Quasi-regression and the relative importance of the ANOVA components of a function

Christiane Lemieux[1] and Art B. Owen[2]

[1] Department of Mathematics and Statistics, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, T2N 1N4, CANADA
[2] Department of Statistics, Stanford University, Sequoia Hall, 390 Serra Mall, Stanford, CA, 94305, U.S.A.

Abstract. In this paper we use quasi-regression to study high dimensional integrands. The variance of an integrand can be expressed as an infinite sum of squared components of an orthogonal basis. Various sums over subsets of these components have meaningful interpretations, and we develop numerical estimates for such sums. We pay particular attention to certain sums related to the ANOVA components of the integrand, because these are related to the effectiveness of quasi-Monte Carlo integration methods. We find that randomized quasi-Monte Carlo methods can estimate the unknown coefficients more accurately than ordinary Monte Carlo. We illustrate the method on two problems: valuing an Asian option, and finding the expected completion time in a stochastic activity network.

1 Introduction

Let f be a function defined over the unit hypercube $[0,1)^s$. When the dimension s is large, it is well known that it can be extremely difficult to find the integral of f. Yet very accurate results have been reported for some integration problems in large dimensions [5,14,19].

When a method succeeds on a high dimensional integrand, one might suspect that the integrand has some special property that was exploited by the method. One explanation offered is that the *effective dimension* [5,19] of f can be much less than the nominal dimension s. Quasi-Monte Carlo (QMC) methods often have very good equidistribution properties in their low dimensional projections. Such methods are well suited to integrands of low effective dimension. For example Owen [16] shows that in scrambled net quadrature certain coarse low dimensional features are integrated without error, while features that are either of higher dimension or more spiky add a quadrature error similar to that seen in ordinary Monte Carlo.

Caflisch et al. [5] define the effective dimension of f in terms of its *ANOVA decomposition* [8,12,17]. An ANOVA decomposition allows one to determine the relative importance of different subsets of input variables for an integrand. This knowledge can be used to select or tune a QMC method to a given problem, or to a class of similar problems.

In this paper we employ an infinite orthogonal decomposition of square integrable functions over $[0,1)^s$. Sums of squared components in this decomposition can be used to describe the importance of various aspects of f, including the ANOVA components. In practice, lower bounds on these contributions are computed because the ANOVA component magnitude is defined through an infinite sum of squares that we truncate.

The remainder of this paper is organized as follows. Section 2 defines the ANOVA decomposition of a function, quantifies the relative importance of each component, and connects this relative importance to some terms in an orthonormal basis. Section 3 describes our numerical method for estimating relative importance from sample values. We present both a Monte Carlo and a randomized QMC approach. Section 4 contains estimation results for two examples. For an Asian option the results show what changing the strike price or switching to a Brownian bridge representation of the underlying price process can do to the amount of low dimensional structure. For a stochastic activity network, we see how conditional simulation reduces effective dimension. Section 5 discusses related methods and ideas for further research.

2 ANOVA decomposition and orthogonal bases

We assume that $\int_{[0,1)^s} f^2(\boldsymbol{u})d\boldsymbol{u} < \infty$. Then f has an ANOVA decomposition [8,12,17] written as a sum of 2^s components, with one component per subset $I \subseteq \{1,\dots,s\}$. For $\boldsymbol{u} \in [0,1)^s$ write $f(\boldsymbol{u}) = \sum_{I\subseteq\{1,\dots,s\}} f_I(\boldsymbol{u})$, where

$$f_I(\boldsymbol{u}) := \int_{[0,1)^s} f(\boldsymbol{u})d\boldsymbol{u}_{I^c} - \sum_{J\subsetneq I} f_J(\boldsymbol{u}), \tag{1}$$

and $\boldsymbol{u}_{I^c} = (u_j)_{j\notin I}$ represents the vector of variables u_j whose index j is not in I. One can show that for any nonempty subset I, $\int_{[0,1)^s} f_I(\boldsymbol{u})du_j = 0$ for any $j \in I$, and $\int_{[0,1)^s} f_\emptyset(\boldsymbol{u})d\boldsymbol{u} = \mu$, where

$$\mu = \int_{[0,1)^s} f(\boldsymbol{u})d\boldsymbol{u}.$$

The ANOVA decomposition is orthogonal, i.e., for two subsets $I \neq J \subseteq \{1,\dots,s\}$ one has

$$\int_{[0,1)^s} f_I(\boldsymbol{u})f_J(\boldsymbol{u})d\boldsymbol{u} = 0. \tag{2}$$

From (2) we can decompose the variance of f via

$$\sigma^2 := \int_{[0,1)^s} f^2(\boldsymbol{u})d\boldsymbol{u} - \mu^2 = \sum_{I\subseteq\{1,\dots,s\}} \sigma_I^2,$$

where $\sigma_I^2 = \int_{[0,1)^s} f_I^2(\boldsymbol{u})d\boldsymbol{u}$ for $I \neq \emptyset$, and $\sigma_\emptyset^2 = 0$.

For any component f_I, the quantity σ_I^2/σ^2 represents the fraction of the variance of f that is due to f_I and we take this as a measure of the relative importance of f_I. We would like to estimate σ_I^2/σ^2 for different components f_I. Since the variance σ^2 is comparatively easy to estimate using standard procedures, the main task is to estimate the σ_I^2.

As in [1] we introduce basis functions obtained by taking tensor products of a complete orthonormal univariate basis. We write our univariate basis $\{\phi_j(u) \mid j = 0, 1, 2, \ldots\}$, and suppose that $\phi_0(u) \equiv 1$ for all $u \in [0, 1)$. Orthonormality implies that $\int_0^1 \phi_j(u)\phi_k(u)du = 1_{j=k}$.

Our numerical examples are all done using orthogonal polynomials, although the methods extend naturally to other bases. The orthogonal polynomials over $[0, 1)$ are simply rescaled versions of the Legendre polynomials [9] defined on $[-1, 1]$, so that our $\phi_j(u)$ is the j'th Legendre polynomial evaluated at $2u - 1$ [1], and then scaled to have unit norm.

To decompose functions defined over $[0, 1)^s$ we use multivariate basis functions defined as tensor products of ϕ_j. For $\boldsymbol{r} = (r_1, \ldots, r_s)$, where the r_j are non-negative integers, define

$$\phi_{\boldsymbol{r}}(\boldsymbol{u}) = \prod_{j=1}^{s} \phi_{r_j}(u_j).$$

We can then decompose f using the basis $\{\phi_{\boldsymbol{r}}(\boldsymbol{u}) \mid \boldsymbol{r} = (r_1, \ldots, r_s), r_j \geq 0, j = 1, \ldots, s\}$, getting

$$f(\boldsymbol{u}) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_s=0}^{\infty} \beta_{\boldsymbol{r}}\, \phi_{\boldsymbol{r}}(\boldsymbol{u}), \tag{3}$$

where the coefficients $\beta_{\boldsymbol{r}}$ in this expansion are obtained via

$$\beta_{\boldsymbol{r}} = \int_{[0,1)^s} f(\boldsymbol{u})\phi_r(\boldsymbol{u})d\boldsymbol{u}. \tag{4}$$

The following result explains the correspondence between the ANOVA decomposition of f and the one expressed in (3):

Proposition 1. *Let $f : [0, 1)^s \to \mathbb{R}$ be a square-integrable function. For each subset $I \subseteq \{1, \ldots, s\}$, define the component $f_I(\cdot)$ as in (1). Let the coefficients $\beta_{\boldsymbol{r}}$ be defined as in (4). Then for I nonempty, one has*

$$\sigma_I^2 = \sum_{\boldsymbol{r} \in R(I)} \beta_{\boldsymbol{r}}^2, \tag{5}$$

where the set $R(I)$ contains all vectors $\boldsymbol{r} = (r_1, \ldots, r_s)$ whose entries are non-negative integers satisfying $r_j > 0$ if and only if $j \in I$.

This result shows that σ_I^2 can be estimated indirectly by estimating the coefficients $\beta_{\boldsymbol{r}}^2$. In practice, the infinite sum on the right-hand side of (5) must be truncated, which means we then estimate *lower bounds* on σ_I^2. An obvious disadvantage to our approach is that the terms in the sum might decay slowly. If f_I is very spiky then the lower bound won't be close to σ_I^2. A mitigating advantage is that we can sum other subsets of $\beta_{\boldsymbol{r}}^2$ values to learn something about the relative importance of high and low frequency components of f or of f_I.

3 Description of the estimators

To estimate σ_I^2 (or more precisely, a lower bound on σ_I^2), we truncate the sum of values $\beta_{\boldsymbol{r}}^2$ on the right-hand side of (5). For nonnegative integers d and $m \le d$ define

$$\mathcal{R}(I) = \mathcal{R}(I,d,m) := \{\boldsymbol{r} \in R(I) \mid \sum_{j=1}^{s} r_j \le d, r_j \le m \text{ for each } j \in I\}. \quad (6)$$

The truncated version of σ_I^2 is

$$\sigma_{I,d,m}^2 := \sum_{\boldsymbol{r} \in R(I,d,m)} \beta_{\boldsymbol{r}}^2.$$

The parameters d and m are called the *degree* and *order* respectively in [1]. When, as here, ϕ_j is a univariate polynomial of degree j then $\phi_{\boldsymbol{r}}$ is an s-dimensional polynomial of degree $\sum_{j=1}^{s} r_j$. The polynomials with indices in $\mathcal{R}(I,d,m)$ have degree at most d, and are products of $|I|$ nonconstant univariate polynomials each having degree at most m.

To estimate a coefficient $\beta_{\boldsymbol{r}}^2$, we construct an estimator for $\beta_{\boldsymbol{r}}$, take its square, and then apply a bias correction described below. Since the estimation of the coefficients $\beta_{\boldsymbol{r}}$ amounts to an integration problem, either a Monte Carlo or QMC approach can be used to carry out the approximation.

The quantity

$$\gamma(p) := \frac{1}{\sigma^2} \sum_{I:|I| \le p} \sigma_I^2 \quad (7)$$

describes the fraction of variance in f attributable to ANOVA components of dimension p and lower.

If for a threshold $\pi \in [0,1]$, p is the smallest integer with $\gamma(p) \in [\pi, 1]$, then the *effective dimension* of f (in the superposition sense [5,10]) is p. Having a lower bound on $\gamma(p) \ge \pi$ implies that f has an effective dimension of *at most* p. The definition depends on the threshold, which is usually taken at a high value like 0.99.

3.1 Monte Carlo estimators

First we construct an unbiased estimator for $\beta_{\boldsymbol{r}}^2$. Assume that $\boldsymbol{u}_k$ for $k = 1, \ldots, n$ are i.i.d. uniform random points in $[0,1)^s$. The quasi-regression estimator [1]

$$\hat{\beta}_{\boldsymbol{r}}^2 = \left(\frac{1}{n} \sum_{k=1}^{n} f(\boldsymbol{u}_k) \phi_{\boldsymbol{r}}(\boldsymbol{u}_k) \right)^2$$

is biased, because

$$E(\hat{\beta}_{\boldsymbol{r}}^2) = \beta_{\boldsymbol{r}}^2 + \frac{1}{n} \mathrm{Var}(f(\boldsymbol{u}_k)\phi_{\boldsymbol{r}}(\boldsymbol{u}_k)) \geq \beta_{\boldsymbol{r}}^2 \tag{8}$$

with strict inequality outside of trivial cases. Rewriting (8) as

$$E(\hat{\beta}_{\boldsymbol{r}}^2) = \beta_{\boldsymbol{r}}^2 + \frac{1}{n} [E(f^2(\boldsymbol{u}_k)\phi_{\boldsymbol{r}}^2(\boldsymbol{u}_k)) - \beta_{\boldsymbol{r}}^2] \tag{9}$$

allows us to derive the unbiased estimator

$$\hat{\beta}_{\boldsymbol{r},\mathrm{bc}}^2 = \frac{n}{n-1} \left(\hat{\beta}_{\boldsymbol{r}}^2 - \frac{1}{n^2} \sum_{k=1}^{n} \phi_{\boldsymbol{r}}^2(\boldsymbol{u}_k) f^2(\boldsymbol{u}_k) \right),$$

where the subscript "bc" stands for "bias-corrected".

Using $\hat{\beta}_{\boldsymbol{r},\mathrm{bc}}^2$, estimate $\sigma_{I,d,m}^2 = \sum_{\boldsymbol{r} \in \mathcal{R}(I,d,m)} \beta_{\boldsymbol{r}}^2 \leq \sigma_I^2$ without bias as follows:

$$\begin{aligned} \hat{\sigma}_{I,d,m}^2 &= \sum_{\boldsymbol{r} \in \mathcal{R}(I,d,m)} \hat{\beta}_{\boldsymbol{r},\mathrm{bc}}^2 \\ &= \frac{n}{n-1} \left(\sum_{\boldsymbol{r} \in \mathcal{R}(I,d,m)} \hat{\beta}_{\boldsymbol{r}}^2 - \frac{1}{n^2} \sum_{k=1}^{n} S_{I,d,m}^2(\boldsymbol{u}_k) f^2(\boldsymbol{u}_k) \right), \end{aligned} \tag{10}$$

where

$$S_{I,d,m}^2(\boldsymbol{u}) = \sum_{\boldsymbol{r} \in \mathcal{R}(I,d,m)} \phi_{\boldsymbol{r}}^2(\boldsymbol{u}).$$

It is now possible to construct an approximation for the lower bound

$$\gamma_{d,m}(p) = \frac{1}{\sigma^2} \sum_{I:|I| \leq p} \sigma_{I,d,m}^2$$

on the quantity $\gamma(p)$ defined in (7). To do that, we first introduce the function $\mathcal{S}_{p,d,m}^2(\boldsymbol{u}) := \sum_{I:|I| \leq p} S_{I,d,m}^2(\boldsymbol{u})$, which represents the sum of the functions $S_{I,d,m}^2(\boldsymbol{u})$ included in the approximation for $\gamma_{d,m}(p)$. We then define

$$\begin{aligned} \hat{\gamma}_{d,m}(p) &= \frac{1}{\hat{\sigma}^2} \sum_{I:|I| \leq p} \hat{\sigma}_{I,d,m}^2, \\ &= \frac{n}{\hat{\sigma}^2(n-1)} \left(\sum_{I:|I| \leq p} \sum_{\boldsymbol{r} \in \mathcal{R}(I,d,m)} \hat{\beta}_{\boldsymbol{r}}^2 - \frac{1}{n^2} \sum_{k=1}^{n} \mathcal{S}_{p,d,m}^2(\boldsymbol{u}_k) f^2(\boldsymbol{u}_k) \right), \end{aligned}$$

where $\hat{\sigma}^2 = \left[\sum_{k=1}^n (f(\boldsymbol{u}_k) - \hat{\mu})^2\right]/(n-1)$, and $\hat{\mu} = \sum_{k=1}^n f(\boldsymbol{u}_k)/n$. Note that $\hat{\gamma}_{d,m}(p)$ is a biased but consistent estimator of $\gamma_{d,m}(p)$.

Since the quantities σ^2 and σ_I^2 do not depend on $\mu = \int_{[0,1)^s} f(\boldsymbol{u})d\boldsymbol{u}$, one could replace f by $\tilde{f} = f - c$ for some constant c and apply the above estimation procedure to $\tilde{f}$ instead of f. As mentioned in [18], this can improve the quality of estimators whose performance depends on $\beta_{\mathbf{0}} = \mu$; a detailed analysis of the variance of $\hat{\sigma}^2_{I,d,m}$ can show that this is indeed the case here. A natural candidate for c is to take $c = \mu$ [18], which in practice means that we take $c = \hat{\mu}$. In Section 4, we refer to this alternative approximation method as the *centered version* of the estimators $\hat{\sigma}^2_{I,d,m}$, and $\hat{\gamma}_{d,m}(p)$.

Lastly, to estimate the accuracy of the estimators $\hat{\sigma}^2_{I,d,m}$, and $\hat{\gamma}_{d,m}(p)$, a simple solution is to generate R i.i.d. copies of each of them using R random point sets $\{\boldsymbol{u}_{k,r}, k = 1, \ldots, n\}$, $r = 1, \ldots, R$, and then to use the sample variance within this set of R i.i.d. estimators to provide approximate confidence intervals on the true values these estimators try to approximate.

3.2 Randomized quasi-Monte Carlo estimators

Here we replace the Monte Carlo points $\boldsymbol{u}_1$ to $\boldsymbol{u}_n$ by randomized QMC points (see [6,17] for examples) $\boldsymbol{v}_k = \chi(\boldsymbol{u}_k, \Delta)$, $k = 1, \ldots, n$, where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \subsetneq [0,1)^s$ are n equidistributed points [15], and χ is a randomization function that uses a random vector Δ to transform each $\boldsymbol{u}_k$ into a uniformly distributed vector $\boldsymbol{v}_k$ in $[0,1)^s$. For example, a *randomly shifted lattice rule* [6,14,21] is used in Section 4. In this case, $\boldsymbol{u}_1, \ldots \boldsymbol{u}_n$ are found by intersecting an *integration lattice* (i.e., a discrete subset of $\mathbb{R}^s$ containing $\mathbb{Z}^s$ and closed under addition and subtraction) with $[0,1)^s$, the vector Δ is randomly and uniformly generated over $[0,1)^s$, and $\chi(\boldsymbol{u}_k, \Delta) = (\boldsymbol{u}_k + \Delta) \bmod 1$. One way to construct an integration lattice is to choose an integer a relatively prime to n and set $\boldsymbol{u}_k = (k-1)(1, a, \ldots, a^{s-1})/n \bmod 1$, for $k = 1, \ldots, n$; this special case is called a *Korobov rule* [13].

To construct unbiased estimators for $\sigma^2_{I,d,m}$ and $\gamma_{d,m}(p)$, we use Q independent randomized QMC rules with points $\boldsymbol{v}_{k,q} = \chi(\boldsymbol{u}_k, \Delta_q), k = 1, \ldots, n$, for $q = 1, \ldots, Q$, where the random vectors Δ_q are independent. Based on these Q randomized rules we construct Q unbiased estimators

$$\hat{\beta}_{\boldsymbol{r},\mathrm{QM,q}} = \frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{v}_{k,q})\phi_{\boldsymbol{r}}(\boldsymbol{v}_{k,q})$$

for $\beta_{\boldsymbol{r}}$. These estimators are combined into the following estimator of $\beta^2_{\boldsymbol{r}}$:

$$\hat{\beta}^2_{\boldsymbol{r},\mathrm{QM}} = \frac{Q}{Q-1}(\bar{\beta}_{\boldsymbol{r},\mathrm{QM}})^2 - \frac{1}{Q(Q-1)}\sum_{q=1}^{Q}(\hat{\beta}_{\boldsymbol{r},\mathrm{QM,q}})^2, \tag{11}$$

where $\bar{\beta}_{\boldsymbol{r},\mathrm{QM}} = \left(\sum_{q=1}^{Q}\hat{\beta}_{\boldsymbol{r},\mathrm{QM,q}}\right)/Q$. It can be shown that $\hat{\beta}^2_{\boldsymbol{r},\mathrm{QM}}$ is an unbiased estimator of $\beta^2_{\boldsymbol{r}}$.

Using $\hat{\beta}^2_{\boldsymbol{r},\mathrm{QM}}$, we now define an unbiased estimator ${}_{\mathrm{QM}}\hat{\sigma}^2_{I,d,m}$ for $\sigma^2_{I,d,m}$:

$$ {}_{\mathrm{QM}}\hat{\sigma}^2_{I,d,m} = \frac{Q}{Q-1} \sum_{\boldsymbol{r}\in\mathcal{R}(I,d,m)} (\bar{\beta}_{\boldsymbol{r},\mathrm{QM}})^2 - \frac{1}{Q(Q-1)} \sum_{\boldsymbol{r}\in\mathcal{R}(I,d,m)} \sum_{q=1}^{Q} (\hat{\beta}_{\boldsymbol{r},\mathrm{QM,q}})^2. \tag{12} $$

In addition, we can define an estimator ${}_{\mathrm{QM}}\hat{\gamma}_{d,m}(p)$ for the lower bound $\gamma_{d,m}(p)$ on $\gamma(p)$ defined in (7):

$$ \begin{aligned} {}_{\mathrm{QM}}\hat{\gamma}_{d,m}(p) &= \frac{1}{\hat{\sigma}^2_{\mathrm{QM}}} \sum_{I:|I|\le p} {}_{\mathrm{QM}}\hat{\sigma}^2_{I,d,m} \\ &= \frac{Q}{(Q-1)\hat{\sigma}^2_{\mathrm{QM}}} \left(\sum_{I:|I|\le p} \sum_{\boldsymbol{r}\in\mathcal{R}(I,d,m)} (\bar{\beta}_{\boldsymbol{r},\mathrm{QM}})^2 - \frac{1}{Q^2} \sum_{q=1}^{Q} (\hat{\beta}_{\boldsymbol{r},\mathrm{QM,q}})^2 \right), \end{aligned} $$

where

$$ \hat{\sigma}^2_{\mathrm{QM}} = \frac{1}{nQ} \sum_{q=1}^{Q} \sum_{k=1}^{n} f^2(\boldsymbol{v}_{k,q}) - \frac{2}{Q(Q-1)} \sum_{1\le q_1<q_2\le Q} \hat{\mu}_{\mathrm{QM},q_1} \hat{\mu}_{\mathrm{QM},q_2}, $$

and $\hat{\mu}_{\mathrm{QM,q}} = \sum_{k=1}^{n} f(\boldsymbol{v}_{k,q})/n$, $q = 1, \ldots, Q$. It can be easily shown that $\hat{\sigma}^2_{\mathrm{QM}}$ is an unbiased estimator of σ^2.

Centered versions of the above estimators can be defined similarly as in the Monte Carlo case, i.e., by replacing f by $\tilde{f} = f - \hat{\mu}_{\mathrm{QM}}$, where $\hat{\mu}_{\mathrm{QM}} = \sum_{q=1}^{Q} \hat{\mu}_{\mathrm{QM,q}}/Q$. Also, to estimate the accuracy of ${}_{\mathrm{QM}}\hat{\gamma}_{d,m}(p)$, we construct R i.i.d. copies and use the sample variance within the set of R i.i.d. estimators of $\gamma_{d,m}(p)$ thus obtained to construct approximate confidence intervals on the true value $\gamma_{d,m}(p)$. We use R replicates in order to get an especially simple and direct variance estimate. We do not explore alternatives such as the jackknife based on subsets of the original Q random vectors.

Note that using the Monte Carlo method to estimate the $\beta_{\boldsymbol{r}}$ as in the previous subsection avoids *aliasing* problems. That is, in general the estimation of the regression coefficients included in a model (in our case, the $\beta_{\boldsymbol{r}}$ for $\boldsymbol{r} \in \mathcal{R}(I,d,m)$) can be affected by the effects that are not included in the model (in our case, the $\phi_{\boldsymbol{r}}$ for $\boldsymbol{r} \notin \mathcal{R}(I,d,m)$). However, this aliasing problem typically takes place when a deterministic point set is used for the estimation. When a random point set is used (as in Monte Carlo), it can be shown that the aliasing becomes negligible as the number of points goes to infinity [11]. For a randomized QMC point set, the adverse effects of aliasing (which could affect the accuracy of the unbiased estimator) can be reduced by making sure that the point set P_n has a small *discrepancy*, and by taking n sufficiently large. See [11] for more details on the connection between aliasing and low-discrepancy point sets.

4 Examples

In this section, we provide numerical results where the estimators for $\sigma^2_{I,d,m}$ and $\gamma_{d,m}(p)$ described in the preceding section are computed in different problems. We look at two examples considered in [14], where randomized lattice rules were used with success to estimate the integral $\mu = \int_{[0,1)^s} f(\boldsymbol{u})d\boldsymbol{u}$.

For each problem, the randomized QMC estimators are based on a Korobov rule defined by $n = 1021$ and $a = 76$. We use $Q = 15$ random shifts and then construct $R = 10$ copies of each estimator ${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$ in order to compute confidence intervals. For the Monte Carlo estimators, we use n points and $R = 150$, so that the total number of points used is the same for both types of estimators. Results are presented in tables in which each column corresponds to a particular type of estimator of $\gamma_{d,m}(p)$. Each entry in these tables is accompanied by a value in parentheses that represents the half-length of a 95% confidence interval for $\gamma_{d,m}(p)$ based on this estimator. The centered versions of the estimators are denoted by "cent." We vary the value of d in order to study the relative importance of the low and high order subcomponents of f_I for $|I| \leq p$. The estimates $\hat{\gamma}_{d,m}(p)$ with the largest value of d are used as lower bounds for $\gamma(p)$. All our estimators are based on (modified) Legendre polynomials [1].

4.1 Asian option pricing

The problem here is to evaluate an *Asian option* under the Black-Scholes model [7]. An Asian option is a financial contract that has an expiration date T, a *strike price* K, and it depends on the price of an underlying asset whose value at time t is denoted by $S(t)$, for $0 \leq t \leq T$. Assuming the Black-Scholes model implies that $S(t)/S(0)$ has a lognormal distribution with parameters $(\mu t, \sigma\sqrt{t})$, where μ is the mean return on the asset and σ is its volatility. The Asian option depends on the average value taken by the underlying asset over a predetermined period of time; in our experiments, we assume that the average is taken at s equally spaced times $t_1, \ldots, t_s$, where $t_1 = T/s$ and $t_s = T$. In this case, the final value of the option at the expiration date is given by

$$C(T) = \max\left(0, \frac{1}{s}\sum_{j=1}^{s} S(t_j) - K\right).$$

The quantity of interest is $\mu = E(e^{-rT}C(T))$, where the expectation is taken under the risk-neutral measure, and r represents the risk-free rate in the economy. For this problem, the function that is integrated in order to get μ is s-dimensional and is given explicitly in [14].

In all our experiments, we use the parameters $T = 1$ year, $S(0) = 50$, $s = 32$, $r = 0.05$, and $\sigma = 0.3$. In Table 1, we give four different estimators for $\gamma_{d,m}(p)$, for $p = 1, 2$, and $d = m = 2, 4, 8$.

Table 1. Estimation of $\gamma_{d,m}(p)$ for the Asian-option problem, with $K = 45$

		Monte Carlo		Rand. Shif. LR	
$d = m$	p	$\hat{\gamma}_{d,m}(p)$	$\hat{\gamma}_{d,m}(p)$ (cent.)	$_{\text{QM}}\hat{\gamma}_{d,m}(p)$	$_{\text{QM}}\hat{\gamma}_{d,m}(p)$ (cent.)
2	1	0.82 (1.2e-2)	0.82 (6.4e-3)	0.82 (1.9e-3)	0.82 (1.8e-3)
	2	0.94 (2.1e-2)	0.94 (1.3e-2)	0.93 (5.8e-3)	0.93 (5.4e-3)
4	1	0.85 (1.3e-2)	0.85 (7.0e-3)	0.85 (2.6e-3)	0.85 (2.5e-3)
	2	0.97 (3.1e-2)	0.98 (1.9e-2)	0.96 (8.7e-3)	0.97 (8.1e-3)
8	1	0.86 (1.4e-2)	0.86 (7.5e-3)	0.86 (2.9e-3)	0.86 (2.8e-3)
	2	1.01 (6.1e-2)	1.01 (3.3e-2)	0.97 (8.4e-3)	0.97 (9.3e-3)

As we can see from the results in Table 1 for the case $K = 45$ and when $d = m = 8$, about 97% of the variance of the underlying function is explained by one and two-dimensional components. This explains the success of lattice rules having good low-dimensional components as an integration method on this problem [14]. In addition, we estimated $\gamma_{1,1}(1)$ (which is denoted by σ_L^2/σ^2 in [18])) to be about 0.82 for this function, which means its linear part explains about 82% of the variance; for linear functions, it is known that estimators based on randomly shifted lattice rules of rank 1 (e.g., such as Korobov rules), have a variance equal to σ^2/n^2. Consequently, components of f that explain a large part of the variance can be integrated very accurately by such rules, which explains the overall good results of randomly shifted lattice rules on this particular problem.

The results for the case $d = m = 2$ indicate that the low order terms of the approximation already capture about 93% of the variance of the function. We also see that increasing d from 4 to 8 does not significantly change the estimates $\hat{\gamma}_{d,m}(2)$.

In Table 2, we give results similar to those in Table 1 except that now the strike price is $K = 55$, which means the option is "out-of-the-money". This implies that the function f has a value of zero for a large part of the domain $[0,1)^s$, with a spike when the variables u_j approach 1.

Compared to the case where $K = 45$, the lower bound on $\gamma(2)$ is now smaller; more precisely, it went from 0.97 to 0.90, which means that the one and two-dimensional components of f now explain about 9/10 of the variance $\sigma^2 = \text{Var}(f)$. Table 3 contains results for the same problem as Table 2, except that a *Brownian bridge technique* [4,5] is used to generate the Brownian motion that drives the price process $S(t)$. This technique has the goal of reducing the effective dimension of f. Hence we expect $\hat{\gamma}_{8,8}(p)$ to be larger in Table 3 than in Table 2, which is indeed what we get since $\hat{\gamma}_{8,8}(1)$ increases from 0.55 to 0.80, and $\hat{\gamma}_{8,8}(2)$ from 0.90 to 0.99.

Table 2. Estimation of $\gamma_{d,m}(p)$ for the Asian-option problem, with $K = 55$

		Monte Carlo		Rand. Shif. LR	
$d = m$	p	$\hat{\gamma}_{d,m}(p)$	$\hat{\gamma}_{d,m}(p)$ (cent.)	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$ (cent.)
2	1	0.52 (7.7e-3)	0.52 (6.1e-3)	0.52 (2.6e-3)	0.52 (2.6e-3)
	2	0.85 (2.1e-2)	0.85 (1.7e-2)	0.85 (8.3e-3)	0.85 (7.9e-3)
4	1	0.55 (8.6e-3)	0.55 (6.8e-3)	0.54 (2.8e-3)	0.54 (2.8e-3)
	2	0.90 (2.9e-2)	0.90 (2.4e-2)	0.89 (1.1e-2)	0.89 (1.1e-2)
8	1	0.56 (9.8e-3)	0.55 (7.7e-3)	0.55 (2.8e-3)	0.55 (2.8e-3)
	2	0.92 (5.2e-2)	0.93 (4.2e-2)	0.89 (1.2e-2)	0.90 (1.1e-2)

Increasing d has roughly the same effect in the setting of Table 2 as what was observed when $K = 45$. However, we can see in Table 3 that when the Brownian bridge technique is applied, the low order terms become relatively less important; for example, $\hat{\gamma}_{2,2}(2)$ is only 0.81 (which is even slightly smaller than *without* using the Brownian bridge), whereas $\hat{\gamma}_{4,4}(2)$ is about 0.94, and $\hat{\gamma}_{8,8}(2)$ is almost 1.

In each of the three tables presented in this subsection, the estimators ${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$ appear to be more accurate than the Monte Carlo equivalent $\hat{\gamma}_{d,m}(p)$.

Table 3. Estimation of $\gamma_{d,m}(p)$ for the Asian-option problem, with $K = 55$ and using the Brownian bridge technique

		Monte Carlo		Rand. Shif. LR	
$d = m$	p	$\hat{\gamma}_{d,m}(p)$	$\hat{\gamma}_{d,m}(p)$ (cent.)	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$ (cent.)
2	1	0.67 (1.0e-2)	0.67 (7.3e-3)	0.67 (1.4e-3)	0.67 (1.5e-3)
	2	0.81 (1.7e-2)	0.81 (1.3e-2)	0.81 (4.6e-3)	0.81 (4.4e-3)
4	1	0.77 (1.5e-2)	0.77 (1.1e-2)	0.78 (1.7e-3)	0.78 (1.7e-3)
	2	0.95 (3.1e-2)	0.95 (2.6e-2)	0.94 (1.6e-2)	0.94 (1.5e-2)
8	1	0.80 (1.7e-2)	0.80 (1.3e-2)	0.80 (2.9e-3)	0.80 (2.8e-3)
	2	1.00 (6.7e-2)	1.00 (5.4e-2)	0.99 (3.7e-2)	0.99 (3.6e-2)

4.2 Stochastic activity network

This problem is taken from [3]. A *stochastic activity network* (SAN) is represented by a directed acyclic graph $(\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is a set of nodes that contains one source and one sink, and $\mathcal{A}$ is a set of edges corresponding to the activities. Each activity k in $\mathcal{A}$ has a random length V_k. We assume each activity has a nonzero probability of being larger than 0. The *completion time* of the network is denoted by T and is equal to the length of the longest path from the source to the sink.

The quantity of interest here is $\mu = P(T \leq t_0)$, for some threshold t_0, and the dimension of this problem is $s = |\mathcal{A}|$, the number of activities. In [3], the authors use *Conditional Monte Carlo* (CMC) as a variance (and dimension) reduction technique for this problem. The idea is to define a *uniformly directed cutset* $\mathcal{L} \subseteq \mathcal{A}$, which has the property that each path from the source to the sink contains exactly one edge in $\mathcal{L}$. The integrand is then the conditional probability $P[T \leq t_0 \mid \{V_j, j \notin \mathcal{L}\}]$ instead of being the indicator function $1_{T \leq t_0}$. In our experiments, we used the same parameters as in Network 1 of [3] that contains $|\mathcal{A}| = 13$ edges, and we took $t_0 = 90$.

Table 4. Estimation of $\gamma_{d,m}(p)$ for the stochastic activity network

		Monte Carlo		Rand. Shif. LR	
$d = m$	p	$\hat{\gamma}_{d,m}(p)$	$\hat{\gamma}_{d,m}(p)$ (cent.)	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$ (cent.)
4	1	0.46 (2.0e-2)	0.47 (9.6e-3)	0.47 (5.0e-3)	0.47 (4.7e-3)
	2	0.52 (4.7e-2)	0.52 (1.4e-2)	0.52 (4.1e-3)	0.52 (4.2e-3)
	3	0.51 (7.0e-2)	0.52 (1.9e-2)	0.51 (2.1e-2)	0.52 (6.0e-3)
6	1	0.57 (2.2e-2)	0.59 (1.2e-2)	0.59 (4.8e-3)	0.59 (4.3e-3)
	2	0.64 (6.4e-2)	0.66 (1.8e-2)	0.66 (5.5e-3)	0.66 (6.9e-3)
	3	0.62 (0.14)	0.67 (3.5e-2)	0.58 (9.2e-2)	0.66 (2.2e-2)
8	1	0.63 (2.4e-2)	0.64 (1.3e-2)	0.64 (4.4e-3)	0.64 (3.8e-3)
	2	0.71 (9.1e-2)	0.73 (2.3e-2)	0.73 (1.2e-2)	0.73 (7.7e-3)
	3	0.69 (0.22)	0.76 (5.4e-2)	0.51 (0.20)	0.73 (4.4e-2)

In Table 4, we give four different estimators for $\gamma_{d,m}(p)$, for $p = 1, 2, 3$, and $d = m = 4, 6, 8$. These results show that even if the dimension is relatively small here, the components f_I with $|I| \leq 3$ explain approximately only 73% of the variance of f. However, when the problem is modified through the use of CMC with a uniformly directed cutset $\mathcal{L}$ containing five edges, the dimension s reduces from thirteen to eight since we only need to generate the

random durations of the activities that are *not* in $\mathcal{L}$; we refer the reader to [3] for more details on how to use CMC for this problem. As we see in Table 5, one-dimensional components now explain about 90% of the variance instead of 64% without CMC. These results are consistent with those reported in [14], where randomly shifted lattice rules were used to estimate μ; it was noted there that with the use of CMC, estimators based on lattice rules had an empirical variance apparently in $O(n^{-2})$, and were more successful in this case than without CMC (although they were still improving upon MC without CMC).

For this problem the high order terms bring a significant improvement to the approximation, whether we use CMC or not; for a fixed value of p, the estimates $\hat{\gamma}_{d,m}(p)$ roughly increase by at least 0.17 when d goes from 4 to 8.

Finally it seems that with or without CMC, the lattice rule estimators are more accurate than the Monte Carlo ones. For example, when $p = 2$ and $d = 8$ the confidence interval based on the lattice rule is about seven times smaller than in the MC case for the standard estimators, and about three times smaller for the centered versions. Also, the centered estimators are clearly more accurate than the standard ones for this problem. For example, when $p = 3$ and $d = 8$ the standard estimators are practically useless because of their large variance, whereas the centered ones remain reasonably accurate.

Table 5. Estimation of $\gamma_{d,m}(p)$ for the stochastic activity network, when conditional Monte Carlo is used

		Monte Carlo		Rand. Shif. LR	
$d = m$	p	$\hat{\gamma}_{d,m}(p)$	$\hat{\gamma}_{d,m}(p)$ (cent.)	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$	${}_{\text{QM}}\hat{\gamma}_{d,m}(p)$ (cent.)
4	1	0.74 (5.9e-2)	0.70 (1.4e-2)	0.71 (3.8e-3)	0.72 (3.2e-3)
	2	0.80 (0.12)	0.72 (1.9e-2)	0.74 (5.3e-3)	0.74 (6.9e-3)
	3	0.81 (0.15)	0.72 (2.2e-2)	0.74 (4.6e-2)	0.74 (7.2e-3)
6	1	0.87 (6.9e-2)	0.84 (2.1e-2)	0.86 (3.9e-3)	0.87 (2.9e-3)
	2	0.90 (0.16)	0.87 (2.8e-2)	0.90 (1.1e-2)	0.90 (8.3e-3)
	3	0.89 (0.26)	0.87 (3.8e-2)	0.87 (0.18)	0.90 (1.5e-2)
8	1	0.91 (8.0e-2)	0.88 (2.4e-2)	0.90 (3.9e-3)	0.91 (2.6e-3)
	2	0.96 (0.21)	0.90 (3.8e-2)	0.96 (2.6e-2)	0.94 (1.2e-2)
	3	0.96 (0.39)	0.91 (6.2e-2)	0.85 (0.43)	0.94 (2.4e-2)

5 Conclusion and discussion

We have resolved the variance of an integrand f into a sum of squared components corresponding to various basis functions in an infinite expansion. These squared coefficients can be estimated and then summed over various subsets of indices. We have been mostly interested in sums corresponding to the lower frequency or smoother parts of the ANOVA components of f, as these are known to be related to the accuracy of integration rules applied to f. We have seen that the estimators based on randomized lattice rules were often more accurate than those constructed with Monte Carlo.

Our method differs from the one used by Sobol' and his collaborators [2,22]. They directly estimate the variance of each ANOVA component f_I of f. They avoid truncation and its associated bias, but their method does not separate high and low order subcomponents of f_I.

Unbiased variance estimates formed by subtraction are widely used but they almost inevitably have some probability of producing negative variance estimates. Sometimes we find that increasing p corresponds to a decrease in $\hat{\gamma}$ even though it cannot decrease γ. Similarly we find some $\hat{\gamma}$'s slightly larger than 1. These correspond to some negative variance estimates, with the interpretation that the true variance is small compared to its sampling uncertainty. For a thorough discussion of variance estimation see [20].

Some aspects that could be investigated further would be to try other bases in addition to Legendre polynomials, to apply our method to other practical problems, and to compare it empirically with Sobol's method. For example, truncation of a Haar like basis may provide more realistic descriptions of scrambled net accuracy than would the infinite untruncated sums. Also, we could perform the truncation by using other or additional parameters than the degree d and the order m.

6 Acknowledgments

We are pleased to acknowledge the helpful comments of two reviewers. This work has been supported by an FCAR-Québec postdoctoral scholarship to the first author, and has largely been realized while she was visiting the second author in the Statistics Department of Stanford University. This work was also supported by NSF grants DMS-9704495 and DMS-0072445.

References

1. J. An and A.B. Owen. Quasi-regression. *Journal of Complexity*, 2001. To appear.
2. G. E. B. Archer, A. Saltelli, and I. M. Sobol. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statist. Comput. Simul.*, 58:99–120, 1997.

3. A. N. Avramidis and J. R. Wilson. Integrated variance reduction strategies for simulation. *Operations Research*, 44:327–346, 1996.
4. R. E. Caflisch and B. Moskowitz. Modified Monte Carlo methods using quasi-random sequences. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, number 106 in Lecture Notes in Statistics, pages 1–16, New York, 1995. Springer-Verlag.
5. R. E. Caflisch, W. Morokoff, and A. B. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance*, 1(1):27–46, 1997.
6. R. Cranley and T. N. L. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Num. Anal.*, 13(6):904–914, 1976.
7. D. Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, second edition, 1996.
8. B. Efron and C. Stein. The jackknife estimator of variance. *Annals of Statistics*, 9:586–596, 1981.
9. G. B. Folland. *Fourier Analysis and its Applications.* Wadsworth and Brooks, 1992.
10. F. J. Hickernell. A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, 67:299–322, 1998.
11. F.J. Hickernell and M.-Q. Liu. Uniform designs limit aliasing. Submitted for publication, 2001.
12. W. Hoeffding. A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19:293–325, 1948.
13. N. M. Korobov. The approximate computation of multiple integrals. *Dokl. Akad. Nauk SSSR*, 124:1207–1210, 1959. In Russian.
14. P. L'Ecuyer and C. Lemieux. Variance reduction via lattice rules. *Management Science*, 46:1214–1235, 2000.
15. H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics.* SIAM, Philadelphia, 1992.
16. A. B. Owen. Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal of Numerical Analysis*, 34(5):1884–1910, 1997.
17. A. B. Owen. Latin supercube sampling for very high-dimensional simulations. *ACM Transactions of Modeling and Computer Simulation*, 8(1):71–102, 1998.
18. A. B. Owen. Assessing linearity in high dimensions. *Annals of Statistics*, 8:1–19, 2000.
19. S. Paskov and J. Traub. Faster valuation of financial derivatives. *Journal of Portfolio Management*, 22:113–120, 1995.
20. S. R. Searle, G. Casella, and C. E. McCulloch. *Variance Components.* Wiley, 1992.
21. I. H. Sloan and S. Joe. *Lattice Methods for Multiple Integration.* Clarendon Press, Oxford, 1994.
22. I.M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414, 1993.