Sobol' Indices: an Introduction and Some Recent Results

Art B. Owen

Stanford University

Ilya Meerovich Sobol'

At MCM 2001, Salzburg

Known for Sobol' sequences
and Sobol' indices

Every time I read one of his papers,
I wish I'd read it earlier

Outline

1) ANOVA

originated in agriculture, used in medicine & industry

2) Global sensitivity

ANOVA-based measures of variable importance

3) Estimation

Pick-freeze methods

4) Use in quasi-Monte Carlo

effective dimension and mean dimension

5) Extensions

Generalized Sobol' indices

New results

- Sobol' indices and Shapley value
- Explaining extremes, not just variance
- New sampling algorithms

You're invited to

MCQMC 2016

At Stanford University

August 14-19, 2016

Overlapping topics with UQ2016

mcqmc2016.stanford.edu

Black box functions

Suppose $y = f(\boldsymbol{x})$ computes

- electrical properties of a semi-conductor, or
- lift and drag of a plane's wing, or
- projections under a climate change model, or
- predicted effects of a malaria eradication strategy,
- etc.

We want to understand f. Usually there is no closed form, just code. Often the code is slow.

- which inputs are most important?
- which interactions (if any) are important?

The main use of Sobol' indices is quantifying importance of variables

Global sensitivity analysis

For books giving context and uses see:

Fang, Li & Sudijanto (2010), Saltelli, Chan & Scott (2009), Saltelli, Ratto & Andres (2008), Cacuci, Ionescu-Bujor & Navon (2005), Saltelli, Tarantola & Campolongo (2004), Santner, Williams & Notz (2003)

Many scientific communities participate, many terms:

FANOVA DACE FAST SAMO MASCOT UCM HDMR NPUA UQ

Kriging approach

Sacks, Welch, Mitchell, Wynn, Ylvisaker, Currin, Morris, Yu, Kleijnen, Koehler, O'Hagan, Kennedy, Stein, Ginsbourger, Roustant, $\cdots$

Derivative based

Sobol', Kucherenko, Shah, Rodriguez-Fernandez, Pantelides, Iooss, Gamboa, Popelin, Lamboni

Survey Iooss & Lemaitre (2015)

How important is importance?

This talk is mostly about ways to define and estimate importance of variables.

Attention can then be focussed on the most important variables.

One risk: measuring importance may be a very expensive first step.

Surrogate

It may be necessary to do this on a surrogate, sampling $\tilde{f} \approx f$ from, e.g., kriging or polynomials.

E.g., Iooss, Van Dorpe & Devictor (2006), Wang, Lu, Tang (2013)

Or directly Oakley & O'Hagan (2004), Chen, Jin, Sujianto (2005), Marrel, Iooss, Laurent, Roustant (2009)

ANOVA: starting with potatoes

Fisher & MacKenzie (1923)

Studies in crop variation II: The manurial response of different potato varieties

Hypothetical potato yields, Y_{ij}

Four varieties, and 3 fertilizer levels

Yield (kg)	V_1	V_2	V_3	V_4
F_1	109.0	110.9	94.2	125.9
F_2	104.9	113.4	110.1	138.0
F_3	151.8	160.9	111.9	145.0

Potatoes continued

The average yield is $\bar{Y}_{\bullet\bullet} = 123.0$. (index $= \bullet$ for average)

Q: Did fertilizer F_i raise or lower the yield?

A: Subtract 123 from row i and average: $\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$

For fertilizer i : $\frac{1}{J}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{\bullet\bullet})$

F_1	F_2	F_3
-13.0	-6.4	19.4

For variety j : $\frac{1}{I}\sum_{i=1}^{I}(Y_{ij} - \bar{Y}_{\bullet\bullet})$

V_1	V_2	V_3	V_4
-1.1	5.4	-17.6	13.3

These are the 'main effects' for fertilizer and variety respectively.

By construction they sum to zero. $\bar{Y}_{\bullet\bullet}$ is the 'grand mean'.

ANOVA for potatoes

$$\begin{bmatrix} 109.0 & 110.9 & 94.2 & 125.9 \\ 104.9 & 113.4 & 110.1 & 138.0 \\ 151.8 & 160.9 & 111.9 & 145.0 \end{bmatrix} = \begin{bmatrix} 123 & 123 & 123 & 123 \\ 123 & 123 & 123 & 123 \\ 123 & 123 & 123 & 123 \end{bmatrix}$$

$$+ \begin{bmatrix} -13.0 & -13.0 & -13.0 & -13.0 \\ -6.4 & -6.4 & -6.4 & -6.4 \\ 19.4 & 19.4 & 19.4 & 19.4 \end{bmatrix} + \begin{bmatrix} -1.1 & 5.4 & -17.6 & 13.3 \\ -1.1 & 5.4 & -17.6 & 13.3 \\ -1.1 & 5.4 & -17.6 & 13.3 \end{bmatrix}$$

$$+ \begin{bmatrix} 0.1 & -4.5 & 1.8 & 2.6 \\ -10.6 & -8.6 & 11.1 & 8.1 \\ 10.5 & 13.1 & -12.9 & -10.7 \end{bmatrix}$$

The last term is the 'interaction'.

Extensions of ANOVA

From $I \times J$ tables to $I \times J \times K \times \cdots \times Z$

e.g. 2^d designs in industrial statistics Box, Hunter, Hunter (2005)

$$f \in L^2[0,1]^d$$

Embedded N^d grid as $N \to \infty$

Hoeffding (1948), Sobol' (1969), Efron & Stein (1981) and others surveyed in Takemura (1983)

Further generalization

Any d independent inputs: $L^2(\prod_{j=1}^d \mathcal{X}_j)$

Also $d = \infty$ via martingales. O (1997) (Latin supercubes)

ANOVA for $L^2[0,1]^d$

Hoeffding (1948) for U-statistics

Sobol' (1969) for QMC

Efron & Stein (1981) for jackknife

$$
\begin{aligned}
f(\boldsymbol{x}) &= f_{()}() + \sum_{j=1}^{d} f_{(j)}(x_j) + \sum_{j<k} f_{(j,k)}(x_j, x_k) + \cdots + f_{(1,2,\dots,d)}(x_1,\dots,x_d)\\
&= f_{()}() + \sum_{r=1}^{d} \sum_{1\leqslant j_1<j_2<\cdots<j_r\leqslant d} f_{(j_1,j_2,\dots,j_r)}(x_{j_1}, x_{j_2},\dots,x_{j_r})
\end{aligned}
$$

More simply

$$
f(\boldsymbol{x}) = \sum_u f_u(\boldsymbol{x})
$$

Sum over all $u \subseteq \mathcal{D} = \{1,2,\dots,d\}$

Notation

For $u \subseteq \mathcal{D} \equiv \{1, \ldots, d\}$

$$|u| = \mathbf{card}(u)$$

$$-u = u^c = \{1, 2, \ldots, d\} - u$$

$v \subset u$ **strict** subset i.e. $\subsetneq$

If $u = \{j_1, j_2, \ldots, j_{|u|}\}$ then $\boldsymbol{x}_u = (x_{j_1}, \ldots, x_{j_{|u|}})$ and $\mathrm{d}\boldsymbol{x}_u = \prod_{j \in u} \mathrm{d}x_j$

Dependence

$f_u(\boldsymbol{x})$ is a function of $\boldsymbol{x}$ that only depends on $\boldsymbol{x}_u$

$f_u(\boldsymbol{x}) + f_v(\boldsymbol{x})$ is well defined

Recursive definition

Overall mean	$\mu \equiv f_{\varnothing}(\boldsymbol{x}) = \int f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$
Main effect j	$f_{\{j\}}(\boldsymbol{x}) = \int \bigl(f(\boldsymbol{x}) - f_{\varnothing}(\boldsymbol{x})\bigr)\,\mathrm{d}\boldsymbol{x}_{-\{j\}}$
Interaction u	$f_u(\boldsymbol{x}) = \int \bigl(f(\boldsymbol{x}) - \sum_{v \subset u} f_v(\boldsymbol{x})\bigr)\,\mathrm{d}\boldsymbol{x}_{-u}$
	$= \int f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{-u} - \sum_{v \subset u} f_v(\boldsymbol{x})$

ANOVA properties

$$j \in u \implies \int_0^1 f_u(\boldsymbol{x})\,\mathrm{d}x_j = 0 \qquad \text{induction on } |u|$$

$$u \neq v \implies \int f_u(\boldsymbol{x}) f_v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0 \qquad \text{integrate over } j \in (u-v) \cup (v-u)$$

$$\& \quad \int f_u(\boldsymbol{x}) g_v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0$$

Variances

$$\mathrm{Var}(f) \equiv \int (f(\boldsymbol{x}) - \mu)^2\,\mathrm{d}\boldsymbol{x} = \sum_{u \subseteq \mathcal{D}} \sigma_u^2$$

$$\sigma_u^2 = \sigma_u^2(f) = \begin{cases} \int f_u(\boldsymbol{x})^2\,\mathrm{d}\boldsymbol{x} & u \neq \varnothing \\ 0 & u = \varnothing. \end{cases}$$

ANOVA for dependent inputs

- Stone (1984)

Retains $\int f_u(\boldsymbol{x})f_v(\boldsymbol{x})w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0$ for $u \subset v$

- Hooker (1997)

Applies to machine learning functions

- Chastaing, Gamboa & Prieur (2012,2015)

New estimation methods for generalized indices

- Kucherenko, Tarantola & Annoni (2012)

Use Gaussian copula

The challenges

- Conceptual. How should one define importance here? Should it ever be negative?

What if support of x_j depends on x_k? E.g., input space has ‘holes’.

- Computational.

New work from Song, Nelson & Staum (2015) using Shapley value looks super promising.

Sobol's decomposition

Sobol' (1969) obtained the same decomposition.

"Decomposition into summands of different dimension"

Sobol' and Hoeffding

Hoeffding (1948)	**Analysis**	Break f into pieces, one for each f_u
Sobol' (1969)	**Synthesis**	Assemble parts of f to make f_u

In more detail

Sobol' used a complete orthonormal basis of $L^2[0,1]^d$

(tensor product of Haar wavelets).

Then he gathered terms for each $u \subseteq \{1,\dots,d\}$.

Thanks to A. Chouldechova for translation.

Variable importance

How important is $\boldsymbol{x}_u$?

Larger σ_u^2 means that $f_u(\boldsymbol{x})$ contributes more.

We also want to count σ_v^2 for $v \subset u$.

Sobol's (1993) importance measures

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2$$ v contained in u

$$\overline{\tau}_u^2 = \sum_{v \cap u \neq \varnothing} \sigma_v^2$$ v touches u, so interactions count

Large $\underline{\tau}_u^2$ means $\boldsymbol{x}_u$ important

Small $\overline{\tau}_u^2$ means $\boldsymbol{x}_u$ unimportant can be frozen Sobol'

Normalized versions

Normalized versions are analogues of R^2, proportion of variance explained.

Partial/closed sensitivity	$\dfrac{\underline{\tau}_u^2}{\sigma^2}$	S_i for $u = \{i\}$
Total sensitivity	$\dfrac{\overline{\tau}_u^2}{\sigma^2}$	S_i^{tot} for $u = \{i\}$

The denominator is easier to estimate, so focus on numerators.

Units

Normalized sensitivities are dimensionless (% of variance).

If we need an answer in meters or volts or €, etc. then we are back to the numerator.

Interpretation

$$\mathbb{E}(f(\boldsymbol{x}) \mid \boldsymbol{x}_u) = \sum_{v \subseteq u} f_v(\boldsymbol{x})$$

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2 = \mathrm{Var}(\mathbb{E}(f(\boldsymbol{x}) \mid \boldsymbol{x}_u))$$

If you control $\boldsymbol{x}_u$ you control $\underline{\tau}_u^2$ of the variance in f.

If $\underline{\tau}_u^2$ is large, then $\boldsymbol{x}_u$ are important.

$$\overline{\tau}_u^2 = \sigma^2 - \underline{\tau}_{-u}^2$$

$\overline{\tau}_u^2$ includes the interaction between $\boldsymbol{x}_u$ and $\boldsymbol{x}_{-u}$

If an adversary controls $\boldsymbol{x}_u$ then the most 'damage' they can do is $\overline{\tau}_u^2$.

If $\overline{\tau}_u^2$ is small then $\boldsymbol{x}_u$ is **not** important. Can be 'frozen'. Sobol' (1990/1993)

Unimportance is important

1) It lets you focus on the key inputs.

2) Potential for faster code.

Factor sparsity

Often **most** variables are unimportant, e.g., Box & Meyer (1986)

Also: they cannot **all** be relatively important.

Examples

$d = 4$ and $u = \{1, 2\}$

$$\underline{\tau}^2_{\{1,2\}} = \sigma^2_{\{1\}} + \sigma^2_{\{2\}} + \sigma^2_{\{1,2\}}$$

$$\begin{aligned}\overline{\tau}^2_{\{1,2\}} = \sigma^2_{\{1\}} &+ \sigma^2_{\{2\}} + \sigma^2_{\{1,2\}} \\ &+ \sigma^2_{\{1,3\}} + \sigma^2_{\{1,4\}} + \sigma^2_{\{2,3\}} + \sigma^2_{\{2,4\}} \\ &+ \sigma^2_{\{1,3,4\}} + \sigma^2_{\{2,3,4\}} + \sigma^2_{\{1,2,3,4\}}\end{aligned}$$

Brute force estimation

Naive approach for $\underline{\tau}_u^2$:

1) Sample $\boldsymbol{x}_i \in [0,1]^d$ and get $y_i = f(\boldsymbol{x}_i)$ for $i = 1, \dots, n$.

2) Get statistical machine learning estimates of $\hat{f}_v(\boldsymbol{x})$.

3) Put $\hat{\sigma}_v^2 = \int \hat{f}_u(\boldsymbol{x})^2 \,\mathrm{d}\boldsymbol{x}$, $u \neq \varnothing$.

4) Sum: $\widehat{\underline{\tau}}_u^2 = \sum_{v \subseteq u} \hat{\sigma}_v^2$.

This is expensive and has many biases.

Pick-freeze

Sobol' has a much better way, using pick-freeze identities (next).

Hybrid points

$\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$ takes $\boldsymbol{x}_u$ from $\boldsymbol{x}$ and $\boldsymbol{z}_{-u}$ from $\boldsymbol{z}$

For $\boldsymbol{y} = \boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$

$$y_j = \begin{cases} x_j, & j \in u \\ z_j, & j \notin u. \end{cases}$$

Example

$$\boldsymbol{x} = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$$

$$\boldsymbol{z} = (0.9, 0.8, 0.7, 0.6, 0.5, 0.4)$$

$$\boldsymbol{x}_{\{1,2,4,5\}}{:}\boldsymbol{z}_{\{3,6\}} = (0.1, 0.2, 0.7, 0.4, 0.5, 0.4)$$

Fixing methods

Evaluate f at two points:

freeze: repeat some components

pick: independent draws for others

Recall f_u does not depend on $\boldsymbol{x}_{-u}$

Therefore $f_u(\boldsymbol{x}) = f_u(\boldsymbol{x}_u{:}\boldsymbol{x}_{-u}) = f_u(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\ \forall \boldsymbol{z} \in [0,1]^d$

Sobol' (1990/3) used the identities:

$$\underline{\tau}_u^2 = \int f(\boldsymbol{x}) f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} - \mu^2$$

$$\overline{\tau}_u^2 = \frac{1}{2}\int ((f(\boldsymbol{x}) - f(\boldsymbol{x}_{-u}{:}\boldsymbol{z}_u))^2\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

Identity for $\underline{\tau}_u^2$

$$\iint f(\boldsymbol{x})f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

$$= \sum_v \sum_w \iint f_v(\boldsymbol{x})f_w(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} \qquad \text{(ANOVA)}$$

$$= \sum_{v\subseteq\mathcal{D}} \iint f_v(\boldsymbol{x})f_v(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} \qquad \text{(orthogonality)}$$

$$= \sum_{v\subseteq {\color{red}u}} \iint f_v(\boldsymbol{x})f_v(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} \qquad \text{(line integrals over } z_j\text{)}$$

$$= \sum_{v\subseteq u} \int f_v(\boldsymbol{x})^2\,\mathrm{d}\boldsymbol{x} \qquad (f_v \text{ only depends on } \boldsymbol{x}_v \text{ and } v\subseteq u)$$

$$= \mu^2 + \sum_{v\subseteq u} \sigma_u^2$$

$$\equiv \mu^2 + \underline{\tau}_u^2.$$

Identity for $\overline{\tau}_u^2$

$$
\begin{aligned}
&\frac{1}{2}\iint \bigl(f(\boldsymbol{x}) - f(\boldsymbol{x}_{-u}{:}\boldsymbol{z}_u)\bigr)^2 \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}\\
&= \frac{1}{2}\Bigl(\sigma^2 + \mu^2 - 2\bigl(\underline{\tau}_{-u}^2 + \mu^2\bigr) + \sigma^2 + \mu^2\Bigr)\\
&= \sigma^2 - \underline{\tau}_{-u}^2\\
&= \overline{\tau}_u^2.
\end{aligned}
$$

Sobol's identities are like tomography:

global integrals reveal internal structure.

Computation

$\overline{\tau}_u^2$ and $\underline{\tau}_u^2$ can be done via $2d$ dimensional integrals.

Monte Carlo or quasi-Monte Carlo approaches.

Also: Poincaré inequalities let one estimate bounds using derivatives:

Lamboni, Iooss, Popelin, Gamboa (2012), Roustant, Fruth, Iooss, Kuhnt (2014),

Kucherenko, Iooss (2014)

MC or QMC estimation

$$\widehat{\underline{\tau}}_u^2 = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) - \Bigl(\frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)\Bigr)^2$$

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

$\widehat{\underline{\tau}}_u^2$ needs $(\boldsymbol{x}_i, \boldsymbol{z}_{i,-u}) \in [0,1]^{d+|-u|} = [0,1]^{2d-|u|}$

$\widehat{\overline{\tau}}_u^2$ needs $(\boldsymbol{x}_i, \boldsymbol{z}_{i,u}) \in [0,1]^{d+|u|}$

Bias

The subtraction in $\widehat{\underline{\tau}}_u^2$ introduces an annoying bias for MC or randomized QMC.

Even better

$$\underline{\tau}_u^2 = \iint f(\boldsymbol{x})\bigl(f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) - f(\boldsymbol{z})\bigr)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

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

This avoids subtracting $\hat{\mu}^2$. It is unbiased: $\mathbb{E}\bigl(\widehat{\underline{\tau}}_u^2\bigr) = \underline{\tau}_u^2$

Mauntz (2002), Saltelli (2002), Kucherenko, Feil, Shah, Mauntz (2011)

Improved statistical efficiency

$$\widehat{\underline{\tau}}_u^2 = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) - \Biggl(\frac{1}{n}\sum_{i=1}^n \frac{f(\boldsymbol{x}_i)+f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u})}{2}\Biggr)^2$$

From Monod, Naud & Makowski (2006)

Janon, Klein, Lagnoux, Nodet & Prieur (2012)

prove efficiency in a class of estimators $\cdots$ that does not include the unbiased one above.

(Either one could be better for given f)

Simplified Saint-Venant flood model

Lamboni, Iooss, Popelin, Gamboa, (2012)

Overflow in meters at a dyke

$$S = Z_v + H - H_d - C_b, \quad \text{where}$$

$$H = \left(\frac{Q}{BK_s\sqrt{(Z_m - Z_v)/L}} \right)^{3/5} \qquad \text{(max annual river height)}$$

Q	Maximal annual flow	m^3/s	Gumbel$(1013, 558) \cap [500, 3000]$
K_s	Strickler coefficient	$m^{1/3}/s$	$\mathcal{N}(30, 8) \cap [15, \infty)$
Z_v	River downstream level	m	Triangle$(49, 50, 51)$
Z_m	River upstream level	m	Triangle$(54, 55, 56)$
H_d	Dyke height	m	$\mathbf{U}[7, 9]$
C_b	Bank level	m	Triangle$(55, 55.5, 56)$
L	Length of river stretch	m	Triangle$(4990, 5000, 5010)$
B	River width	m	Triangle$(295, 300, 305)$

Reduced from a Navier-Stokes model; Usually we don't see a formula.

The cost model

$$C_p = 1_{S>0m}$$ (flood cost)

$$+ 1_{S\leqslant 0m}(0.2 + 0.8(1 - e^{-1000m^4/S^4}))$$ (dyke maintenance)

$$+ 0.05 \min(H_d m^{-1}, 8)$$ (investment cost, from construction)

in millions of Euros

$\overline{\tau}^2_{\{j\}}/\sigma^2$ for the flood model

$\overline{\tau}^2/\sigma^2$	Q	K_s	Z_v	Z_m	H_d	C_b	L	B
Height H	0.72	0.29	0.0078	0.0077	0	0	7.4×10^{-7}	0.00021
Overflow S	0.35	0.14	0.19	0.0038	0.28	0.036	3.6×10^{-7}	0.00010
Cost C_p	0.48	0.25	0.23	0.0077	0.17	0.039	6.8×10^{-7}	0.00019

From $n = 100{,}000$ runs

Q	Maximal annual flow	m^3/s	Gumbel$(1013, 558) \cap [500, 3000]$
K_s	Strickler coefficient	$m^{1/3}/s$	$\mathcal{N}(30, 8) \cap [15, \infty)$
Z_v	River downstream level	m	Triangle$(49, 50, 51)$
Z_m	River upstream level	m	Triangle$(54, 55, 56)$
H_d	Dyke height	m	$\mathbf{U}[7, 9]$
C_b	Bank level	m	Triangle$(55, 55.5, 56)$
L	Length of river stretch	m	Triangle$(4990, 5000, 5010)$
B	River width	m	Triangle$(295, 300, 305)$

Related ANOVA quantities

1) Superset importance. Used to find parsimonious models.
2) Shapley value. Maybe a more 'fair' importance measure.
3) Effective and mean dimension. Used in quasi-Monte Carlo (QMC) integration.

Superset importance

Statistical and machine learning prediction.

The data are $(\boldsymbol{x}, y)$ pairs. We have a model to predict y by $f(\boldsymbol{x})$.

For parsimony

Remove interaction f_u and all super-effects. Squared error:

$$\Upsilon_u^2 = \sum_{v \supseteq u} \sigma_v^2$$

Liu & O (2006)

Small Υ_u^2 means deleting f_u and f_v for $v \supseteq u$ costs little.

Relevant to Hooker (2004)'s simplifications of black box functions.

For superset importance

After some algebra:

$$\Upsilon_u^2 \equiv \sum_{v \supseteq u} \sigma_v^2 = \frac{1}{2^{|u|}} \iint \Bigl(\sum_{v \subseteq u} (-1)^{|u-v|} f(\boldsymbol{x}_{-v} {:} \boldsymbol{z}_v) \Bigr)^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{z}$$

Mean of a square of differences $\cdots$ always get a non-negative estimate.

From Liu & O (2006)

Generalizes $\overline{\tau}_u^2$ formula from 2 terms to $2^{|u|}$ terms.

As a design

Use n repeats of a $2^{|u|} \times 1^{d-|u|}$ factorial randomly embedded in the unit cube.

Does best in comparisons by Fruth, Roustant, Kuhnt (2012)

Shapley value

15 million Francs

Shapley's (1953) value can be used to quantify the contribution of members to a team.

We need to know what each subset of the team would have accomplished.

Example from Bank of International Settlement

Team	Output value in Swiss Francs
$\varnothing$	0
A	4,000,000
B	4,000,000
C	4,000,000
A,B	9,000,000
A,C	10,000,000
B,C	11,000,000
A,B,C	15,000,000

Q: How should we split the CHF 15,000,000 earned by A, B, C among them?

15 million Francs

Shapley's (1953) value can be used to quantify the contribution of members to a team.

We need to know what each subset of the team would have accomplished.

Example from Bank of International Settlement

Team	Output value in Swiss Francs
$\varnothing$	0
A	4,000,000
B	4,000,000
C	4,000,000
A,B	9,000,000
A,C	10,000,000
B,C	11,000,000
A,B,C	15,000,000

Q: How should we split the CHF 15,000,000 earned by A, B, C among them?

A: Shapley says: A gets CHF 4,500,000, B gets CHF 5,000,000, C gets CHF 5,500,000

Shapley setup

Let team $u \subseteq \mathcal{D} \equiv \{1, 2, \dots, d\}$ create value **val**(u).

Total value is **val**$(\mathcal{D})$.

We attribute ϕ_j of this to $j \in \mathcal{D}$.

Shapley axioms

Efficiency	$\sum_{j=1}^{d} \phi_j = \mathbf{val}(\mathcal{D})$
Dummy	If $\mathbf{val}(u \cup \{i\}) = \mathbf{val}(u)$, all u then $\phi_i = 0$
Symmetry	If $\mathbf{val}(u \cup \{i\}) = \mathbf{val}(u \cup \{j\})$, all $u \cap \{i, j\} = \varnothing$ then $\phi_i = \phi_j$
Additivity	If games **val**, **val**$'$ have values ϕ, ϕ' then **val** + **val**$'$ has value $\phi_j + \phi'_j$

Unique solution

$$\phi_j = \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} (\mathbf{val}(u + j) - \mathbf{val}(u))$$

For variable importance

Let variables $x_1, x_2, \ldots, x_d$ be team members trying to explain f.

The value of any subset u is how much can be explained by $\boldsymbol{x}_u$.

Choose **val**$(u) \equiv \underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2$.

Shapley value

$$\phi_j = \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} (\underline{\tau}_{u+j}^2 - \underline{\tau}_u^2)$$

Shapley $\neq$ Sobol'

The Shapley value ϕ_j is **not** $\underline{\tau}^2_{\{j\}}$ or $\overline{\tau}^2_{\{j\}}$. After some algebra O (2014),

$$\phi_j = \sum_{u \subseteq \{1,\dots,d\}, j \in u} \frac{\sigma^2_u}{|u|}.$$

Bracketing

$$\sigma^2_{\{j\}} = \boldsymbol{\underline{\tau}^2_{\{j\}}} \leqslant \boldsymbol{\phi_j} \leqslant \boldsymbol{\overline{\tau}^2_{\{j\}}} = \sum_{u:j\in u} \sigma^2_u$$

Shapley seems like a more reasonable allocation.

Sobol' indices are easier to compute than Shapley and they provide bounds.

Bracketing holds for any 'totally monotone game' (where the analogue of $\sigma^2_u \geqslant 0$).

Shapley for dependent $\boldsymbol{x}$

Song, Nelson & Staum (2015)

$$\underline{\tau}_u^2 = \mathrm{Var}(\mathbb{E}(f(\boldsymbol{x}) \mid \boldsymbol{x}_u))$$

$$\phi_j = \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} (\underline{\tau}_{u+j}^2 - \underline{\tau}_u^2)$$

$$\geqslant 0$$

Works even if

1) Support of x_2 depends on x_1

2) $x_1 = x_2$ or x_3 is constant

3) $\boldsymbol{x}$ Gaussian

Special cases

$$\phi_j = \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} (\underline{\tau}^2_{u+j} - \underline{\tau}^2_u)$$

When $d = 2$

$$\begin{aligned}\frac{\phi_1}{\sigma^2} &= \frac{1}{2}\Bigl(1 + \frac{\mathrm{Var}(\mathbb{E}(Y\,|\,x_1)) - \mathrm{Var}(\mathbb{E}(Y\,|\,x_2))}{\sigma^2}\Bigr)\\ &= \frac{1}{2}\Bigl(1 + \frac{\mathbb{E}(\mathrm{Var}(Y\,|\,x_2)) - \mathbb{E}(\mathrm{Var}(Y\,|\,x_1))}{\sigma^2}\Bigr).\end{aligned}$$

$\boldsymbol{x} \sim \mathcal{N}(0, \Sigma)$, $f(\boldsymbol{x}) = \boldsymbol{x}^{\mathsf{T}}\beta$

$$\underline{\tau}^2_u = \bigl(\beta_u + \Sigma_{uu}^{-1}\Sigma_{u,-u}\beta_{-u}\bigr)^{\mathsf{T}}\Sigma_{uu}\bigl(\beta_u + \Sigma_{uu}^{-1}\Sigma_{u,-u}\beta_{-u}\bigr)$$

Issue

Neither requires nor provides effects f_u.

Quasi-Monte Carlo (QMC)

QMC context

$$\mu = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

Decomposition

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) = \frac{1}{n}\sum_{u\subseteq\{1,\dots,d\}} f_u(\boldsymbol{x}_i) = \mu + \sum_{|u|>0}\frac{1}{n}\sum_{i=1}^{n} f_u(\boldsymbol{x}_i)$$

Sum of $2^d - 1$ integration errors.

For small cardinality $|u|$ the integration problem is easier.

Favorable when f is dominated by its low dimensional components.

More precise statements require Koksma-Hlawka or similar inequalities.

See Niederreiter (1992), Dick & Pillichshammer (2010)

Effective dimension

Caflisch, Morokoff & O (1997)

f has low effective dimension if it is dominated by $s \ll d$ dimensional parts

Truncation sense

$$f(\boldsymbol{x}) \approx \sum_{u \subseteq \{1,2,\dots,s\}} f_u(\boldsymbol{x})$$

Superposition sense

$$f(\boldsymbol{x}) \approx \sum_{u:|u| \leqslant s} f_u(\boldsymbol{x})$$

These make integration (and some other) problems easier.

Smoothness also matters.

Measures of dimensionality

For $u \neq \emptyset$, let

$$\lfloor u \rfloor = \min\{j \mid j \in u\}$$
$$\lceil u \rceil = \max\{j \mid j \in u\}$$
$$\text{span}(u) = \lceil u \rceil - \lfloor u \rfloor + 1$$

Effective dimension

$$\min s \geqslant 1 \quad \text{with} \quad \sum_{\lceil u \rceil \leqslant s} \sigma_u^2 \geqslant 0.99\sigma^2$$ Truncation sense $u \subseteq \{1, 2, \dots, s\}$

$$\min s \geqslant 1 \quad \text{with} \quad \sum_{|u| \leqslant s} \sigma_u^2 \geqslant 0.99\sigma^2$$ Superposition sense

$$\min s \geqslant 1 \quad \text{with} \quad \sum_{\text{span}(u) \leqslant s} \sigma_u^2 \geqslant 0.99\sigma^2$$ Successive dimensions sense

Superposition & Truncation Caflisch, Morokoff & O (1997)

Successive dimensions L'Ecuyer & Lemieux (2000)

Dimension moments

$$\sum_u |u|\sigma_u^2 = (1+\epsilon)\sigma^2 \implies \sum_u (|u|-1)\sigma_u^2 = \epsilon\sigma^2$$

Then k-fold interactions contribute $\leqslant \epsilon \times \dfrac{\sigma^2}{k-1}, \quad k \geqslant 2$

So f is nearly additive.

$\lceil u \rceil$ is highest coordinate $j \in u$

	Mean	Mean square
Superposition	$\frac{1}{\sigma^2}\sum_u \lvert u\rvert\sigma_u^2$	$\frac{1}{\sigma^2}\sum_u \lvert u\rvert^2\sigma_u^2$
Truncation	$\frac{1}{\sigma^2}\sum_u \lceil u\rceil\sigma_u^2$	$\frac{1}{\sigma^2}\sum_u \lceil u\rceil^2\sigma_u^2$
Successive	$\frac{1}{\sigma^2}\sum_u \mathrm{span}(u)\sigma_u^2$	$\frac{1}{\sigma^2}\sum_u \mathrm{span}(u)^2\sigma_u^2$

Superposition moments in Liu & O (2006)

For mean dimension

$$\sum_{j=1}^{d} \overline{\tau}_j^2 = \sum_{j=1}^{d} \sum_{v\cap\{j\}\neq\varnothing} \sigma_v^2$$
$$= \sum_{v} \sigma_v^2 \sum_{j=1}^{d} 1_{v\cap\{j\}\neq\varnothing}$$
$$= \sum_{v} |v|\sigma_v^2$$

Much easier to estimate than effective dimension s.

Generalizes to $\sum_{u} |u|^k \sigma_u^2$ for $k \geqslant 1$

Liu & O (2006)

Example

Kuo, Schwab, Sloan (2012) consider quadrature for

$$f(\boldsymbol{x}) = \frac{1}{1+\sum_{j=1}^{500} x_j/j!}, \quad \boldsymbol{x} \in [0,1]^{500}$$

$R = 50$ replicated estimates of $\sum_u |u|\sigma_v^2/\sigma^2$ using $n = 10{,}000$
had mean 1.0052 and standard deviation 0.0058.

Upshot

$f(\boldsymbol{x})$ is nearly additive

mean dimension between 1.00356 and 1.00684

(± 2 standard errors)

Generalized Sobol' indices

Arbitrary pick-freeze

For any $u, v \subseteq \mathcal{D}$, let

$$\Theta_{uv} = \iint f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{z}$$

Unbiased estimate

$$\widehat{\Theta}_{uv} = \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) f(\boldsymbol{x}_{i,v}{:}\boldsymbol{z}_{i,-v})$$

Generalized Sobol' index

$$\sum_{u \subseteq \mathcal{D}} \sum_{v \subseteq \mathcal{D}} \Omega_{uv} \Theta_{uv} = \mathrm{tr}(\Omega^{\mathsf{T}} \Theta) \qquad \Theta, \Omega \in \mathbb{R}^{2^d \times 2^d}$$

Choosing Ω lets us estimate many things.

O (2013) JUQ

NXOR

$$\mathrm{XOR}(u,v) = u \triangle v = u \cup v - u \cap v \qquad \text{(exclusive OR)}$$

$$\mathrm{NXOR}(u,v) = \mathrm{XOR}(u,v)^c \qquad \text{(**not** exclusive OR)}$$

$\mathrm{NXOR}(u,v)$ has indices j in **both** u and v or in **neither** u or v

Θ via $\underline{\tau}^2$

$$\Theta_{uv} \equiv \iint f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v}) \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

$$\vdots$$

$$= \mu^2 + \underline{\tau}^2_{\mathrm{NXOR}(u,v)}$$

$$= \mu^2 + \sum_{w \subseteq \mathrm{NXOR}(u,v)} \sigma^2_w$$

Written XNOR in electrical engineering

Non-uniqueness

$$\int \sum_{j=1}^{d} f(\boldsymbol{x})(f(\boldsymbol{x}) - f(\boldsymbol{x}_j{:}\boldsymbol{z}_{-j}))\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} = \sum_u |u|\sigma_u^2$$

$$\frac{1}{2}\int \sum_{j=1}^{d} \bigl(f(\boldsymbol{x}) - f(\boldsymbol{x}_j{:}\boldsymbol{z}_{-j})\bigr)^2\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} = \sum_u |u|\sigma_u^2$$

Special GSIs

1) Mean squares $\Omega = \lambda\lambda^{\mathsf{T}}$

$$\iint \Bigl(\sum_u \lambda_u f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\Bigr)^2 \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

Nonnegative & fast

2) Bilinear (rank one) $\Omega = \lambda\gamma^{\mathsf{T}}$

$$\iint \Bigl(\sum_u \lambda_u f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\Bigr)\Bigl(\sum_v \gamma_v f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v})\Bigr) \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

Fast

3) Simple

$$\iint \Bigl(\sum_u \lambda_u f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\Bigr) f(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

Only uses one row/col of Θ

4) Contrast

$$\sum_u \sum_v \Omega_{u,v} = 0$$

Free of μ^2

N.B.: Here a contrast can also be a sum of squares.

Squares

For a square (or a sum of squares) tr$(\Omega^\mathsf{T}\widehat{\Theta}) \geqslant 0$.

Also $\mathbb{E}\bigl(\text{tr}(\Omega^\mathsf{T}\widehat{\Theta})\bigr) = \text{tr}(\Omega^\mathsf{T}\Theta)$

Therefore tr$(\Omega^\mathsf{T}\Theta) = 0$ implies $\Pr\bigl(\text{tr}(\Omega^\mathsf{T}\widehat{\Theta}) = 0\bigr) = 1$.

GSIs with sum of squares estimators

$$\overline{\tau}_u^2 \qquad \text{and} \qquad \Upsilon_u^2 \qquad \text{and} \qquad \sum_u |u|\sigma_u^2$$

No sum of squares exists for $\underline{\tau}_u^2$ when $|u| < d$

The coefficient of $\sigma_{\mathcal{D}}^2$ is $\sum_u \lambda_u^2$ Never 0 for nontrivial λ

generally $\sum_u \Omega_{uu}$ i.e., tr(Ω)

Same thing happens in ANOVA tables:

every variance component has a contribution from the measurement error.

Cost of a GSI

$C(\Omega)$ counts the # of function evaluations per $(\boldsymbol{x}, \boldsymbol{z})$ pair.

We can have $\mathrm{tr}(\Omega_1^\mathsf{T}\Theta) = \mathrm{tr}(\Omega_2^\mathsf{T}\Theta)$ but $C(\Omega_1) < C(\Omega_2)$.

Three factor interaction

$$\sigma^2_{\{1,2,3\}} = \underline{\tau}^2_{\{1,2,3\}} - \underline{\tau}^2_{\{1,2\}} - \underline{\tau}^2_{\{1,3\}} - \underline{\tau}^2_{\{2,3\}} + \underline{\tau}^2_{\{1\}} + \underline{\tau}^2_{\{2\}} + \underline{\tau}^2_{\{3\}} - \underline{\tau}^2_{\varnothing}$$

$C(\Omega) = 9$ evaluations

$$\begin{aligned} f(\boldsymbol{x})\bigl(&f(x_1,x_2,x_3,z_4,z_5) \\ &- f(x_1,x_2,z_3,z_4,z_5) - f(x_1,z_2,x_3,z_4,z_5) - f(z_1,x_2,x_3,z_4,z_5) \\ &+ f(x_1,z_2,z_3,z_4,z_5) + f(z_1,x_2,z_3,z_4,z_5) + f(z_1,z_2,x_3,z_4,z_5) \\ &- f(z)\bigr) \end{aligned}$$

$C(\Omega) = 6$ evaluations

$$\bigl(f(\boldsymbol{z}) - f(x_1,z_2,z_3,z_4,z_5)\bigr)\times$$
$$\bigl(f(z_1,z_2,z_3,x_4,x_5) - f(z_1,x_2,z_3,x_4,x_5) - f(z_1,z_2,x_3,x_4,x_5) + f(z_1,x_2,x_3,x_4,x_5)\bigr)$$

N.B. The bilinear version is invariant under $f \to f + c$

See O (2013) JUQ for this and generalizations

Bilinear, with $O(d)$ evaluations

Suppose $\lambda_u = 0$ for $|u| \notin \{0, 1, d-1, d\}$. Same for $\gamma_v = 0$.

Then the rule

$$\sum_u \sum_v \lambda_u \gamma_v \iint f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v}) \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

takes $O(d)$ computation $\cdots$ not $O(d^2)$.

Bilinear example

$$\int \Bigl[df(\boldsymbol{z}) - \sum_{j=1}^{d} f(\boldsymbol{x}_j{:}\boldsymbol{z}_{-j})\Bigr]\Bigl[(d-2)f(\boldsymbol{z}) - \sum_{j=1}^{d} f(\boldsymbol{x}_{-j}{:}\boldsymbol{z}_j)\Bigr]\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

Reduces to

$$\sum_{u:|u|=2} \sigma_u^2$$

This is a sum of $\binom{d}{2}$ ANOVA components but it only takes $2d+1$ evaluations of f.

O (2013) JUQ

$O(d)$ pairs, with $k \neq j$

For $j \neq k$, let j represent $\{j\}$ and $-j$ represent $-\{j\}$ etc.

All the XORs

Every u and v is j or k or $-j$ or $-k$.

$\mathcal{D} \equiv \{1, 2, \dots, d\}$.

XOR	$\varnothing$	j	k	$-j$	$-k$	$\mathcal{D}$
$\varnothing$	$\varnothing$	j	k	$-j$	$-k$	$\mathcal{D}$
j	j	$\varnothing$	$\{j,k\}$	$\mathcal{D}$	$-\{j,k\}$	$-j$
$-j$	$-j$	$\mathcal{D}$	$-\{j,k\}$	$\varnothing$	$\{j,k\}$	j
$\mathcal{D}$	$\mathcal{D}$	$-j$	$-k$	j	k	$\varnothing$

All the NXORs

$$
\begin{array}{c} \text{NXOR} \\ \varnothing \\ j \\ -j \\ \mathcal{D} \end{array}
\quad
\begin{array}{cccccc} \varnothing & j & k & -j & -k & \mathcal{D} \end{array}
$$

$$
\begin{array}{c|cccccc}
\text{NXOR} & \varnothing & j & k & -j & -k & \mathcal{D} \\
\hline
\varnothing & \mathcal{D} & -j & -k & j & k & \varnothing \\
j & -j & \mathcal{D} & -\{j,k\} & \varnothing & \{j,k\} & j \\
-j & j & \varnothing & \{j,k\} & \mathcal{D} & -\{j,k\} & -j \\
\mathcal{D} & \varnothing & j & k & -j & -k & \mathcal{D}
\end{array}
$$

For $|u|$ and $|v|$ in $\{0,1,d-1,d\}$.

We can estimate the corresponding $\underline{\tau}^2_{\mathrm{NXOR}(u,v)}$ with $O(d)$ cost per $(\boldsymbol{x},\boldsymbol{z})$ pair.

Saltelli (2002) already noticed this (or at least most of it).

What we can get

After some algebra we can get unbiased estimates of

$$\sum_u |u|\sigma_u^2 \qquad \sum_{|u|=1} \sigma_u^2$$

$$\sum_u |u|^2\sigma_u^2 \qquad \sum_{|u|=2} \sigma_u^2$$

at cost $2d + 2$. (Some parts can be gotten at $C = d + 1$)

O (2013) JUQ

Initial and final segments

Suppose that $x_1, x_2 \cdots x_d$ are used in that order. E.g. time steps in a Markov chain

First j variables

$$(0,j] \equiv \begin{cases} \{1,2,\dots,j\}, & 1 \leqslant j \leqslant d \\ \varnothing, & j = 0 \end{cases}$$

Last $d-j$ variables

$$(j,d] \equiv \begin{cases} \{j+1,\dots,d\}, & 0 \leqslant j \leqslant d-1 \\ \varnothing & j = d \end{cases}$$

There are $2d+1$ of these subsets.

Enumeration

NXOR	$\varnothing$	$(0,j]$	$(0,k]$	$(j,d]$	$(k,d]$	$\mathcal{D}$
$\varnothing$	$\mathcal{D}$	$(j,d]$	$(k,d]$	$(0,j]$	$(0,k]$	$\varnothing$
$(0,j]$	$(j,d]$	$\mathcal{D}$	$-(j,k]$	$\varnothing$	$(j,k]$	$(0,j]$
$(j,d]$	$(0,j]$	$\varnothing$	$(j,k]$	$\mathcal{D}$	$-(j,k]$	$(j,d]$
$\mathcal{D}$	$\varnothing$	$(0,j]$	$(0,k]$	$(j,d]$	$(k,d]$	$\mathcal{D}$

WLOG $j < k$.

Effect of recent variables

Recall, first and last elements of $u \neq \varnothing$:

$$\lfloor u \rfloor = \min\{j \mid j \in u\}$$
$$\lceil u \rceil = \max\{j \mid j \in u\}$$

Recency weighted variance components

$$\sum_{u \subseteq \mathcal{D}} (\lfloor u \rfloor - 1)\sigma_u^2, \quad \text{and,}$$
$$\sum_{u \subseteq \mathcal{D}} (d - \lceil u \rceil)\sigma_u^2.$$

Another measure of how fast $f()$ forgets its initial conditions.

Weighting by $\lfloor u \rfloor (d - \lceil u \rceil + 1)$ also possible.

Test functions

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

$$\int g_j = 0 \quad \int g_j^2 = 1 \quad \text{and} \quad \int g_j^4 < \infty.$$

$$\sigma_u^2 = \prod_{j\in u} \tau_j^2 \times \prod_{j\notin u} \mu_j^2$$

$$g(x) = \sqrt{12}(x - 1/2)$$

Min function

$$f(\boldsymbol{x}) = \min_{1\leqslant j\leqslant d} x_j$$

$$\underline{\tau}_u^2 = \frac{|u|}{(d+1)^2(2d - |u| + 2)}$$

Liu and O. (2006)

$$\sigma^2_{\{1,2,3\}}$$

Product function $\Longrightarrow$ numerically same estimate for simple or bilinear.

Therefore bilinear is better because of lower cost.

For $\min(x)$ and $d = 6$ the bilinear estimator was about 5 times as efficient as the simple one based on $n = 1{,}000{,}000$ $(\boldsymbol{x}, \boldsymbol{z})$ pairs.

$$\Upsilon^2_{\{1,2,3,4\}}$$

Product function with $d = 8$ and $\mu_j = 1$ and $\tau = (4, 4, 3, 3, 2, 2, 1, 1)/4$.

Square beats bilinear:

Measure	Value	R^2	Square's efficiency
$\Upsilon^2_{\{1,2,3,4\}}$	0.558	0.034	14.7
$\Upsilon^2_{\{5,6,7,8\}}$	0.0024	0.000147	2710.0

Hard to beat a sum of squares when the true effect is small.

Lower index $\underline{\tau}_u^2$

No sum of squares is available.

Contrast

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

Simple estimator (bias adjusted)

$$\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) - \biggl(\frac{1}{2n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) + f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u})\biggr)^2$$

The contrast has an advantage on small $\underline{\tau}_u^2$.

The simple estimator sometimes beats it on large ones.

Shapley again

$$\phi_j = \sum_{u:u\cap\{j\}\neq\varnothing} |u|^{-1}\sigma_u^2$$

This is a -1^{st} moment.

Hard to get a nice formula like the ones for 1^{st}, 2^{nd} etc. moments.

Optimal estimates

Let $\eta^2 = \sum_u \delta_u \sigma_u^2$.

We would like

$$\mathbb{E}\big(\hat{\eta}^2\big) = \eta^2 \quad \text{and,} \quad \mathrm{Var}\big(\hat{\eta}^2\big) \times \text{cost} = \text{minimum.}$$

Using variance components theory

Unfortunately $\mathrm{Var}(\hat{\eta}^2)$ depends on 4'th moments

Fortunately There is a theory of **MIN**imum **Q**uadratic **N**orm **UN**biased **E**stimates (MINQUE)*

Unfortunately They do not appear to be available for crossed random effects

Fortunately The computed case gives us more options, e.g., quadrature.

*C. R. Rao (1970s)

Optimality is still an open problem

GSIs so far

Just use 2 inputs, $\boldsymbol{x}$ and $\boldsymbol{z}$

What about 3?

$$\boldsymbol{x}, \quad \boldsymbol{y}, \quad \boldsymbol{z}$$

Small Sobol' indices

When $\overline{\tau}_u^2$ is small it is easy to estimate (sum of squares)

When $\underline{\tau}_u^2$ is small it is hard to estimate.

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

$$\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) - f(\boldsymbol{z}_i)) \qquad \text{Mauntz (2002), Saltelli (2002)}$$

$$\frac{1}{n}\sum_{i=1}^{n} (f(\boldsymbol{x}_i) - c)(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) - f(\boldsymbol{z}_i)) \qquad \text{Sobol' \& Myshetskaya (2007)}$$

The basic estimate can be dominated by noise in $\hat{\mu}$.

Sobol' & Myshetskaya (2007) found an advantage by centering at c near μ.

Averaging small$\times$small

For small $\underline{\tau}_u^2$

Here it pays to use 3 vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x} \in [0,1]^d$

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

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

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

$$\frac{1}{n}\sum_{i=1}^{n} \bigl(f(\boldsymbol{x}_i) - f(\boldsymbol{z}_{i,u}{:}\boldsymbol{x}_{i,-u})\bigr)\bigl(f(\boldsymbol{x}_{i,u}{:}\boldsymbol{y}_{i,-u}) - f(\boldsymbol{y}_i)\bigr) \qquad \text{(Use 3 vectors)} \quad (*)$$

where $(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i) \overset{\text{iid}}{\sim} \mathbf{U}[0,1]^{3d}$ for $i = 1, \dots, n$.

* Double centering should work if $\boldsymbol{x}_u$ is really unimportant. Tiny×tiny

An example

For Sobol's g-function from Sobol' & Myshetskaya (2007) and smallest $\underline{\tau}_{\{j\}}$

Efficiencies

Maunt-Saltelli	Oracle centered	Double oracle	Three vector
1	518	74	4256

These account for varying numbers of function evaluations.

These numbers are for the smallest $\underline{\tau}^2$ and are the most extreme from O (2013)

Some theory

In a limit where n is fixed and $\overline{\tau}_u^2 = O(\epsilon^2)$, $\mathrm{Var}(\hat{\underline{\tau}}_u^2)$ is

Maunt-Saltelli	Oracle centered	Double oracle	Three vector
$O(\epsilon^2)$	$O(\epsilon^2)$	$O(1)$	$O(\epsilon^4)$

Sensitivity at the extremes

Sobol' indices measure the importance of variable subsets.

They are derived from the ANOVA, an L^2 quantity.

Gary Tang (Stanford aero/astro) asked:

> How should Sobol' indices be adapted if we're interested in variables that drive the function to its most extreme values? Maybe an L^p approach with $p > 2$ would work.

This is not verbatim.

Why focus on extremes?

Many physical processes are sensitive to extremes:

- highest temperature for a chemical reaction
- single heaviest vehicle for road damage
- maximum acceleration/stress for mechanical systems

So $\cdots$ which variables dominate the attainment of extremely high or extremely low values for the function $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_d)$?

Simplest solution

Replace $f(\boldsymbol{x})$ by

- $\exp(kf(\boldsymbol{x}))$, some $k > 0$, or,
- $f(\boldsymbol{x})^4$, or,
- $1_{f(\boldsymbol{x})>72}$, or,
- $\cdots$

Then use usual Sobol' indices.

But

Transformations may introduce unnecessary interactions.

$f(\boldsymbol{x})$ may already be dichotomous (e.g., safe vs. not).

So, transformation is not always suitable.

What we get

1) An L^4 version that shows which variables dominate extremes

2) An L^3 version that distinguishes variables dominating high vs low extremes

O, Dick & Chen (2014) Information and Inference

Piston cycle time C

$$C = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0}{T_0} \frac{T_a}{V^2}}}, \quad \text{where}$$

$$V = \frac{S}{2k}\left(\sqrt{A^2 + 4k\frac{P_0 V_0}{T_0} T_a} - A\right) \quad \text{and} \quad A = P_0 S + 19.62M - \frac{kV_0}{S},$$

Variable	Range	Description
M	$[30, 60]$	piston weight (kg)
S	$[0.005, 0.020]$	piston surface area (m^2)
V_0	$[0.002, 0.010]$	initial gas volume (m^3)
k	$[1000, 5000]$	spring coefficient (N/m)
P_0	$[90{,}000, 110{,}000]$	atmospheric pressure (N/m^2)
T_a	$[290, 296]$	ambient temperature (K)
T_0	$[340, 360]$	filling gas temperature (K)

Kenett, Zacks & Amberti (2014), Surjanovic & Bingham (2014)

Spoiler!

	Wt	Area	Vol	Spring	Pres	Amb	Fill
$10^2\underline{\tau}_j^{(2)}$	$\mathbf{0.073}$	$\mathbf{1.088}$	$\mathbf{0.626}$	$\mathbf{0.040}$	0.001	-0.002	-0.002
$10^5\underline{\tau}_j^{(3)}$	-0.096	$\mathbf{8.931}$	$-\mathbf{3.830}$	-0.270	-0.219	-0.206	-0.210
$10^5\underline{\tau}_j^{(4)}$	0.074	$\mathbf{2.258}$	$\mathbf{0.805}$	-0.006	0.000	0.007	0.007

These are estimates (explaining negative Sobol indices)

This function is about 94% additive (in L^2)

Area is most important for extreme positive values

Volume is most important for extreme negative values

Bold means above 5 standard errors. Others are less than 2 s.e.

Approaches

We could generalize

1) Hoeffding's (1947) analysis
2) Sobol's (1969) synthesis
   a) Fourier version
   b) Walsh version
3) Sobol's (1990/1993) pick-freeze identity

For L^2, these all coincide.

For L^p, analysis becomes cumbersome but synthesis and pick-freeze go through.

Precursor 1

Analysis of skewness Wang (2001) for X_{ij}

Define $\bar{X}_{i\bullet}$ $\bar{X}_{\bullet j}$ $\bar{X}_{\bullet\bullet}$ by averaging over index with $\bullet$

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij}-\bar{X}_{\bullet\bullet})^3 = J\sum_{i=1}^{I}(\bar{X}_{i\bullet}-\bar{X}_{\bullet\bullet})^3 + I\sum_{j=1}^{J}(\bar{X}_{ij}-\bar{X}_{i\bullet})^3 + 3\sum_{j=1}^{J}(\bar{X}_{i\bullet}-\bar{X}_{\bullet\bullet})\sum_{j=1}^{J}(\bar{X}_{ij}-\bar{X}_{i\bullet})^2.$$

These terms have interpretations in biology. They can be negative. I and J are treated asymmetrically. (It allows us to have J_i obs at level i.)

Precursor 2

Median polish Tukey (1977), Siegel (1983)

Essentially (but not exactly) an additive L_1 fit

Try to make every row and column have **median** 0 instead of mean 0.

Alternately adjust rows/columns to have median zero

Analysis

$$\mu = \arg\min_m \int |f(\boldsymbol{x}) - m|^p \, \mathrm{d}\boldsymbol{x}$$

$$f_{\{j\}}(\boldsymbol{x}) = \arg\min_m \int |f(\boldsymbol{x}) - \mu - m|^p \, \mathrm{d}\boldsymbol{x}_{-\{j\}}$$

Or

$f_{[u]}(\boldsymbol{x}) \equiv \sum_{v \subseteq u} f_v(\boldsymbol{x})$ minimizes $\int (f(\boldsymbol{x}) - g(\boldsymbol{x}))^2 \, \mathrm{d}\boldsymbol{x}$
over functions $g(\cdot)$ that depend only on $\boldsymbol{x}_u$.

For $p \neq 2$ such projection becomes difficult.

And so

we don't generalize the analysis approach.

Sobol's decomposition

Let $\phi_k(x)$, $k \in \mathbb{I}$ be a complete orthonormal basis of $L^2[0,1]$. Sobol' (1969) used Haar

Assume that $\phi_0(x) \equiv 1$ and define $\mathbb{I}_* = \mathbb{I} - \{0\}$.

For $\boldsymbol{k} = (k_1, \dots, k_d) \in \mathbb{I}^d$ define

$$\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^{d} \phi_{k_j}(x_j)$$

Synthesis

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{I}^d} \beta_{\boldsymbol{k}} \phi_{\boldsymbol{k}}(\boldsymbol{x}), \qquad \beta_{\boldsymbol{k}} = \int f(\boldsymbol{x}) \bar{\phi}_{\boldsymbol{k}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

$$f_u(\boldsymbol{x}) = \sum_{\boldsymbol{k}_u \in \mathbb{I}_*^{|u|}} \beta_{\boldsymbol{k}_u : 0_{-u}} \phi_{\boldsymbol{k}_u : 0_{-u}}(\boldsymbol{x})$$

u are 'active' variables

$$\sigma_u^2 = \int f_u(\boldsymbol{x})^2 \, \mathrm{d}\boldsymbol{x} = \sum_{\boldsymbol{k}_u \in \mathbb{I}_*^{|u|}} |\beta_{\boldsymbol{k}_u : 0_{-u}}|^2, \quad u \neq \varnothing$$

Pick-freeze

For $\boldsymbol{x}, \boldsymbol{z} \in [0,1]^d, \quad \boldsymbol{y} = \boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$ means

$$y_j = \begin{cases} x_j, & j \in u \\ z_j, & j \notin u. \end{cases}$$

We glue together part of $\boldsymbol{x}$ and part of $\boldsymbol{z}$ to form $\boldsymbol{y} = \boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$.

Recall

$$\mu^2 + \underline{\tau}_u^2 = \int f(\boldsymbol{x}) f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{z}$$

$$\overline{\tau}_u^2 = \frac{1}{2} \int (f(\boldsymbol{x}) - f(\boldsymbol{x}_{-u}{:}\boldsymbol{z}_u))^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{z}$$

We will generalize the first one.

Generalized synthesis (Fourier)

For L^p we will take p functions $f_0, f_1, \dots, f_{p-1}$ and define a p-fold inner product. First,

$$f_j(\boldsymbol{x}) = \sum_{k \in \mathbb{Z}^d} \hat{f}_j(\boldsymbol{k}) e^{2\pi i \boldsymbol{k}^\mathsf{T} \boldsymbol{x}}, \quad j = 0, 1, \dots p-1$$

successor $\quad j+ \equiv j + 1 \bmod p$

predecessor $\quad j- \equiv j - 1 \bmod p$

We also use

$$\{x\} \equiv x - \lfloor x \rfloor = x \bmod 1$$

interpreted componentwise for vectors

Fourier multilinear

$$\langle f_0, f_1, \dots, f_{p-1}\rangle_p = \int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f_j\bigl(\{(-1)^j(\boldsymbol{x}_j - \boldsymbol{x}_{j+})\}\bigr)\,\mathrm{d}\boldsymbol{x}_0 \cdots \mathrm{d}\boldsymbol{x}_{p-1}$$

For even p, eg $p = 4$

uses $\{\boldsymbol{x}_0 - \boldsymbol{x}_1\}$ $\{(-1)(\boldsymbol{x}_1 - \boldsymbol{x}_2)\}$ $\{\boldsymbol{x}_2 - \boldsymbol{x}_3\}$ $\{(-1)(\boldsymbol{x}_3 - \boldsymbol{x}_0)\}$

i.e. $\{\boldsymbol{x}_0 - \boldsymbol{x}_1\}$ $\{\boldsymbol{x}_2 - \boldsymbol{x}_1\}$ $\{\boldsymbol{x}_2 - \boldsymbol{x}_3\}$ $\{\boldsymbol{x}_0 - \boldsymbol{x}_3\}$

For odd p, eg $p = 3$

uses $\{\boldsymbol{x}_0 - \boldsymbol{x}_1\}$ $\{\boldsymbol{x}_2 - \boldsymbol{x}_1\}$ $\{\boldsymbol{x}_2 - \boldsymbol{x}_0\}$

For $p = 2$

uses $\{\boldsymbol{x}_0 - \boldsymbol{x}_1\}$ twice getting

$$\int f_0(\{\boldsymbol{x}_0 - \boldsymbol{x}_1\})f_1(\{\boldsymbol{x}_0 - \boldsymbol{x}_1\})\,\mathrm{d}\boldsymbol{x}_0\,\mathrm{d}\boldsymbol{x}_1$$

Fourier multilinear

$$\langle f_0, f_1, \ldots, f_{p-1} \rangle_p = \int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f_j \big(\{ (-1)^j (\boldsymbol{x}_j - \boldsymbol{x}_{j+}) \} \big) \, \mathrm{d}\boldsymbol{x}_0 \cdots \mathrm{d}\boldsymbol{x}_{p-1}$$

Diagonality on Fourier basis functions

$$\langle \phi_{\boldsymbol{k}_0}, \phi_{\boldsymbol{k}_1}, \ldots, \phi_{\boldsymbol{k}_{p-1}} \rangle_p = \begin{cases} 1, & \boldsymbol{k}_j = (-1)^j \boldsymbol{k}_0, \quad j = 1, \ldots, p-1 \\ 0, & \text{else.} \end{cases}$$

Consequences of diagonality

Recall $f_u(\boldsymbol{x}) = \sum_{\boldsymbol{k}_u \in \mathbb{Z}_*^{|u|}} \hat{f}(\boldsymbol{k}_u{:}0_{-u}) e^{2\pi i \boldsymbol{k}_u^\mathsf{T} \boldsymbol{x}_u}$

$\langle f_{u_0}, f_{u_1}, \dots, f_{u_{p-1}} \rangle_p = 0$ unless $u_0 = u_1 = \cdots = u_{p-1}$

Define

$\sigma_p(f) \equiv \langle f, f, \dots, f \rangle_p$ then $\sigma_p(f) = \sum_{u \subseteq \mathcal{D}} \sigma_p(f_u)$

Also

$$\sigma_p(f) = \cdots = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \hat{f}(\boldsymbol{k})^{\lceil p/2 \rceil} \hat{f}(-\boldsymbol{k})^{\lfloor p/2 \rfloor}$$

For even p

$$\sigma_p(f) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \hat{f}(\boldsymbol{k})^{p/2} \overline{\hat{f}(\boldsymbol{k})}^{p/2} = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} |\hat{f}(\boldsymbol{k})|^p.$$

Fourier importance for even $p \geqslant 2$

$$\underline{\tau}_u^{[p]} + \mu^p \equiv \int \prod_{j=0}^{p-1} f\Bigl(\{(-1)^j(\boldsymbol{x}_u^{(j)} - \boldsymbol{x}_u^{(j+)})\}{:}\boldsymbol{y}_{-u}^{(j)}\Bigr) \prod \mathrm{d}\boldsymbol{x}^{(j)} \prod \mathrm{d}\boldsymbol{y}^{(j)}$$

$$\underline{\tau}_u^{[p]} = \sum_{v \subseteq u} \sigma_p(f_v)$$

$$\sigma_p(f_u) = \sum_{\boldsymbol{k}_u \in \mathbb{Z}_*^{|u|}} |\hat{f}(\boldsymbol{k}_u{:}0_{-u})|^p$$

Interpretation

For equal L^2 norm, larger L^p is a measure of sparsity

So important $\sigma_p(f_u)$ have sparser signal

Does not favor any part of the spectrum

Depends on the basis

Walsh multilinear

The same process can be carried out in the Walsh basis.

The key property is again diagonality.

Sparsity in the Walsh basis measures different things than Fourier sparsity.

Major open problem

Can it be done in a modern wavelet basis?

Diagonality does not appear to hold.

Generalizing the pick-freeze identity

$$\underline{\tau}_u^2 + \mu^2 = \int f(\boldsymbol{x}) f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$

$$\underline{\tau}_u^{(p)} + \mu^2 \equiv \int \prod_{k=1}^{p} f\bigl(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}^{(k)}\bigr)\,\mathrm{d}\boldsymbol{x} \prod_{k=1}^{p} \mathrm{d}\boldsymbol{z}^{(k)}$$

One copy of $\boldsymbol{x}_u$ with p copies of $\boldsymbol{z}_{-u}$

For even $p \geqslant 2$

Monotonicity: $u \subseteq v \implies \underline{\tau}_u^{(p)} \leqslant \underline{\tau}_v^{(p)}$

Nonnegativity: $\underline{\tau}_u^{(p)} \geqslant 0 = \underline{\tau}_\varnothing^{(p)}$

Mobius relations

$$\sigma_u^{[p]} = \sum_{v \subseteq u} (-1)^{|u-v|} \underline{\tau}_v^{[p]}$$ Fourier

$$\sigma_{u,\mathrm{wal}}^{[p]} = \sum_{v \subseteq u} (-1)^{|u-v|} \underline{\tau}_v^{[p]}$$ Walsh

$$\sigma_u^{(p)} = \sum_{v \subseteq u} (-1)^{|u-v|} \underline{\tau}_v^{(p)}$$ Sobol' pick-freeze

Piston cycle time C

$$C = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0}{T_0} \frac{T_a}{V^2}}},$$ where

$$V = \frac{S}{2k}\left(\sqrt{A^2 + 4k\frac{P_0 V_0}{T_0} T_a} - A\right)$$ and $$A = P_0 S + 19.62M - \frac{kV_0}{S},$$

Variable	Range	Description
M	$[30, 60]$	piston weight (kg)
S	$[0.005, 0.020]$	piston surface area (m^2)
V_0	$[0.002, 0.010]$	initial gas volume (m^3)
k	$[1000, 5000]$	spring coefficient (N/m)
P_0	$[90{,}000, 110{,}000]$	atmospheric pressure (N/m^2)
T_a	$[290, 296]$	ambient temperature (K)
T_0	$[340, 360]$	filling gas temperature (K)

Kenett, Zacks & Amberti (2014), Surjanovic & Bingham (2014)

Integrals

$$\mu^2 + \underline{\tau}_j^{(2)} =$$

$$\int f\bigl(\{x_j^{(0)} - x_j^{(1)}\}{:}\boldsymbol{z}_{-j}^{(0)}\bigr) f\bigl(\{x_j^{(0)} - x_j^{(1)}\}{:}\boldsymbol{z}_{-j}^{(1)}\bigr)$$

$$\mu^3 + \underline{\tau}_j^{(3)} =$$

$$\int f\bigl(\{x_j^{(0)} - x_j^{(1)}\}{:}\boldsymbol{z}_{-j}^{(0)}\bigr) f\bigl(\{x_j^{(2)} - x_j^{(1)}\}{:}\boldsymbol{z}_{-j}^{(1)}\bigr) f\bigl(\{x_j^{(2)} - x_j^{(0)}\}{:}\boldsymbol{z}_{-j}^{(2)}\bigr), \quad \text{and}$$

$$\mu^4 + \underline{\tau}_j^{(4)} =$$

$$\int f\bigl(\{x_j^{(0)} - x_j^{(1)}\}{:}\boldsymbol{z}_{-j}^{(0)}\bigr) f\bigl(\{x_j^{(2)} - x_j^{(1)}\}{:}\boldsymbol{z}_{-j}^{(1)}\bigr) f\bigl(\cdots\bigr) f\bigl(\{x_j^{(0)} - x_j^{(3)}\}{:}\boldsymbol{z}_{-j}^{(3)}\bigr)$$

Quadrature approach

We study $\bar{f} = f - \mu$

From $n = 7^6 = 117{,}649$ points of a fully scrambled $(0,6,7)$ Faure net in base 7:

$\mu \doteq 0.4624964$ standard error 9.8×10^{-7}

For $\underline{\hat{\tau}}_j^{(3)}$

$$\frac{1}{n}\sum_{i=1}^{n} \bar{f}\bigl(\{x_{i0} - x_{i1}\}{:}\boldsymbol{z}^{(0)}_{\pi_0(i),-j}\bigr) \times \bar{f}\bigl(\{x_{i2} - x_{i1}\}{:}\boldsymbol{z}^{(1)}_{\pi_1(i),-j}\bigr) \times \bar{f}\bigl(\{x_{i2} - x_{i0}\}{:}\boldsymbol{z}^{(2)}_{\pi_2(i),-j}\bigr)$$

Latin supercube sampling

Five randomized Faure nets, $\boldsymbol{x}_i \in [0,1]^4$, $\boldsymbol{z}_i^{(0)}$, $\boldsymbol{z}_i^{(1)}$, $\boldsymbol{z}_i^{(2)}$, $\boldsymbol{z}_i^{(3)} \in [0,1]^6$

4 random permutations $\pi_0, \dots, \pi_3$

10 replicates

Pick-freeze results

	Wt	Area	Vol	Spring	Pres	Amb	Fill
$10^2\underline{\tau}_j^{(2)}$	$\mathbf{0.073}$	$\mathbf{1.088}$	$\mathbf{0.626}$	$\mathbf{0.040}$	0.001	-0.002	-0.002
$10^5\underline{\tau}_j^{(3)}$	-0.096	$\mathbf{8.931}$	$\mathbf{-3.830}$	-0.270	-0.219	-0.206	-0.210
$10^5\underline{\tau}_j^{(4)}$	0.074	$\mathbf{2.258}$	$\mathbf{0.805}$	-0.006	0.000	0.007	0.007

This function is about 94% additive (in L^2)

Area is most important for extreme positive values

Volume is most important for extreme negative values

Bold means above 5 standard errors. Others are less than 2 s.e.

Thanks

- Josef Dick, Su Chen, Rhea Liu co-authors
- Gary Tang (posed the extremes problem)
- Alex Chouldechova (translated Sobol's decomposition)
- National Science Foundation of the U.S. DMS-0906056
- Bertrand Iooss for answering questions about Saint-Venant
- Fabio Nobile + Jim Berger + Clayton Webster
- Nicole Giacomini
- SIAM and its staff members