New results for the functional ANOVA and Sobol' indices

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

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

The cost model

$$
\begin{aligned}
C_p &= 1_{S>0m} && \text{(flood cost)}\\
&\quad + 1_{S\leqslant 0m}(0.2 + 0.8(1 - e^{-1000m^4/S^4})) && \text{(dyke maintenance)}\\
&\quad + 0.05\min(H_d m^{-1}, 8) && \text{(investment cost, from construction)}
\end{aligned}
$$

in millions of Euros

Quibble

A discontinuity at $S = 0$ would be better.

Variable importance

Which of these variables is most important?

How important is any given subset of variables taken together?

How should one define/measure variable importance?

Similar problems come up in black box models for

- aerospace,
- semiconductor manufacturing,
- climate modeling, etc.

Why do they care?

The ultimate goal may be to optimize something (a mean or variance or max or min) by choosing levels for those variables under your control.

Understanding a function is an intermediate goal, not an ultimate goal.

Yet $\cdots$ variable importance is widely applied.

Similar question

Why compute R^2?

Sobol' indices yield many analogues of R^2.

Quantitative measures of importance aid discussions.

Quasi-Monte Carlo

QMC sometimes gives accurate estimates for high dimensional integrals.

Despite the curse of dimensionality.*

When it happens, we usually find that the integrand was dominated by low order interactions among the variables.

That understanding motivates trying to lower the 'effective dimension' of our integrands. (Analogous to variance reduction.)

* which never actually said that **all** high dim problems are insoluble.

ANOVA

Fisher (1923) to Hoeffding (1948) to Sobol' (1967) to Efron & Stein (1981)

then back to Sobol' (1990/3)

Recall

$$X_{ij} = \bar{X}_{\bullet\bullet} + (\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet}) + (\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}) + (X_{ij} - \bar{X}_{i\bullet} - \bar{X}_{\bullet j} + \bar{X}_{\bullet\bullet})$$

Take a grand average $\bar{X}_{\bullet\bullet}$

Subtract it out and average over j, getting $\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet}$

Similarly get $\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}$

Generally: subtract away all lower order effects and average out the extra variables.

Extensions of ANOVA

From functions on $\{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\}$ to

$$\{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \times \cdots \times \{1, 2, \ldots, Z\}$$

To functions in $L^2[0, 1]^d$.

To L^2 functions of d arbitrary independent inputs.

Also $d = \infty$ works via martingales.

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

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

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

$v \subset u$ **strict** 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 $j \in u$.

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

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

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

$$u \neq v \implies \int f_u(\boldsymbol{x}) f_v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0$$

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

Variances

$$\mathrm{Var}(f) \equiv \int (f(\boldsymbol{x}) - \mu)^2\,\mathrm{d}\boldsymbol{x} = \sum_{u \subseteq \{1,\dots,d\}} \sigma_u^2$$

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

Sobol's decomposition

Let $\phi_0, \phi_1, \phi_2 \ldots$ be a complete orthonormal basis of $L^2[0,1]$ with $\phi_0(x) \equiv 1$.

Sobol' (1969) expanded $f(\boldsymbol{x})$ in a tensor product basis (Haar wavelets).

He grouped the terms into 2^d subsets depending on which inputs are 'active'.

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

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 importance measures

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

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

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

Normalized versions: $\dfrac{\underline{\tau}_u^2}{\sigma^2}$ and $\dfrac{\overline{\tau}_u^2}{\sigma^2}$

More derived importance measures

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.

Mean dimension

$$\sum_{u \subseteq \mathcal{D}} \frac{\sigma_u^2}{\sigma^2} \times |u|$$

Measures 'dimensionality' of f. Liu & O (2006)

Higher dimensionality makes for harder numerical handling.

Many quadrature problems have mean dimension near 1

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.

Hybrid points

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

Sobol' (1990/3) used the identities:

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

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

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

$$
\begin{aligned}
&\iint f(\boldsymbol{x})f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} \\
&= \sum_{v\subseteq\mathcal{D}} \iint f_v(\boldsymbol{x})f_v(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} && \text{(orthogonality)} \\
&= \sum_{v\subseteq u} \iint f_v(\boldsymbol{x})f_v(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z} && \text{(line integrals)} \\
&= \mu^2 + \sum_{v\subseteq u}\sigma_u^2 \\
&\equiv \mu^2 + \underline{\tau}_u^2.
\end{aligned}
$$

Bias adjustment

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

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

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

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

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

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 estimates are like tomography: integrals reveal internal structure.

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

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

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, though it is hard to quantify near perfect additivity.

(The difficulty seems to be in forming the ratio.)

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 $\Omega = \lambda\lambda^{\mathsf{T}}$

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

Nonnegative

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

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

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.

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) tr$(\Omega^{\mathsf{T}}\widehat{\Theta}) \geqslant 0$.

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

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

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 $|u| < d$

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

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.

Targeting one variance component

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

A simple (contrast) GSI

$$f(\boldsymbol{x}) \sum_{u \subseteq \{1,2,3\}} \lambda_u f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}), \qquad \lambda_u = (-1)^{3-|u|}$$

$$\begin{array}{cc} & \begin{array}{cccccccc} \varnothing & 1 & 2 & 3 & 12 & 13 & 23 & 123 \end{array} \\ \lambda_u & \left[\begin{array}{cccccccc} -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \end{array} \right] \end{array}$$

Cost is $8 + 1 = 9$ function evaluations.

A bilinear GSI

For $u, v \subseteq w \equiv \{1, 2, 3\}$

$$
\begin{array}{cc}
\mathrm{NXOR}(u,v+w^c) & \begin{array}{cccccccc} \varnothing & 1 & 2 & 3 & 12 & 13 & 23 & 123 \end{array} \\
\begin{array}{c} \varnothing \\ 1 \end{array} & \begin{bmatrix} 123 & 23 & 13 & 12 & 3 & 2 & 1 & \varnothing \\ 23 & 123 & 3 & 2 & 13 & 12 & \varnothing & 1 \end{bmatrix}
\end{array}
$$

$$\Omega = \lambda\gamma^{\mathsf{T}}$$

$$
\begin{array}{cc}
 & \begin{array}{cccccccc} \varnothing & 1 & 2 & 3 & 12 & 13 & 23 & 123 \end{array} \\
\begin{array}{c} \lambda \\ \gamma \end{array} & \begin{bmatrix} 1 & 0 & -1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{array}
$$

Coefficients of $\underline{\tau}_u^2$ are in

$$[(123) - (13) - (12) + (1)] \quad - \quad [(23) - (3) - (2) + \varnothing]$$

Cost is 6. (no duplication)

Simple vs. bilinear for $d = 5$

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

Versus

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

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

More generally

Simple estimator at cost $2^{|w|} + 1_{|w|<d}$

$$\sigma_w^2 = \sum_{u \subseteq w} (-1)^{|w-v|} \Theta_{u,\mathcal{D}}$$

Bilinear for $w_1 \subseteq w$ and $w_2 = w - w_1$

$$\sigma_w^2 = \sum_{u_1 \subseteq w_1} \sum_{u_2 \subseteq w_2} (-1)^{|u_1|+|u_2|} \Theta_{u_1, u_2 + w^c}$$

Bilinear cost is $2^{|w_1|} + 2^{|w_2|} \approx 2^{|w|/2+1}$.

$$C_{\text{bilinear}} \approx 2\sqrt{C_{\text{simple}}}$$

Superset importance

Let w be a nonempty subset of $\mathcal{D}$ for $d \geqslant 1$.

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

Choose $w_1 \subseteq w$ and put $w_2 = w - w_1$. Then

$$\Upsilon_w^2 = \sum_{u_1 \subseteq w_1} \sum_{u_2 \subseteq w_2} (-1)^{|u_1|+|u_2|} \Theta_{w^c+u_1, w^c+u_2}$$

$$\text{compare} \quad \sigma_w^2 = \sum_{u_1 \subseteq w_1} \sum_{u_2 \subseteq w_2} (-1)^{|u_1|+|u_2|} \Theta_{u_1, w^c+u_2}.$$

Lower cost than a square estimator but probably much higher variance.

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

$$
\begin{array}{c c}
\mathrm{XOR} & \begin{array}{cccccc} \varnothing & j & k & -j & -k & \mathcal{D} \end{array} \\
\begin{array}{c} \varnothing \\ j \\ -j \\ \mathcal{D} \end{array} &
\begin{bmatrix}
\varnothing & j & k & -j & -k & \mathcal{D} \\
j & \varnothing & \{j,k\} & \mathcal{D} & -\{j,k\} & -j \\
-j & \mathcal{D} & -\{j,k\} & \varnothing & \{j,k\} & j \\
\mathcal{D} & -j & -k & j & k & \varnothing
\end{bmatrix}
\end{array}
$$

All the NXORs

$$
\begin{array}{cc}
\mathrm{NXOR} & \begin{array}{cccccc} \varnothing & j & k & -j & -k & \mathcal{D} \end{array} \\
\begin{array}{c} \varnothing \\ j \\ -j \\ \mathcal{D} \end{array} &
\begin{bmatrix}
\mathcal{D} & -j & -k & j & k & \varnothing \\
-j & \mathcal{D} & -\{j,k\} & \varnothing & \{j,k\} & j \\
j & \varnothing & \{j,k\} & \mathcal{D} & -\{j,k\} & -j \\
\varnothing & j & k & -j & -k & \mathcal{D}
\end{bmatrix}
\end{array}
$$

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 $C = d+1$)

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

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

WLOG $j < k$.

Effect of recent variables

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

$$\sum_{j=1}^{d-1} \bigl(\Theta_{\mathcal{D},(j,d]} - \Theta_{\mathcal{D},\varnothing}\bigr) = \sum_{u \subseteq \mathcal{D}} (\lfloor u \rfloor - 1)\sigma_u^2, \quad \text{and,}$$
$$\sum_{j=1}^{d-1} \bigl(\Theta_{\mathcal{D},(0,j]} - \Theta_{\mathcal{D},\varnothing}\bigr) = \sum_{u \subseteq \mathcal{D}} (d - \lceil u \rceil)\sigma_u^2.$$

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

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

Simple estimator (bias adjusted)

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

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 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x} \in [0,1]^d$

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

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

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

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

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

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

Double oracle wins on large effects.

Conclusions

Sums of squares are very good.

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

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

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)