ANOVA, global sensitivity, Sobol' indices and all that

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

Watch for talks in honour of Sobol'
and session on sensitivity analysis

Outline

1) ANOVA

originated in agriculture, used in medicine & industry

2) Effective dimension

ANOVA helps explain QMC successes

3) Global sensitivity

ANOVA-based measures of variable importance

4) Estimation

of sums of squared ANOVA variances via integrals

Emphasis

This tutorial will look at these topics from the MCQMC point of view:

- as sources of integration problems
- as a way to understand integration
- and related approximation ideas

It omits many thousands of references.

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: $\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$

Fix i & average $Y_{ij} - \bar{Y}_{\bullet\bullet}$ over j for fertilizer i

F_1	F_2	F_3
-13.0	-6.4	19.4

Fix j & average $Y_{ij} - \bar{Y}_{\bullet\bullet}$ over i for variety j

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)

To functions in $L^2[0,1]^d$. Embedded K^d as $K \to \infty$

Hoeffding (1948), Sobol' (1969) and others surveyed in Takemura (1983)

To L^2 functions of d arbitrary independent inputs.

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

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

Goes back to Hoeffding (1948) for U-statistics (skillful reading may be required)

& Efron & Stein (1981) for jackknife

$$f(\boldsymbol{x}) = f_{()}() + \sum_{j=1}^{d} f_{(j)}(x_j) + \sum_{j<k} f_{(j,k)}(x_j, x_k) + \cdots + f_{(1,2,\dots,d)}(x_1, \dots, x_d)$$

$$= f_{()}() + \sum_{r=1}^{d} \sum_{1 \leqslant j_1 < j_2 < \cdots < j_r \leqslant d} f_{(j_1, j_2, \dots, j_d)}(x_{j_1}, x_{j_2}, \dots, x_{j_d})$$

More simply

$$f(\boldsymbol{x}) = \sum_{u} f_u(\boldsymbol{x})$$

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

Notation

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

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

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

$$v \subset u \qquad \textbf{strict} \text{ subset i.e. } \subsetneq$$

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

Recursive definition

For $u \subseteq \{1,\dots,d\}$, $f_u(\boldsymbol{x})$ only depends on x_j for $j \in u$.

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

Dependence

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

$f_u(\boldsymbol{x}) + f_v(\boldsymbol{x})$ makes sense because they're on the same domain

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

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

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

Hybrid points

We often make a new point from parts of two other points.

For $u \subseteq \mathcal{D}$ we glue together $\boldsymbol{x}_u$ and $\boldsymbol{z}_{-u}$ to form $\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$.

Specifically

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

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

Works for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^d, \mathbb{R}^d, \mathbb{N}^d$, or, arbitrary d-fold Cartesian products,

Sobol's decomposition

Let ϕ_i for $i \in \mathbb{I}$ be a complete orthonormal basis of $L^2[0,1]$ with $\phi_0(x) \equiv 1$.

Sobol' (1969) used Haar wavelets. Fourier, or Walsh bases also work.

For $L^2[0,1]^d$

Tensor product basis: $\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^d \phi_{k_j}(x_j)$, $\boldsymbol{k} \in \mathbb{I}^d$

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{I}^d} \beta_{\boldsymbol{k}}\phi_{\boldsymbol{k}}(\boldsymbol{x}) \quad \text{where} \quad \beta_{\boldsymbol{k}} = \int \phi_{\boldsymbol{k}}(\boldsymbol{x}) f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

For Sobol' (1969)

$$f_u = \sum_{\boldsymbol{k}_u \in \mathbb{I}_*^{|u|}} \beta_{\boldsymbol{k}_u:\boldsymbol{0}_{-u}} \prod_{j\in u} \phi_{k_j}(x_j), \quad \mathbb{I}_* \equiv \mathbb{I} - \{0\}$$

f_u has terms for which $\boldsymbol{x}_u$ is 'active'.

Sobol' has a **synthesis** not an **analysis** for this decomposition.

Thanks to A. Chouldechova for translation.

QMC basics

$$\mu = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \quad \text{and} \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i), \quad \boldsymbol{x}_i \in [0,1]^d$$

Koksma-Hlawka inequality

$$|\hat{\mu} - \mu| \leqslant D_n^*(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) V_{\mathrm{HK}}(f)$$

$$D_n^* = O(n^{-1}\log(n)^{d-1}) \quad \text{is attainable}$$

Low dimensional projections of QMC points typically have much smaller discrepancy.

Or, the asymptote sets in earlier.

QMC and Koksma-Hlawka

$$
\begin{aligned}
|\hat{\mu} - \mu| &\leqslant \sum_{|u|>0} \left| \frac{1}{n} \sum_{i=1}^{n} f_u(\boldsymbol{x}_i) - \int f_u(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_u \right| \\
&= \sum_{|u|>0} \left| \frac{1}{n} \sum_{i=1}^{n} f_u(\boldsymbol{x}_i) \right| \\
&\leqslant \sum_{|u|>0} D_n^*(\boldsymbol{x}_{1,u}, \dots, \boldsymbol{x}_{n,u}) V_{\mathrm{HK}}(f_u) \\
&= \sum_{|u|>0} \Bigl(|u|\text{-dim discrep.} \Bigr) \times \Bigl(|u|\text{-dim variation} \Bigr)
\end{aligned}
$$

Favorable case

For every u, either $V_{\mathrm{HK}}(f_u)$ or $D_n^*(\boldsymbol{x}_{i,u})$ is small.

Eg: $V_{\mathrm{HK}}(f_u)$ small whenever $|u|$ large.

Randomized QMC and V_{HK}

$\boldsymbol{x}_i \sim \mathbf{U}[0,1]^d$ individually, and $D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)$ small

Smooth f_u contributes $O(n^{-3/2}\log(n)^{(|u|-1)/2})$ to $\sqrt{\mathbb{E}((\hat{\mu}-\mu)^2)}$

There is a smoothing effect for low order f_u Griebel, Sloan & Kuo (2010,2013)

RQMC survey: L'Ecuyer & Lemieux (2002)

QMC for non-smooth: X. Wang and co-authors

Effective dimension

f has low effective dimension if it is dominated by low dimensional coordinate projections

Use σ_u as a proxy for $V_{\mathrm{HK}}(f_u)$

Crude but we can measure it.

V_{HK} requires more smoothness than σ_u^2

Note smoothing effect for small $|u|$ Griebel, Sloan & Kuo (2010,2013)

Measures of dimensionality

For $u \neq \varnothing$, let

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

Effective dimension

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

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

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

Superposition & Truncation Caflisch, Morokoff & O (1997)

Successive dimensions L'Ecuyer & Lemieux (2000)

Additional notions of subset size may be useful.

Dimension moments

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

If $\sum_u \frac{\sigma_u^2 \lvert u\rvert}{\sigma^2} = 1 + \epsilon$, then f is nearly additive.

Superposition moments in Liu & O (2006)

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?
- how can we get a much faster approximation $\widetilde{f}$ for f?

Given a surrogate $\widetilde{f}$ we may then optimize, integrate, visualize etc.

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 disjoint scientific communities are involved.

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

Derivative based

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

Simplified Saint-Venant flood model

Overflow in meters (Lamboni, Iooss, Popelin, Gamboa, 2012) 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

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 importance measures

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2 \qquad v \text{ contained in } u \text{ (strict)}$$

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

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

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

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

Normalized versions: $\dfrac{\underline{\tau}_u^2}{\sigma^2}$ (partial/closed sensitivity) and $\dfrac{\overline{\tau}_u^2}{\sigma^2}$ (total sensitivity)

Examples

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

$$\underline{\tau}_u^2 \equiv \sum_{v \subseteq u} \sigma_v^2 = \sigma_{\{1\}}^2 + \sigma_{\{2\}}^2 + \sigma_{\{1,2\}}^2$$

$$\begin{aligned}\overline{\tau}_u^2 \equiv \sum_{v \cap u \neq \varnothing} \sigma_v^2 &= \sigma_{\{1\}}^2 + \sigma_{\{2\}}^2 + \sigma_{\{1,2\}}^2 + \sigma_{\{1,3\}}^2 + \sigma_{\{1,4\}}^2 \\ &\quad + \sigma_{\{2,3\}}^2 + \sigma_{\{2,4\}}^2 + \sigma_{\{1,3,4\}}^2 + \sigma_{\{2,3,4\}}^2 + \sigma_{\{1,2,3,4\}}^2\end{aligned}$$

Superset importance

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

Liu & O (2006)

Small Υ_u^2 means deleting f_u and f_v for $v \supseteq u$ (to stay hierarchical) makes little difference.

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

Estimation of $\underline{\tau}_u^2$ and $\overline{\tau}_u^2$

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) Statistical machine learning estimate $\hat{f}_v(\boldsymbol{x})$ for all necessary v.
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.

Sobol' has a much better way.

Fixing methods

Evaluate f at two points:

repeat some components

independent draws for the others.

Recall

$$\boldsymbol{y} = \boldsymbol{x}_u{:}\boldsymbol{z}_{-u} \text{ has } \quad y_j = \begin{cases} x_j, & j \in u \\ z_j, & j \notin u. \end{cases}$$

Sobol' (1990/3) used the identities:

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

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

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

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

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

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

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

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

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

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

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

Sobol's identities are like tomography: global integrals reveal internal structure.

MC or QMC estimation

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

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

Integrals over $[0,1]^{2d}$ or less

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

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

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: $\mathbb{E}\bigl(\widehat{\underline{\tau}}_u^2\bigr) = \underline{\tau}_u^2$

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

Improved statistical efficiency

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

From Monod, Naud & Makowki (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)

$\overline{\tau}^2_{\{j\}}$ 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$(295, 300, 305)$

For mean dimension

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

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

$$= \sum_{v} |v| \sigma_v^2$$

Estimator from Liu & O (2006)

Generalizes to $\sum_{v} |v|^k \sigma_v^2$ for $k \geqslant 1$.

Example

Kuo, Schwab, Sloan (2012) consider quadrature for

$$f(\boldsymbol{x}) = \frac{1}{1 + \sum_{j=1}^{d} x_j^{\alpha}/j!}, \quad 0 < \alpha \leqslant 1.$$

For $\alpha = 1$ and $d = 500$

$R = 50$ replicated estimates of $\sum_v |v|\sigma_v^2/\sigma^2$ 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)

For superset importance

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

Mean of a square of differences $\cdots$ better than differences of means of squares.

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)

Generalized Sobol' indices

What can be attained via fixing methods?

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

Generalized Sobol' index

$$\sum_{u\subseteq\mathcal{D}}\sum_{v\subseteq\mathcal{D}}\Omega_{uv}\Theta_{uv} = \mathrm{tr}(\Omega^{\mathsf{T}}\Theta)$$

$\Theta \in \mathbb{R}^{2^d\times 2^d}$ is the "Sobol' matrix". $\Omega \in \mathbb{R}^{2^d\times 2^d}$ has coefficients.

Redundant (but useful)

We have a 2^{2d} dimensional space of estimators $\cdots$

for a 2^d dimensional space of estimands:

$$\delta_\varnothing\mu^2 + \sum_{|u|>0}\delta_u\sigma_u^2$$

NXOR

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

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

$$\begin{aligned}\Theta_{uv} &\equiv \iint f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} \\ &= \mu^2 + \underline{\tau}^2_{\mathrm{NXOR}(u,v)} \\ \widehat{\Theta}_{uv} &= \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) f(\boldsymbol{x}_{i,v}{:}\boldsymbol{z}_{i,-v})\end{aligned}$$

Use $\mathrm{tr}(\Omega^{\mathsf{T}}\widehat{\Theta})$

often written XNOR

Special GSIs

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

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

Nonnegative

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

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

Fast

3) Simple

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

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.

Cost of a GSI

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

Recipe

1) Count the rows u that are needed for some $f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})$
2) add the columns (where u appears as the needed 'v')
3) subtract any doubly counted subsets

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

Squares

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

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

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

GSIs with sum of squares estimators

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

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

Can show that the coefficient of $\sigma_{\mathcal{D}}^2 = \sum_u \lambda_u^2$

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

Same thing happens in ANOVA tables:

every variance component has a contribution from the measurement error.

Three factor interaction

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

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

9 evaluations versus 6:

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

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

See O (2013) J. Uncertainty Quantification 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 $O(d^2)$.

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

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

All the XORs

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

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

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

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

O (2013) J. Uncertainty Quantification

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

$$
\begin{array}{c}
\text{NXOR} \\
\varnothing \\
(0,j] \\
(j,d] \\
\mathcal{D}
\end{array}
\quad
\begin{array}{cccccc}
\varnothing & (0,j] & (0,k] & (j,d] & (k,d] & \mathcal{D} \\
\end{array}
$$

$$
\begin{bmatrix}
\mathcal{D} & (j,d] & (k,d] & (0,j] & (0,k] & \varnothing \\
(j,d] & \mathcal{D} & -(j,k] & \varnothing & (j,k] & (0,j] \\
(0,j] & \varnothing & (j,k] & \mathcal{D} & -(j,k] & (j,d] \\
\varnothing & (0,j] & (0,k] & (j,d] & (k,d] & \mathcal{D}
\end{bmatrix}
$$

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 $\implies$ 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.

GSIs so far

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

What about 3?

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

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

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

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

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

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

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

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

Simulations: On small effects the new estimator beats both oracles.

Double oracle wins on large effects.

ANOVA grand challenge

What if $\boldsymbol{x} \sim w$ but w is not a product measure? **Very hard**. See:

- 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)

New estimation methods for generalized indices

- Kucherenko, Tarantola & Annoni (2012)

Use Gaussian copula

Sampling designs

- Stein (1987) LHS for dependent data
- Petelet, Iooss, Asserin & Loredo (2010) Linearly constrained LHS

Conclusions

Sums of squares are very good.

Bilinear estimators $\lambda^{\mathsf{T}}\widehat{\Theta}\gamma$ work well, especially when $1^{\mathsf{T}}\gamma = 1^{\mathsf{T}}\lambda = 0$.

Recent work

Extensions from L^2 to L^p with O & Chen & Dick (2014) IMAIAI

Comparison to Shapley value O (2014) JUQ to appear

Further work

1) Pursue variance inequalities
2) Replace plain MC by Quasi-Monte Carlo and/or
3) Find nice confidence intervals for ratios of means over U-statistics
4) Variance reductions

Thanks

- Bertrand Iooss for answering questions about Saint-Venant
- Sergei Kucherenko for extensive discussions
- National Science Foundation of the U.S. DMS-0906056
- Ronald Cools, Dirk Nuyens, and the local organizing committee

Optimal estimates

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

We would like

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

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)