Better estimation of small Sobol' sensitivity indices

ART B. OWEN
Stanford University

September 2012

Abstract

A new method for estimating Sobol' indices is proposed. The new method makes use of 3 independent input vectors rather than the usual 2. It attains much greater accuracy on problems where the target Sobol' index is small, even outperforming some oracles which adjust using the true but unknown mean of the function. The new estimator attains a better rate of convergence than the old one in a small effects limit. When the target Sobol' index is quite large, the oracles do better than the new method.

This work was supported by the U.S. National Science Foundation under grant DMS-0906056.

Author's address: A. B. Owen, Department of Statistics, Stanford University.

1 Introduction

Let f be a deterministic function on $[0,1]^d$ for $d \geqslant 1$. Sobol' sensitivity indices, derived from a functional ANOVA, are used to measure the importance of subsets of input variables. There are two main types of index, but one of them is especially hard to estimate in cases where that index is small.

The problematic index can be represented as a covariance between outcomes of f evaluated at two random input points, that share some but not all of their components. A natural estimator then is a sample covariance based on pairs of random d-vectors of this type. Sobol' and Myshetskaya (2007) report a numerical experiment where enormous efficiency differences are obtained depending on how one estimates that covariance. The best gains arise from applying some centering strategies to those pairs of function evaluations.

This article introduces a new estimator for the Sobol' index, based on three input vectors, not two. The new estimator makes perhaps surprising use of randomly generated centers. The random centering adds to the cost of every simulation run and might be thought to add noise. But in many examples that noise must be strongly negatively correlated with the quantity it adjusts because (in those examples) the random centering greatly increases efficiency. The new estimate is not always most efficient. In particular when the index to be estimated is large, the new estimate is seen to perform worse than some oracles that one could approximate numerically.

The motivation behind Sobol' indices, is well explained in the text by Saltelli et al. (2008). These indices have been applied to problems in industry, science and public health. For a recent mathematical account of Sobol' indices, see Owen (2012).

The outline of this article is as follows. Section 2 introduces the ANOVA decomposition, Sobol' indices, and our notation. Section 3 presents the original estimator of the Sobol' indices and the four improved estimators we consider here. Section 4 considers some numerical examples. For small Sobol' indices, the newly proposed estimator is best, beating two oracles. For very large indices, the best performance comes from an oracle that uses the true function mean twice. Section 5 presents some theoretical support for the estimator on product functions which are commonly used for numerical examples. It generalizes the estimator to a wider class of methods and shows that the proposed estimator minimizes a proxy for the variance, in the product function context. Section 6 considers a setting where all contributions of the variables of interest are proportional to a number ε. In the 'total insensitivity' limit $\varepsilon \to 0$, the new estimator attains a variance of $O(\varepsilon^4)$ which yields a fixed relative precision as $\varepsilon \to 0$. Two of the old estimators attain $O(\varepsilon^2)$ and the fourth is $O(1)$. These estimators are unable to attain fixed relative precision at bounded cost. Total insensitivity is one of seven asymptotes considered there. In other limits, especially those with large effects for the variables of interest, a different estimator has minimal variance, but the effect is in the constant, not the rate of convergence. Section 7 supplies some of the proofs needed for the aysmptotic results. Section 8 has a discussion. It considers why it is valuable to identify unimportant variables, why unimportant ones might outnumber important ones, and why total insensitivity is a natural limit for applications.

2 Background

Here we review the Analysis of Variance (ANOVA) decomposition of $L^2[0,1]^d$ and then the Sobol' indices that are based on the ANOVA. For some history and additional details see Liu and Owen (2006). We begin with notation.

The point $\boldsymbol{x} \in [0,1]^d$ has components $x_1, \ldots, x_d$. For $u \subseteq \mathcal{D} \equiv \{1, 2, \ldots, d\}$ we use $|u|$ for its cardinality. The symbol $\boldsymbol{x}_u$ represents a point in $[0,1]^{|u|}$ containing all the x_j for $j \in u$. We use $-u$ or u^c depending on typographical readability, to denote the complement of u in $\mathcal{D}$. Given two points $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^d$

and the subset $u \subseteq \mathcal{D}$ the hybrid point $\boldsymbol{z} = \boldsymbol{x}_u{:}\boldsymbol{y}_{-u}$ is the one with $z_j = x_j$ for $j \in u$ and $z_j = y_j$ for $j \notin u$.

For $d \geqslant 1$, let $f \in L^2[0,1]^d$. Then f has a finite variance $\sigma^2 = \int (f(\boldsymbol{x}) - \mu)^2 \,\mathrm{d}\boldsymbol{x}$ where $\mu = \int f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$ is the mean of $f(\boldsymbol{x})$ for random $\boldsymbol{x} \sim \mathbf{U}[0,1]^d$. The region of integration here is $[0,1]^d$ and integrals with respect to $\mathrm{d}\boldsymbol{x}_u = \prod_{j\in u} \mathrm{d}x_j$ are over $[0,1]^{|u|}$.

The ANOVA decomposition has the form

$$f(\boldsymbol{x}) = \sum_{u \subseteq \mathcal{D}} f_u(\boldsymbol{x}) \tag{1}$$

where $f_u(\boldsymbol{x})$ only depends on $\boldsymbol{x}$ through $\boldsymbol{x}_u$. That is $f_u(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) = f_u(\boldsymbol{x})$ for all $\boldsymbol{y} \in [0,1]^d$. Many decompositions satisfy (1). The ANOVA has $f_\varnothing(\boldsymbol{x}) = \int f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \mu$, $f_{\{j\}}(\boldsymbol{x}) = \int (f(\boldsymbol{x}) - \mu) \,\mathrm{d}\boldsymbol{x}_{-\{j\}}$ and in general

$$f_u(\boldsymbol{x}) = \int \Bigl(f(\boldsymbol{x}) - \sum_{v \subset u} f_v(\boldsymbol{x}) \Bigr) \,\mathrm{d}\boldsymbol{x}_{-u} = \int f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}_{-u} - \sum_{v\subset u} f_v(\boldsymbol{x})$$

where $\subset$ denotes the proper subset relation.

In statistics, the quantity $f_{\{j\}}$ is called the main effect of variable x_j and for $|u| > 1$, $f_u(\boldsymbol{x})$ is the interaction among x_j for $j \in u$. The variance of $f_u(\boldsymbol{x})$ for $\boldsymbol{x} \sim \mathbf{U}[0,1]^d$, is denoted σ_u^2.

Several properties of the ANOVA follow readily from its definition. First $\int f_u(\boldsymbol{x}) \,\mathrm{d}x_j = 0$ whenever $j \in u$. It follows that $\int f_u(\boldsymbol{x}) f_v(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = 0$ for $u \neq v$ by integrating first over some $j \in (u-v) \cup (v-u)$. Next, the variance decomposes as $\sigma^2 = \sum_{u \subseteq \mathcal{D}} \sigma_u^2$. This fact is the origin the name ANOVA. If $u \neq \varnothing$ then $\sigma_u^2 = \int f_u(\boldsymbol{x})^2 \,\mathrm{d}\boldsymbol{x}$, while $\sigma_\varnothing^2 = 0$.

Given a function $f \in L^2[0,1]^d$ we might seek to approximate it by one that only depends on $\boldsymbol{x}_u$. The closest approximation to f (in L^2) is the conditional expectation $\mathbb{E}(f(\boldsymbol{x}) \mid \boldsymbol{x}_u) = \sum_{v \subseteq u} f_v(\boldsymbol{x}) \equiv f_{\lfloor u \rfloor}(\boldsymbol{x})$. If this approximation is very close to f then $\mathbb{E}\bigl(f_{\lfloor u \rfloor}(\boldsymbol{x})^2\bigr)$ will be almost as large as σ^2 and then the variables $\boldsymbol{x}_u$ can be considered important. One of Sobol's indices is

$$\underline{\tau}_u^2 \equiv \sum_{v \subseteq u} \sigma_v^2 \tag{2}$$

which is exactly $\mathbb{E}\bigl(f_{\lfloor u \rfloor}(\boldsymbol{x})^2\bigr)$. A second Sobol' index is

$$\overline{\tau}_u^2 = \sum_{v \cap u \neq \varnothing} \sigma_v^2. \tag{3}$$

These satisfy $0 \leqslant \underline{\tau}_u^2 \leqslant \overline{\tau}_u^2 \leqslant \sigma^2$ and $\underline{\tau}_u^2 = \sigma^2 - \overline{\tau}_{-u}^2$. These indices provide two measures of the importance of the variables in subset u. The larger measure includes interactions between variables in u and variables in u^c, while the smaller measure excludes those interactions. The quantity $\overline{\tau}_u^2$, or more commonly $\overline{\tau}_u^2/\sigma^2$, is the total sensitivity index for the set u. By contrast, $\underline{\tau}_u^2$ or $\underline{\tau}_u^2/\sigma^2$ are partial sensitivity indices.

Most estimation strategies for Sobol' indices are based on the identities

$$\underline{\tau}_u^2 = \int f(\boldsymbol{x})f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} - \mu^2, \quad \text{and} \tag{4}$$

$$\overline{\tau}_u^2 = \frac{1}{2}\int \bigl(f(\boldsymbol{x}) - f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})\bigr)^2\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}. \tag{5}$$

When $\overline{\tau}_u^2$ is small, then (5) leads to a very effective Monte Carlo strategy based on

$$\widehat{\overline{\tau}}_u^2 = \frac{1}{2n}\sum_{i=1}^n \bigl(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u})\bigr)^2$$

for $(\boldsymbol{x}_i, \boldsymbol{y}_i) \overset{\text{iid}}{\sim} \mathbf{U}[0,1]^{2d}$. This estimator is a sum of squares, hence nonnegative, and it is unbiased. If the true $\overline{\tau}_u^2 = 0$, then $\widehat{\overline{\tau}}_u^2 = 0$ with probability one. More generally, if the true $\overline{\tau}_u^2$ is small, then the estimator averages squares of typically small quantities. We assume throughout that $\int f(\boldsymbol{x})^4\,\mathrm{d}\boldsymbol{x} < \infty$ so that the variance of this and our other estimators is finite.

It is more difficult to estimate small $\underline{\tau}_u^2$. The natural way to estimate $\underline{\tau}_u^2$ is via

$$\widehat{\underline{\tau}}_u^2 = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u}) - \hat\mu^2. \tag{6}$$

The simplest estimator of μ is $\hat\mu = (1/n)\sum_{i=1}^n f(\boldsymbol{x}_i)$ but Janon et al. (2012) have recently proved that it is better to use $\hat\mu = (1/2n)\sum_{i=1}^n (f(\boldsymbol{x}_i) + f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u}))$. This estimate was used in a crop modeling context by Monod et al. (2006).

3 The estimators

The problem with (6) is that it has very large variance when $\mu^2 \gg \underline{\tau}_u^2$. Although $\underline{\tau}_u^2$ is invariant with respect to shifts replacing $f(\boldsymbol{x})$ by $f(\boldsymbol{x}) - c$ for a constant c, the variance of (6) can be strongly affected by such shifts. Sobol' (1993) (originally published as Sobol' (1990)) recommends shifting f by an amount close to μ, which while not necessarily optimal, should be reasonable.

An approximation to μ can be obtained by Monte Carlo or quasi-Monte Carlo (QMC) sampling prior to estimation of $\underline{\tau}_u^2$. In our simulations we suppose that an oracle has supplied μ and then we compare estimators that do and do not make use of the oracle.

Another estimator of $\underline{\tau}_u^2$ was considered independently by Saltelli (2002) and the Masters thesis of Mauntz (2002) under the supervision of S. S. Kucherenko and C. Pantelides. Results of Mauntz's thesis have appeared in Sobol et al. (2007) and Kucherenko et al. (2011). This estimator, called correlated sampling by Sobol' and Myshetskaya (2007) replaces $f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,u})$ by $f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,u}) - f(\boldsymbol{y}_i)$ in (6) and then it is no longer necessary to subtract $\hat\mu^2$. Indeed the method can be viewed as subtracting the estimate $n^{-2}\sum_{i=1}^n\sum_{i'=1}^n f(\boldsymbol{x}_i)f(\boldsymbol{y}_{i'})$ from the sample mean of $f(\boldsymbol{x})f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})$. That estimator is called "Correlation 1" below.

Sobol' and Myshetskaya (2007) find that even the correlated sampling method has increased variance when μ is large. They propose another estimator replacing the first $f(\boldsymbol{x}_i)$ by $f(\boldsymbol{x}_i)-c$ for a constant c near μ. Supposing that an oracle has supplied $c=\mu$ we call the resulting method "Oracle 1" because it makes use of the true μ one time. One could also make use of the oracle's μ in both the left and right members of the cross moment pair. We call this estimator "Oracle 2" below. The fourth method to compare is a new estimator, called "Correlation 2", that uses two random offsets. Instead of replacing $f(\boldsymbol{x}_i)$ by $f(\boldsymbol{x}_i)-\mu$ it draws a third variable $\boldsymbol{z}\sim \mathbf{U}[0,1]^d$ and is based on the identity

$$\begin{aligned}&\iiint \bigl(f(\boldsymbol{x})-f(\boldsymbol{z}_u{:}\boldsymbol{x}_{-u})\bigr)\bigl(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})-f(\boldsymbol{y})\bigr)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z}\\ &=(\mu^2+\underline{\tau}_u^2)-\mu^2-\mu^2+\mu^2=\underline{\tau}_u^2.\end{aligned}\tag{7}$$

We compare the following estimators

$$\frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u})-f(\boldsymbol{y}_i)) \qquad \text{(Correlation 1)}$$

$$\frac{1}{n}\sum_{i=1}^n (f(\boldsymbol{x}_i)-f(\boldsymbol{z}_{i,u}{:}\boldsymbol{x}_{i,-u})(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u})-f(\boldsymbol{y}_i)) \qquad \text{(Correlation 2)}$$

$$\frac{1}{n}\sum_{i=1}^n (f(\boldsymbol{x}_i)-\mu)(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u})-f(\boldsymbol{y}_i)) \qquad \text{(Oracle 1)}$$

$$\frac{1}{n}\sum_{i=1}^n (f(\boldsymbol{x}_i)-\mu)(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u})-\mu) \qquad \text{(Oracle 2)}$$

where $(\boldsymbol{x}_i,\boldsymbol{y}_i,\boldsymbol{z}_i)\overset{\text{iid}}{\sim}\mathbf{U}[0,1]^{3d}$ for $i=1,\dots,n$. Not all components of these vectors are necessary to estimate $\underline{\tau}_u^2$ for a single u, but many applications seek $\underline{\tau}_u^2$ for several sets u at once, so it is simpler to write them this way. The cost is assumed to be largely in evaluating f, not in producing the inputs $(\boldsymbol{x}_i,\boldsymbol{y}_i,\boldsymbol{z}_i)$.

Correlation 2 requires $4n$ function values to generate an estimate of $\underline{\tau}_u^2$. The other estimators require $2n$ or $3n$. If $\underline{\tau}_u^2$ has already been estimated and we then want to estimate $\underline{\tau}_{u'}^2$ for a subset $u'\neq u$, then we can reuse the $f(\boldsymbol{x}_i)$ and $f(\boldsymbol{y}_i)$ values but require $2n$ additional function values. The other estimators all require just n additional function value for u'. Our efficiency comparisons are based on computing a single $\underline{\tau}_u^2$, but can be modified to handle multiple sets. The properties of these estimators are given in Table 1. We ignore the possibility that even more sample reuse might be achievable for special collections of subsets $u_1,u_2,\dots,u_K$.

The intuitive reason why "Correlation 2" should be effective on small indices is as follows. If the variables in the set u are really unimportant then $f(\boldsymbol{x})$ will be determined almost entirely by the values in $\boldsymbol{x}_{-u}$. Then both $f(\boldsymbol{x}_i)-$

Table 1: Estimators of $\underline{\tau}_u^2$ with expected value and number of function values required per sample, for the first subset u and also for additional subsets.

Name	Expectation	Cost (first u)	Cost (addl. u)
Original (6)	$\mu^2 + \underline{\tau}_u^2$	2	1
Correlation 1	$\underline{\tau}_u^2$	3	1
Correlation 2	$\underline{\tau}_u^2$	4	2
Oracle 1	$\underline{\tau}_u^2$	3	1
Oracle 2	$\underline{\tau}_u^2$	2	1

$f(\boldsymbol{z}_{i,u}{:}\boldsymbol{x}_{i,-u})$ and $f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u}) - f(\boldsymbol{y})$ should be small values, even smaller than by centering at μ, and so the estimator takes a mean product of small quantities.

We do not compare the original estimator (6). The bias correction makes it more complicated to describe the accuracy of this method. Also that estimator had extremely bad performance in Sobol' and Myshetskaya (2007).

4 Examples

4.1 g function

This is the example used by Sobol' and Myshetskaya (2007). It has $d = 3$ and

$$f(\boldsymbol{x}) = \prod_{j=1}^{3} \frac{|4x_j - 2| + 2 + 3a}{1 + a_j}.$$

This function has $\mu = 27$ and $\sigma^2_{\{1\}} = 0.0675$, $\sigma^2_{\{2\}} = 0.27$, $\sigma^2_{\{3\}} = 1.08$, $\sigma^2_{\{1,2\}} = 0.000025$, $\sigma^2_{\{1,3\}} = 0.0001$, $\sigma^2_{\{2,3\}} = 0.0004$, and $\sigma^2_{\{1,2,3\}} \doteq 3.7 \times 10^{-8}$. The smallest and therefore probably the most difficult $\underline{\tau}_u^2$ to estimate is $\underline{\tau}^2_{\{1\}} = \sigma^2_{\{1\}}$. That is the one that they measure.

They report numerical values of $\widehat{\underline{\tau}}^2_{\{1\}}/\underline{\tau}^2_{\{1\}}$ for the four estimates in Table 1 (exclusive of the new "Correlation 2" estimate) based on $n = 256{,}000$ samples. The original estimator gave a value 2.239 times as large as the true $\underline{\tau}^2_{\{1\}}$. The others ranged from 0.975 to 1.104 times the true value. They did not use the oracle for μ, but centered their estimator instead on $c = 26.8$ to investigate a somewhat imperfect oracle.

The four estimators we consider here are all simply sample averages. As a result we can measure their efficiency by just estimating their variances. The efficiencies of these methods, using "Correlation 1" as the baseline are given by

$$E_{\text{corr 2}} = \frac{3}{4}\frac{\text{Var(corr 1)}}{\text{Var(corr 2)}}, \quad E_{\text{orcl 1}} = \frac{\text{Var(corr 1)}}{\text{Var(orcl 1)}}, \quad \text{and} \quad E_{\text{orcl 2}} = \frac{3}{2}\frac{\text{Var(corr 1)}}{\text{Var(orcl 2)}}$$

where the multiplicative factors accounts for the unequal numbers of function

Table 2: Relative efficiencies of 4 estimators of $\underline{\tau}_u^2$ for the g-function, rounded to the nearest integer. Relative indices $\underline{\tau}_u^2/\sigma^2$ rounded to three places.

Set u	$\underline{\tau}_u^2/\sigma^2$	Corr 1	Corr 2	Orcl 1	Orcl 2
$\{1\}$	0.048	1	4256	518	74
$\{2\}$	0.190	1	1065	525	297
$\{3\}$	0.762	1	267	556	1329
$\{1,2\}$	0.238	1	774	503	364
$\{1,3\}$	0.809	1	243	529	1306
$\{2,3\}$	0.952	1	194	473	1261

calls required by the methods for a single set u. For additional sets the efficiency multipliers 1/2, 1, and 1 could be used instead.

The efficiencies of the four estimators are compared in Table 2 based on $n = 1{,}000{,}000$ function evaluations. This is far more than one would ordinarily use to estimate the indices themselves, but we are interested in their sampling variances here. We consider all non-empty index sets except $u = \{1, 2, 3\}$ because $\underline{\tau}_{\{1,2,3\}}^2 = \sigma^2$ which can be estimated more directly. The table contains one moderately small index $\underline{\tau}_{\{1\}}^2$, (the one Sobol' and Myshetskaya (2007) studied). On the small effect, the new Correlation 2 estimator is by far the most efficient, outperforming both oracles. Inspecting the table, it is clear that it pays to use subtraction in both left and right sides of the estimator and that the smaller the effect $\underline{\tau}_u^2$ is, the better it is to replace the oracle's μ with a correlation based estimate.

4.2 Other product functions

It is convenient to work with functions of the form

$$f(\boldsymbol{x}) = \prod_{j=1}^{d} (\mu_j + \tau_j g_j(x_j)) \tag{8}$$

where $\int_0^1 g(x)\,\mathrm{d}x = 0$, $\int_0^1 g(x)^2\,\mathrm{d}x = 1$, and $\int_0^1 g(x)^4\,\mathrm{d}x < \infty$. For this function $\sigma_u^2 = \prod_{j\in u}\tau_j^2 \prod_{j\notin u}\mu_j^2$. Taking $g(x) = \sqrt{12}(x - 1/2)$, $d = 6$, $\mu = (1, 1, 1, 1, 1, 1)$ and $\tau = (4, 4, 2, 2, 1, 1)/4$ and sampling $n = 1{,}000{,}000$ times lead to the results in Table 3. The results are not as dramatic as for the g-function, but they show the same trends. The smaller $\underline{\tau}_u^2$ is, the more improvement comes from the new estimator. On the smallest indices it beats both oracles.

The improvements for the g-function are much larger than for the product studied here. For the purposes of Monte Carlo sampling the absolute value cusp in the g-function makes no difference. The g-function has the same moments as the product function with $\mu_j = 3$ and $\tau_j = 1/(\sqrt{3}a_j)$. Computing the g function estimates with the product function code (as a check) yields the same magnitude of improvement seen in Table 2.

Table 3: Relative efficiencies of 4 estimators of $\underline{\tau}_u^2$ for the product function (8). Relative indices $\underline{\tau}_u^2/\sigma^2$ rounded to three places.

Set u	$\underline{\tau}_u^2/\sigma^2$	Corr 1	Corr 2	Orcl 1	Orcl 2
$\{1\}$	0.165	1	0.74	1.13	1.23
$\{2\}$	0.165	1	0.73	1.14	1.24
$\{3\}$	0.041	1	1.69	1.15	0.54
$\{4\}$	0.041	1	1.67	1.15	0.54
$\{5\}$	0.010	1	5.45	1.16	0.20
$\{6\}$	0.010	1	5.58	1.16	0.20
$\{1,2\}$	0.495	1	0.75	1.21	1.86
$\{3,4\}$	0.093	1	1.23	1.16	0.94
$\{5,6\}$	0.012	1	2.94	1.16	0.38

5 Product functions

The best unbiased estimator of $\underline{\tau}_u^2$ is the one that minimizes the variance after making an adjustment for the number of function calls. In this section we look at that variance for product functions. We consider a more general estimator than Correlation 2

5.1 More general centering

The estimators in Section 4 are all formed by taking pairs $f(\boldsymbol{x})$ and $f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})$, subtracting centers from them, and averaging the product of those two centered values. Where they differ is in how they are centered.

We can generalize this approach to a spectrum of centering methods.

Theorem 1. *Let v and v' be two subsets of u^c and let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{z}$ be independent $\mathbf{U}[0,1]^d$ random vectors. Then*

$$\mathbb{E}\Bigl(\bigl(f(\boldsymbol{x}) - f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v})\bigr)\bigl(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}_{v'}{:}\boldsymbol{w}_{-v'})\bigr)\Bigr) = \underline{\tau}_u^2. \tag{9}$$

Proof.

$$\begin{aligned}&\mathbb{E}\Bigl(\bigl(f(\boldsymbol{x}) - f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v})\bigr)\bigl(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}_{v'}{:}\boldsymbol{w}_{-v'})\bigr)\Bigr)\\ &= (\mu^2 + \underline{\tau}_u^2) - (\mu^2 + \underline{\tau}_\varnothing^2) - (\mu^2 + \underline{\tau}_{u\cap v}^2) + (\mu^2 + \underline{\tau}_\varnothing^2)\\ &= \underline{\tau}_u^2,\end{aligned}$$

because $u \cap v = \varnothing$ and $\underline{\tau}_\varnothing^2 = 0$. □

As a result of Theorem 1, we may estimate $\underline{\tau}_u^2$ by

$$\frac{1}{n}\sum_{i=1}^n \bigl(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i,v}{:}\boldsymbol{z}_{i,-v})\bigr)\bigl(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u}) - f(\boldsymbol{y}_{i,v'}{:}\boldsymbol{w}_{i,-v'})\bigr) \tag{10}$$

where $(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{w}_i, \boldsymbol{z}_i) \overset{\text{iid}}{\sim} \mathbf{U}(0,1)^{4d}$.

The new estimate (10) uses up four independent vectors, not the three used in the Correlation 2 estimator, so we should check that it really is a generalization.

First, suppose that $v' = u^c$. Then the only part of the vector $\boldsymbol{w}$ that is used in (10) is $\boldsymbol{w}_{-v'} = \boldsymbol{w}_u$. Because (10) does not use $\boldsymbol{y}_u$ the needed parts of $\boldsymbol{y}$ and $\boldsymbol{w}$ fit within the same vector. That is we can sample $\boldsymbol{y}$ as before and use $\boldsymbol{y}_u$ for $\boldsymbol{w}_u$. As a result when $v' = u^c$ we only need three vectors as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\bigl(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i,v}{:}\boldsymbol{z}_{i,-v})\bigr)\bigl(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u}) - f(\boldsymbol{y}_i)\bigr). \tag{11}$$

If we take $v = u^c$ too, then (11) reduces to the Correlation 2 estimator.

At first, it might appear that the Oracle 2 estimator arises by taking $v = v' = \varnothing$, but this is not what happens, even when $\mu = 0$. A more appropriate generalization of the oracle estimators is based on the identity

$$\underline{\tau}_v^2 = \mathbb{E}\Bigl(\bigl(f(\boldsymbol{x}) - \mu_v(\boldsymbol{x}_v)\bigr)\bigl(f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) - \mu_{v'}(\boldsymbol{z}_{v'})\bigr)\Bigr)$$

where $\mu_v(\boldsymbol{x}_v) = \mathbb{E}(f(\boldsymbol{x}) \mid \boldsymbol{x}_v)$ and $v, v' \subseteq u^c$. To turn this identity into a practical estimator requires estimation of these conditional expectations. For $v = v' = \varnothing$ the conditional expectations become the unconditional expectation, which is simply the integral of f. For other v and v', such estimation requires some sort of nonparametric regression, with bias and variance expressions that complicate the analysis of the resulting estimate. As a result methods fixing subsets of the input variables are much easier to study.

5.2 General centering on product functions

The Correlation 2 estimator has $v = v' = u^c$, so it holds constant all of the variables in $\boldsymbol{x}_{-u}$. From Theorem 1, we see that this is just one choice among many and it raises the question of which variables should be held fixed in a Monte Carlo estimate of $\underline{\tau}_u^2$. The result is that we find taking $v = v' = u^c$ to be a principled choice. That is, match all of the variables outside the set of interest.

We can get some insight by considering functions of product form. Even there, the resulting variance formulas become cumbersome, but simplified versions yield some insight. We can write a product function as

$$f(\boldsymbol{x}) = \prod_{j=1}^{d} h_j(x_j) \tag{12}$$

where $h_j(x) = \mu_j + \tau_j g_j(x)$ with $\int_0^1 g_j(x)^p \,\mathrm{d}x$ taking values 0, 1, γ_j and κ_j for $p = 1$, 2, 3, and 4 respectively. In statistical terms, the random variable $h_j(x)$ has skewness γ_j/τ_j^3 and kurtosis $\kappa_j/\tau_j^4 - 3$ if $x \sim \mathbf{U}[0,1]$ and $\tau_j > 0$. We will suppose that all $\tau_j \geqslant 0$ and that all $\kappa_j < \infty$.

Proposition 1. *Let $\widehat{\underline{\tau}}_u^2$ be given by* (10)*, where f is given by the product model* (12)*. Then, for $v, v' \subseteq u^c$,*

$$n\mathrm{Var}(\widehat{\underline{\tau}}_u^2) = \mathbb{E}\bigl(Q_v(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{w})Q_{uv'}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{w})\bigr) - \underline{\tau}_u^4$$

for $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{w} \overset{\text{iid}}{\sim} \mathbf{U}[0,1]^d$ where

$$Q_v = \prod_{j=1}^d h_j^2(x_j) + \prod_{j\in v} h_j^2(x_j)\prod_{j\notin v} h_j^2(z_j) - 2\prod_{j\in v} h_j^2(x_j)\prod_{j\notin v} h_j(x_j)h_j(z_j), \quad \textit{and}$$

$$Q_{uv'} = \prod_{j\in u} h_j^2(x_j)\prod_{j\notin u} h_j^2(y_j) + \prod_{j\in v'} h_j^2(y_j)\prod_{j\notin u} h_j^2(w_j) - 2\prod_{j\in u^c\cap v'} h_j^2(y_j)\prod_{j\in u\cap v'^c} h_j(x_j)h_j(w_j)\prod_{j\in u^c\cap v'^c} h_j(y_j)h_j(w_j)$$

Proof. We need the expected square of the quantity inside the expectation in equation (9). First we expand

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v}) = \prod_{j=1}^d h_j(x_j) - \prod_{j\in v} h_j(x_j)\prod_{j\notin v} h_j(z_j).$$

Squaring this term yields Q_v, and similarly, squaring

$$f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}_{v'}{:}\boldsymbol{w}_{-v'}) = \prod_{j\in u} h_j(x_j)\prod_{j\notin u} h_j(y_j) - \prod_{j\in v'} h_j(y_j)\prod_{j\notin v'} h_j(w_j)$$

yields $Q_{uv'}$, after using $u \cap v' = \varnothing$. □

Using Proposition 1 we can see what makes for a good estimator in the product function setting. The quantities Q_v and $Q_{uv'}$ should both have small variance and their correlation should be small. The latter effect is very complicated depending on the interplay among u, v and v', and one might expect it to be of lesser importance. So we look at $\mathbb{E}(Q_v^2)$ for insight as to which indices should be in v. Then we suppose that it will usually be best to take the same indices for both v and v'.

Theorem 2. *Let $\widehat{\underline{\tau}}_u^2$ be given by* (10)*, where f is given by the product model* (12) *and let Q_v be as defined in Proposition 1. Then Q_v is minimized over $v \subseteq u^c$ by taking $v = u^c$.*

Proof. Let $\mu_{4j} = \int_0^1 h_j(x)^4\,\mathrm{d}x$ and $\mu_{2j} = \int_0^1 h_j(x)^2\,\mathrm{d}x$. It is elementary that $\mu_{4j} \geqslant \mu_{2j}^2$. Expanding $\mathbb{E}(Q_u^2)$ and gathering terms yields,

$$\prod_{j=1}^d \mu_{4j} + \prod_{j=1}^d \mu_{4j} + 4\prod_{j\in v}\mu_{4j}\prod_{j\notin v}\mu_{2j}^2 + 2\prod_{j\in v}\mu_{4j}\prod_{j\notin v}\mu_{2j}^2 - 4\prod_{j\in v}\mu_{4j}\prod_{j\notin v}\mu_{2j}^2 - 4\prod_{j\in v}\mu_{4j}\prod_{j\notin v}\mu_{2j}^2$$

$$= 2\prod_{j=1}^{d}\mu_{4j} - 2\prod_{j\in v}\mu_{4j}\prod_{j\notin v}\mu_{2j}^2.$$

We minimize this expression by taking the largest possible set $v \subseteq u^c$, that is $v = u^c$. □

From Theorem 2, we see that the Correlation 2 estimator minimizes $\mathbb{E}(Q_u^2)$ for product functions among estimators of the form (10).

6 Small effect asymptotics

Here we consider efficiency of our four estimators for small effects. To study small effects we let

$$f(\boldsymbol{x}) = \mu + \varepsilon_1 f_1(\boldsymbol{x}_u) + \varepsilon_2 f_2(\boldsymbol{x}_{-u}) + \varepsilon_{12} f_{12}(\boldsymbol{x}), \tag{13}$$

where $\mu \in \mathbb{R}$, f_1 is defined on $[0,1]^{|u|}$, with $0 < \int |f_1(\boldsymbol{x}_u)|^4 \,\mathrm{d}\boldsymbol{x}_u < \infty$, f_2 is defined on $[0,1]^{d-|u|}$ with $0 < \int |f_2(\boldsymbol{x}_{-u})|^4 \,\mathrm{d}\boldsymbol{x}_{-u} < \infty$, and f_{12} is defined on $[0,1]^d$ with $0 < \int |f_{12}(\boldsymbol{x})|^4 \,\mathrm{d}\boldsymbol{x} < \infty$. Each of ε_1, ε_2 and ε_{12} is either 1 or a common value $\varepsilon \geqslant 0$, leading to these eight functions:

$$\begin{aligned}
f_{000}(\boldsymbol{x}) &= \mu + f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{x}_{-u}) + f_{12}(\boldsymbol{x})\\
f_{001}(\boldsymbol{x}) &= \mu + f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{x}_{-u}) + \varepsilon f_{12}(\boldsymbol{x})\\
f_{010}(\boldsymbol{x}) &= \mu + f_1(\boldsymbol{x}_u) + \varepsilon f_2(\boldsymbol{x}_{-u}) + f_{12}(\boldsymbol{x})\\
f_{011}(\boldsymbol{x}) &= \mu + f_1(\boldsymbol{x}_u) + \varepsilon f_2(\boldsymbol{x}_{-u}) + \varepsilon f_{12}(\boldsymbol{x})\\
f_{100}(\boldsymbol{x}) &= \mu + \varepsilon f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{x}_{-u}) + f_{12}(\boldsymbol{x})\\
f_{101}(\boldsymbol{x}) &= \mu + \varepsilon f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{x}_{-u}) + \varepsilon f_{12}(\boldsymbol{x})\\
f_{110}(\boldsymbol{x}) &= \mu + \varepsilon f_1(\boldsymbol{x}_u) + \varepsilon f_2(\boldsymbol{x}_{-u}) + f_{12}(\boldsymbol{x}), \quad \text{and}\\
f_{111}(\boldsymbol{x}) &= \mu + \varepsilon f_1(\boldsymbol{x}_u) + \varepsilon f_2(\boldsymbol{x}_{-u}) + \varepsilon f_{12}(\boldsymbol{x}).
\end{aligned} \tag{14}$$

We will study the four estimators in the limit as $\varepsilon \to 0$. The last seven functions above all give limits as $\varepsilon \to 0$. The first one, f_{000}, could be any function in $L^4[0,1]^d$ for which the parts f_1, f_2, and f_{12} are all nonzero. The seven limits make various components of this generic function shrink towards zero. The ones that shrink are indicated by a 1 in the triple index to f.

We do not use a term $\varepsilon\mu$. Small norms on f_1, f_2 and f_{12} will sometimes arise in nature. A small mean μ ordinarily arises when one has applied numerical techniques such as quasi-Monte Carlo integration to center the function f. To study small μ we can insert $\mu = 0$ into results for any of the seven limits. The value of μ will only play a role in the asymptotic variance for the Correlation 1 estimator. For $\mu = 0$, the Correlation 1 estimator becomes identical to the Oracle 1 estimator.

For small effects, we focus on f_{101} in the $\varepsilon \to 0$ limit. This limit corresponds to small main effects for $\boldsymbol{x}_u$ and a small interaction between $\boldsymbol{x}_u$ and its complement $\boldsymbol{x}_{-u}$. This is a 'total insensitivity' limit of primary interest for the study of small effects. With only a little more work we can obtain results for all seven limits. For instance, the related limit given by f_{100} is a 'partial insensitivity' limit where $\boldsymbol{x}_u$ has small main effect but potentially large interaction with $\boldsymbol{x}_{-u}$. Some of the other limits are useful for studying large $\underline{\tau}_u^2$.

The variances for our four estimates depend on several quantities derived from f_{000}. We write them as expectations with respect to independent random vectors $\boldsymbol{x}, \boldsymbol{y} \sim \mathbf{U}[0,1]^d$. The first group are variances:

$$\sigma_1^2 = \mathbb{E}\big(f_1(\boldsymbol{x}_u)^2\big), \qquad \sigma_2^2 = \mathbb{E}\big(f_2(\boldsymbol{x}_{-u})^2\big), \qquad \text{and,} \qquad \sigma_{12}^2 = \mathbb{E}\big(f_{12}(\boldsymbol{x})^2\big).$$

Next are various fourth moments, $\kappa_1 = \mathbb{E}\big(f_1(\boldsymbol{x}_u)^4\big)$,

$$\begin{aligned}
\kappa_{1,12} &= \mathbb{E}\big(f_1(\boldsymbol{x}_u)^2 f_{12}(\boldsymbol{x})^2\big) = \mathbb{E}\big(f_1(\boldsymbol{x}_u)^2 f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2\big)\\
\kappa_{12,12} &= \mathbb{E}\big(f_{12}(\boldsymbol{x})^2 f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2\big),\\
\kappa_{1,1,2,12} &= \mathbb{E}\big(f_1(\boldsymbol{x}_u)^2 f_2(\boldsymbol{x}_{-u}) f_{12}(\boldsymbol{x})\big) = \mathbb{E}\big(f_1(\boldsymbol{x}_u)^2 f_2(\boldsymbol{y}_{-u}) f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})\big)\\
\kappa_{2,12,12,12} &= \mathbb{E}\big(f_2(\boldsymbol{x}_{-u}) f_{12}(\boldsymbol{x}) f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2\big)\\
&= \mathbb{E}\big(f_2(\boldsymbol{x}_{-u}) f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2\big), \quad \text{and}\\
\kappa_{2,2,12,12} &= \mathbb{E}\big(f_2(\boldsymbol{x}_{-u}) f_2(\boldsymbol{y}_{-u}) f_{12}(\boldsymbol{x}) f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})\big).
\end{aligned}$$

The names of these fourth moments are just mnemonics. They are not part of a systematic nomenclature for all possible nonzero fourth moments, simply shorthand for the ones that actually appear in our variance expressions. Of these $\kappa_{1,1,2,12}$ and $\kappa_{2,12,12,12}$ can be positive or negative or zero. The others are all positive. Finally, some third moments appear: $\gamma_1 = \mathbb{E}\big(f_1(\boldsymbol{x}_u)^3\big)$ and $\gamma_{1,12,12} = \mathbb{E}\big(f_1(\boldsymbol{x}_u) f_{12}(\boldsymbol{x})^2\big) = \mathbb{E}\big(f_1(\boldsymbol{x}_u) f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2\big)$.

The Correlation 2 estimator satisfies

$$\mathbb{E}((\hat{\underline{\tau}}_u^2(f))^2) = 3\sigma_1^4\varepsilon_1^4 + 3\sigma_{12}^4\varepsilon_{12}^4 + 6\sigma_1^2\sigma_{12}^2\varepsilon_1^2\varepsilon_{12}^2 + \kappa_1\varepsilon_1^4 + \kappa_{1,12}\varepsilon_1^2\varepsilon_{12}^2 + \kappa_{12,12}\varepsilon_1^2\varepsilon_{12}^4$$

when $n = 1$. For $n > 1$, we may subtract ε_1^4 and divide by n, getting an expression of fourth degree in ε_1 and ε_{12} for $\mathrm{Var}(\hat{\underline{\tau}}_u^2; \text{Corr 2})$. That makes Correlation 2 well suited to settings with small ε_1 and ε_{12}. Comparable expressions for all four estimators in all seven asymptotes are given by Theorem 3.

Theorem 3. *Let f be defined as at equation (13). If $n = 1$, then $\mathbb{E}((\hat{\underline{\tau}}_u^2)^2; \text{Corr 1})$, $\mathbb{E}((\hat{\underline{\tau}}_u^2)^2; \text{Corr 2})$, $\mathbb{E}((\hat{\underline{\tau}}_u^2)^2; \text{Orcl 1})$, and $\mathbb{E}((\hat{\underline{\tau}}_u^2)^2; \text{Orcl 2})$ are given by the linear combinations of moments of f_{000} recorded in Table 4.*

Proof. See Section 7. □

Theorem 4. *Let f_{000} through f_{111} be as given at equation (14). For any such functions, the variances $\mathrm{Var}(\hat{\underline{\tau}}_u^2; \text{Corr 1})$, $\mathrm{Var}(\hat{\underline{\tau}}_u^2; \text{Corr 2})$, $\mathrm{Var}(\hat{\underline{\tau}}_u^2; \text{Orcl 1})$, and $\mathrm{Var}(\hat{\underline{\tau}}_u^2; \text{Orcl 2})$ applied to f_{001} through f_{111} attain the rates given in Table 5 as $\varepsilon \to 0$ for fixed $n \geqslant 1$.*

Table 4: The left column lists moments of f_{000} defined at (14). The top row names four estimators of $\underline{\tau}_u^2$. The main body of the table gives the coefficients of each moment in $\mathbb{E}((\hat{\underline{\tau}}_u^2(f_{000}))^2)$ for each estimator (blank for zero). The right three columns gives the order of magnitude in $(\varepsilon_1, \varepsilon_2, \varepsilon_{12})$ for each coefficient when applied to f defined by (13). A blank denotes $O(1)$.

	Corr 1	Corr 2	Orcl 1	Orcl 2	Rates		
σ_1^4	1	3	1		ε_1^4		
σ_2^4				1		ε_2^4	
σ_{12}^4	1	3	1				ε_{12}^4
$\sigma_1^2\sigma_{12}^2$	2	6	2		ε_1^2		ε_{12}^2
$\sigma_2^2\sigma_{12}^2$	2		2	2		ε_2^2	ε_{12}^2
$\sigma_1^2\sigma_2^2$	2		2	2	ε_1^2	ε_2^2	
κ_1	1	1	1	1	ε_1^4		
$\kappa_{1,12}$	2	2	2	2	ε_1^2		ε_{12}^2
$\kappa_{12,12}$	1	1	1	1			ε_{12}^4
$\kappa_{1,1,2,12}$	2		2	4	ε_1^2	ε_2	ε_{12}
$\kappa_{2,12,12,12}$	2		2	3		ε_2	ε_{12}^3
$\kappa_{2,2,12,12}$				4		ε_2^2	ε_{12}^2
$\mu^2\sigma_1^2$	2				ε_1^2		
$\mu^2\sigma_{12}^2$	2						ε_{12}^2
$\mu\gamma_1$	2				ε_1^3		
$\mu\gamma_{1,12,12}$	2				ε_1		ε_{12}^2

Proof. All the variances decrease proportionally to $1/n$ and so it suffices to consider $n = 1$. It is also enough to show that $\mathbb{E}((\hat{\underline{\tau}}_u^2)^2)$ attains the rates shown in Table 5 because $\mathrm{Var}(\hat{\underline{\tau}}_u^2) \leqslant \mathbb{E}((\hat{\underline{\tau}}_u^2)^2)$. By Theorem 3, the value of $\mathbb{E}((\hat{\underline{\tau}}_u^2)^2)$ is given by the moments and coefficients in Table 4.

All four estimators contain the term $\kappa_{12,12} > 0$ which decays as ε_{12}^4 for f. Therefore unless $\varepsilon_{12} \to 0$, the mean square has rate $O(1)$ for all four estimators. Similarly, all four estimators contain the term κ_1 which decays as $O(\varepsilon_1^4)$ for f. Therefore $\varepsilon_1 \to 0$ is necessary for the rate to be better than $O(1)$. These two observations cover five of the seven asymptotes.

In the limit with $\varepsilon_1 = \varepsilon_2 = \varepsilon_{12} = \varepsilon \to 0$, Correlation 2 and both oracles attain $O(\varepsilon^4)$ because all their non-zero coefficients are multiplied by fourth order polynomials in $(\varepsilon_1, \varepsilon_2, \varepsilon_{12})$. The limit for Correlation 1 is only $O(\varepsilon^2)$ due to the presence of $2\mu^2(\sigma_1^2 + \sigma_{12}^2)$.

The remaining limit is the total insensitivity limit with $\varepsilon_1 = \varepsilon_{12} = \varepsilon \to 0$ and $\varepsilon_2 = 1$. Correlation 2 attains $O(\varepsilon^4)$ in this limit because it only has nonzero coefficients for terms which are $O(\varepsilon_1^4)$, $O(\varepsilon_{12}^4)$, and $O(\varepsilon_1^2\varepsilon_{12}^2)$. Oracle 2 has rate $O(1)$ due to the σ_2^2 term. Correlation 1 and Oracle 1 both attain the rate $O(\varepsilon^2)$;

their error is dominated by $2\sigma_2^2(\sigma_1^2+\sigma_{12}^2)$ which is of this rate. □

Remark 1. Theorem 4 proves that $\mathrm{Var}(\hat{\underline{\tau}}_u^2)$ attains the displayed convergence rates by showing that $\mathbb{E}((\underline{\tau}_u^2)^2)$ attains them. That leaves open the possibility that $\mathrm{Var}(\hat{\underline{\tau}}_u^2) = \mathbb{E}((\hat{\underline{\tau}}_u^2)^2) - \mathbb{E}(\hat{\underline{\tau}}_u^2)^2$ attains a strictly better rate than the ones tabulated. This cannot happen, apart from one very peculiar exception described next. To convert a mean square to a variance we subtract $\mathbb{E}(\hat{\underline{\tau}}_u^2)^2 = \sigma_1^4\varepsilon_1^4$. All four expected mean squares contain the term $\kappa_1(f) = (\sigma_1^4 + \mathrm{Var}(f_1(\boldsymbol{x}_u)^2))\varepsilon_1^4$. Subtracting $\sigma_1^4\varepsilon_1^4$ from the expected mean squares makes an adjustment of $O(\varepsilon_1^4)$. The remaining variance still contains the term $\mathrm{Var}(f_1(\boldsymbol{x}_u)^2)\varepsilon_1^4$. Thus the convergence rate does not change if $\mathrm{Var}(f_1(\boldsymbol{x}_u)^2) > 0$. There exist functions $f_1 \neq 0$ with $\mathrm{Var}(f_1(\boldsymbol{x}_u)^2) = 0$ (e.g, f_1 takes only values ± 1). Since we seek rates that holds for all functions f, we use the ones in Table 5.

Remark 2. Correlation 1 and Oracle 1 both use $3n$ function evaluations. From Table 4, the variance of Correlation 1 minus that of Oracle 1 is

$$2\mu^2\sigma_1^2\varepsilon_1^2 + 2\mu^2\sigma_{12}^2\varepsilon_{12}^2 + 2\mu\gamma_1\varepsilon_1^3 + 2\mu\gamma_{1,12,12}\varepsilon_1\varepsilon_{12}^2.$$

On balance this result favors Oracle 1. Two of the terms are sure to be positive and the other two could be either positive or negative. In limits with $\varepsilon_1 \to 0$, the possibly negative terms vanish, leaving Oracle 1 at least as good as Correlation 1. This argument in favor of Oracle 1 supposes that we have already spent the effort to get a good estimate of μ.

6.1 Seven asymptotes

For small effects, the most important asymptote is for f_{101}. In the f_{101} asymptote (total insensitivity) the methods attain different rates. If we consider rela-

Table 5: Rates of convergence of $\mathrm{Var}(\hat{\underline{\tau}}_u^2)$ as $\varepsilon \to 0$ for four estimators and seven asymptotes.

	$\varepsilon_1 \downarrow 0$	$\varepsilon_2 \downarrow 0$	$\varepsilon_{12} \downarrow 0$	Corr 1	Corr 2	Orcl 1	Orcl 2
f_{001}			✓	1	1	1	1
f_{010}		✓		1	1	1	1
f_{011}		✓	✓	1	1	1	1
f_{100}	✓			1	1	1	1
f_{101}	✓		✓	ε^2	ε^4	ε^2	1
f_{110}	✓	✓		1	1	1	1
f_{111}	✓	✓	✓	ε^2	ε^4	ε^4	ε^4

tive errors then

$$\frac{\sqrt{\mathrm{Var}(\hat{\underline{\tau}}_u^2)}}{\underline{\tau}_u^2} = \begin{cases} O(\varepsilon^{-1}) & \text{(Corr. 1)} \\ O(1) & \text{(Corr. 2)} \\ O(\varepsilon^{-1}) & \text{(Orcl. 1)} \\ O(\varepsilon^{-2}) & \text{(Orcl. 2)} \end{cases}$$

as $\varepsilon \to 0$, because $\underline{\tau}_u^2 = O(\varepsilon^2)$. The Correlation 2 estimator is the best in this limit, justifying the recommendation to use it on small effects.

Next we briefly consider the remaining asymptotes in numerical order. The comparisons are summarized in Table 6.

The f_{001} asymptote corresponds to a limit in which the interaction is negligible but both main effects are large. In this case, the limiting variance for Correlation 1 is larger than that for Oracle 1 by a term proportional to $\mu^2\sigma_1^2$. The other estimators satisfy

$$n\mathrm{Var}(\hat{\underline{\tau}}_u^2) \to \mathrm{Var}(f_1(\boldsymbol{x}_u)^2) + \begin{cases} 3\sigma_1^4 & \text{(Corr. 2)} \\ \sigma_1^4 + 2\sigma_1^2\sigma_2^2 & \text{(Orcl. 1)} \\ \qquad 2\sigma_1^2\sigma_2^2 + \sigma_2^4 & \text{(Orcl. 2).} \end{cases}$$

The ranking of methods depends on the relative sizes of σ_1^4, σ_2^4 and $V = \mathrm{Var}(f_1(\boldsymbol{x}_u)^2)$. The best estimate is the one with the smallest value of cost times variance, that is $4\big(V + 3\sigma_1^4\big)$ (Corr. 2) or $3(V + \sigma_1^4 + 2\sigma_1^2\sigma_2^2)$ (Orcl. 1) or $2\big(V + 2\sigma_1^2\sigma_2^2 + 2\sigma_2^4\big)$ (Orcl. 2). Any of these could be best: Correlation 2 has no dependence on σ_2^4 while the others have positive coefficients. Oracle 2 is best when σ_1^2 is very large and also when V is very large. Finally, for large values of σ_2^2/σ_1^2 there is an interval of positive V values on which Oracle 1 is best.

The f_{010} asymptote is one in which $\boldsymbol{x}_{-u}$ becomes unimportant except perhaps through an interaction with $\boldsymbol{x}_u$. Oracle 2 has the smallest coefficients for every component in this limit apart from the term $2(\mu^2(\sigma_1^2 + \sigma_{12}^2) + \mu(\gamma_1 + \gamma_{1,12,12}))$ appearing only in Correlation 1, which can be either positive or negative. That negative value gives Correlation 1 the advantage in some problems where μ has the opposite sign to $\gamma_1 + \gamma_{1,12,12}$ while Oracle 2 has the advantage otherwise.

The f_{011} asymptote corresponds to small f_2 and f_{12}, or equivalently large f_1. In this limit,

$$n\mathrm{Var}(\hat{\underline{\tau}}_u^2) \to \mathrm{Var}(f_1(\boldsymbol{x}_u)^2) + \begin{cases} \sigma_1^4 & \text{(Corr. 1)} \\ 3\sigma_1^4 & \text{(Corr. 2)} \\ \sigma_1^4 & \text{(Orcl. 1)} \\ 0 & \text{(Orcl. 2)} \end{cases}$$

and so the Oracle 2 estimator has an advantage when $\underline{\tau}_u^2$ is large. The advantage is all the stronger because Oracle 2 uses fewer function evaluations than the other estimators.

Table 6: Best methods for estimating $\underline{\tau}_u^2$ as $\varepsilon \to 0$ in seven asymptotes. The last column is the measure by which best beats second best. For four of the asymptotes the best method depends on problem details.

Function	Description	Best method	Measure
f_{001}	Small interaction	depends	
f_{010}	Small $\underline{\tau}_{-u}^2$	depends	
f_{011}	Large $\underline{\tau}_u^2$	Orcl. 2	constant
f_{100}	Small $\underline{\tau}_u^2$	depends	
f_{101}	Small $\overline{\tau}_u^2$	Corr. 2	rate
f_{110}	Large interaction	Orcl. 2	constant
f_{111}	Large μ	depends	

In the f_{100} limit $\boldsymbol{x}_u$ has small main effect but might interact with $\boldsymbol{x}_{-u}$. In this limit Correlation 1 is always worse than Oracle 1 (by $2\mu^2\sigma_{12}^2$). In some cases Oracle 2 is best (large σ_{12}^2/σ_2^2) in others Correlation 2 is best. It is hard to ascertain whether Oracle 1 is ever the best. The difficulty is that $\kappa_{2,12,12,12}$ and $\kappa_{2,2,12,12}$ contribute.

The f_{110} asymptote corresponds to the case where f is dominated by the interaction between $\boldsymbol{x}_u$ and $\boldsymbol{x}_{-u}$ while neither main effect is large. Then

$$n\mathrm{Var}(\hat{\underline{\tau}}_u^2) \to \mathrm{Var}(f_1(\boldsymbol{x}_u)^2) + \begin{cases} \sigma_1^4 + 3\sigma_2^4 & \text{(Corr. 1)} \\ 3\sigma_{12}^4 & \text{(Corr. 2)} \\ \sigma_{12}^4 & \text{(Orcl. 1)} \\ 0 & \text{(Orcl. 2)} \end{cases}$$

so that this limit favors Oracle 2.

The f_{111} asymptote corresponds to problems with large value of μ. All methods attain the same rate except Correlation 1, which is much worse. This case reinforces the importance of centering.

7 Proof of Theorem 3

We need the following moments:

$$\begin{aligned} &\mathbb{E}\bigl(f(\boldsymbol{x})^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}))^2\bigr), && \text{(Correlation 1)} \\ &\mathbb{E}\bigl((f(\boldsymbol{x}) - f(\boldsymbol{z}_u{:}\boldsymbol{x}_{-u}))^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}))^2\bigr), && \text{(Correlation 2)} \\ &\mathbb{E}\bigl((f(\boldsymbol{x}) - \mu)^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}))^2\bigr), && \text{(Oracle 1)} \\ &\mathbb{E}\bigl((f(\boldsymbol{x}) - \mu)^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - \mu)^2\bigr). && \text{(Oracle 2)} \end{aligned}$$

for $f = f_{000}$.

Correlation 1 for $f = f_{000}$

$$\begin{aligned}
&\mathbb{E}\big(f(\boldsymbol{x})^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}))^2\big)\\
&= \mathbb{E}\big((\mu + f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{x}_{-u}) + f_{12}(\boldsymbol{x}))^2(f_1(\boldsymbol{x}_u) - f_1(\boldsymbol{y}_u) + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f_{12}(\boldsymbol{y}))^2\big)\\
&= \mathbb{E}\big((\mu^2 + f_1(\boldsymbol{x}_u)^2 + f_2(\boldsymbol{x}_{-u})^2 + f_{12}(\boldsymbol{x})^2 + 2\mu f_1(\boldsymbol{x}_u) + 2f_2(\boldsymbol{x}_{-u})f_{12}(\boldsymbol{x}))\\
&\qquad \times (f_1(\boldsymbol{x}_u)^2 + f_1(\boldsymbol{y}_u)^2 + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2 + f_{12}(\boldsymbol{y})^2)\big)\\
&= \mu^2\sigma_1^2 + \mu^2\sigma_1^2 + \mu^2\sigma_{12}^2 + \mu^2\sigma_{12}^2\\
&\quad + \kappa_1 + \sigma_1^4 + \kappa_{1,12} + \sigma_1^2\sigma_{12}^2\\
&\quad + \sigma_2^2\sigma_1^2 + \sigma_2^2\sigma_1^2 + \sigma_2^2\sigma_{12}^2 + \sigma_2^2\sigma_{12}^2\\
&\quad + \kappa_{1,12} + \sigma_{12}^2\sigma_1^2 + \kappa_{12,12} + \sigma_{12}^4\\
&\quad + 2\mu\gamma_1 + 2\mu\gamma_{1,12,12}\\
&\quad + 2\kappa_{1,1,2,12} + 2\kappa_{2,12,12,12}\\
&= \sigma_1^4 + \sigma_{12}^4 + 2\sigma_1^2\sigma_{12}^2 + 2\sigma_2^2\sigma_{12}^2 + 2\sigma_1^2\sigma_2^2 + \kappa_1 + 2\kappa_{1,12} + \kappa_{12,12}\\
&\quad + 2\kappa_{1,1,2,12} + 2\kappa_{2,12,12,12} + 2\mu^2\sigma_1^2 + 2\mu^2\sigma_{12}^2 + 2\mu\gamma_1 + 2\mu\gamma_{1,12,12}.
\end{aligned}$$

At the second equality sign most of the cross terms within each square can be omitted because they involve one or more mean zero variables not present in the other quadratic factor. At subsequent steps a few more terms vanish.

Correlation 2 for $f = f_{000}$

$$\begin{aligned}
&\mathbb{E}\big((f(\boldsymbol{x}) - f(\boldsymbol{z}_u{:}\boldsymbol{x}_{-u}))^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}))^2\big)\\
&= \mathbb{E}\big((f_1(\boldsymbol{x}_u) - f_1(\boldsymbol{z}_u) + f_{12}(\boldsymbol{x}) - f_{12}(\boldsymbol{z}_u{:}\boldsymbol{x}_{-u}))^2(f_1(\boldsymbol{x}_u) - f_1(\boldsymbol{y}_u) + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f_{12}(\boldsymbol{y}))^2\big)\\
&= \mathbb{E}\big((f_1(\boldsymbol{x}_u)^2 + f_1(\boldsymbol{z}_u)^2 + f_{12}(\boldsymbol{x})^2 + f_{12}(\boldsymbol{z}_u{:}\boldsymbol{x}_{-u})^2)\\
&\quad \times\ (f_1(\boldsymbol{x}_u)^2 + f_1(\boldsymbol{y}_u)^2 + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2 + f_{12}(\boldsymbol{y})^2)\big)\\
&= \kappa_1 + \sigma_1^4 + \kappa_{1,12} + \sigma_1^2\sigma_{12}^2\\
&\quad + \sigma_1^4 + \sigma_1^4 + \sigma_1^2\sigma_{12}^2 + \sigma_1^2\sigma_{12}^2\\
&\quad + \kappa_{1,12} + \sigma_{12}^2\sigma_1^2 + \kappa_{12,12} + \sigma_{12}^4\\
&\quad + \sigma_{12}^2\sigma_1^2 + \sigma_{12}^2\sigma_1^2 + \sigma_{12}^4 + \sigma_{12}^4\\
&= 3\sigma_1^4 + 3\sigma_{12}^4 + 6\sigma_1^2\sigma_{12}^2 + \kappa_1 + 2\kappa_{1,12} + \kappa_{12,12}.
\end{aligned}$$

Oracle 1 for $f = f_{000}$

$$\begin{aligned}
&\mathbb{E}\big((f(\boldsymbol{x}) - \mu)^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f(\boldsymbol{y}))^2\big)\\
&= \mathbb{E}\big((f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{x}_{-u}) + f_{12}(\boldsymbol{x}))^2(f_1(\boldsymbol{x}_u) - f_1(\boldsymbol{y}_u) + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - f_{12}(\boldsymbol{y}))^2\big)\\
&= \mathbb{E}\big((f_1(\boldsymbol{x}_u)^2 + f_2(\boldsymbol{x}_{-u})^2 + f_{12}(\boldsymbol{x})^2 + 2f_1(\boldsymbol{x}_{-u})f_{12}(\boldsymbol{x}))\\
&\quad \times (f_1(\boldsymbol{x}_u)^2 + f_1(\boldsymbol{y}_u)^2 + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2 + f_{12}(\boldsymbol{y})^2)\big)
\end{aligned}$$

$$
\begin{aligned}
&= \kappa_1 + \sigma_1^4 + \kappa_{1,12} + \sigma_1^2\sigma_{12}^2 \\
&\quad + \sigma_2^2\sigma_1^2 + \sigma_2^2\sigma_1^2 + \sigma_2^2\sigma_{12}^2 + \sigma_2^2\sigma_{12}^2 \\
&\quad + \kappa_{1,12} + \sigma_{12}^2\sigma_1^2 + \kappa_{12,12} + \sigma_{12}^4 \\
&\quad + 2\kappa_{1,1,2,12} + 2\kappa_{2,12,12,12} \\
&= \sigma_1^4 + \sigma_{12}^4 + 2\sigma_1^2\sigma_{12}^2 + 2\sigma_2^2\sigma_{12}^2 + 2\sigma_1^2\sigma_2^2 + \kappa_1 + 2\kappa_{1,12} + \kappa_{12,12} + 2\kappa_{1,1,2,12} + 2\kappa_{2,12,12,12}.
\end{aligned}
$$

Oracle 2 for $f = f_{000}$

$$
\begin{aligned}
&\mathbb{E}\big((f(\boldsymbol{x}) - \mu)^2(f(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}) - \mu)^2\big) \\
&= \mathbb{E}\big((f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{x}_{-u}) + f_{12}(\boldsymbol{x}))^2(f_1(\boldsymbol{x}_u) + f_2(\boldsymbol{y}_{-u}) + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u}))^2\big) \\
&= \mathbb{E}\big((f_1(\boldsymbol{x}_u)^2 + f_2(\boldsymbol{x}_{-u})^2 + f_{12}(\boldsymbol{x})^2 + 2f_2(\boldsymbol{x}_{-u})f_{12}(\boldsymbol{x})) \\
&\quad \times (f_1(\boldsymbol{x}_u)^2 + f_2(\boldsymbol{y}_{-u})^2 + f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})^2 + 2f_2(\boldsymbol{y}_{-u})f_{12}(\boldsymbol{x}_u{:}\boldsymbol{y}_{-u})\big) \\
&= \kappa_1 + \sigma_1^2\sigma_2^2 + \kappa_{1,12} + 2\kappa_{1,1,2,12} \\
&\quad + \sigma_1^2\sigma_2^2 + \sigma_2^4 + \sigma_2^2\sigma_{12}^2 \\
&\quad + \kappa_{1,12} + \sigma_2^2\sigma_{12}^2 + \kappa_{12,12} + \kappa_{2,12,12,12} \\
&\quad + 2\kappa_{1,1,2,12} + 2\kappa_{2,12,12,12} + 4\kappa_{2,2,12,12} \\
&= \sigma_2^4 + 2\sigma_2^2\sigma_{12}^2 + 2\sigma_1^2\sigma_2^2 + \kappa_1 + 2\kappa_{1,12} + \kappa_{12,12} + 4\kappa_{1,1,2,12} + 3\kappa_{2,12,12,12} + 4\kappa_{2,2,12,12}.
\end{aligned}
$$

8 Discussion

This paper has emphasized the estimation of small Sobol' indices. Small $\overline{\tau}_u^2$ and small $\underline{\tau}_u^2$ have different interpretations. The former means that $\boldsymbol{x}_u$ has little impact on f even counting interactions with $\boldsymbol{x}_{-u}$. It might then be reasonable to freeze the value of $\boldsymbol{x}_u$ at a default value while devoting more attention to $\boldsymbol{x}_{-u}$. See Sobol et al. (2007). Small $\underline{\tau}_u^2$ implies that $\boldsymbol{x}_u$ acting alone are unimportant but leaves open the possibility of an important interaction between $\boldsymbol{x}_u$ and $\boldsymbol{x}_{-u}$.

The usual estimator of $\overline{\tau}_u^2$ is quite good when true value is small. This work has provided a good estimator of $\underline{\tau}_u^2$ for cases with small $\underline{\tau}_u^2$. The simulations of Sobol's g-function in Table 2 showed efficiency gains of hundreds to thousands for Correlation 2 versus Correlation 1. The gains were greatest for small effects; for larger $\underline{\tau}_u^2$, Oracle 2 gained even more efficiency than Correlation 2. For the product function in Table 3, the same trends are apparent, but the magnitudes of the differences are smaller.

The new Correlation 2 estimator is the most accurate of our four estimates of $\underline{\tau}_u^2$ when $\sigma_2^2 = \underline{\tau}_{-u}^2$ is very large, and hence $\sigma_2^2 + \sigma_{12}^2 = \overline{\tau}_u^2$ is relatively small. We see this from Table 4 in which Correlation 2 is the only estimator for which σ_2^2 makes no contribution to the variance. In Table 5 this effect shows up as a better rate of convergence in the total insensitivity limit.

The simulations show better performance for Correlation 2 when $\underline{\tau}_u^2$ is small, and the theory shows it is better when $\overline{\tau}_u^2$ is small. For product functions $\underline{\tau}_u^2$ and $\overline{\tau}_u^2$ tend to be large or small together.

From a purely mathematical point of view, no function in $L^2[0,1]^d$ or $L^4[0,1]^d$ might seem more likely than any other to be the one under study. But real world applications often have some special features. The phenomenon of "factor sparsity" has long been used to motivate statistical experimental designs in science and engineering. Box and Meyer (1986) include a discussion and some history of factor sparsity. Under factor sparsity, a small number of input variables will be relatively important while most variables are relatively unimportant. At extreme levels, factor sparsity becomes a tautology: we cannot have more than k inputs each of which is individually responsible for more than $1/k$ of the function's variance. Thus it seems reasonable to expect many or even most of the factors to be relatively unimportant and perhaps a few to be important, especially in high dimensional settings.

A second concept from statistical modeling is that of hierarchy. Standard advice when building a predictive model is to only consider interactions among variables with main effects, and to only consider joint effects among variables x_j for $j \in u$ when all proper subsets $v \subset u$ are also included. Cox (1984) states that as a general principle large interactions are more likely to arise between factors having large main effects. On those grounds we should consider functions with small $\underline{\tau}_u^2$ but large $\overline{\tau}_u^2$ to be unusual in applications, though of course possible.

The ideas behind factor sparsity and hierarchy have recently been embodied in models for weighted Sobolev spaces. Hickernell (1996) introduces weights to downplay high order interactions and Sloan and Woźniakowski (1998) introduces weighted spaces with monotonically diminishing variable importance.

The assumptions behind hierarchical models are better described by the total insensitivity limit in which $\overline{\tau}_u^2$ is also small when $\underline{\tau}_u^2$ is. It is in this limit that the Correlation 2 estimator does very well. In the partial insensitivity limit the differences were not as stark.

Sometimes we may know on physical grounds or from prior experimentation which indices are small and which are large. In other settings some preliminary investigations with small n will be needed to decide which effects are likely to be small and which large before following up with Correlation 2 or Oracle 2.

This paper has focussed on Monte Carlo sampling because it is much easier to study. Whichever estimator one uses, there are likely to be gains from quasi-Monte Carlo Dick and Pillichshammer (2010) and randomized quasi-Monte Carlo Lemieux (2009) sampling. Indeed, Sobol' (1993) reports such gains from QMC. Both of these techniques become more effective on integrands of small norm (though many other factors are also important) and so we might expect Oracle 2 to do best there on large effects and Correlation 2 to do best on small effects.

Acknowledgments

I thank Sergei Kucherenko for helpful discussions and three anonymous reviewers for valuable comments.

References

Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. Technometrics, 28(1):11–18.

Cox, D. (1984). Interaction. International Statistical Review, 52(1):1–24.

Dick, J. and Pillichshammer, F. (2010). Digital sequences, discrepancy and quasi-Monte Carlo integration. Cambridge University Press, Cambridge.

Hickernell, F. J. (1996). Quadrature error bounds with applications to lattice rules. SIAM Journal of Numerical Analysis, 101(5):1995–2016.

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2012). Asymptotic normality and efficiency of two Sobol' index estimators. Technical report, INRIA.

Kucherenko, S., Feil, B., Shah, N., and Mauntz, W. (2011). The identification of model effective dimensions using global sensitivity analysis. Reliability Engineering & System Safety, 96(4):440–449.

Lemieux, C. (2009). Monte Carlo and quasi-Monte Carlo Sampling. Springer, New York.

Liu, R. and Owen, A. B. (2006). Estimating mean dimensionality of analysis of variance decompositions. Journal of the American Statistical Association, 101(474):712–721.

Mauntz, W. (2002). Global sensitivity analysis of general nonlinear systems. Master's thesis, Imperial College. Supervisors: C. Pantelides and S. Kucherenko.

Monod, H., Naud, C., and Makowki, D. (2006). Uncertainty and sensitivity analysis for crop models. In Wallach, D., Makowski, D., and Jones, J. W., editors, Working with dynamic crop models: evaluation, analysis, parametrization and examples, pages 55–99. Elsevier.

Owen, A. B. (2012). Variance components and generalized Sobol' indices. Technical report, Stanford University.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications, 145:280–297.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). Global Sensitivity Analysis. The Primer. John Wiley & Sons, Ltd, New York.

Sloan, I. H. and Woźniakowski, H. (1998). When are quasi-Monte Carlo algorithms efficient for high dimensional integration? Journal of Complexity, 14:1–33.

Sobol', I. M. (1990). On sensitivity estimation for nonlinear mathematical models. Matematicheskoe Modelirovanie, 2(1):112–118. (In Russian).

Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiment, 1:407–414.

Sobol', I. M. and Myshetskaya, E. E. (2007). Monte Carlo estimators for small sensitivity indices. Monte Carlo methods and their applications, 13(5–6):455–465.

Sobol, I. M., Tarantola, S., Gatelli, D., Kucherenko, S. S., and Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. Reliability Engineering & System Safety, 92(7):957–960.