinct monomials of degree n or less in s variables. Let $\mathbf{x}_1, \cdots, \mathbf{x}_L$ be points in H_s such that the matrix $(\phi_j(\mathbf{x}_i))$ is nonsingular. (That this can be done is shown, e.g., in [4].) Then for any f defined on H_s, there is a unique polynomial of degree n or less which coincides with f at $\mathbf{x}_1, \cdots, \mathbf{x}_L$; call it P_f. Let $\mathbf{X}_1$ be a random variable uniformly distributed on H_s, and set

$$Q(f) = 2^s(f(\mathbf{X}_1) - P_f(\mathbf{X}_1)) + \int_{H_s} P_f ;$$

Q is an s.q.f. of degree n. (Formally, in the notation of the Definition, there are $L + 1$ random variables $\mathbf{X}_i$, the first of which is $\mathbf{X}_1$ while the last L are concentrated at the points $\mathbf{x}_1, \cdots, \mathbf{x}_L$ respectively; $A_1 \equiv 1$, while the last L A_i's are expressions involving the $\mathbf{X}_i$, arising from $P_f(\mathbf{X}_1)$ and $\int_{H_s} P_f$.)

The s.q.f.'s constructed in this proof are not apt to be practically useful.

The next theorem shows how stochastic quadrature formulas can be used to obtain approximations with errors of the order desired. We make the following notational conventions:

Bold-face letters $\mathbf{i}$ and $\mathbf{j}$ will denote ordered n-tuples of (not necessarily distinct) integers between 1 and s. If $\mathbf{i} = (i^1, i^2, \cdots, i^n)$, $\mathbf{x}$ is a point in s-space, and f is a function of s variables, then we set

$$f^{(\mathbf{i})} = \frac{\partial^n f}{(\partial x^{i^1}) \cdots (\partial x^{i^n})} ; \qquad \mathbf{x}^{\mathbf{i}} = x^{i^1} \cdot x^{i^2} \cdot \cdots \cdot x^{i^n} .$$

THEOREM 2. *If Q is a k-point stochastic quadrature formula of degree $n - 1$, and $f \in D_s^n$, then*

(2)
$$\sigma(Q_M(f)) \sim C(f)M^{-(n/s+1/2)} ,$$

where

(3)
$$C(f) = \frac{1}{2^{s/2}n!} \left(\sum m_{\mathbf{i}\mathbf{j}} \int_{H_s} f^{(\mathbf{i})}f^{(\mathbf{j})} \right)^{1/2} ,$$

and

(4)
$$m_{\mathbf{i}\mathbf{j}} = m\left(\left(Q(x^{\mathbf{i}}) - \int_{H_s} x^{\mathbf{i}} \right)\left(Q(x^{\mathbf{j}}) - \int_{H_s} x^{\mathbf{j}} \right) \right) .$$

Here the sum in (3) is over all n-tuples $\mathbf{i}$ and $\mathbf{j}$ of integers from 1 to s, and "$g(M) \sim h(M)$" means $g(M)/h(M) \to 1$ as $M \to \infty$.

Note that in terms of the number $N = kM$ of function evaluations in Q_M,

(5)
$$\sigma(Q_M(f)) \sim C(f)k^{(n/s+1/2)}N^{-(n/s+1/2)} .$$

Proof. Let $K_1, \cdots, K_M$ be the subcubes into which H_s is partitioned, enumerated somehow. Then, since Q is applied to the different K's independently,

$$(6) \qquad \sigma^2(Q_M(f)) = \sum_{r=1}^{M} \sigma^2(Q_{K_r}(f)) \ .$$

Let $\mathbf{C}_r$ be the center of K_r, and, for $\mathbf{x} \in K_r$, set $\boldsymbol{\delta} = \mathbf{x} - \mathbf{C}_r$. Then, since $f \in D_s{}^n$, for $\mathbf{x} \in K_r$ we can write

$$(7) \qquad \begin{aligned} f(\mathbf{x}) = f(\mathbf{C}_r) + \sum_1 f^{(\mathbf{i})}(\mathbf{C}_r)\boldsymbol{\delta}^{\mathbf{i}} + \frac{1}{2!} \sum_2 f^{(\mathbf{i})}(\mathbf{C}_r)\boldsymbol{\delta}^{\mathbf{i}} + \cdots \\ \cdots + \frac{1}{n!} \sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\boldsymbol{\delta}^{\mathbf{i}} + o(M^{-n/s}) \end{aligned}$$

where $\sum_L$, $L = 1, 2, \cdots, n$, is over all L-tuples $\mathbf{i}$ of integers from 1 to s.

Since Q is of degree $n - 1$, we have

$$\begin{aligned} \sigma^2(Q_{K_r}(f)) &= \sigma^2\left(Q_{K_r}(f) - \int_{K_r} f\right) \\ &= \frac{1}{(n!)^2} \sigma^2\left(Q_{K_r}\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\boldsymbol{\delta}^{\mathbf{i}}\right) - \int_{K_r}\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\boldsymbol{\delta}^{\mathbf{i}}\right) + R_r\right) \end{aligned}$$

where $R_r = o(M^{-n/s-1})$.

Applying the affine transformation taking K_r onto H_s, to the last integral, and noting that the integrand is a homogeneous polynomial of degree n, we have

$$\int_{K_r}\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\boldsymbol{\delta}^{\mathbf{i}}\right) = (M^{-1/s})^{n+s} \int_{H_s}\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\mathbf{x}^{\mathbf{i}}\right),$$

and similarly

$$Q_{K_r}\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\boldsymbol{\delta}^{\mathbf{i}}\right) = (M^{-1/s})^{n+s} Q\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\mathbf{x}^{\mathbf{i}}\right).$$

Therefore, setting $t_r = \sigma^2(Q_{K_r}(f))$, we have

$$\begin{aligned} t_r &= \frac{1}{(n!)^2}(M^{-2/s})^{n+s}\sigma^2\left(Q\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\mathbf{x}^{\mathbf{i}}\right) - \int_{H_s}\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\mathbf{x}^{\mathbf{i}}\right) + R_r{}'\right) \\ &= \frac{1}{(n!)^2}(M^{-2/s})^{n+s}\sigma^2\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\right)Q(\mathbf{x}^{\mathbf{i}}) - \int_{H_s}\mathbf{x}^{\mathbf{i}}\right) + R_r{}'\right), \end{aligned}$$

where $R_r{}' = o(1)$.

Since, for any random variable X, $\sigma^2(X) = m(X^2) - m^2(X)$, and $m(Q(\mathbf{x}^{\mathbf{i}}) - \int_{H_s}\mathbf{x}^{\mathbf{i}}) = 0$ by hypothesis,

$$\begin{aligned} (M^{2/s})^{n+s}(n!)^2 t_r &= m\left(\sum_n f^{(\mathbf{i})}(\mathbf{C}_r)\left(Q(\mathbf{x}^{\mathbf{i}}) - \int_{H_s}\mathbf{x}^{\mathbf{i}}\right)\right)^2 + o(1) \\ &= \sum m_{\mathbf{i}\mathbf{j}} f^{(\mathbf{i})}(\mathbf{C}_r) f^{(\mathbf{j})}(\mathbf{C}_r) + o(1) \ , \end{aligned}$$

where the last sum is as in (3).

Summing, we have:

$$(M^{2/s})^{n+s}(n!)^2\sigma^2(Q_M(f)) = \frac{M}{2^s}\sum m_{ij}\left(\sum_{r=1}^{M}\frac{2^s}{M}f^{(i)}(\mathbf{C}_r)f^{(j)}(\mathbf{C}_r)\right) + o(M).$$

Since $2^s/M$ is the (s-dimensional) volume of K_r, the inner sum is, for each $\mathbf{i}$ and $\mathbf{j}$, a Riemann sum, and so can be written

$$\int_{H_s} f^{(i)}f^{(j)} + o(1);$$

and the theorem follows.

Though Theorem 2 gives an explicit (asymptotic) expression for the error of the quadrature formula $\mathbf{Q}_M$, it does not provide a practical a priori error estimate, since in most cases no useful estimate of the integrals appearing in (3) will be available. However by a modification of the integration procedure a very useful a posteriori error estimate can be obtained:

If we apply Q twice, independently, to each subcube K_r, denoting the results $Q'_{K_r}(f)$ and $Q''_{K_r}(f)$, and replace $Q_M(f)$ by

$$Q_M^*(f) = \sum_{r=1}^{M}\frac{1}{2}\left(Q'_{K_r}(f) + Q''_{K_r}(f)\right),$$

it is clear that $\sigma^2(Q_M^*) = \frac{1}{2}\sigma^2(Q_M)$. We can then at the same time calculate the quantity

(8)
$$\sum_{r=1}^{M}\left(\frac{Q'_{K_r}(f) - Q''_{K_r}(f)}{2}\right)^2$$

which I shall call $(\sigma^*)^2$. Since

$$m(\sigma^*)^2 = \frac{1}{4}\sum_{r=1}^{M}m(Q'_{K_r}(f) - Q''_{K_r}(f))^2$$

$$= \frac{1}{2}\sum_{r=1}^{M}\sigma^2(Q_{K_r}(f)),$$

σ^* may be taken as an estimate of $\sigma(Q_M^*(f))$. In numerical experiments (with the very simplest Q's) reported in [2], σ^* turned out to be a very good estimate, for values of M which were not impracticably large. In practice, if one uses Q_M^* for a succession of different values of M, σ^* can be regarded as quite reliable when $M^{(n/s+1/2)}\sigma^*$ remains approximately constant.

3. Quadratures of Degree 2. In [1] an s-dimensional s.q.f. of degree zero with $k = 1$ was proposed; in [2] one of degree 1 with $k = 2$ (and equal coefficients) was found. For higher degrees the situation is more complex. It follows from a theorem of A. H. Stroud [5] that

$$k \geq \binom{n + s}{[n/2]}$$

("[·]" denoting the greatest-integer function); and so, for $n \geq 2$, k cannot be independent of s. I shall construct, in the case $n = 2$, some equal-coefficients quadratures. Equal-coefficients formulas have several advantages; they are apt to be computationally simpler, the coefficients are necessarily nonnegative, and they

have certain benefits as regards round-off and other random errors.

For Q to have equal coefficients and to be of degree 0 or more, it must be of the form

$$(9) \qquad Q(f) = \frac{2^s}{k} \sum_{i=1}^{k} f(\mathbf{X}_i) \, .$$

The condition that $Q(P) = \int_A P$ for $P(\mathbf{x}) = x^1, x^2, \cdots, x^s, x^1x^1, x^1x^2, \cdots, x^sx^s$ is equivalent to:

$$\sum_{i=1}^{k} X_i^{\,j} = 0 \, , \qquad \jmath = 1, \cdots, s \, ,$$

$$(10) \qquad \sum_{i=1}^{k} X_i^{\,j} X_i^{\,L} = 0 \, , \qquad 1 \leqq j \neq L \leqq s \, ,$$

$$\sum_{i=1}^{k} (X_i^{\,j})^2 = k/3 \, , \qquad j = 1, \cdots, s \, .$$

Defining the k-vectors $\mathbf{V}_j = (X_1^{\,j}, \cdots, X_k^{\,j})$, $j = 1, 2, \cdots, s$, and $\mathbf{V}_0 = (1/\sqrt{3})\,(1, 1, \cdots, 1)$, we see that (10) is just the condition that these vectors be orthogonal, and each of Euclidean length $(k/3)^{1/2}$.

The last requirement for Q to be an s.q.f. of degree 2 is that

$$m\!\left(\frac{2^s}{k} \sum_{i=1}^{k} f(\mathbf{X}_i) \right) = \int_{H_s} f$$

for any $f \in L^2(H_s)$. Now let $g_i(\mathbf{x})$, defined for all $\mathbf{x}$ in s-dimensional real space R^s, be the probability density function of $\mathbf{X}_i$. Then

$$(11) \qquad m\!\left(\frac{2^s}{k} \sum_{i=1}^{k} f(\mathbf{X}_i) \right) = \int_{R^s} f(\mathbf{x}) \left(\frac{2^s}{k} \sum_{i=1}^{k} g_i(\mathbf{x}) \right) d\mathbf{x} \, ;$$

so that the requirement on Q is equivalent to the condition

$$(12) \qquad \begin{aligned} \frac{1}{k} \sum_{i=1}^{k} g_i(\mathbf{x}) &= 2^{-s} \, , \qquad \mathbf{x} \in H_s \, , \\ &= 0 \quad \, , \qquad \mathbf{x} \notin H_s \, . \end{aligned}$$

Thus (9), (10), and (12) together are the necessary and sufficient conditions for Q to be a equal-coefficients s.q.f. of degree $\geqq 2$.

A consequence of (12) is that some one of the random variables $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k$ must be able to take on the value $(1, 1, \cdots, 1)$; say it is $\mathbf{X}_1$.** Then the vectors $\mathbf{V}_1, \cdots, \mathbf{V}_s$ must all be able to have a projection of length 1 on the k-vector $\mathbf{e}_1 = (1, 0, \cdots, 0)$. The projection of $\mathbf{e}_1$ on the subspace of R^k orthogonal to $\mathbf{V}_0$ is $\mathbf{e}' = ((k-1)/k, -1/k, \cdots, -1/k)$, which is of length $((k-1)/k)^{1/2}$. Therefore each $\mathbf{V}_i$, $1 \leq i \leq s$, must be able to have a projection of length $(k/(k-1))^{1/2}$ on $\mathbf{e}'$. This can occur with the lengths of the $\mathbf{V}_i$ shortest, and so with k minimal (since the length

---

** Strictly speaking, (12) need hold only almost everywhere; so all that can be said is that for any $\epsilon > 0$, some $\mathbf{X}_i$ must be able to take values within distance ϵ of $(1, 1, \cdots, 1)$. Modifying the argument which follows, to take account of this, would complicate its form without changing it essentially.

of each $\mathbf{V}_i$ is $(k/3)^{1/2}$) when $\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_s$ are arranged symmetrically about $\mathbf{e}'$. In this arrangement, which is the same as the configuration in R^s consisting of $(1, 1, \cdots, 1)$ and the natural basis vectors $(1, 0, \cdots, 0), (0, 1, 0, \cdots, 0), \cdots,$ $(0, \cdots, 0, 1)$, the cosine of the angle between each $\mathbf{V}_i$ and $\mathbf{e}'$ is $1/\sqrt{s}$. Thus the length of the projection of $\mathbf{V}_i$ on $\mathbf{e}'$ is at most $(k/3s)^{1/2}$ and therefore

$$(k/3s)^{1/2} \geqq (k/(k-1))^{1/2}$$

which proves:

THEOREM 3. *If*

$$Q(f) = \frac{2^s}{k} \sum_{i=1}^{k} f(\mathbf{X}_i)$$

is an s.q.f. of degree 2 or more for H_s, then $k \geqq 3s + 1$.

(The above proof deals only with the case that the $\mathbf{X}_i$ have absolutely continuous distributions; it can easily be extended to the general case.)

I shall first construct a formula of the desired type in the case $s = 1$, with $k = 4$. Here it is sufficient to find four 1-dimensional real random variables $X_1, \cdots, X_4$ such that the vector $\mathbf{V}_1 = (X_1, \cdots, X_4)$ and the respective probability densities $g_1, \cdots, g_4$ (assumed to exist) satisfy:

$$\mathbf{V}_1 \cdot \mathbf{V}_0 = 0, \qquad \|\mathbf{V}_1\|^2 = 4/3,$$
$$g_1(t) + \cdots + g_4(t) = 2, \qquad -1 \leqq t \leqq 1,$$
$$= 0, \qquad \text{otherwise}.$$

I shall use the following notation: S is the (3-dimensional) subspace of R^4 which is orthogonal to $\mathbf{V}_0$; $\Pi_{i,t}$ is the plane which is the intersection of S with the hyperplane $X_i = t$ in R^4; Σ is the sphere of radius $(4/3)^{1/2}$ and center at the origin in S; and $c_{i,t}$ is the circle $\Sigma \cap \Pi_{i,t}$. Then $\mathbf{V}_1 \in \Sigma$, and letting P denote the probability density function of $\mathbf{V}_1$ on Σ,

$$\text{(13)} \qquad \text{Prob } (X_i \in [t - \delta, t + \delta]) = \int_{D_i} P,$$

where D_i is the portion of Σ lying between $\Pi_{i, t-\delta}$ and $\Pi_{i, t+\delta}$.

We now estimate the integral in (13): If $\mathbf{x} = (x^1, x^2, x^3, x^4)$ is in $\Pi_{i,0}$, then $x^i = 0$ and, since $\Pi_{i,0} \subset S$, the sum of the three other coordinates of $\mathbf{x}$ is zero. Therefore the 4-vector whose ith coordinate is $+1$ and whose other three coordinates are $-1/3$ is orthogonal to $\Pi_{i,0}$, and so to $\Pi_{i,t}$. This vector lies in S, and the ratio of its length to the absolute value of its ith coordinate is $(4/3)^{1/2}$. It follows that the distance of $\Pi_{i,t}$ from the origin is $t(4/3)^{1/2}$, and the distance from $\Pi_{i,t-\delta}$ to $\Pi_{i,t+\delta}$ is $2\delta(4/3)^{1/2}$. Thus the radius of $c_{i,t}$ is $(4(1 - t^2)/3)^{1/2}$, and a normal to Σ at any point of $c_{i,t}$ makes, with the normal to $\Pi_{i,t}$, an angle whose sine is $(1 - t^2)^{1/2}$. Therefore the area of D_i is, to within $o(\delta)$,

$$2\pi(4(1 - t^2)/3)^{1/2} \frac{2\delta(4/3)^{1/2}}{(1 - t^2)^{1/2}} = \frac{16\pi}{3} \delta$$

if $|t| < 1$, and is zero if $|t| > 1$.

Now let $m = m(P, c_{i,t})$ denote the mean of P over $c_{i,t}$. Then

$$\text{Prob } (X_i \in [t - \delta, t + \delta]) = (16\pi/3)\delta m + o(\delta)$$

and therefore $g_i(t) = (8\,\pi/3)\,m(P, c_{i,t})$. The condition on the g_i is seen to be satisfied if $m(P, c_{i,t}) \equiv 3/(16\,\pi)$; and this is certainly true if P is constant, i.e., if $\mathbf{V}_1$ is uniformly distributed on Σ. This implies

THEOREM 4. *Let* $\mathbf{Y} = (Y^1, Y^2, Y^3)$ *be a random variable uniformly distributed on the sphere in* R^3 *with center at the origin and radius* $(4/3)^{1/2}$. *Let* $A = (a_{ij})$ *be a* 4×4 *orthogonal matrix with* $a_{4,j} = 1/2, j = 1, \cdots, 4$.

Set

(14) $$X_j = a_{1,j}Y^1 + a_{2,j}Y^2 + a_{3,j}Y^3, \quad j = 1, \cdots, 4.$$

Then

$$Q(f) = \frac{1}{2} \sum_{j=1}^{4} f(X_j)$$

is an s.q.f. of degree 2 for H_1.

Proof. In the notation of the discussion above, the transformation (14) can be written $\mathbf{V}_1 = \mathbf{Z}A$, where $\mathbf{Z} = (Y^1, Y^2, Y^3, 0)$. Since $\mathbf{V}_1\mathbf{V}_1^T = \mathbf{Z}\mathbf{Z}^T = \mathbf{Y}\mathbf{Y}^T$, and

$$\sum_{j=1}^{4} X_j = \sum_{j=1}^{4} \sum_{i=1}^{3} a_{i,j}Y^i = \sum_{i=1}^{3} Y^i \sum_{j=1}^{4} a_{ij} = 0,$$

the transformation is an isometry of R^3 into the subspace S of R^4. Since $\mathbf{Y}$ is uniformly distributed on the sphere $(Y^1)^2 + (Y^2)^2 + (Y^3)^2 = 4/3$, $\mathbf{V}_1$ is uniformly distributed on Σ, which is what was required for Q to be of degree ≥ 2. That Q is not of degree 3 or higher can be seen directly by setting $f(x) = x^3$ and noting that $(X_1, \cdots, X_4)$ can take the value $(2\sqrt{2}/3, -\sqrt{2}/3, -\sqrt{2}/3, 0)$.

Remarks. (a) Matrices A of the form required in the theorem may be constructed by taking a basis for R^4 containing the vector $(1/2, \cdots, 1/2)$, and orthonormalizing to generate the remaining three rows of A. (b) The use of points uniformly distributed on a sphere is numerically practical; for methods of generating pseudo-random points with that distribution see, e.g., [6] and [7].

By regarding an integral over H_s as an s-fold iterated integral, and applying the Q of Theorem 4 to each axis, we may construct s.q.f.'s of degree 2 for any $s > 1$; they will have $k = 4^s$. A more efficient method of generalizing the result of Theorem 4 is as follows:

Let $k \geq 3s + 1$ be an integer to be specified later. Let $\mathbf{W}_1, \cdots, \mathbf{W}_{3s+1}$ be an orthonormal set of vectors in R^k, with $\mathbf{W}_{3s+1} = (1/\sqrt{k})\,(1, \cdots, 1)$. For $L = 1, 2, \cdots, s$, let S_L be the subspace of R^k spanned by $\mathbf{W}_{3L-2}, \mathbf{W}_{3L-1}$, and $\mathbf{W}_{3L}$; and let Σ_L be the sphere in S_L centered at the origin and with radius $(k/3)^{1/2}$. Let A be the matrix whose rows are $\mathbf{W}_1, \cdots, \mathbf{W}_{3s+1}$.

Let $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_s$ be independent 3-dimensional random variables, each uniformly distributed on the sphere in R^3 with center at the origin and radius $(k/3)^{1/2}$. For $L = 1, 2, \cdots, s$, let $\mathbf{Z}_L$ be the $(3s + 1)$-vector $((3s + 1)$-tuple) all of whose entries are zero except for the 3L-2nd, 3L-1st and 3Lth, those being Y_L^1, Y_L^2, and Y_L^3 respectively. Let $\mathbf{Z}_0$ be the $(3s + 1)$-vector $(0, 0, \cdots, 0, (k/3)^{1/2})$. Set

(15) $$\mathbf{V}_L = \mathbf{Z}_L A . \qquad L = 0, 1, \cdots . s \,,$$

and set

(16) $$\mathbf{X}_i = (V_1{}^i, V_2{}^i, \cdots, V_s{}^i) \,, \qquad i = 1, \cdots, k \,.$$

The **V**'s and **X**'s are then related as in the discussion leading to Theorem 3, and so the condition that

$$Q(f) = \frac{2^s}{k} \sum_{i=1}^{k} f(\mathbf{X}_i)$$

be an s.q.f. of degree ≥ 2 is just that the **V**'s be orthogonal and all of length $(k/3)^{1/2}$, and that Eq. (12) hold. The geometrical conditions are satisfied, since

$$\mathbf{V}_L \mathbf{V}_M{}^T = \mathbf{Z}_L A A^T \mathbf{Z}_m{}^T = \mathbf{Z}_L \mathbf{Z}_m{}^T = k/3 \,, \qquad L = m$$
$$= 0 \,, \qquad L \neq m$$

and it remains to check condition (12).

Let $\mathbf{t} = (t^1, \cdots, t^s)$ be any point of R^s and let $\delta_1, \cdots, \delta_s$ be positive numbers. Since the **Y**'s are independent, so are the $\mathbf{V}_1, \cdots, \mathbf{V}_L$; and so

$$\text{Prob } (|X_i{}^1 - t^1| \leq \delta_1, |X_i{}^2 - t^2| \leq \delta_2, \cdots, |X_i{}^s - t^s| \leq \delta_s)$$
$$= \prod_{j=1}^{s} \text{Prob } (|X_i{}^j - t^j| \leq \delta_j) = \prod_{j=1}^{s} \text{Prob } (|V_j{}^i - t^j| \leq \delta_j) \,.$$

Now $\mathbf{V}_j = Y_j{}^1 \mathbf{W}_{3j-2} + Y_j{}^2 \mathbf{W}_{3j-1} + Y_j{}^3 \mathbf{W}_{3j}$, and so is in S_j. In fact, since $A A^T = I^i$, it is uniformly distributed on Σ_j. Therefore

$$\text{Prob } (|V_j{}^i - t^j| \leq \delta_j) = \frac{a}{4\pi k/3}$$

where $a = a(t^j; i, j, \delta_j)$ is the area of the portion of Σ_j lying between the hyperplanes $x^i = t^j - \delta_j$ and $x^i = t^j + \delta_j$ in R^k.

Now let $\Pi_t = \Pi_t(i, j)$ be the intersection of the hyperplane $x^i = t$ with S_j. If $\theta = \theta(i, j)$ is the angle between the ith coordinate vector in R^k and a normal to Π_t in S_j, then $|t/\cos \theta|$ is the distance of Π_t from the origin. Then $a(t) = 0$ (for small δ_j) if $|t/\cos \theta| > (k/3)^{1/2}$; and when $a(t) \neq 0$ Π_t cuts Σ_j in a circle of radius $(k/3 - t^2/\cos^2 \theta)^{1/2}$, and the normal to Σ_j at any point of this circle makes, with the normal to Π_t, an angle whose sine is $(1 - 3t^2/k \cos^2 \theta)^{1/2}$. Also, the distance between $\Pi_{t-\delta}$ and $\Pi_{t+\delta}$ is $2\delta/\cos \theta$. Thus, to within $o(\delta_j)$,

$$a = \frac{2\delta_j}{\cos \theta} \left(1 - \frac{3(t^j)^2}{k \cos^2 \theta} \right)^{-1/2} (2\pi) \left(\frac{k}{3} - \frac{(t^j)^2}{\cos^2 \theta} \right)^{1/2}$$
$$= 4\pi \delta_j (k/3)^{1/2} (\cos \theta)^{-1} \,;$$

and

$$\frac{1}{2\delta_j} \text{Prob } (|V_j{}^i - t^j| \leq \delta_j) = \frac{(3/k)^{1/2}}{2 \cos \theta} + o(1) \,.$$

Since

$$g_i(\mathbf{t}) = \lim_{\delta_1, \cdots \delta_s \to 0} \prod_{j=1}^{s} \left(\frac{1}{2\delta_j} \text{Prob } (|V_j{}^i - t^j| \leq \delta_j) \right),$$

it follows that

$$(17) \qquad g_i(\mathbf{t}) = 2^{-s} \left(\frac{3}{k} \right)^{s/2} \prod_{j=1}^{s} (\cos \theta(i,j))^{-1}$$

if, for $j = 1, 2, \cdots, s$, $|t^j| \leqq (k/3)^{1/2} \cos \theta(i,j)$; and $g_i(\mathbf{t}) = 0$ otherwise. Condition (12) then implies that for each i, $g_i(\mathbf{x}) = 0$ if $\mathbf{x} \nsubseteq H_s$; it follows that for all i and j, we must have $\cos \theta(i,j) \leqq (3/k)^{1/2}$. Therefore by (17)

$$(18) \qquad g_i(\mathbf{t}) \geqq 2^{-s}, \qquad i = 1, 2, \cdots, k.$$

Now for (12) to hold, there must be equality in (18) for every i; and this can happen only if

$$(19) \qquad \cos \theta(i,j) = (3/k)^{1/2}; \qquad 1 \leqq i \leqq k, 1 \leqq j \leqq s.$$

So we see that (19) is a necessary and sufficient condition for (12).

To calculate $\cos \theta_{ij}$, we note that Π_0 is the set of all vectors of R^k of the form $a\mathbf{W}_{3j-2} + b\mathbf{W}_{3j-1} + c\mathbf{W}_{3j}$ satisfying the equation

$$a\mathbf{W}_{3j-2}^i + b\mathbf{W}_{3j-1}^i + c\mathbf{W}_{3j}^i = 0;$$

so that the vector

$$W_{3j-2}^i \mathbf{W}_{3j-2} + W_{3j-1}^i \mathbf{W}_{3j-1} + W_{3j}^i \mathbf{W}_{3j}$$

is normal to Π_0 (and so to Π_t). The unit normal to the hyperplane $x^i = t$ in R^k is just the vector whose ith component is 1 and whose other components are zero; thus by taking inner products it is seen that

$$(20) \qquad \cos \theta_{ij} = ((W_{3j-2}^i)^2 + (W_{3j-1}^i)^2 + (W_{3j}^i)^2)^{1/2}.$$

Combining (19) and (20), we have

THEOREM 5. *If*

(1) *s is a positive integer and k is an integer $\geq 3s + 1$.*

(2) *$A = (W_{i,j})$ is a $(3s + 1) \times k$ matrix such that*

    (2a) *$W_{3s+1,j} = k^{-1/2}$, $1 \leq j \leq k$*

    (2b) *The rows of A, regarded as k-vectors, are orthogonal and each is of Euclidean norm 1.*

    (2c) *$W_{i,j}^2 + W_{i+1,j}^2 + W_{i+2,j}^2 = 3/k$ for $1 \leq j \leq k$ and $i = 1, 4, 7, \cdots$, $3s - 2$.*

(3) *$\mathbf{Y}_1 = (Y_1^1, Y_1^2, Y_1^3), \cdots, \mathbf{Y}_s = (Y_s^1, Y_s^2, Y_s^3)$ are independent random variables, each uniformly distributed on the sphere in R^3 with center at the origin and radius $(k/3)^{1/2}$.*

(4) *For $j = 1, \cdots, k$, $\mathbf{X}_j = (X_j^1, \cdots, X_j^s)$ is defined by*

$$X_j^i = Y_i^1 \mathbf{W}_{3i-2,j} + Y_i^2 \mathbf{W}_{3i-1,j} + Y_i^3 \mathbf{W}_{3i,j}, \qquad i = 1, \cdots, s.$$

Then

$$Q(f) = \frac{2^s}{k} \sum_{j=1}^{k} f(\mathbf{X}_j)$$

is a stochastic quadrature formula of degree 2 for the s-cube H_s.

(It is easy to see, by an example, that it is not of higher degree.)

It remains to be seen for which values of k do matrices A exist satisfying the requirements of the theorem.

4. Hadamard Matrices. If we multiply the matrix A by $k^{1/2}$, we obtain a matrix $A_1 = (a_{ij})$ whose rows are orthogonal and each of Euclidean norm k, and which satisfies the additional condition

$$(21) \quad a_{3s+1,j} = 1 \; ; \quad a_{i,j}^2 + a_{i+1,j}^2 + a_{i+2,j}^2 = 3 \; , \quad i = 1, 4, \cdots, 3s - 2 \; ;$$
$$1 \leqq j \leqq k \; .$$

A class of square matrices satisfying all these conditions, and which is in fact defined by the condition that the rows be orthogonal and that

$$a_{i,j} = \pm 1 \; , \quad \text{all } i \text{ and } j$$

is the class of Hadamard matrices, which has been studied extensively in other connections (see, e.g., [8] − [11]). (Any Hadamard matrix may be converted into one in which all entries in the bottom row are +1, merely by multiplying each column of the original matrix by −1 whenever its bottom entry is −1.) We thus have

THEOREM 6. *If s is a positive integer, and k the least integer $\geqq 3s + 1$ such that a $k \times k$ Hadamard matrix exists, then there exists a k-point, equal-coefficients, stochastic quadrature formula of degree 2 for H_s.*

Proof. Let M be the $k \times k$ Hadamard matrix, with all entries in the kth row equal to +1. Let A be the matrix which consists of the bottom $3s + 1$ rows of $k^{-1/2} M$. Then A satisfies all the requirements of Theorem 5.

It is known that the order k of a Hadamard matrix must be 1 or 2 or a multiple of 4. It is conjectured that Hadamard matrices of order $4n$ exist for all integers $n \geqq 1$; it is known that such exist for all $n \leqq 50$ except possibly for $n = 47$. Thus for $s = 1, 5, 9, \cdots$, 65 equal-coefficients s.q.f.'s of degree 2 exist with $k = 3s + 1$, which is the least possible; and for $1 \leqq s \leqq 65$ they exist with $k \leqq 3s + 4$. The question of whether there exists, for every s, an equal-coefficients s.q.f. of degree 2 with $k = 3s + 1$ is left open. In particular, for $s = 2$, Hadamard matrices provide an 8-point quadrature but the question of the existence of a 7-point formula is equivalent to the question of the existence of a 7×7 orthogonal matrix (a_{ij}) satisfying:

$$a_{1,j}^2 + a_{2,j}^2 + a_{3,j}^2 = a_{4,j}^2 + a_{5,j}^2 + a_{6,j}^2 = 3/7 \; , \quad a_{7,j}^2 = 1/7 \; , \quad j = 1, 2, \cdots, 7 \; ;$$

which I have been unable to resolve.

5. Some Formulas of Degree 3. If

$$Q(f) = \sum_{i=1}^{k} A_i f(\mathbf{X}_i)$$

is an s.q.f. of degree $2n$ for H_s, then

$$Q'(f) = \sum_{i=1}^{k} A_i \left(\frac{f(\mathbf{X}_i) + f(-\mathbf{X}_i)}{2} \right)$$

is a ($2k$-point) s.q.f. of degree $\geqq 2n + 1$ for H_s. By this device each of the formulas of degree 2 constructed in the last two sections can be converted into an equal-co-efficients formula of degree 3. I do not know how close the number of points in the resulting formula is to the minimum number of points needed for an equal-co-efficients s.q.f. of degree 3.

6. Comments. There is an apparent paradox in the way stochastic formulas are used. The random variables in terms of which the formula is defined are replaced, in practice, by well-defined mathematical rules for calculating "pseudo-random" numbers. Once this is done in any specific manner, the quadrature formula becomes deterministic, and Bahvalov's lower bound for the error, $\Omega(N^{-n/s})$, applies. Yet the lower error estimate of Theorem 2 — $O(N^{-n/s-1/2})$ — is actually used, and found, in practice, to be quite valid ([1], [2]). The answer is as follows: Bahvalov showed that for each s and each n there is a constant $C_{n,s}$ such that, if Q is any deterministic N-point quadrature formula there is an $f \in D_s{}^n$ (f dependent on Q) whose nth order derivatives are bounded by 1 on H_s, and such that

$$(22) \qquad \left| Q(f) - \int_{H_s} f \right| \geqq C_{n,s} N^{-n/s} .$$

The functions constructed by Bahvalov to satisfy (22) are, naturally, of very special form; and if Q and Q' are two different realizations of a single s.q.f. by means of 2 different methods of calculating pseudo-random numbers, different f's would be required to attain Bahvalov's lower bound for Q and for Q'. In plain, there is little reason to suspect that an integrand one comes across in practice, and determines to belong to $D_s{}^n$, is in fact one of the extreme members of $D_s{}^n$—relative to the particular numerical quadrature formula one is going to apply—for which the error is at, or near, the quantity on the right in (22). It might indeed happen, and could then only be regarded as a piece of extremely bad luck. That it should happen for a specific Q_M for each of several different values of M is a possibility; but one reasonably proceeds on the assumption that it will not happen.

Similarly, while 3σ is usually taken as an upper bound on the error of a prob-abilistic quadrature, one may, by bad luck, have taken a sample which deviates from the mean (i.e., the integral) by 30σ. Indeed, for any given s.q.f. Q and number M, one can construct f's such that $Q_M(f)$ is not at all close to being normally dis-tributed, so that deviations of 30σ may not be terribly unlikely. One can only fall back on the intuition that such f's, like the f's satisfying (22), simply do not occur in numerical problems; and the final justification for this intuition is in the good results achieved in actual calculation.

Classically, there have been two main approaches to efficient numerical quad-rature. The first was to find formulas of maximum degree using a fixed number of points, or when this could not be done (as was generally the case with multiple integrals) to find formulas of as high a degree as could be managed using as few points as could be managed. In one dimension this approach produced the "Gaus-sian" quadrature formulas, see, e.g., [12, Chapters 7, 9, and 10]. [5] contains a sum-mary of work along this line for multiple quadrature, and an extensive list of refer-ences. Formulas of this type are generally applied to the integration of very smooth functions—the idea is that with each increase in the effort of calculation (which is

measured roughly by the number of points used in the quadrature formula), a formula of higher degree will be used, and more of the smoothness of the integrand will come into play.

The second approach has been to find those formulas using a given number of points which minimize the quadrature error, or some functional related to it, over a certain function space; the function space is usually defined in terms of a condition on the smoothness of its members. A discussion of some such formulas is found in [12, Chapter 8]; S. L. Sobolev (see [13], and the references listed there), and R. E. Barnhill [15] have worked on such formulas for multiple integrals. The idea here is to take maximum advantage, for a given level of calculation effort, of the particular amount of smoothness the integrand in question does have. The degree of the formula used does not, generally, increase as the numbers of points is increased.

The first approach has been the one more generally used, and successful, in one-dimensional quadrature. There is some reason to suppose that the second will be relatively more important in the higher-dimensional case. It seems that the functions of several variables that turn up for numerical integration in physical problems are quite commonly not very smooth, being discontinuous or having discontinuous or unbounded derivatives of fairly low order. Functions arising in practical problems are generally piecewise smooth; and in one dimension the "singularities," which occur at isolated points, can often be removed by various devices. In $s > 1$ dimensions the singularities will occur along manifolds of dimension $s - 1$ or lower, and unless these manifolds are of very simple form (planar, in fact) there is not likely to be any practicable way of smoothing the function for integration. (For this reason the $\frac{1}{2}$ added by stochastic quadrature formulas to the convergence exponent n/s for functions in $D_s{}^n$ is important; for n often cannot be raised.)

Theorem 2, together with Bahvalov's results, shows that stochastic quadrature formulas can give better results, for functions of a given degree of smoothness (i.e. in the class $D_s{}^n$, for a given n) than any deterministic formulas; the formulas derived above certainly do this for sufficiently large N. This suggests that the second approach to efficient quadrature may best be carried on in the context of stochastic quadrature formulas.

The classical (first) approach enters into the present one in the attempt to find s.q.f.'s of given degree with least possible k. We wish to minimize k because the calculation effort in using an s.q.f. is roughly proportional to k and to M; but it also depends on other factors, such as the difficulty of generating pseudo-random points imitating the distribution properties of the random variables involved. These questions remain for future investigation.

In practice, one of the main advantages of stochastic quadrature formulas is the availability of a good a posteriori error estimate. A disadvantage is the restriction of the number of points used to numbers of the form $2^s k$, $3^s k$, $4^s k$, $\cdots$, $(2^s(2k)$, $3^s(2k)$, $\cdots$ when the a posteriori error estimate is being established). For even moderate values of s, there are not many numbers of this form that are small enough to make the calculation practicable.

As the cost of computation goes down, it becomes practical to use very large numbers of points in quadrature formulas. Present day computers have begun to make numerical multiple integration practical without extensive adaptation of

techniques to each particular problem. Further decreases in cost, by one or two orders of magnitude, should be accomplished by machines currently being planned; and it should become possible to do numerical multiple integration in a routine manner.

National Bureau of Standards
Numerical Analysis Section
Washington, D. C. 20234

1. S. HABER, "A modified Monte-Carlo quadrature," *Math. Comp.*, v. 20, 1966, pp. 361–368. MR **35** #1178.
2. S. HABER, "A modified Monte-Carlo quadrature. II," *Math. Comp.*, v. 21, 1967, pp. 388–397.
3. N. S. BAHVALOV, "On approximate calculation of multiple integrals," *Vestnik Moscow Univ.*, v. 4, 1959, pp. 3–18. (Russian)
4. P. J. DAVIS, "A construction of nonnegative approximate quadratures," *Math. Comp.*, v. 21, 1967, pp. 578–582. MR **36** #5584.
5. A. H. STROUD, "Quadrature methods for functions of more than one variable," *Ann. New York Acad. Sci.*, v. 86, 1960, pp. 776–791. MR **22** #10179.
6. J. M. COOK, "Rational formulae for the production of a spherically symmetric probability distribution," *MTAC*, v. 11, 1957, pp. 81–82. MR **19**, 466.
7. M. E. MULLER, "A note on a method for generating points uniformly on N-dimensional spheres," *Comm. ACM*, v. 2, 1959, pp. 19–20.
8. J. HADAMARD, "Resolution d'une question relative aux determinants," *Bull. Sci. Math.*, (2), v. 17, 1893, pp. 240–246.
9. R. E. A. C. PALEY, "On orthogonal matrices," *J. Math. and Phys.*, v. 12, 1933, pp. 311–320.
10. J. WILLIAMSON, "Note on Hadamard's determinant theorem," *Bull. Amer. Math. Soc.*, v. 53, 1947, pp. 608–613. MR **8**, 559.
11. E. SPENCE, "A new class of Hadamard matrices," *Glasgow Math. J.*, v. 8, 1967, pp. 59–62. MR **35** #1496.
12. V. I. KRYLOV, *Approximate Calculation of Integrals*, Fizmatgiz, Moscow, 1959; English transl., Macmillan, New York, 1962. MR **22** #2002; MR **26** #2008.
13. S. L. SOBOLEV, "On the order of convergence of cubature formulae," *Dokl. Akad. Nauk SSSR*, v. 162, 1965, pp. 1005–1008 = *Soviet Math. Dokl.*, v. 6, 1965, pp. 808–811. MR **31** #3776.
14. S. HABER, "A combination of Monte-Carlo and classical methods for evaluating multiple integrals," *Bull. Amer. Math. Soc.*, v. 74, 1968, pp. 683–686.
15. R. E. BARNHILL, "An error analysis for numerical multiple integration. II," *Math. Comp.*, v. 22, 1968, pp. 286–292. MR **37** #6027.