Statistica Sinica **13**(2003), 1-17

THE DIMENSION DISTRIBUTION AND QUADRATURE TEST FUNCTIONS

Art B. Owen

Stanford University

Abstract: This paper introduces the dimension distribution for a square integrable function f on $[0,1]^s$. The dimension distribution is used to relate several definitions of the effective dimension of a function. Functions of low effective dimension can be easy to integrate numerically. Many commonly considered quadrature test functions are sums or products of univariate functions, and as a result have particularly simple dimension distributions. Recently some high dimensional isotropic integrals have been successfully treated by quasi-Monte Carlo methods. We show numerically that one such function in 25 dimensions is very nearly a superposition of functions of 3 or fewer variables, explaining the success of QMC on that problem. A new result shows that certain isotropic polynomials of degree n generate integrands that are exact superpositions of functions of n or fewer variables.

Key words and phrases: Discrepancy, effective dimension, isotropic integrand, quasi-Monte Carlo.

1. Introduction

The analysis of variance (ANOVA) is a tool devised for the statistical analysis of designed experiments. It has recently found widespread use in the study of quasi-Monte Carlo (QMC) integration methods, where it is applied to various notions of the effective dimension of an integrand. The ANOVA has also been used in sensitivity analysis of high complexity computer codes to identify important subsets of input variables. In both of these settings we have a function f defined on an s dimensional product domain, which we suppose is $[0,1]^s$ possibly after some transformation.

This paper introduces a probability measure $\mu(u)$ on nonempty subsets $u \subseteq \{1,\ldots,s\}$, in which $\mu(u)$ is proportional to the variance contribution to f of the subset u of input variables of f. If U is a random μ distributed subset, then its cardinality, denoted $|U|$ is a random variable. The distribution $\nu(\cdot)$ of the random variable $|U|$ is the dimension distribution of f. Caflisch, Morokoff, and Owen (1997) defined two notions of the effective dimension of an integral which may be defined through quantiles of this distribution. Here we present some more easily computable measures of the effective dimension of a function based on the mean and variance of $|U|$. Simple probability bounds, such as those of Markov and Chebychev can then be applied to get bounds on the original quantile based definitions.

The main contribution of this paper is to use the dimension distribution to analyze test functions commonly used in QMC problems. Many test problems are constructed as sums or products of one dimensional functions. The dimension distribution of such test problems is very easy to analyze, and this provides insight into which test problems should be hard for QMC and which should be relatively easy.

As Tezuka (2002) notes, good results for QMC methods have been reported for two classes of functions: functions of low effective dimension, and isotropic functions. The isotropic functions arise as expectations of functions of $\|X\|$ where X is an s dimensional spherical Gaussian random vector. For such s dimensional isotropic integration problems, some techniques of Sobol' (2001) for sensitivity analysis can be used to study the dimension distribution. One result is that we are able to show that some isotropic functions studied in the literature are of low effective dimension. In particular a 25 dimensional function studied in Papageorgiou and Traub (1997) is shown to be very nearly a superposition of functions involving 3 or fewer variables.

The ANOVA decomposition of $[0, 1]^s$ has been studied by many authors: Hoeffding (1948) used it in the study of U-statistics; Sobol' (1969) used it in the study of quadrature methods, calling it the "decomposition into summands of different dimensions"; Efron and Stein (1981) used it to prove their famous lemma on jackknife variances; Takemura (1983) gives an historical account; Owen (1997a) presents a continuous space version of the nested ANOVA; Hickernell (1996) presents a reproducing kernel Hilbert space version.

Section 2 introduces the ANOVA decomposition and some sensitivity coefficients of Sobol' (2001), and then the dimension distribution is presented in Section 3. Section 4 explains why QMC can be much better than MC on integrands of low effective dimension. Certain classes of functions with simple dimension distributions are presented in Section 5 and numerical results for examples considered in the literature are presented in Section 6. In particular Section 5.3 describes how to get information about the dimension of isotropic functions through a sensitivity analysis technique. Further, Proposition 3 there shows that certain isotropic polynomials of degree $n < s$ give rise to s dimensional integrands that are exactly superpositions of functions depending on n or fewer variables. Section 6.6 shows that a specific 25 dimensional isotropic integrand is very nearly a superposition of 3 dimensional functions. Some concluding remarks are made in Section 7.

2. ANOVA Decomposition

Let f be a function on $[0,1]^s$ with $\int f(x)^2 dx < \infty$. Here and elsewhere, integrals without explicit domains are expectations over $[0,1]^s$ with respect to the uniform distribution $U[0,1]^s$. The expectation of f is $I = I(f) = \int f(x)dx$, and the variance of f is $\sigma^2 = \sigma^2(f) = \int (f(x) - I)^2 dx$. To avoid trivialities, we assume that $\sigma^2 > 0$.

ANOVA is a tool for describing the dependence of f on subsets of the components in X. Let $u \subseteq \{1, \ldots, s\}$ denote such a subset. We use $|u|$ for the cardinality of u and $-u$ for the complementary set $\{1, \ldots, s\} - u$. A generic point of $[0,1]^s$ is written $X = (X^1, \ldots, X^s)$ and X^u denotes the $|u|$-vector of components X^j for $j \in u$.

We write $[0,1]^u$ for the domain of X^u. The integral over $X^u \in [0,1]^u$ of a function $g(X)$, is a real valued function that depends on X only through X^{-u}. For instance $\int_{[0,1]^{\{2\}}} x^1 + x^2 dx^{\{2\}} = x^1 + 1/2$.

In the ANOVA decomposition, each square integrable function f is written as a sum

$$f(X) = \sum_{u \subseteq \{1,\ldots,s\}} f_u(X), \tag{1}$$

where $f_u(X)$ depends on X only through X^u. The ANOVA terms are defined by:

$$f_u(X) = \int_{\mathcal{X}^{-u}} \Big(f(X) - \sum_{v \subsetneq u} f_v(X) \Big) dX^{-u} \tag{2}$$

$$= \int_{\mathcal{X}^{-u}} f(X) dX^{-u} - \sum_{v \subsetneq u} f_v(X). \tag{3}$$

Using standard conventions $f_\emptyset(X) = I(f)$ for all X. In (2) one first subtracts from f what can be attributed to proper subsets of u and then averages over the values of X^{-u}. Equation (3) is simpler for some manipulations, but for $f > 0$ with $I \gg \sigma$, equation (2) may be preferable numerically.

For the decomposition in (2) one can show by induction on $|u|$ that, for $j \in u$ and any X, $\int_0^1 f_u(X) dX^j = 0$. Then it follows that distinct ANOVA terms belong to orthogonal spaces: if $\int f^2(x)dx < \infty$, $\int g^2(x)dx < \infty$ and $u \neq v$, then $\int f_u(x) g_v(x) dx = 0$.

From the orthogonality of ANOVA terms it is easy to show that the variance of f may be written $\sigma^2 = \sum_{|u|>0} \sigma_u^2$ where $\sigma_u^2 = \sigma_u^2(f) = \int f_u(x)^2 dx$ when $|u| > 0$ and $\sigma_\emptyset^2(f) = 0$. We use σ_u^2 to measure the importance of f_u. Normalized versions σ_u^2/σ^2 are called global sensitivity indices in Sobol' (2001).

Sobol' (2001) presents two ways to quantify the importance of a subset u of variables to a function f, using certain totals of σ_u^2. We write these as

$$\underline{\tau}_u^2 = \sum_{v\subseteq u} \sigma_v^2, \quad \text{and,} \tag{4}$$

$$\overline{\tau}_u^2 = \sum_{v\cap u\neq\emptyset} \sigma_v^2. \tag{5}$$

The quantity $\underline{\tau}_u^2$ describes the effects on f of X^u alone while $\overline{\tau}_u^2$ includes all the effects of variables X^u acting in conjunction with others. It is easy to show that $0 \le \underline{\tau}_u^2 \le \overline{\tau}_u^2 \le \sigma^2$ and that $\underline{\tau}_u^2 + \overline{\tau}_{-u}^2 = \sigma^2$. If $\overline{\tau}_u^2$ is small then, as Sobol' (2001) describes, the variables X^u may be considered inessential, and perhaps be fixed at default values.

Theorem 1 below shows how to express $\underline{\tau}_u^2$ and $\overline{\tau}_u^2$ directly in terms of certain integrals. Sobol' and Levitan (1999) describe QMC methods for evaluating these quantities.

Theorem 1. *Define $f(x^u, y^{-u})$ to be $f(z)$ where $z^u = x^u$ and $z^{-u} = y^{-u}$. Then*

$$\underline{\tau}_u^2 = \int_{[0,1]^{2s-|u|}} f(x^u, y^{-u}) f(x^u, z^{-u})\, dx^u dy^{-u} dz^{-u} - I^2$$

$$\overline{\tau}_u^2 = \frac{1}{2}\int_{[0,1]^{s+|u|}} \Big(f(y^u, x^{-u}) - f(z^u, x^{-u}) \Big)^2 dx^{-u} dy^u dz^u.$$

Proof. These are Theorems 2 and 3 of Sobol' (2001), respectively.

3. Dimension Distribution and Effective Dimension

Caflisch, Morokoff, and Owen (1997) define the effective dimension of a function in two senses. The function f has effective dimension d in the superposition sense if $\sum_{|u|\le d} \sigma_u^2 \ge 0.99\sigma^2$ and it has effective dimension d in the truncation sense if $\sum_{u\subseteq\{1,\dots,d\}} \sigma_u^2 \ge 0.99\sigma^2$.

The idea of effective dimension appears in Paskov and Traub (1995) where they remark that their functions are not determined by just a small number of leading input variables. Sloan and Wozniakowski introduce classes of functions in which the importance of each input variable X^j decays as j increases. Such functions can have very small truncation dimension relative to their nominal dimension.

The definitions above capture two notions in which f is almost d dimensional. As Caflisch, Morokoff, and Owen (1997) remark, the choice of 99'th percentile is arbitrary. Hickernell (1998) makes the threshold quantile a parameter in the definition. Owen (1998a) shows by a martingale argument that any square integrable function on $[0,1]^\infty$ necessarily has finite effective dimension in the truncation sense for any threshold less than 100 percent.

Now consider choosing a subset $U \subseteq \{1,\ldots,s\}$ at random with $\Pr(U=u) = \mu(u) = \sigma_u^2/\sigma^2$. The size of this random U may be measured either through its cardinality or through the maximum of its elements' indices. These then become random variables through which a dimension distribution of f may be defined.

Definition 1. The dimension distribution of f (in the superposition sense) is the probability distribution of $|U|$ when $\Pr(U=u) = \sigma_u^2/\sigma^2$. It has probability mass function $\nu(d) = \nu_{\mathrm{S}}(d) = \sum_{|u|=d} \sigma_u^2/\sigma^2$, $d = 1,\ldots,s$.

Definition 2. The dimension distribution of f in the truncation sense is the probability distribution of $\max\{j \mid j \in U\}$ when $\Pr(U=u) = \sigma_u^2/\sigma^2$. It has probability mass function $\nu_{\mathrm{T}}(d) = \sum_{\max\{j|j\in u\}=d} \sigma_u^2/\sigma^2$, $d = 1,\ldots,s$.

The function f has superposition dimension d if the 99'th percentile of ν is at most d. Similarly f has truncation dimension d if the 99'th percentile of ν_{T} is at most d. The mean dimension of f in the superposition sense is

$$D = D_{\mathrm{S}} = \frac{\sum_{|u|>0} \sigma_u^2 |u|}{\sum_{|u|>0} \sigma_u^2}. \tag{6}$$

The mean dimension of f in the truncation sense is

$$D_{\mathrm{T}} = \frac{\sum_{|u|>0} \sigma_u^2 \max\{j \mid j \in u\}}{\sum_{|u|>0} \sigma_u^2}. \tag{7}$$

These mean dimensions are the expectations of $|U|$ and $\max_{j\in U}$ respectively. The mean dimension is often simpler to compute than are quantiles of the dimension distribution. Variances of these two dimension distributions can also be defined in straightforward ways.

4. Effective Dimension and Quasi-Monte Carlo

A low superposition dimension can help to explain why QMC rules work better than MC on problems of high nominal dimension. Here we outline the connection, using notions of (t,m,s)-nets and elementary intervals in base b from Niederreiter (1992), and a new notion called the active dimension of an elementary interval.

Definition 3. Let $s \geq 1$ and $b \geq 2$ be integers. An elementary interval in base b is a subinterval of $[0,1)^s$ of the form

$$E = \prod_{j=1}^{s} \left[\frac{a_j}{b^{k_j}}, \frac{a_j+1}{b^{k_j}}\right) \tag{8}$$

for integers k_j and a_j satisfying $k_j \geq 0$ and $0 \leq a_j < b^{k_j}$.

An elementary interval with some $k_j = 0$ can be considered less than fully s-dimensional. We take the number of nonzero k_j as a definition of the active dimension of an elementary interval:

Definition 4. The active dimension of the elementary interval E in (8) is $d = d(E) = \sum_{j=1}^{s} 1_{k_j>0}$.

Definition 5. Let $t \leq m$ be a nonnegative integer. A finite sequence of b^m points from $[0,1)^s$ is a (t,m,s)-net in base b if every elementary interval in base b of volume b^{t-m} contains exactly b^t points of the sequence.

Let us say that the set E is "balanced by a (t,m,s)-net" if it is guaranteed by Definition 5 to have $n\text{Vol}(E)$ points of the net in it where $n = b^m$ is the number of points of the net and $\text{Vol}(E)$ is the s dimensional volume of E. The set E in Definition 3 has s dimensional volume $\text{Vol}(E) = b^{-\sum_{j=1}^{s} k_j}$. It is balanced by the net in Definition 5 if and only if $m \geq t + \sum_{j=1}^{s} k_j$.

It is well known that low dimensional coordinate projections of QMC rules tend to be very well equidistributed while higher dimensional projections tend not to be unless n is large. An elementary interval of active dimension d has $\sum_{j=1}^{s} k_j \geq d$, and so no such interval is balanced by the net unless $n \geq b^{t+d}$. For (t,m,s)-nets there is no equidistribution among projected points X_i^u for $u \subseteq \{1,\ldots,s\}$ beyond that implied by equidistribution of projected points X_i^v with $v \subset u$, unless $n \geq b^{t+|u|}$.

Now let $f(x) = \sum_u f_u(x)$ where f_u depends on x only through x^u. Then the quadrature error in a QMC rule $X_1,\ldots,X_n$ satisfies the bound

$$\Bigl|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - I\Bigr| \leq \sum_{|u|>0} D_{n,|u|}(X_1^u,\ldots,X_n^u)\|f_u\|, \tag{9}$$

where $D_{n,|u|}$ is a discrepancy for n points in $[0,1)^{|u|}$ and $\|f_u\|$ is a compatible norm. There are many choices of discrepancy and corresponding norms. See for example Hickernell (1998). Formula (9) appears in Caflisch, Morokoff, and Owen (1997) and several versions are presented in Hickernell (1998).

The bound in (9) can be used to explain why QMC can be so much better than MC on functions of low superposition dimension. Suppose that the QMC rule has much smaller projected discrepancies $D_{n,|u|}$ than MC typically has when u is small and that the integrand f has small higher dimensional parts as measured by $\|f_u\|$ when $|u|$ is large. Then all terms in (9) are small for QMC, while the terms for small $|u|$ can be large for MC. It is possible that the values of $D_{n,|u|}$ for small $|u|$ can appear to follow an asymptote of $n^{-1+\epsilon}$ for practical sample sizes n, even when the full s dimensional discrepancy does not. In such cases the error

(9) will have empirical behavior that appears to be $O(n^{-1+\epsilon})$ for practical sample sizes, even though a worst case analysis could show that the rate $O(n^{-1+\epsilon})$ has to set in when $n \geq \exp(s)$.

5. Additive, Multiplicative, and Isotropic Functions

Multivariate functions defined in a simple way through one or more univariate functions are widely used as test functions for quadrature. Their simplicity often allows one to compute the desired integral in closed form. The ANOVA can also simplify for such functions. Here we consider several cases.

5.1. Additive functions

Additive functions of the form $f(x) = \sum_{j=1}^{s} g_j(x^j)$ have a particularly simple ANOVA decomposition. Letting $\mu_j = \int_0^1 g_j(z)dz$ and $\gamma_j^2 = \int_0^1 (g_j(z) - \mu_j)^2 dz$ we easily find that $f_\emptyset = \sum_{j=1}^{s} \mu_j$, $f_{\{j\}}(x) = g_j(x^j) - \mu_j$, and $f_u(x) = 0$, for $|u| > 1$, and so $\sigma_u^2 = 0$ unless $|u| = 1$ in which case $\sigma_{\{j\}}^2 = \gamma_j^2$. The mean dimension of f is one in the superposition sense. In the truncation sense it is $\sigma^{-2} \sum_{j=1}^{s} j\gamma_j^2$.

5.2. Multiplicative functions

Functions of the product form

$$f(x) = \prod_{j=1}^{s} g_j(x^j) \tag{10}$$

are widely used as test functions. Letting μ_j and σ_j^2 be as in Section 5.1 we find that

$$\begin{aligned} f_u(x) &= \prod_{j\in u} [g_j(x^j) - \mu_j] \prod_{j\notin u} \mu_j, \quad \text{and,} \\ \sigma_u^2 &= \prod_{j\in u} \gamma_j^2 \prod_{j\notin u} \mu_j^2. \end{aligned} \tag{11}$$

Hickernell (1996) has this result explicitly and it also seems clear that Sobol' (1993) used it in some examples.

The denominator in both mean dimensionalities is simply the variance of f. For product functions it may be written

$$\sigma^2 = \prod_{j=1}^{s} (\mu_j^2 + \gamma_j^2) - \prod_{j=1}^{s} \mu_j^2. \tag{12}$$

Proposition 1. *For the product function* (10)*, the mean dimension in the superposition sense is*

$$D = \frac{\sum_{j=1}^{s} \gamma_j^2/(\gamma_j^2 + \mu_j^2)}{1 - \prod_{j=1}^{s} \mu_j^2/(\gamma_j^2 + \mu_j^2)}. \tag{13}$$

Proof. Consider s independent random variables $j_1, \ldots, j_s \in \{0,1\}$ with $\Pr(j_r = 1) = \gamma_r^2/(\gamma_r^2 + \mu_r^2) \equiv p_r$. Let $U = \{r \mid j_r = 1\}$ be a random subset of $\{1, \ldots, s\}$ determined by the j_r. Then σ_u^2/σ^2 is equal to the conditional probability $\Pr(U = u \mid U \neq \emptyset)$, as may be verified by direct calculation from (11) and (12). Finally D may be written

$$\begin{aligned} E(|U| \mid U \neq \emptyset) &= \sum_{r=1}^{s} \Pr(j_r = 1 \text{ and } U \neq \emptyset) / \Pr(U \neq \emptyset) \\ &= \Big(\sum_{r=1}^{s} p_r \Big) \Big/ \Big(1 - \prod_{r=1}^{s} (1 - p_r) \Big), \end{aligned}$$

as required.

Many test functions have $\mu_j = 0$ for some or even all j. The assumption that $\sigma^2 > 0$ rules out the possibility that $\mu_j = \gamma_j = 0$. Suppose that $\mu_j = 0$ for $j = 1, \ldots, r$ where $1 \leq r \leq s$, and that $\mu_j \neq 0$ if $j > r$. Then (13) simplifies to, $D = r + \sum_{j=r+1}^{s} \gamma_j^2/(\gamma_j^2 + \mu_j^2)$, and the consequence of $\mu_j = 0$ is generally to increase the mean dimension.

Markov's inequality bounds the fraction of σ^2 attributable to d and higher dimensional ANOVA components by $\nu([d,s]) \leq D/d$. A related inequality, resembling Chebychev's, is $\nu([d,s]) \leq d^{-2} \sum_{j=1}^{s} \nu(j) j^2$, which uses the mean square of the dimension distribution.

Proposition 2. *For the product function* (10), *the mean square dimension in the superposition sense is*

$$\frac{\sum_{j=1}^{s} p_j(1 - p_j) + \left(\sum_{j=1}^{s} p_j\right)^2}{1 - \prod_{j=1}^{s}(1 - p_j)}, \tag{14}$$

where $p_j = \gamma_j^2/(\gamma_j^2 + \mu_j^2)$.

Proof. As in Proposition 1, introduce s independent random variables $j_1, \ldots, j_s \in \{0,1\}$ with $\Pr(j_r = 1) = \gamma_r^2/(\gamma_r^2 + \mu_r^2) \equiv p_r$. The calculation proceeds as there, except that we compute $E(|U|^2 \mid U \neq \emptyset)$.

The following sampling scheme can be used to generate a random set U with probability proportional to σ_u^2. For $j \in 1, \ldots, s$ let j be in U with probability $p_j = \gamma_j^2/(\gamma_j^2 + \mu_j^2)$, making all decisions independently. If the resulting U is not empty accept it. Otherwise, keep generating such sets U until the first non-empty one, and accept that one.

5.3. Isotropic functions

A class of isotropic test functions due to Capstick and Keister (1996) and Keister (1996) have been used as quadrature test functions. Papageorgiou and

Traub (1997) report good results for quasi-Monte Carlo on these functions, and Papageorgiou (2001) introduces a radial discrepancy measure for them. Novak, Ritter, Schmitt, and Steinbauer (1997) apply an interpolatory rule to some such functions. The isotropic integrals take the form

$$\begin{aligned}\int_{\mathbb{R}^s} h(\|z\|)e^{-\|z\|^2}dz &= \pi^{s/2}\int_{\mathbb{R}^s}\frac{e^{-\frac{1}{2}\|z\|^2}}{(2\pi)^{s/2}}h(\|z\|/\sqrt{2})dz \\ &= \pi^{s/2}\int_{[0,1]^s} h\left(\sqrt{\frac{1}{2}\sum_{j=1}^{s}[\Phi^{-1}(x^j)]^2}\right)dx,\end{aligned} \tag{15}$$

where $\|x\| = \sqrt{x'x}$ and $\Phi(x)$ is the standard normal distribution function. It is convenient to multiply the s dimensional version of this function by $\pi^{-s/2}$. That is we take the integrand to be

$$f(x) = \pi^{-s/2}h\left(\sqrt{\frac{1}{2}\sum_{j=1}^{s}[\Phi^{-1}(x^j)]^2}\right) \tag{16}$$

on $[0,1]^s$. This scale factor $\pi^{-s/2}$ does not change σ_u^2/σ^2 so it does not change the effective dimension of the integrand.

The value of these radially symmetric integrals can be expressed as univariate integrals with respect to a $\chi^2_{(s)}$ distribution:

$$\int_{[0,1]^s} f(x)dx = \int_0^\infty h(\sqrt{z/2}\,)g_s(z)dz, \tag{17}$$

where $g_s(z) = e^{-z/2}z^{s/2-1}/(\Gamma(s/2)2^{s/2})$ for $0 < z < \infty$ is the chi-squared density function on s degrees of freedom.

The formulas of Sobol' can be translated using chi-squared variables. From (4) we can write $(\underline{\tau}_u^2 + I^2)(f)$ as a triple integral

$$\int\int\int h\Big(\sqrt{(x+y)/2}\Big)h\Big(\sqrt{(x+z)/2}\Big)g_{|u|}(x)g_{s-|u|}(y)g_{s-|u|}(z)dxdydz \tag{18}$$

over $(0,\infty)^3$ and, after some rearrangement,

$$\underline{\tau}_u^2 = \int_0^\infty\left[\int_0^\infty\Big[h\Big(\sqrt{(x+y)/2}\Big) - I(f)\Big]g_{s-|u|}(y)dy\right]^2 g_{|u|}(x)dx. \tag{19}$$

By symmetry σ_u^2 depends on u only through $|u|$. Letting $\kappa_{|u|}^2$ be this common value, we may write

$$\underline{\tau}_u^2 = \sum_{j=1}^{|u|}\binom{|u|}{j}\kappa_j^2, \tag{20}$$

allowing us to solve

$$\kappa_r^2 = \underline{\tau}^{\,2}_{\{1,\dots,r\}} - \sum_{k=1}^{r-1}\binom{r}{k}\kappa_k^2 \tag{21}$$

for increasingly large r, and then to compute

$$\nu(r) = \frac{1}{\sigma^2}\binom{s}{r}\kappa_r^2. \tag{22}$$

If the function $h(\sqrt{z/2})$ used to construct the isotropic function is a low order polynomial in the chi-squared variable z, then f is necessarily of low effective dimension. Such an f cannot have $\sigma_u^2 > 0$ for any u with $|u|$ larger than the degree of the polynomial as shown below.

Proposition 3. *Let $f(x)$ be an isotropic integral of the form* (16) *for which $h(\|x\|/\sqrt{2}) = p(\|x\|^2)$ for a polynomial p of degree n where $0 \le n < s$. Let $u \subseteq \{1,\dots,s\}$ with $|u| > n$. Then $f_u(x) = 0$.*

Proof. It suffices to consider monomials $p(z) = z^n$ because more general polynomials are linear combinations of monomials.

Then the integrand on the unit cube is $f(x) = [\sum_{j=1}^s \Phi^{-1}(x^j)^2]^n$. Expanding the polynomial we find that f is a sum of products of squares of $\Phi^{-1}(x^j)$ for which each product has at most n factors.

The practical import of Proposition 3 is that if $h(\sqrt{z/2})$ is close to a low order polynomial in z, then f has low effective dimension. The natural definition of close is with respect to a weight function such as $\exp(-z)z^{\alpha-1}$ on $[0,\infty)$. The orthogonal polynomials for this weight function are the associated Laguerre polynomials (Szego (1975)). Proposition 3 only uses the degree of the polynomial, so it applies to whatever orthogonal polynomials we might consider.

6. Example Functions

6.1. Hellekalek's example

Hellekalek (1988) describes some numerical computations with the function $f(x) = \prod_{j=1}^s g(x^j)$ where, for $z \in [0,1]$, $g(z) = z^\alpha - 1/(\alpha+1)$. Here z^α denotes the α'th power of the scalar z. For this function $\mu_j = 0$ for all j and $\gamma_j^2 = \alpha^2/[(2\alpha+1)(\alpha+1)^2]$. The value of α ranges from 1 to 3.

This function has $\sigma_u^2 = \sigma^2 1_{u=\{1,\dots,s\}}$. Thus it is fully s dimensional. Its dimension distribution is a pointmass with $\nu(d) = 1_{d=s}$. The variance of f decreases exponentially fast as s increases, so numerical work is perhaps better done on $\prod_{j=1}^s \gamma_j^{-1} \times f$.

For this test function, quasi-Monte Carlo integration did not produce a meaningful improvement over Monte Carlo integration. Hellekalek remarks that this lack of improvement is surprising given results on discrepancies of quasi-Monte Carlo points. It is less surprising from an effective dimension view. The integrand is of full dimension s and the (t, m, s)-nets applied do not balance any elementary intervals of dimension s, unless $n \geq b^{t+s}$.

6.2. Sobol's examples

Sobol' (2001) considers product functions in which $g_j(x^j) = (|4x^j - 2| + a_j)/(1 + a_j)$ for a parameter a_j. Elementary calculus gives $\mu_j = 1$ and $\gamma_j^2 = 1/[3(1 + a_j)^2]$. Sobol' remarks that for $a_j = 0$ the value of g_j varies from 0 to 2 and is important, whereas for $a_j = 3$ the value of g_j varies from 0.75 to 1.25 and is unimportant.

Sobol' considers $s = 8$ and considers a function in which the first two variables are important with $a_1 = a_2 = 0$, and the final six are not important, having $a_j = 3$ for $j \geq 3$. It is simple to compute σ_u^2 for all 255 nonempty subsets of $\{1, \ldots, 8\}$.

Table 1. ANOVA probabilities of $|U|$ and $\max(U)$ for Sobol's example with two important and six unimportant variables.

d	$\nu(d)$	$\nu_{\mathrm{T}}(d)$
1	9.45×10^{-1}	0.467
2	5.44×10^{-2}	0.519
3	1.38×10^{-4}	0.00225
4	1.48×10^{-7}	0.00225
5	8.54×10^{-11}	0.00226
6	2.77×10^{-14}	0.00226
7	4.80×10^{-18}	0.00226
8	3.47×10^{-22}	0.00226

Table 1 shows the entire dimension distribution for Sobol's example in both senses. The function is nearly a superposition of univariate functions because $\nu(1) = 0.945$. Similarly the ANOVA contribution of variables taken three or more at a time accounts for only 1.38×10^{-4} of the variance of f. The mean dimension of Sobol's example function is 1.055 in the superposition sense. The mean dimension in the truncation sense is 1.580.

Sobol' (1994) investigates the function $f(x) = \prod_{j=1}^{s}(j + 2x^j)/(j + 1)$. Here $\mu_j = 1$ and $\gamma_j^2 = 1/[3(j + 1)^2]$. Each input variable x^{j+1} is less important than x^j. For dimension $s = 100$, numerical computations based on (13) and (14) give a mean dimension of 1.085 and a mean square dimension of 1.263.

Chebychev's inequality for a random variable Y with finite variance is $\Pr(|Y - E(Y)| > \alpha\sqrt{\mathrm{Var}\,(Y)}) \leq 1/\alpha^2$ for $\alpha > 0$. Applied to the dimension distribution of f with $\alpha = 13$ we find that $\Pr(|U| > 1.085 + 13 \times (1.263 - 1.085^2)^{1/2} \le 1/13^2$ so that $\Pr(|U| > 4.89) \le 0.00592$. For $s = 100$, at least 99.418% of the variation in this function is from its ANOVA components of dimension 4 and smaller.

6.3. Owen's examples

Owen (1997b) and Owen (1998b) include a discussion of functions $f(x) = 12^{s/2} \prod_{j=1}^{s} (X^j - 1/2)$. This function has $I = 0$, and $\sigma^2 = 1$ and its dimension distribution has $\nu(s) = 1$.

Owen (1997b) studied scrambled $(0, s)$-sequences in base b. For $1 \le \lambda < b$ and $m \ge 0$, the first $n = \lambda b^m$ points $X_1, \ldots, X_n$ of such a sequence comprise a $(\lambda, 0, m, s)$-net in base b, and the scrambled net estimate of I is $\hat{I}_n = (1/n) \sum_{i=1}^{n} f(X_i)$. This estimate has $E(\hat{I}_n) = I$ and $\mathrm{Var}\,(\hat{I}_n) = O(n^{-3} \log(n)^{(s-1)})$ along the sequence $n = \lambda b^m \to \infty$. For f above, $\mathrm{Var}\,(\hat{I}_n)$ can be computed exactly. Owen (1998b) develops the approximation $\mathrm{Var}\,(\hat{I}_n) \doteq 1/n$ for $n \le b^s$ and

$$\mathrm{Var}\,(\hat{I}_n) \doteq \frac{(\log n)^{s-1}}{n^3} \frac{\lambda^2}{(s-1)!} \left(\frac{b^2 - 1}{\log b} \right)^{s-1} \tag{23}$$

for $n = \lambda b^m \ge b^s$. These test functions are fully s dimensional and the improvement of QMC over MC sets in at $n \ge b^s$. For further discussion see Owen (1997b).

6.4. Roos and Arnold's examples

Roos and Arnold (1963) studied integrands: $f_1(X) = (1/s) \sum_{j=1}^{s} |4X^j - 2|$, $f_2(X) = \prod_{j=1}^{s} |4X^j - 2|$, and $f_3(X) = \prod_{j=1}^{s} (\pi/2) \sin(\pi X^j)$. The function f_1 is additive, so it is purely one dimensional. It has $\sigma^2_{\{j\}} = 1/3$ and $\nu(d) = 1_{d=1}$.

The functions f_2 and f_3 both have product form, and both have values $\mu_j = \mu$ and $\gamma_j = \gamma$ independent of j. For such functions $\sigma_u^2 = \gamma^{2|u|} \mu^{2(s-|u|)}$ and $\sigma^2 = (\mu^2 + \sigma^2)^s - \mu^{2s}$. Both f_2 and f_3 have $\mu = 1$, so that $\sigma_u^2 = \gamma^{2|u|}$ and $\sigma^2 = (1 + \sigma^2)^s - 1$. The mean dimension s, $\gamma^2 (1 + \gamma^2)^{s-1} / [(1 + \gamma^2)^s - 1]$, is approximately $s\gamma^2/(1 + \gamma^2)$ for large s. For f_2 we find $\gamma^2 = 1/3$ while for f_3 we find $\gamma^2 = \pi^2/8 - 1 \doteq 0.2337$.

Test problems f_2 and f_3 have mean dimensionality that grows nearly linearly with s. The argument in the proof of Proposition 1 shows that the dimension distribution for f_2 is the binomial distribution with parameters s and $\gamma^2/(1+\gamma^2)$, conditioned to be nonzero.

Table 2 shows the dimension distribution of f_2 and f_3 when $s = 10$. The function f_2 has little structure beyond dimension 6 and f_3 has little beyond dimension 5.

Table 2. The dimension distribution $\nu(d)$ is shown for two of Roos and Arnold's example functions: f_2 and f_3 with $s = 10$.

d	f_2	f_3
1	0.199	0.326
2	0.298	0.343
3	0.265	0.214
4	0.155	0.0874
5	0.0619	0.0245
6	0.0172	0.00477
7	0.00327	0.000637
8	0.000409	0.0000559
9	0.0000303	0.0000029
10	0.00000101	0.0000000678

6.5. Genz's examples

Widely used families of test functions were proposed by Genz (1984). All except the first and third functions are of product form:

$$f_2(X) = \prod_{j=1}^{s} \Big(a_j^{-2} + (X^j - u_j)^2 \Big) \qquad \text{(Product Peak);}$$

$$f_4(X) = \exp\Big(- \sum_{j=1}^{s} a_j^2 (X^j - u^j)^2 \Big) \qquad \text{(Gaussian);}$$

$$f_5(X) = \exp\Big(- \sum_{j=1}^{s} a_j |X^j - u^j| \Big) \qquad (C_0);$$

$$f_6(X) = \exp\Big(- \sum_{j=1}^{s} a_j X^j \Big) 1_{X^1 > u_1} 1_{X^2 > u_2} \qquad \text{(Discontinuous).}$$

The values u_j and a_j are parameters providing different members of each family of functions. To generate test functions the u_j are sampled from the $U(0, 1)$ distribution. The a_j are positive variables also generated at random. It is thought that larger values of a_j will make the problem harder.

For product functions the effective dimension tends to be higher for larger values of $p_j = \gamma_j^2/(\gamma_j^2 + \mu_j^2)$. If the p_j are all near 0 then the effective dimension is nearly 1 and conversely if the p_j are all near 1 the effective dimension is nearly s.

For f_4, f_5, and f_6 as $a_j \to \infty$ we find $p_j \to 1$. Thus if all of the a_j are large enough then these functions are of essentially full dimension s. For f_2, as $a_j \to \infty$ we find using symbolic computation in Mathematica that

$$p_j \to \frac{4 - 15\, u_j + 15\, u_j^2}{9 - 45\, u_j + 90\, u_j^2 - 90\, u_j^3 + 45\, u_j^4} \in \left[\frac{4}{9}, \frac{5}{9} \right],$$

so the hard cases for f_2 are only of approximate dimension $s/2$.

Suppose that one wants to test quadrature rules on a set of problems of varying nominal dimension s with nearly constant mean dimension D. Then one can arrange for p_j to be close to D/s. For functions f_4 and f_5 (and also f_2 if $D/s \le 4/9$) one can sample $u_j \sim U(0,1)$, then solve $\gamma_j^2/(\gamma_j^2+\mu_j^2) = D/s$ for a_j.

For f_6 the variables u_1 and u_2 affect how spiky the function is. Because $\lim_{a_1\to 0} p_1 = 1-u_1$ the parameter u_1 (respectively u_2) makes a small difference to the effective dimension of f_6 when a_1 (respectively a_2) is small. But for large values of a_1 the value of u_1 makes little difference to p_1.

6.6. Keister's examples

Capstick and Keister (1996) consider the isotropic integral with $h(z) = \cos(z)$. In this case, the integral (15) can be expressed in terms of some special functions and evaluated by Mathematica (Wolfram (1999)). Papageorgiou and Traub (1997) consider the case $s = 25$. For $s = 25$, Mathematica computes the integral of (15) to be approximately $-1.3569140978979188\times 10^6$ and, after scaling by $\pi^{s/2}$, we find that $I \doteq -0.82828337794550358$ and $\sigma^2 \doteq 0.048330759722972670$.

If $Z \sim \chi^2_{(25)}$ then $3.97 \le Z \le 75.87$ with probability greater than $1-10^{-6}$, and $1.40 \le \sqrt{Z/2} \le 6.16$, also with probability greater than $1-10^{-6}$. The cosine function does not go through even one full period between 1.4 and 6.16 and so it is quite nearly a low order polynomial over this range. In view of Proposition 3 we might therefore expect low effective dimension for this problem.

The value of $\underline{\tau}_u^2$ for $|u| = 1,\dots,5$ was computed several ways. Because (19) has a simple iterated form the first methods tried were midpoint rules, using n equispaced quantiles of the $\chi^2_{s-|u|}$ distribution for the inner integral and n equispaced quantiles of the $\chi^2_{|u|}$ distribution for the outer integration. The computational cost grows as n^2. Varying n, the error appeared to be $O(1/n)$, probably due to the oscillatory nature of the integrand. Table 3 shows values computed with $n = 100,000$ evaluations in each of the inner and outer integrals.

Table 3 also shows values obtained by using a scrambled $(0,14,3)$-net in base 3 (having $n = 3^{14} = 4,782,969$ evaluations) on the three dimensional integral representation (18). Finally Mathematica with (19) and 25 requested digits of accuracy was used.

All three methods of computing $\underline{\tau}_u^2$ agree closely for this function. The methods differ when it comes to estimating κ_r^2 particularly for larger values of r. The reason is that for larger r, the binomial coefficients in (21) and (22) become larger. As a consequence, only the first few κ_j^2 could be computed accurately.

Table 4 shows estimates of κ_r^2, $r = 1,\dots,5$, from the three methods used in Table 3. The table includes some negative values, as might be expected when small quantities are estimated as differences of larger quantities subject to numerical error. The values from Mathematica were computed with 25 digits of accuracy requested. Because it is not certain that this accuracy was attained, a comparison with the scrambled net answers is made below. That comparison leaves little doubt about the value of $\nu(d)$ for d from 1 to 3.

Table 3. Shown are estimates of $100 \times \underline{\tau}_u^2$ for the isotropic function (17) with $h(\cdot) = \cos(\cdot)$ from three methods described in the text. The methods are an iterated midpoint rule, a scrambled net, and a combination symbolic-numeric calculation done with Mathematica.

$10^3 \underline{\tau}_u^2$	1	2	3	4	5
Mid	1.36964	2.78103	4.23437	5.73039	7.26980
Net	1.37021	2.78166	4.23504	5.73108	7.27051
Math.	1.37026	2.78172	4.23510	5.73114	7.27057

Table 4. Shown are estimates of $\kappa_j^2 = \sigma_u^2$ for $|u| = j$, for the isotropic function (17) with $h(\cdot) = \cos(\cdot)$ as computed by methods described in the text.

	$10^3\kappa_1^2$	$10^5\kappa_2^2$	$10^7\kappa_3^2$	κ_4^2	κ_5^2
Mid	1.36964	4.17463	2.07260	5.11450E−7	−5.00549E−7
Net	1.37021	4.12332	6.98281	3.12985E−8	−2.71509E−8
Math.	1.37026	4.11968	7.28342	3.13553E−9	6.51294E−12

Table 5 shows estimates of the dimension distribution of f for $s = 25$ based on the values computed by Mathematica. Based on these estimates, over 99% of the variance in f comes from ANOVA components of dimension 3 or less. So this function is of effective dimension 3 using the definition of Caflisch, Morokoff, and Owen (1997). Only about 3×10^{-8} of the variance is estimated to come from components of dimension 6 or more.

The scrambled net computations of κ_j^2 in Table 4 match the Mathematica ones reasonably well for $j \le 3$, confirming that $\nu([1, 3])$ is very close to 1. It is also reasonable, though not certain, that most of the remainder of the variance of f comes from ANOVA components of dimension 4, but it is hard to be sure that $\nu(5) \doteq 7.16 \times 10^{-6}$, because a small relative error in $\nu(4)$ could imply a large one in $\nu(5)$.

Table 5 also contains values for dimension $s = 80$. Once again it appears that most of the variance of f comes from components of dimension 3 or less. In this instance however most of the variance comes from components of dimension 2 instead of dimension 1. In both of these examples the low effective dimension of the integrand explains why the quasi-Monte Carlo methods are effective.

Table 5. Shown are the values $\nu(r)$ representing the fraction of the variance of the isotropic function described in the text, that is of dimension r for $r = 1, \ldots, 5$. The final column gives the fraction of variance for dimensions higher than 5. The examples here have nominal dimensions $s = 25$ and $s = 80$. The values here were computed using Mathematica with 25 digits of accuracy requested.

s	$\nu(1)$	$\nu(2)$	$\nu(3)$	$\nu(4)$	$\nu(5)$	$\nu([6, s])$
25	0.7088	0.2557	0.03466	0.0008207	7.160E-6	2.999E-8
80	0.06250	0.9107	0.02554	0.001012	0.0002258	7.825E-6

7. Discussion

Effective dimension provides one method for describing the difficulty of quadrature problems. The ANOVA distribution allows us to compute aspects of the effective dimension for test problems commonly used in quadrature.

Low effective dimension for f is not sufficient to imply that QMC will be much better than MC, as easily constructed "spiky" test functions show. The low dimensional parts f_u must also be such that QMC rules will work well on them. This consideration appears in the bound (9) through the norms applied to f_u.

Acknowledgements

I thank Shu Tezuka and Anargyros Papageorgiou for helpful discussions on effective dimension and isotropic functions. This work was supported by NSF grant DMS-0072445.

References

Caflisch, R. E., Morokoff, W. and Owen, A. B. (1997). Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension, *J. Comput. Finance* **1**, 27-46.

Capstick, S. and Keister, B. D. (1996). *J. Comput. Phys.* **123**, 267-273.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance, *Ann. Statist.* **9**, 586-596.

Genz, A. (1984). Testing multidimensional integration routines, In Tools, Methods and Languages for Scientific and Engineering Computation, North-Holland, Amsterdam, 81-94.

Hellekalek, P. (1998). On the assessment of random and quasi-random point sets, In *Random and Quasi-Random Point Sets* (Edited by P. Hellekalek and G. Larcher), 49-108. Springer, New York.

Hickernell, F. J. (1996). Quadrature error bounds and figures of merit for quasi-random points, *SIAM J. Numer. Anal.* **33**, 1995-2016. Corrected printing of Sections 3-6 in ibid. **34** (1997), 853-866.

Hickernell, F. J. (1998). Lattice rules: how well do they measure up? In *Random and Quasi-Random Point Sets* (Edited by P. Hellekalek and G. Larcher), 109-168. Springer, New York.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.* **19**, 293-325.

Keister, B. D. (1996). Multidimensional quadrature algorithms, *Comput. Phys.* **10**, 119-122.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*, S.I.A.M., Philadelphia, PA.

Novak, E., Ritter, K., Schmitt, R. and Steinbauer, A. (1997). On a recent interpolatory method for high dimensional integration, Technical report, University of Erlangen.

Owen, A. B. (1997a). Monte Carlo variance of scrambled equidistribution quadrature, *SIAM J. Numer. Anal.* **34**, 1884-1910.

Owen, A. B. (1997b). Scrambled net variance for integrals of smooth functions, *Ann. Statist.* **25**, 1541-1562.

Owen, A. B. (1998a). Latin supercube sampling for very high dimensional simulations, *ACM Trans. Modeling and Computer Simulation* **8**, 71-102.

Owen, A. B. (1998b). Scrambling Sobol' and Niederreiter-Xing points, *J. Complexity* **14**, 466-489.

Papageorgiou, A. (2001). Fast convergence of quasi-Monte Carlo for a class of isotropic integrals, *Math. Comput.* **70**, 297-306.

Papageorgiou, A. and Traub, J. (1997). Faster evaluation of multidimensional integrals, *Comput. Phys.*, 574-578.

Paskov, S. and Traub, J. (1995). Faster valuation of financial derivatives, *J. Portfolio Management* **22**, 113-120.

Roos, P. and Arnold, L. (1963). Numerische experimente zur mehrdimensionalen quadratur, *Österreich. Akad. Wiss. Math.-Natur.Kl. S.-B. II* **172**, 271-286.

Sloan, I. H. and Wozniakowski, H. (1998). When are quasi-Monte Carlo algorithms efficient for high dimensional integration? *J. Complexity* **14**, 1-33.

Sobol', I. M. (1969). *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow (In Russian).

Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Model. and Comput. Experiment* **1**, 407-414.

Sobol', I. M. (1994). *A Primer for the Monte Carlo Method*, CRC Press.

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulation* **55**, 271-280.

Sobol', I. M. and Levitan, Y. L. (1999). On the use of variance reducing multipliers in Monte Carlo computations of a global sensitivity index, *Comput. Phys. Comm.* **117**, 52-61.

Szego, G. (1975). *Orthogonal Polynomials*. 4th edition. American Mathematics Society, Providence RI.

Takemura, A. (1983). Tensor analysis of ANOVA decomposition, *J. Amer. Statist. Assoc.* **78**, 894-900.

Tezuka, S. (2002). Quasi-Monte Carlo: discrepancy between theory and practice, In *Monte Carlo and Quasi-Monte Carlo Methods* 2000 (Edited by K. T. Fang, F. J. Hickernell and H. Niederreiter), 124-140. Springer-Verlag, Berlin.

Wolfram, S. (1999). *The Mathematica Book*. 4th edition. Cambridge University Press.

Department of Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305, U.S.A.

E-mail: owen@stat.stanford.edu

(Received January 2002; accepted July 2002)