The Annals of Statistics
1988, Vol. 16, No. 2, 895–914

A QUASIRANDOM APPROACH TO INTEGRATION IN BAYESIAN STATISTICS[1]

BY J. E. H. SHAW

University of Warwick

Practical Bayesian statistics with realistic models usually gives posterior distributions that are analytically intractable, and inferences must be made via numerical integration. In many cases, the integrands can be transformed into periodic functions on the unit d-dimensional cube, for which quasirandom sequences are known to give efficient numerical integration rules. This paper reviews some relevant theory, defines new criteria for identifying suitable quasirandom sequences and suggests some extensions to the basic integration rules. Various quasirandom methods are then compared on the sort of integrals that arise in Bayesian inference and are shown to be much more efficient than Monte Carlo methods.

1. Introduction: Numerical integration in Bayesian statistics. The central role of numerical integration in practical Bayesian statistics has been emphasised in, for example, Smith, Skene, Shaw, Naylor and Dransfield (1985). Many important summaries, such as posterior parameter moments, predictive distributions and evaluations of marginal densities, can be expressed via integrals of the form

$$S_R(q) = \int_R q(\theta)m(\theta)\,d\theta, \tag{1.1}$$

where θ is a d-dimensional parameter indexing the plausible models, m is the unnormalised posterior density, i.e., "likelihood $\times$ prior," q is some function (possibly vector- or matrix-valued) and R is a region of the parameter space Θ. If R is irrelevant or is obvious from the context (often $R = \Theta = \mathbb{R}^d$), then it will be omitted.

For example, $c = S(1)^{-1}$ is the normalising constant, $cS(\theta) = E(\theta)$ is the posterior parameter mean vector, and $cS(\theta\theta^{\mathrm{T}}) - c^2S(\theta)S(\theta)^{\mathrm{T}} = V(\theta)$ is the posterior parameter variance–covariance matrix.

With realistic likelihoods and genuine prior information, the integrals are usually analytically intractable and must be estimated numerically, usually in practice by an approximation of the form

$$\hat{S}_R(q) = \sum_{i=1}^{n} w_i m(\theta_i)q(\theta_i). \tag{1.2}$$

The θ_i are called *nodes* (or *points*) and the w_i are *weights*. The nodes and weights together constitute an *integration rule*. Quantities such as $E(\theta)$ and $V(\theta)$ can similarly be estimated by replacing each S by the corresponding $\hat{S}$.

Received September 1986; revised July 1987.
[1]Research partly supported by the Science and Engineering Research Council.
AMS 1980 *subject classifications.* Primary 62F15; secondary 10K05, 65D30.
Key words and phrases. Numerical integration, quasirandom sequences, Bayesian statistics.

As likelihood evaluations may be computationally expensive, we want n to be small, and $\hat{S}(q)$ to be based on the same sequence of nodes $(\boldsymbol{\theta}_i)$ for all q. We shall assume that the sequence of weights (w_i) is to be the same for all q; this is the case for almost all numerical integration schemes in current practice. Some other desirable properties of numerical integration rules for use in Bayesian inference are discused in Shaw (1987a). In particular, the weights should all be positive: If any weights were negative, then, for example, the estimated normalising constant $\hat{S}(1)^{-1}$ might be negative. More general surveys of numerical integration may be found in the books by Stroud (1971) and Davis and Rabinowitz (1984).

The current paper illustrates how "quasirandom sequences" can be incorporated in a general automatic strategy for efficient high-dimensional numerical integration of the type of functions arising in Bayesian statistics. Much of the paper is a review of existing theory, but some new ideas of potential value for statisticians are given in Sections 4 and 6. Section 2 describes how, in many cases, numerical integration can be simplified by adaptively transforming the parameter space Θ to the unit hypercube. Section 3 reviews rational, irrational and irregular quasirandom sequences of points in the unit hypercube and their "discrepancy." Section 4 defines some new criteria for identifying useful rational sequences and tabulates various quasirandom integration rules found by these criteria. Section 5 describes some well-known irrational and irregular sequences. Section 6 suggests some ways of extending the basic integration rules defined by the quasirandom sequences of Sections 4 and 5. Section 7 compares the efficiencies of the different quasirandom integration methods on two artificial but relevant examples and summarises experience gained from applying the methods to genuine statistical problems.

2. Adaptive integration rules. For a given posterior distribution, good integration rules usually have to be found iteratively: Information from one iteration helps construct the rule for the next iteration. The initial integration rule may be determined, for example, by the behaviour of the posterior density in the region of the mode.

We shall assume that the parameter space Θ is $\mathbb{R}^d$ and that the first and second moments of the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ exist, where $\mathbf{y}$ denotes the data. Then a useful simplification in defining good integration rules is to find a transformation B: $\boldsymbol{\theta} \mapsto \mathbf{x}$ such that $\mathbf{x}$ has zero mean and identity variance–covariance matrix. For example, we could use the Cholesky decomposition of the precision matrix $[V(\boldsymbol{\theta})]^{-1}$,

$$\mathbf{x} = B(\boldsymbol{\theta}) = C(\boldsymbol{\theta} - E(\boldsymbol{\theta})),$$

where C is the upper-triangular matrix such that

$$C^{\mathrm{T}}C = [V(\boldsymbol{\theta})]^{-1}.$$

We could equally well choose C to be lower-triangular. In practice, an approximate transformation of this kind can be found using estimates of $E(\boldsymbol{\theta})$ and $V(\boldsymbol{\theta})$ from the previous iteration, or from several previous iterations combined. We may therefore, without essential loss of generality, concentrate on *standardised*

posterior densities $p(\mathbf{x}|\mathbf{y})$ that have mean approximately 0 and variance–covariance matrix approximately the identity.

For standardised posterior distributions with few parameters (say $d \leq 6$), rescaled Cartesian product Gauss–Hermite integration rules are often highly efficient. See Naylor and Smith (1982, 1983) and Smith, Skene, Shaw, Naylor and Dransfield (1985) for a description and examples. However, the number of nodes increases exponentially with increasing d. Even for small d, Gauss–Hermite integration is only efficient if the integrand is smooth (i.e., all kth partial derivatives are continuous and bounded for some large k). Therefore, we still need efficient methods for cases where d is large, or q or m are badly behaved.

Monte Carlo integration has been widely used in both such cases, see Hammersley and Handscomb (1964) for a detailed discussion and Stewart (1985) and van Dijk and Kloek (1985) for recent statistically oriented examples and references. A particular variant of Monte Carlo integration, *importance sampling*, allows integrals over infinite regions to be estimated. The nodes $\mathbf{x}_i$ are generated independently from an importance sampling distribution

$$G(\mathbf{x}) = \Pr(\mathbf{X} \leq \mathbf{x}),$$

and the corresponding weights are defined by

$$w_i = \left|\frac{dG}{d\mathbf{x}}\right|^{-1} \text{ evaluated at } \mathbf{x}_i$$

$$= g(\mathbf{x}_i)^{-1}.$$

$\hat{S}(q)$ is then an unbiased estimator of $S(q)$, and its variance is small if $g(\mathbf{x})$ is roughly proportional to the absolute value of the integrand. As the name suggests, G should be chosen to concentrate the nodes in "important" regions of the parameter space.

Unfortunately, it can be difficult in practice to find an appropriate importance sampling distribution G, especially as we would prefer to use the same nodes and weights for many different integrands. Shaw (1986a) addressed this problem by introducing a flexible class of importance sampling distributions defined implicitly by

$$\begin{aligned} &X_j = c_j\left[A_j f_j(U_j) - (1 - A_j) f_j(1 - U_j) + b_j\right], \qquad j = 1, \ldots, d, \\ &\mathbf{X} = (X_1, \ldots, X_d), \end{aligned} \tag{2.1}$$

where $0 \leq A_j \leq 1$, f_j is a monotonic increasing function on $(0,1)$ such that $f_j(u) \to -\infty$ as $u \to 0+$, $\mathbf{U} = (U_1, \ldots, U_d)$ is uniformly distributed on the unit d-dimensional cube $C_d = [0,1)^d$ and b_j and c_j are constants determined by f_j and A_j. Integration rules using such importance sampling distributions are easy to define. Let $(\mathbf{u}_i)$ be a sequence of n points $\mathbf{u}_i = (u_{i1}, \ldots, u_{in}) \in C_d$, and let (v_i) be a discrete probability distribution on $(\mathbf{u}_i)$, so that $(\mathbf{u}_i)$ are the nodes and (v_i) the weights of an integration rule on C_d. Then the nodes of the corresponding integration rule on $\mathbb{R}^d$ are $(\mathbf{x}_i)$, where the jth coordinate of the ith node is

$$x_{ij} = c_j\left[A_j f_j(u_{ij}) - (1 - A_j) f_j(1 - u_{ij}) + b_j\right], \tag{2.2}$$

with corresponding weight

$$w_i = v_i \prod_{j=1}^{d} \left(c_j \left[A_j f_j'(u_{ij}) - (1 - A_j) f_j'(1 - u_{ij}) \right] \right), \tag{2.3}$$

where f_j' denotes the derivative of f_j. Typically, $v_i = 1/n$ for all i, although an exception is discussed in Section 6.3.

Thus the problem of numerically integrating $q(\theta)m(\theta)$ over $\mathbb{R}^d$ can be reduced at each iteration to the equivalent problem of numerically integrating $q(B^{-1}G^{-1}\mathbf{u})m(B^{-1}G^{-1}\mathbf{u})$ over C_d. If G is sufficiently heavy-tailed, then the transformed integrands will tend to 0 at the boundary of C_d, and can therefore be extended to a periodic function on $\mathbb{R}^d$, where "periodic" will here always mean "continuous, multivariate, with period 1 in each variable." The remainder of this paper therefore concerns integration rules that are efficient on C_d, particularly for well-behaved periodic integrands. The gain in efficiency over straightforward Monte Carlo methods helps compensate for the fact that just one importance sampling distribution is used for several integrands. The choice of A_j and f_j in (2.1) to transform from C_d to $\mathbb{R}^d$ at a given iteration is considered in a separate paper [Shaw (1986a)].

3. Quasirandom sequences and discrepancy. A *quasirandom sequence* is any infinite sequence $s = (\mathbf{u}_i | i \in \mathbb{N})$ of points in $C_d = [0,1)^d$ generated by some algorithm. Our interest is in using s_n, the first n elements of s, as the nodes of an integration rule for C_d, with weights $v_i = 1/n$, $i = 1, \ldots, n$. Various studies from a number theoretic perspective, see Niederreiter (1978) for a detailed survey, suggest algorithms for "good" sequences, as well as considering various definitions of "goodness."

Quasirandom sequences will be called *regular* if

$$\mathbf{u}_{i+1} = \mathbf{u}_i + \boldsymbol{\alpha} \pmod{\mathbf{1}} \tag{3.1}$$

(i.e., the jth coordinate in $\mathbf{u}_{i+1}$ is the fractional part of the jth coordinate in $\mathbf{u}_i + \boldsymbol{\alpha}$), or *irregular* if generated by any other form of algorithm. Regular sequences will be called *rational* if each $\alpha_j \in \mathbb{Q}$, *irrational* if each $\alpha_j \notin \mathbb{Q}$ and *mixed* otherwise. These names are not standard.

Criteria for good integration rules can be derived from the *discrepancy* of the sequence $s = (\mathbf{u}_i)$. Using a similar notation to Section 1, define

$$S_R(q) = \int_R q(\mathbf{u})\, d\mathbf{u},$$

$$\hat{S}_R(q; s_n) = \frac{1}{n} \sum_{\mathbf{u}_i \in R} q(\mathbf{u}_i),$$

for $R \subset C_d$, so that $\Delta = \hat{S}_R(q; s_n) - S_R(q)$ is the error in using nodes $\mathbf{u}_i$ and weights $v_i = 1/n$, $i = 1, \ldots, n$, to integrate q over R. In the particular case $q = 1$ and $R = [\mathbf{0}, \mathbf{x}) = [0, x_1) \times \cdots \times [0, x_d)$, Δ is called the *local discrepancy* of the sequence s_n at $\mathbf{x}$.

There are several measures of *global discrepancy*; see, for example, Kuipers and Niederreiter (1974), Chapter 2. In particular, the L^∞ *starred discrepancy*, D_n^*, is just the one-sample Kolmogorov–Smirnov test statistic defined by

$$D_n^* = \sup_{\mathbf{u} \in C_d} \left| \hat{S}_{[\mathbf{0},\mathbf{u})}(1; s_n) - S_{[\mathbf{0},\mathbf{u})}(1) \right|. \tag{3.2}$$

If $\lim_{n\to\infty} D_n^* = 0$, then s is said to be *equidistributed*. This discrepancy measure is important theoretically since if q: $C_d \to \mathbb{R}$ is of bounded variation $V(q)$ in the sense of Hardy and Krause [see Niederreiter (1978), pages 966–967], then the integration error is bounded:

$$\left| \hat{S}_{C_d}(q; s_n) - S_{C_d}(q) \right| \leq V(q) D_n^*. \tag{3.3}$$

Another bound on the integration error, valid for all Riemann-integrable q, is given in Hlawka (1971). Unfortunately, all such bounds tend in practice to be grossly pessimistic. Note also that D_n^* is impossibly costly to evaluate if d or n is large.

In the next two sections, we shall consider some particular quasirandom sequences with low discrepancy and/or other desirable properties.

4. Rational sequences. If $\boldsymbol{\alpha}$ in (3.1) is a member of $\mathbb{Q}^d$, then we can write

$$\alpha_j = \frac{k_j}{n}, \qquad k_j \in \mathbb{N},\ j = 1, \ldots, d, \tag{4.1}$$

where

$$\text{g.c.d.}(k_1, k_2, \ldots, k_d, n) = 1.$$

Then clearly $\mathbf{u}_i = \mathbf{u}_{n+i}$, but $\mathbf{u}_i \neq \mathbf{u}_{k+i}$ for $k < n$. We shall therefore consider the finite rational sequence $s_n = (\mathbf{u}_i | i = 1, \ldots, n)$, and call n the period of $s = (\mathbf{u}_i)$. This is a multivariate analogue of the one-dimensional trapezium rule, which is well known to be efficient for integrating periodic functions; see Hammersley and Handscomb (1964), formulae 5.6.9–5.6.13, and Walley and Fearn (1979). However, we still need criteria for choosing suitable n and $\mathbf{k}$.

The standard criterion, "ρ," arises as follows. Let

$$M(\mathbf{k}) = \{\mathbf{m} | \mathbf{m} \in \mathbb{Z}^d, \mathbf{m} \neq \mathbf{0}, \mathbf{m}.\mathbf{k} \equiv 0 \pmod{n}\},$$

and

$$r(\mathbf{m}) = \prod_{j=1}^{d} \max(1, |m_j|). \tag{4.2}$$

Thus $M(\mathbf{k})$ is the set of all integer vectors $\mathbf{m} = (m_1, \ldots, m_d)$, excluding $(0, \ldots, 0)$, such that $\Sigma m_i k_i$ is a multiple of n, and $r(\mathbf{m})$ is the product of the nonzero coordinates of $\mathbf{m}$. Then for $\mathbf{k} \in \mathbb{Z}^d$ define

$$\rho(\mathbf{k}, n) = \min_{\mathbf{m} \in M(\mathbf{k})} r(\mathbf{m}), \tag{4.3}$$

so that ρ is a measure of the distance from the origin to the nearest point of $M(\mathbf{k})$. From results in Section 4 of Niederreiter (1978), it follows that the

discrepancy of s_n satisfies an inequality of the form

$$D_n^* \leq \frac{d}{n} + \frac{c(d,n)}{\rho(\mathbf{k},n)},$$

where $c(d, n)$ is independent of $\mathbf{k}$. Therefore, for given d and n, we should look for values of $\mathbf{k}$ that yield large ρ. If ρ is sufficiently large, then $\mathbf{k}$ is often called a *good lattice point modulo n* (although the literature contains several conflicting definitions of "sufficiently"). This criterion, due to Hlawka, yields integration rules that are particularly efficient for periodic functions whose multiple Fourier expansion has coefficients tending rapidly to zero. See Zaremba (1968) and Niederreiter (1978) for details.

We could use more natural measures of distance than (4.2). Thus an alternative criterion is that

$$\nu(\mathbf{k},n) = \min_{\mathbf{m}\in M(\mathbf{k})} \sum_{j=1}^{d} |m_j| \tag{4.4}$$

be large, since ν is an upper bound on the minimum number of parallel $(d-1)$-dimensional hyperplanes covering s_n. For suppose that $\mathbf{m}$ is such that $\mathbf{m}.\mathbf{k} \equiv 0 \pmod{n}$, and let

$$ik_j = a_{ij}n + b_{ij},$$

where $a_{ij} \in \mathbb{N}$ and $b_{ij} \in \{0, 1, \ldots, n-1\}$. Then

$$\begin{aligned} n\mathbf{m}.\mathbf{u}_i &= \sum_{j=1}^{d} m_j b_{ij} = \sum_{j=1}^{d} m_j(ik_j - a_{ij}n) \\ &= i\mathbf{m}.\mathbf{k} - n\sum_{j=1}^{d} m_j a_{ij} \\ &\equiv 0 \pmod{n}. \end{aligned} \tag{4.5}$$

Therefore $\mathbf{m}.\mathbf{u}_i$ is an integer. Also, since $0 \leq u_{ij} < 1$ for $j = 1, 2, \ldots, d$, we have:

(1) If all $m_j \leq 0$, then $\Sigma m_j < \mathbf{m}.\mathbf{u}_i \leq 0$;
(2) if all $m_j \geq 0$, then $0 \leq \mathbf{m}.\mathbf{u}_i < \Sigma m_j$; and
(3) if some m_j are positive and some negative, then $\Sigma_{m_j<0} m_j < \mathbf{m}.\mathbf{u}_i < \Sigma_{m_j>0} m_j$.

In case (1), the integer $\mathbf{m}.\mathbf{u}_i$ can take only $\Sigma|m_j|$ possible values $(\Sigma m_j + 1, \Sigma m_j + 2, \ldots, 0)$. Similarly, $\mathbf{m}.\mathbf{u}_i$ has only $\Sigma|m_j|$ and $\Sigma|m_j| - 1$ possible values in cases (2) and (3), respectively. Thus s_n is covered by at most $\Sigma|m_j|$ hyperplanes of the form $\mathbf{m}.\mathbf{x}$ = constant. This gives ν the natural interpretation described previously, and also implies that if $\mathbf{m} \in M(\mathbf{k})$ is such that $\Sigma|m_j|$ is small, then numerical integration with nodes s_n will tend to be inaccurate for functions that are badly behaved in the direction $\mathbf{m}$.

Although ρ is important in identifying rational sequences with low discrepancy, it appears that ν must also be large. For example, the two-dimensional rational sequence $\mathbf{k} = (1, 27)$, $n = 125$, $d = 2$, with $\nu = 15$, has lower discrepancy ($D^* = 0.026048$) than the sequence $\mathbf{k} = (1, 33)$, $n = 125$, $d = 2$, $\nu = 11$, $D^* = 0.027200$, even though it has smaller ρ (27 compared to 28).

TABLE 1
Examples of recommended rational sequences

Rule			Criterion							
k	n	d	ρ_2	ρ_3	ρ_4	ρ_5	ν_2	ν_3	ν_4	ν_5
121	555	3	121	24			30	14		
61	388	3	61	18			24	12		
47	252	3	47	12			20	10		
36	155	3	36	14			15	9		
17	78	3	17	6			12	6		
7	38	3	7	4			8	6		
5	18	3	5	2			6	4		
3	14	3	3	2			4	4		
188	857	4	116	19	18		31	12	11	
109	390	4	56	18	8		24	10	8	
69	226	4	15	12	8		16	8	8	
32	533	6	32	8	4	2	25	8	6	6
23	328	5	23	8	3	3	20	8	6	6
23	246	5	23	6	3	3	18	6	6	6
15	124	5	15	4	4	2	12	6	6	6
7	60	4	7	4	1		8	6	4	
32	325	6	32	6	4	2	15	7	6	6
19	394	7	19	10	3	3	14	8	6	6
12	211	7	12	4	2	2	13	6	5	5
11	171	6	11	6	2	2	12	6	5	5
13	98	7	13	2	1	1	14	4	4	4
9	70	6	9	2	1	1	10	4	4	4
6	49	7	6	2	1	1	7	4	4	4
4	25	5	4	2	1	1	5	4	4	4
4	29	7	4	1	1	1	5	3	3	3
23	610	10	23	4	3	2	18	6	6	6
32	425	8	32	6	1	1	15	7	4	4
10	237	13	10	3	1	1	11	5	4	4
17	342	18	17	2	1	1	18	4	4	4
16	391	22	16	2	1	1	17	4	4	4
13	322	22	13	2	1	1	14	4	4	4
10	121	11	10	2	1	1	11	4	4	4
6	91	12	6	2	1	1	7	4	4	4
5	54	9	5	2	1	1	6	4	4	4
9	230	22	9	2	1	1	10	4	4	4
9	188	23	9	2	1	1	8	4	4	4
4	95	18	4	2	1	1	5	4	4	4
4	53	13	4	1	1	1	5	3	3	3
10	341	30	10	2	1	1	11	4	4	4
5	198	30	5	2	1	1	6	4	4	4
4	115	22	4	2	1	1	5	4	4	4
2	19	9	2	1	1	1	3	3	3	3
4	149	37	4	1	1	1	5	3	3	3
2	47	23	2	1	1	1	3	3	3	3

Parameters to form good rational sequences with period n and increment

$$\alpha = \left(\frac{1}{n}, \frac{k}{n}, \frac{k^2}{n} \bmod n, \ldots, \frac{k^{d-1}}{n} \bmod n\right).$$

See Section 4 for a description of the criteria ρ_i and ν_i. The table is ordered so that given an upper bound on n and a lower bound on d, the recommended rule is the first one satisfying the bounds.

In high dimensions, n must, unfortunately, be very large for ρ or ν to identify good rational sequences. For example, the table of recommended sequences for $d = 9$ in Hua and Wang (1981) has $\rho \geq 2$ only for $n > 85{,}000$. However, the integrands we are interested in, which arise from statistical models, are typically very well behaved in all but a few dimensions [Shaw (1986a)]. It makes more sense therefore to demand just that ρ and ν be large for all projections of s_n onto low-dimensional faces of C_d, and define criteria ρ_i and ν_i to be the minima over all $\mathbf{k}^*$ of $\rho(\mathbf{k}^*, n)$ and $\nu(\mathbf{k}^*, n)$, respectively, where $\mathbf{k}^*$ omits $d - i$ coordinates from $\mathbf{k}$. For example, if a given rule has $d = 7$ and $\nu_3 = 8$, then projecting its nodes onto any three-dimensional "face" of C_7 will define a rule with $d = 3$ and $\nu \geq 8$.

Computer searches to identify such s_n have been carried out. To obtain results in reasonable time, most of the searches were restricted to $\mathbf{k}$ of the form $(1, k, k^2, \ldots, k^{d-1}) \bmod \mathbf{n}$, the resulting rules being then identified as triples (k, n, d). Table 1 gives some recommended rules and their properties; note that if (k, n, d) is good by the preceding criteria, then $(k, n, d - 1)$ will also be fairly good, since its values of ρ_i and ν_i for each $i = 1, 2, \ldots, d - 1$ must be at least as large as the corresponding ρ_i and ν_i for the rule (k, n, d). All the rules in Table 1 have k and n coprime, so that $\rho_1 = \nu_1 = n$. Table 1 is ordered so that given an upper bound on n and a lower bound on d, the recommended rule is the first one satisfying the bounds. Thus, for integration over C_{10}, the recommended rule with most points is $(k, n, d) = (23, 610, 10)$, but if fewer nodes are desired, then the next rule satisfying $n < 610$ and $d \geq 10$ is $(10, 237, 13)$. Any ten-dimensional projection of this rule could be used; the simplest being just $(10, 237, 10)$. If 237 nodes are still too many, then ten-dimensional projections of the rules $(10, 121, 11)$, $(6, 91, 12)$, $(4, 53, 13)$ or $(2, 47, 23)$ could be used.

Similar rules, but with restricted n and different criteria, have been studied by Korobov, whose results are reviewed in Stroud (1971), Section 6.3. See also Haber (1983).

5. Irrational and irregular sequences. Davis and Rabinowitz (1984), Section 5.9.3, suggest several possible values for irrational $\boldsymbol{\alpha}$ in (3.1). These include

$$\boldsymbol{\alpha} = \left(\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_d}\right), \tag{5.1}$$

where p_j are different primes,

$$\boldsymbol{\alpha} = \left(\xi, \xi^2, \ldots, \xi^d\right), \tag{5.2}$$

where $\xi = p^{1/(d+1)}$ for some prime p, and

$$\boldsymbol{\alpha} = \left(2\cos\frac{2\pi}{p}, 2\cos\frac{4\pi}{p}, \ldots, 2\cos\frac{2\pi d}{p}\right), \tag{5.3}$$

where $p \geq 2d + 3$ is a prime such that either

(i) 2 has order $p - 1 \bmod p$ (i.e., $2^{p-1} \equiv 1 \bmod p$, $2^k \not\equiv 1 \bmod p$ for $1 \leq k < p - 1$), or

(ii) 2 has order $(p - 1)/2 \bmod p$ and $p \equiv 7 \bmod 8$.

Thus p in (5.3) has possible values $5, 7, 11, 13, 19, 23, 29, 37, 47, \ldots$. This restriction on p is not mentioned by Davis and Rabinowitz, but is necessary for the α_i to be independent over $\mathbb{Q}$ (and hence for the sequence to be equidistributed); see, for example, Hua and Wang (1965). Irrational sequences with increments given by (5.1)–(5.3) are all equidistributed, but (5.3) generally appears to give the lowest discrepancy in finite subsequences.

The irregular quasirandom sequences most used in statistics are *pseudorandom numbers*. Taking a subjective view of probability, a good pseudorandom number generator is a quasirandom sequence whose successive points, to accuracy (say) 10^{-10}, are independent realisations of a random variable. From this perspective, the philosophical objections to Monte Carlo methods expressed by, for example, Zaremba (1968) seem irrelevant. There is, of course, the practical difficulty of generating pseudorandom sequences, although many pesudorandom number generators such as that given by Wichmann and Hill (1982) appear to be adequate for all practical purposes. The primary objection remains, however, that other quasirandom sequences tend more rapidly to equidistribution than do pseudorandom sequences, so are more efficient for numerically integrating smooth functions over C_d.

The following three irregular sequences have been used successfully for quasirandom integration over C_d [e.g., Warnock (1972)].

Haber sequences:

$$\mathbf{u}_i = \left(\frac{i(i+1)}{2}\sqrt{p_1}, \ldots, \frac{i(i+1)}{2}\sqrt{p_d} \right) (\mathrm{mod}\, 1), \tag{5.4}$$

where the p_j are prime (usually p_j is the jth prime). Haber (1970) suggests that this sequence should be better than more frequently used pseudorandom sequences for Monte Carlo integration, since the $2d$-dimensional sequence with ith node $(\mathbf{u}_i, \mathbf{u}_{n+i})$ is equidistributed for $n = 1, 2, \ldots$. Analogous sequences to (5.4) can obviously be based on (5.2) or (5.3) rather than on (5.1).

Halton sequences:

$$\mathbf{u}_{i+1} = \left(\phi_{p_1}(i), \phi_{p_2}(i), \ldots, \phi_{p_d}(i) \right), \tag{5.5}$$

where the p_j are pairwise coprime (usually they are chosen to be the first d primes), and $\phi_p(i)$ is the *radical inverse function* of i, obtained by writing i to base p and "reflecting about the decimal point." Thus 15 (base 10) = 120 (base 3), so $\phi_3(15) = 0.021 = \frac{2}{9} + \frac{1}{27} = \frac{7}{27}$.

Hammersley sequences:

$$\mathbf{u}_{i+1} = \left(\frac{i}{n}, \phi_{p_1}(i), \ldots, \phi_{p_{d-1}}(i) \right), \qquad i = 0, \ldots, n-1, \tag{5.6}$$

with ϕ_p as given previously. Warnock (1972) has $\mathbf{u}_i$ rather than $\mathbf{u}_{i+1}$ in definitions (5.5) and (5.6), but our definitions, with $\mathbf{u}_1 = \mathbf{0}$ in both cases, are more common.

The LP_τ sequences, defined in Sobol' (1967), are also based on the radical inverse function. Particular LP_τ sequences satisfying further criteria are described in Sobol' (1976); we shall call these restricted sequences *Sobol'*

TABLE 2
Discrepancies D^* of some two-dimensional sequences.

Sequence	Number of points n			
	32	64	128	256
Irrational (5.1)	0.075593	0.056624	0.047269	0.024452
Irrational (5.2)	0.147048	0.063998	0.051435	0.035319
Irrational (5.3)	0.232526	0.120083	0.069723	0.039034
Haber	0.175382	0.109077	0.116184	0.064437
Halton	0.104167	0.052083	0.036651	0.018760
Hammersley	0.097656	0.053711	0.029541	0.016052
Sobol'	0.089844	0.053711	0.025146	0.014587
Rational (k, n, d)	0.084961	0.041748	0.023071	0.012451
k =	7	19	47	75

sequences. An algorithm to generate Sobol' sequences may be found in Sobol' (1979); the number of nodes should be a power of 2.

The discrepancies D^* of some two-dimensional sequences are given in Table 2, for $n = 32$, 64, 128 and 256 points. Of the irrational sequences, (5.3) has the highest D^*, but it is decreasing more rapidly than the D^* for sequences (5.1) and (5.2). Of the irregular sequences, the Haber sequence has relatively high D^*, but the Halton, Hammersley and Sobol' sequences are shown to have only slightly higher discrepancies than the rational sequences (7, 32, 2), (19, 64, 2), (47, 128, 2) and (75, 256, 2), and have the advantage of being defined for arbitrarily large n. Note also that the often used "Cartesian product" set of nodes $\{\frac{1}{32}, \frac{3}{32}, \ldots, \frac{31}{32}\} \times \{\frac{1}{32}, \frac{3}{32}, \ldots, \frac{31}{32}\}$ has $D^* = (32 + 31)/32^2 = 0.061523$, much higher than the D^* for any of the 256-point sequences in Table 2 apart from the Haber sequence.

6. Some possible extensions.

6.1. *Randomised quasirandom rules.* Quasirandom sequences $s^{(1)} = (\mathbf{u}_i^{(1)})$, $s^{(2)} = (\mathbf{u}_i^{(2)}), \ldots$ can be combined in many ways to give new sequences $s = (\mathbf{u}_i)$. For example, suppose $s^{(1)}$ is a pseudorandom sequence. Then for $i = 1, 2, \ldots$, we can use $\mathbf{u}_{2i-1}^{(1)}$ to define a pseudorandom permutation σ_i of $\{1, 2, \ldots, d\}$, and define s by

$$\mathbf{u}_{ni+j-n} = \mathbf{u}_{2i}^{(1)} + \sigma_i\left(\mathbf{u}_{ni+j-n}^{(2)}\right) \pmod{\mathbf{1}}, \qquad i = 1, 2, \ldots, \; j = 1, 2, \ldots, n.$$

We are thus taking successive configurations of n points and applying pseudorandom offsets and permutations of the axes. An integral calculated numerically using the first $N = mn$ points as nodes is then the average of m subestimates, and the "between subestimate" variance, which should be much less than the "within subestimate" variance, can be used to assess the accuracy of the overall integral estimate. This is particularly useful if $s^{(2)}$ is a rational sequence with period n.

The sequence $s^{(1)}$ could be *any* quasirandom sequence that does not interact significantly with $s^{(2)}$; choosing a good pseudorandom sequence is just the easiest way to ensure this. Cranley and Patterson (1976) consider combining several rational sequences $s^{(1)}$, $s^{(2)}$, etc.

6.2. *Augmented rational sequences.* If s_n is a rational sequence with period n, then its centroid $\mathbf{c}$ is at $n\mathbf{u}_1 \pmod{\mathbf{1}}$. Therefore the $(n+1)$-point sequence $s'_{n+1} = s_n$, $(\mathbf{1} - \mathbf{c})$ has centroid $\mathbf{0.5} = (0.5, \ldots, 0.5)$, and with weights $v_i = 1/(n+1)$ will integrate exactly any linear function on C_d. This might seem a desirable property, and a similar idea with a particular irregular sequence appears in Arvidsen and Johnsson (1982). However, the analogy with the trapezium rule is destroyed in our case, and the resulting integration rules are, in fact, considerably less accurate for periodic integrands than the original rules given by the unaugmented rational sequences.

Similarly the $2n$-point sequence obtained by combining s_n with its reflection about $\mathbf{0.5}$ has centroid $\mathbf{0.5}$, but is in practice worse than combining two randomised n-point configurations.

6.3. *Nonuniform weights.* For a finite irrational sequence s_n, there may be some advantage in letting the corresponding weights v_i tend to 0 as i tends to 0 or n. This will help reduce any end effects. For example, we could define

$$
\begin{aligned}
v_i &= \frac{i}{r(n-r+1)}, && i = 1, 2, \ldots, r, \\
&= \frac{1}{n-r+1}, && i = r+1, \ldots, n-r, \\
&= \frac{n-i+1}{r(n-r+1)}, && i = n-r+1, \ldots, n.
\end{aligned} \tag{6.1}
$$

The case $n = 2r - 1$ with $\mathbf{u}_r = \mathbf{0.5}$ was studied by Haselgrove (1961), see also Sugihara and Murota (1982). Haselgrove derived rules with even higher accuracy for integrating periodic functions, but, unfortunately, their weights are not all positive.

The weights (6.1) could be used with other quasirandom sequences, such as

$$
s_{2m+1} = \mathbf{u}_0 + (-\mathbf{u}_m, -\mathbf{u}_{m-1}, \ldots, \mathbf{0}, \mathbf{u}_1, \ldots, \mathbf{u}_m) \pmod{\mathbf{1}}, \tag{6.2}
$$

where $n = 2m + 1$, $\mathbf{u}_0 \in C_d$ and $(\mathbf{u}_i)$ is the Haber sequence (5.4) or the Halton sequence (5.5). Some examples of rules with weights given by (6.1) are tested in Section 7.

6.4. *Quasirandom spherical rules.* This section briefly describes a transformation of C_d that may be better than (2.1) if the posterior distribution is ellipsoidally symmetric, or at least approximately so, which implies that the standardised posterior distribution will be roughly spherically symmetric. Such posterior distributions often arise from Bayesian analogues of classical statistical procedures, as in Lindley and Smith (1972).

A point $\mathbf{u}$ in C_{2m} can be transformed into a point $\mathbf{x}$ on U_{2m}, the surface of the $(2m)$-dimensional unit sphere, as follows. For $i = 1, 2, \ldots, m$, let

$$r_i = [-\log(u_{2i-1})]^{1/2},$$

$$c_i = \cos(2\pi u_{2i}), \qquad s_i = \sin(2\pi u_{2i}), \tag{6.3}$$

$$y_{2i-1} = c_i r_i, \qquad y_{2i} = s_i r_i,$$

and for $j = 1, 2, \ldots, 2m$, let

$$x_j = \frac{y_j}{\left[\sum_{i=1}^{2m} y_i^2\right]^{1/2}}. \tag{6.4}$$

This transformation is based on the Box–Muller (1958) method for generating pseudorandom Normal deviates; it easily follows that if $\mathbf{u}$ is a uniformly distributed random point in C_{2m}, then $\mathbf{x}$ will be uniformly distributed on U_{2m}. We can also generate a point on the $(2m-1)$-dimensional unit sphere by omitting the last coordinate and summing from 1 to $2m-1$ in the denominator of (6.4). A good sequence of points in C_{2m} can be transformed in this way into a "fairly good" sequence on U_d ($d = 2m$ or $2m-1$), although the transformation [(6.3) and (6.4)] is less smooth than (2.1) since the denominator in formula (6.4) jumps whenever $\mathbf{u}$ crosses the boundary of C_{2m}.

Sibuya (1962) used a similar idea to generate points uniformly on the surface of a sphere, but his method is less appropriate for transforming quasirandom sequences. Note that finding well distributed sequences on U_d is a very difficult problem; see, for example, Beck (1985).

Standard methods can be used to combine quasirandom integration rules on concentric spherical surfaces, by making the radii proportional to the square roots of the zeroes of a Laguerre polynomial [see Stroud (1971), Section 2.8]. The resulting integration rules have been used successfully in up to 20 dimensions, whereas previously known spherical rules as in Stroud (1968) and Keast and Diaz (1983) failed because of negative weights. An example is given in Shaw (1987a).

7. Examples. The methods described previously have been incorporated into computer programs for automatic adaptive integration, and have been extensively used and tested on many posterior densities. Skene, Shaw and Lee (1986) include examples requiring numerical integration in up to 12 dimensions. In this section we consider two artificial examples, so the different quasirandom sequences suggested previously can be tested on problems with known solutions.

7.1. *A well-behaved ten-dimensional posterior distribution.* Suppose we use the importance sampling function (2.1) with $f_j(u) = \log(u)$, $A_j = 0.5$, $b_j = 0$ and $c_j = 1.1633925$. The importance sampling density is then a product of logistic densities, and c_j is chosen so that straightforward Monte Carlo importance sampling will integrate a standard Normal density with minimum mean squared error, as described in the following discussion. We shall compare Monte Carlo

with some of the quasirandom integration rules given previously, when $p(\mathbf{x}\,|\,\mathbf{y})$ is the ten-dimensional standard Normal density.

The accuracy of Monte Carlo integration can be calculated exactly as follows. Consider first one-dimensional importance sampling with just one point; i.e., we shall estimate the integral

$$S(q) = \int_{-\infty}^{\infty} q(x)p(x)\,dx$$

by

$$\hat{S}(q;\ x_0) = \frac{q(x_0)p(x_0)}{g(x_0)},$$

where x_0 is a random point from density g. Then the expected mean squared error of this estimator is

$$\begin{aligned}\text{e.m.s.e.}(\hat{S};\ q) &= \int_{-\infty}^{\infty}\left[\frac{q(x)p(x)}{g(x)} - S(q)\right]^2 g(x)\,dx \\ &= \int_{-\infty}^{\infty}\frac{q(x)^2p(x)^2}{g(x)}dx - S(q)^2 \\ &= I(q) - S(q)^2, \quad \text{say.}\end{aligned}$$

If

$$p(x) = (2\pi)^{-1/2}\exp\left(-\tfrac{1}{2}x^2\right)$$

and

$$g(x) = \frac{2\exp(2x/c)}{c[1 + \exp(2x/c)]^2},$$

so that g corresponds to (2.1) with $f_j = \log$, $A_j = 0.5$ and $b_j = 0$, then it can be shown that

$$I(q) = \frac{c}{4\sqrt{\pi}}\left[2E\left\{q\left(\frac{Z}{\sqrt{2}}\right)^2\right\} + E\left\{q\left(\frac{Z}{\sqrt{2}} - \frac{1}{c}\right)^2 + q\left(\frac{Z}{\sqrt{2}} + \frac{1}{c}\right)^2\right\}\exp\frac{1}{c^2}\right],$$

where E denotes expectation with respect to Z, a random variable having a standard Normal distribution. In particular, e.m.s.e.$(\hat{S};\ 1)$ is minimised at $c = 1.1633925$, giving

$$I(1) = \frac{c}{2\sqrt{\pi}}\left\{\exp\frac{1}{c^2} + 1\right\} = 1.015245, \tag{7.1}$$

$$I(x) = \frac{c}{4\sqrt{\pi}}\left\{1 + \left(1 + \frac{2}{c^2}\right)\exp\frac{1}{c^2}\right\} = 1.015245 \tag{7.2}$$

and

$$I(x^2) = \frac{c}{2\sqrt{\pi}}\left\{\frac{3}{4} + \left(\frac{3}{4} + \frac{3}{c^2} + \frac{1}{c^4}\right)\exp\frac{1}{c^2}\right\} = 2.659350. \tag{7.3}$$

In the ten-dimensional case, since g and p are both products of independent densities, it follows from (7.1)–(7.3) that

$$\text{e.m.s.e.}(\hat{S}; 1) = I(1)^{10} - 1 = 0.163345,$$

$$\text{e.m.s.e.}(\hat{S}; x_1) = I(1)^9 I(x) - 0 = 1.163345,$$

$$\text{e.m.s.e.}(\hat{s}; x_1^2) = I(1)^8 I(x^2) - 1 = 2.001528,$$

$$\text{e.m.s.e.}(\hat{S}; x_1 x_2) = I(1)^8 I(x)^2 - 0 = 1.163345,$$

and the corresponding root e.m.s.e.'s are 0.40416, 1.0786, 1.4148 and 1.0786, respectively.

The first row (MC) in Table 3 repeats these expected mean squared errors, and hence gives the "mean squared error per node" for Monte Carlo with a large number of points. The remaining rows of Table 3 summarise the efficiencies of various quasirandom integration rules compared to straightforward Monte Carlo. The efficiencies (ratio of e.m.s.e. per node for Monte Carlo to e.m.s.e. per node for the given rule) were estimated numerically from $m = 100$ replications of the given n point rule, each replication incorporating a random offset and permutation of the axes, as outlined in Section 6.1. The randomisations used the standard APL pseudorandom number generator, a multiplicative congruential generator with multiplier 7^5 and modulus $2^{31} - 1$.

Rules R1 to R5 are the rational integration rules $(k, n, d) = (4, 53, 13)$, $(6, 91, 12)$, $(10, 121, 11)$, $(10, 237, 13)$ and $(23, 610, 10)$, respectively (see Table 1). Thus the row labelled R1 was obtained as follows:

(1) A permutation of the vector $\mathbf{k} = (1, 4, 16, \ldots, 4^{12} \bmod 53)$ was formed from the pseudorandom vector $(\mathbf{u}_1^{(1)})$: For $j = d, (d-1), \ldots, 3, 2$ in turn the jth and $(1 + \lfloor j u_{1j} \rfloor)$th elements of $\mathbf{k}$ were exchanged, where $\lfloor \cdot \rfloor$ denotes "integer part of."

(2) The first 10 elements of the permuted vector were taken as $(k_1, k_2, \ldots, k_{10})$ in formula (4.1), with $n = 53$, to give an integration rule over C_d.

(3) The pseudorandom vector $(\mathbf{u}_2^{(1)})$ was added to each of the 53 points in the integration rule, and the result was reduced modulo 1 in each dimension.

(4) The resulting nodes $(\mathbf{u}_i | i = 1, 2, \ldots, 53)$ were transformed to $(\mathbf{x}_i | i = 1, 2, \ldots, 53)$ by

$$x_{ij} = 1.1633925\big[0.5 \log(u_{ij}) - 0.5 \log(1 - u_{ij})\big], \qquad j = 1, 2, \ldots, d,$$

with corresponding weights

$$w_i = \frac{1}{53} \prod_{j=1}^{10} \big(1.1633925\big[0.5/u_{ij} + 0.5/(1 - u_{ij})\big]\big),$$

giving an integration rule over $\mathbb{R}^{10}$. This integration rule was used to obtain estimates of the integrals $S(1)$, $S(x_1)$, $S(x_1^2)$ and $S(x_1 x_2)$.

(5) Steps (1) to (5) were repeated another 99 times using the pseudorandom vectors $(\mathbf{u}_i^{(1)})$ for $i = 3, 4, \ldots, 200$, and the ratios of the observed mean squared errors to the theoretical mean squared errors for straightforward Monte Carlo importance sampling were calculated.

TABLE 3
Efficiencies of integration rules (smooth integrand)

Rule	n	$S(1)$	$S(x_1)$	$S(x_1^2)$	$S(x_1x_2)$
MC	1	0.163	1.163	2.002	1.163
R1	53	7	6	5	1
R2	91	11	8	8	2
R3	121	19	13	11	3
R4	237	14	10	10	4
R5	610	29	25	23	7
I1A	201	8	7	6	2
I2A	201	4	3	4	2
I3A	201	4	6	7	2
I3A1	201	5	5	4	1
I3A2	201	3	5	4	1
O1A	201	1	1	1	1
O2A	201	1	2	1	1
O3A	201	3	10	7	2
O4A	201	4	10	8	3
O5A	256	21	14	12	2
I1B	501	9	7	8	2
I2B	501	10	9	8	2
I3B	501	7	6	9	1
I3B1	501	12	10	6	1
I3B2	501	13	5	7	1
O1B	501	1	2	2	1
O2B	501	1	1	1	1
O3B	501	10	12	9	3
O4B	501	13	18	17	3
O5B	512	30	21	19	4
O1B1	501	1	1	1	1
O1B2	501	1	1	1	1
O1B3	501	1	1	1	1
O3B1	501	2	5	3	1
O3B2	501	2	5	4	1
O3B3	501	2	3	3	1
I3C	1001	8	6	6	2
I3C1	1001	10	10	7	2
I3C2	1001	12	6	7	2
O3C	1001	17	20	25	6
O4C	1001	16	19	20	5
O5C	1024	31	23	27	6

n = number of points in the given rule. The first row (MC) gives the expected mean squared error per point in estimating the corresponding integrals $S(1)$, $S(x_1)$, $S(x_1^2)$ and $S(x_1x_2)$ by Monte Carlo. Other rows give approximate relative efficiencies of the named rule compared to Monte Carlo for the corresponding integral. Thus, using a single replication of rule R3 with a random offset and axis permutation, the estimate of $S(1)$ has approximate standard error $\{0.163/(121 \times 19)\}^{1/2} = 0.008$, and the estimate of $S(x_1x_2)$ has approximate standard error $\{1.163/(121 \times 3)\}^{1/2} = 0.06$. See Section 7.1 for a full description.

The other rows of Table 3 were produced similarly. Rules I1A, I2A and I3A use the first 201 points of the irrational sequences (5.1) with p_j the jth prime, (5.2) with $\xi = 3$, and (5.3) with $p = 23$, respectively. Rules I3A1 and I3A2 have the same nodes as I3A, but the weights are given by (6.1) with $r = 50$ (I3A1) or $r = (n + 1)/2$ (I3A2). Rule O1A uses the first 201 points of the Haber sequence (5.4), rule O2A uses the Haber sequence with I3A replacing I1A as the parent sequence and rules O3A and O4A use the Halton sequence (5.5) and the Hammersley sequence (5.6), respectively, with $n = 201$ and p_j the jth prime. Rule O5A uses the first 256 points from the ten-dimensional Sobol' (1976) sequence.

Rules I1B to O5B are the same as I1A to O5A, but with 501 nodes (or 512 nodes in the case of O4C).

Rules O1B1 to O1B3 use the Haber sequence in (6.2), with $m = 250$ and weights given by (6.1) and $r = 1$ (O1B1), $r = 50$ (O1B2) or $r = (n + 1)/2$ (O1B3). Rules O3B1 to O3B3 are similar, but based on the Halton sequence.

Finally, rules I3C, I3C1, I3C2, O3C, O4C and O5C are as for I3B to O5B, but with more nodes.

The estimated efficiencies are sufficiently accurate to demonstrate the advantage of quasirandom rules, particularly those based on suitable rational sequences, on the Sobol' sequence, and, at least for larger n, on the Halton and Hammersley sequences. For example, a single replication of the 121-point rule R3 will estimate the normalising constant $S(1)^{-1} = 1$ with standard error less than 0.01, whereas over 1600 points would be needed with Monte Carlo. This is despite the fact that the importance sampling distribution was scaled to show Monte Carlo at its best.

7.2. *A badly behaved three-dimensional posterior distribution.* Consider the standardised density given by

$$p(\mathbf{x}|\mathbf{y}) = p_1(x_1)p_2(x_2, x_3),$$

where $p_1(x_1)$ is the standard exponential density relocated to have zero mean,

$$p_1(x_1) = \exp(-1 - x_1), \qquad x_1 \geq -1,$$

and p_2 is uniform within three circles, each with radius 1, centred at $(0, \sqrt{\frac{3}{2}})$, $(\sqrt{\frac{9}{8}}, -\sqrt{\frac{3}{8}})$ and $(-\sqrt{\frac{9}{8}}, -\sqrt{\frac{3}{8}})$,

$$p_2(x_2, x_3) = \frac{1}{3\pi}\left\{\psi\left(x_2, x_3, 0, \sqrt{\frac{3}{2}}\right) + \psi\left(x_2, x_3, \sqrt{\frac{9}{8}}, -\sqrt{\frac{3}{8}}\right) + \psi\left(x_2, x_3, -\sqrt{\frac{9}{8}}, -\sqrt{\frac{3}{8}}\right)\right\},$$

where

$$\psi(x_2, x_3, c_2, c_3) = \begin{cases} 1, & \text{if } (x_2 - c_2)^2 + (x_3 - c_3)^2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus p is 0 except within three parallel semiinfinite cylinders, is multimodal, discontinuous, highly skewed and asymmetric with respect to permutations of its

arguments. Posterior distributions qualitatively similar to p have arisen in practice from nonlinear calibration models; an example is given in Naylor and Smith (1986).

Table 4 summarises the efficiency of quasirandom integration rules, using the importance sampling function (2.1) with $f_j(u) = -([1-u]/u)^{0.3}$, $c_j = 1.897896$, $j = 1,2,3$, $b_1 = -0.8$, $b_2 = b_3 = 0$ and $A_1 = 0.1$, $A_2 = A_3 = 0.5$. This function was suggested by a single iteration of the adaptive importance sampling scheme mentioned at the start of Section 7; the constants c_j are again designed for integrating the standard Normal density, see Shaw (1986a) for details. More appropriate importance sampling distributions are easily found, but the purpose of this example is to illustrate the increased efficiency of quasirandom methods over Monte Carlo, even in circumstances that are far from ideal. Note, in particular, that the transformed integrands in C_d are discontinuous across surfaces not parallel to the faces of C_d, and are therefore not of bounded variation. Formula (3.2) therefore gives no upper bound on the numerical integration errors.

The e.m.s.e.'s in row 1 (MC) of Table 4 were calculated numerically. Rules R1 to R8 are the recommended integration rules for $d = 3$ from Table 1, other rows are as for Table 4, described in Section 7.1.

Again the Halton, Hammersley and Sobol' sequences perform well, and the rules R1 to R8 based on rational sequences are also surprisingly efficient given that they were not designed for such badly behaved integrands.

7.3. *Summary of experience with other examples.* The following general recommendations stem from work with many different examples, some of which appear in an unpublished Ph.D. thesis [Shaw (1987b)].

The new criteria (ν, ρ_i and ν_i) defined in Section 4 have identified integration rules that prove efficient in calculating statistically important integrals, the relative efficiencies in Tables 3 and 4 being typical. These rules can be randomised and replicated to obtain realistic estimates of the errors in numerical integration (Section 6.1). This is recommended as a general method for integration when combined with a smooth transformation of the unit d-dimensional cube, such as that given by formula (2.2).

Integration rules based on irrational regular sequences, possibly with nonuniform weights, are asymptotically more efficient than randomised replications of any given integration rule, since a randomised integration rule is just high-dimensional Monte Carlo on a transformed integrand. However, the number of nodes necessary for randomised rational rules to be dominated seems large (see Tables 3 and 4). Rules based on formula (5.3) have been found more useful with quasirandom spherical rules (Section 6.4), presumably because the nonsmooth transformation (6.3) destroys the "trapezium rule-like" property of rational regular sequences.

Rules based on the Haber sequence (5.4) were disappointing, but quasirandom sequences based on the radical inverse function yield generally efficient rules, and can be useful in exploratory analysis as well as for integration, see Shaw (1987a).

It should be emphasised that the choice of parametrisation is critical for high-dimensional numerical integration to be efficient: The ideal posterior

TABLE 4
Efficiencies of integration rules (awkward integrand)

Rule	n	$S(1)$	$S(x_1)$	$S(x_1^2)$	$S(x_1x_2)$
MC	1	3.2	3.3	11.2	4.8
R1	14	2	2	2	1
R2	18	2	2	2	1
R3	38	3	2	2	1
R4	78	2	2	2	2
R5	155	3	3	2	2
R6	252	4	5	2	2
R7	388	5	3	3	3
R8	555	5	4	3	2
I1A	201	2	2	2	1
I2A	201	3	2	2	2
I3A	201	3	2	2	1
I3A1	201	2	2	2	1
I3A2	201	2	1	2	1
O1A	201	1	1	1	2
O2A	201	1	1	1	1
O3A	201	3	3	2	2
O4A	201	4	2	3	2
O5A	256	3	3	2	1
I1B	501	4	3	3	2
I2B	501	3	4	3	2
I3B	501	2	4	2	2
I3B1	501	2	3	1	2
I3B2	501	2	2	2	1
O1B	501	1	1	1	1
O2B	501	1	1	1	1
O3B	501	3	2	3	2
O4B	501	4	3	3	2
O5B	512	2	2	3	2
O1B1	501	1	1	1	1
O1B2	501	1	1	1	1
O1B3	501	1	1	1	1
O3B1	501	2	1	1	1
O3B2	501	1	1	1	1
O3B3	501	1	1	1	1
I3C	1001	2	4	2	2
I3C1	1001	2	2	2	2
I3C2	1001	1	2	1	2
O3C	1001	4	3	3	2
O4C	1001	5	4	4	3
O5C	1024	4	4	3	3

n = number of points in the given rule. The first row (MC) gives the expected squared error per point in estimating the corresponding integrals $S(1)$, $S(x_1)$, $S(x_1^2)$ and $S(x_1x_2)$ by Monte Carlo. Other rows give approximate relative efficiencies of the named rule compared to Monte Carlo for the corresponding integral. See Section 7.2 for a full description.

density is Normal, so that the affinely transformed standardised posterior density (Section 2) is $N(\mathbf{0}, \mathbf{I})$, as in Section 7.1. For example, even though posterior densities arising from variance component problems can be well approximated by products of univariate densities, see Box and Tiao (1973), Section 6.2, parametrising in terms of the logarithms of variances generally makes the posterior more nearly Normal, and numerical integration with the methods described previously are consequently more efficient. Other natural preliminary transformations, such as logits of proportions, are also helpful. Even so, the example of Section 7.2 shows that quasirandom methods are still useful with badly behaved posterior distributions.

Acknowledgments. This paper is an extension of Shaw (1986b) and forms a contribution to the Warwick University Bayesian Statistics Study Year. Part of this research was carried out during 1985 while the author was a research assistant on a Science and Engineering Research Council funded project at Nottingham University to investigate numerical methods in Bayesian statistics.

I am extremely grateful to Professor Adrian Smith and to the (anonymous) referees for detailed comments on earlier versions of this paper.

REFERENCES

ARVIDSEN, N. I. and JOHNSSON, T. (1982). Variance reduction through negative correlation, a simulation study. *J. Statist. Comput. Simulation* **15** 119–127.

BECK, J. (1985). Irregularities of distribution and combinatorics. In *Surveys in Combinatorics 1985. London Math. Soc. Lecture Notes* **103** 25–46. Cambridge Univ. Press, Cambridge.

BOX, G. E. P. and MULLER, M. E. (1958). A note on the generation of random normal deviates. *Ann. Math. Statist.* **29** 610–611.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Mass.

CRANLEY, R. and PATTERSON, T. N. L. (1976). Randomization of number theoretic methods for multiple integration. *SIAM J. Numer. Anal.* **13** 904–914.

DAVIS, P. J. and RABINOWITZ, P. (1984). *Methods of Numerical Integration*, 2nd ed. Academic, Orlando, Fla.

HABER, S. (1970). Numerical evaluation of multiple integrals. *SIAM Rev.* **12** 481–526.

HABER, S. (1983). Parameters for integrating periodic functions of several variables. *Math. Comp.* **41** 115–129.

HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964). *Monte Carlo Methods*. Methuen, London.

HASELGROVE, C. B. (1961). A method for numerical integration. *Math. Comp.* **15** 323–337.

HLAWKA, E. (1971). Discrepancy and Riemann integration. In *Studies in Pure Mathematics* (*Papers Presented to Richard Rado*) (L. Mirsky, ed.) 121–129. Academic, New York.

HUA, L.-K. and WANG, Y. (1965). On numerical integration of periodic functions of several variables. *Sci. Sinica* **14** 964–978.

HUA, L.-K. and WANG, Y. (1981). *Applications of Number Theory to Numerical Analysis*. Springer, Berlin.

KEAST, P. and DIAZ, J. C. (1983). Fully symmetric integration formulas for the surface of the sphere in S dimensions. *SIAM J. Numer. Anal.* **20** 406–419.

KUIPERS, L. and NIEDERREITER, H. (1974). *Uniform Distribution on Sequences*. Wiley, New York.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 1–34.

NAYLOR, J. C. and SMITH, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *Appl. Statist.* **31** 214–225.

NAYLOR, J. C. and SMITH, A. F. M. (1983). A contamination model in clinical chemistry: An illustration of a method for the efficient computation of posterior distributions. *Statistician* **32** 82–87.

NAYLOR, J. C. and SMITH, A. F. M. (1986). An archaeological inference problem. Research Report 06-86, Nottingham Statist. Group, Dept. Math., Univ. Nottingham.

NIEDERREITER, H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc.* **84** 957–1042.

SHAW, J. E. H. (1986a). A class of univariate distributions for use in Monte Carlo studies. Research Report 04-86, Nottingham Statist. Group. Dept. Math., Univ. Nottingham.

SHAW, J. E. H. (1986b). A quasirandom approach to integration in Bayesian statistics. Research Report 05-86, Nottingham Statist. Group. Dept. Math., Univ. Nottingham.

SHAW, J. E. H. (1987a). Aspects of numerical integration and summarisation. Research Report 103, Dept. Statist., Univ. Warwick.

SHAW, J. E. H. (1987b). Numerical integration and display methods for Bayesian inference. Unpublished Ph.D. thesis, Dept. Math., Univ. Nottingham.

SIBUYA, M. (1962). A method for generating uniformly distributed points on n-dimensional spheres. *Ann. Inst. Statist. Math.* **14** 81–85.

SKENE, A. M., SHAW, J. E. H. and LEE, T. D. (1986). Bayesian modelling and sensitivity analysis using interactive computing. *Statistician* **35** 281–288.

SMITH, A. F. M., SKENE, A. M., SHAW, J. E. H., NAYLOR, J. C. and DRANSFIELD, M. (1985). The implementation of the Bayesian paradigm. *Comm. Statist. A—Theory Methods* **14** 1079–1102.

SOBOL' I. M. (1967). The distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Comput. Math. and Math. Phys.* **7** 86–112.

SOBOL', I. M. (1976). Uniformly distributed sequences with an additional uniform property. *U.S.S.R. Comput. Math. and Math. Phys.* **16** 236–242.

SOBOL', I. M. (1979). On the systematic search in a hypercube. *SIAM J. Numer. Anal.* **16** 790–793.

STEWART, L. T. (1985). Multiparameter Bayesian inference using Monte Carlo integration—some techniques for bivariate analysis. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 495–510. North-Holland, Amsterdam.

STROUD, A. H. (1968). Some seventh degree integration formulas for the surface of an n-sphere. *Numer. Math.* **11** 273–276.

STROUD, A. H. (1971). *Approximate Calculation of Multiple Integrals.* Prentice-Hall, Englewood Cliffs, N.J.

SUGIHARA, M. and MUROTA, K. (1982). A note on Haselgrove's method for numerical integration. *Math. Comp.* **39** 549–554.

VAN DIJK, J. K. and KLOEK, T. (1985). Experiments with some alternatives for simple importance sampling in Monte Carlo integration. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 511–530. North-Holland, Amsterdam.

WALLEY, D. and FEARN, T. (1979). Trapezium rule, OK. *Appl. Statist.* **28** 296–297.

WARNOCK, T. T. (1972). Computational investigations of low-discrepancy point sets. In *Applications of Number Theory to Numerical Analysis* (S. K. Zaremba, ed.) 319–343. Academic, New York.

WICHMANN, B. A. and HILL, I. D. (1982). Algorithm ASR183. An efficient and portable pseudo-random number generator. *Appl. Statist.* **31** 188–190. (Correction **33** 123. Remarks **34** 198–200; **35** 89.)

ZAREMBA, S. K. (1968). The mathematical basis of Monte Carlo and quasi-Monte Carlo methods. *SIAM Rev.* **10** 303–314.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM