Robustness of Sobol' indices to distributional uncertainty

SAND2020-2382C

Joseph Hart [1]

with Pierre Gremaud [2]

[1]Sandia National Laboratories[3]
Center for Computing Research
Optimization and UQ Group

[2]North Carolina State University
Department of Mathematics

SIAM UQ
March 26, 2020

[3]Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. SAND No: XXXX

Hierarchical Decomposition

Consider a function $f : \mathbb{R}^p \to \mathbb{R}$ and assume

- $\mathbf{X} = (X_1, X_2, \dots, X_p)$ has some known probability distribution
- $f(\mathbf{X})$ is square integrable

Hierarchical Decomposition

Consider a function $f : \mathbb{R}^p \to \mathbb{R}$ and assume

- $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ has some known probability distribution
- $f(\mathbf{X})$ is square integrable

Then

$$f(\mathbf{X}) = f_0 + \sum_i f_i(X_i) + \sum_{i<j} f_{i,j}(X_i, X_j) + \cdots = f_0 + \sum_{k=1}^{p} \sum_{|u|=k} f_u(\mathbf{X}_u)$$

where

$$f_u(\mathbf{X}_u) = \mathbb{E}(f(\mathbf{X})|\mathbf{X}_u) - \sum_{v \subset u} f_v(\mathbf{X}_u), \qquad u \subset \{1, 2, \ldots, p\}$$

is a function of only $\mathbf{X}_u$.

Hierarchical Decomposition

Consider a function $f : \mathbb{R}^p \to \mathbb{R}$ and assume

- $\mathbf{X} = (X_1, X_2, \dots, X_p)$ has some known probability distribution
- $f(\mathbf{X})$ is square integrable

Then

$$f(\mathbf{X}) = f_0 + \sum_i f_i(X_i) + \sum_{i<j} f_{i,j}(X_i, X_j) + \cdots = f_0 + \sum_{k=1}^{p} \sum_{|u|=k} f_u(\mathbf{X}_u)$$

where

$$f_u(\mathbf{X}_u) = \mathbb{E}(f(\mathbf{X})|\mathbf{X}_u) - \sum_{v \subset u} f_v(\mathbf{X}_u), \qquad u \subset \{1, 2, \dots, p\}$$

is a function of only $\mathbf{X}_u$.

- this hierarchical decomposition exists $\forall f$ such that $f(\mathbf{X}) \in L^2$
- if $X_1, X_2, \dots, X_p$ are statistically independent then

$$\mathbb{E}(f_u(\mathbf{X}_u) f_v(\mathbf{X}_v)) = 0 \quad u \neq v$$

$$\implies \text{Var}(f(\mathbf{X})) = \sum_{k=1}^{p} \sum_{|u|=k} \text{Var}(f_u(\mathbf{X}_u))$$

Sobol' Indices with Independent Variables

Assuming independence, use the decomposition of variance

$$\mathrm{Var}(f(\mathbf{X})) = \sum_{k=1}^{p} \sum_{|u|=k} \mathrm{Var}(f_u(\mathbf{X}_u)),$$

to define the Sobol' indices as [1]

$$S_u = \frac{\mathrm{Var}(f_u(\mathbf{X}_u))}{\mathrm{Var}(f(\mathbf{X}))}, \qquad u \subset \{1, 2, \dots, p\}.$$

and define the k^{th} *total Sobol' index* as

$$T_k = \sum_{u \cap \{k\} \neq \emptyset} S_u$$

[1] Sensitivity estimates for non linear mathematical models. Sobol'. 1993

Sobol' Indices with Independent Variables

Assuming independence, use the decomposition of variance

$$\mathrm{Var}(f(\mathbf{X})) = \sum_{k=1}^{p} \sum_{|u|=k} \mathrm{Var}(f_u(\mathbf{X}_u)),$$

to define the Sobol' indices as [1]

$$S_u = \frac{\mathrm{Var}(f_u(\mathbf{X}_u))}{\mathrm{Var}(f(\mathbf{X}))}, \qquad u \subset \{1, 2, \dots, p\}.$$

and define the k^{th} *total Sobol' index* as

$$T_k = \sum_{u \cap \{k\} \neq \emptyset} S_u$$

- S_k only measures contribution of X_k
- T_k measures the contribution of all interactions involving X_k
- $\{S_k, T_k\}_{k=1}^{p}$ may be estimated via Monte Carlo integration with $N(p+2)$ evaluations of f, where N is the number of Monte Carlo samples

[1] Sensitivity estimates for non linear mathematical models. Sobol'. 1993

Users Guide to Sobol' Indices

1. Define distributions for $X_1, X_2, \ldots, X_p$
2. Draw samples from these distributions
3. Evaluate f at these samples
4. Estimate the Sobol' indices using these evaluations of f

Users Guide to Sobol' Indices

1. Define distributions for $X_1, X_2, \ldots, X_p$
2. Draw samples from these distributions
3. Evaluate f at these samples
4. Estimate the Sobol' indices using these evaluations of f

What if we do not know the distributions of $X_1, X_2, \ldots, X_p$?

- Distributions are typically defined from limited data and/or knowledge.
- They are frequently uncertain.
- How robust are the Sobol' indices to changes in the distributions?

Goal

Goal: Develop a numerical method to assess the robustness of the Sobol' indices to distributional uncertainty that:

- Does not require a priori knowledge about parametric form of the distributions of $X_1, X_2, \ldots, X_p$
- Does not require additional evaluations of f beyond what is computed to estimate Sobol' indices

Assumptions

Assume that,

- X_1, X_2,...,X_p are independent
- X_i is supported on a compact interval Ω_i with PDF ϕ_i, $i = 1, 2, \dots, p$
- ϕ_i is continuous on Ω_i, $i = 1, 2, \dots, p$
- f is bounded on $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_p$

Assumptions

Assume that,

- X_1, X_2,...,X_p are independent
- X_i is supported on a compact interval Ω_i with PDF ϕ_i, $i = 1, 2, \ldots, p$
- ϕ_i is continuous on Ω_i, $i = 1, 2, \ldots, p$
- f is bounded on $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_p$

Idea:

- define a Banach space in which we may vary marginal PDFs
- compute the Fréchet derivative of the Sobol' indices with respect to the marginal PDFs (only requires existing function evaluations)
- define "optimal perturbations" according to the directions of steepest ascent and estimate perturbed Sobol' indices (using existing function evaluations)

Banach Space Containing PDF's

Define V_i as the Banach space of bounded functions on Ω_i with

$$||\psi||_{V_i} = \left|\left| \frac{\psi}{\phi_i} \right|\right|_{L^\infty(\Omega_i)}$$

for $i = 1, 2, \ldots, p$ and $V = V_1 \times V_2 \times \cdots \times V_p$ with norm

$$||(\psi_1, \psi_2, \ldots, \psi_p)||_V = \max_{1 \le i \le p} ||\psi_i||_{V_i}.$$

Banach Space Containing PDF's

Define V_i as the Banach space of bounded functions on Ω_i with

$$||\psi||_{V_i} = \left|\left|\frac{\psi}{\phi_i}\right|\right|_{L^\infty(\Omega_i)}$$

for $i = 1, 2, \dots, p$ and $V = V_1 \times V_2 \times \cdots \times V_p$ with norm

$$||(\psi_1, \psi_2, \dots, \psi_p)||_V = \max_{1 \leq i \leq p} ||\psi_i||_{V_i}.$$

For any $(\psi_1, \psi_2, \dots, \psi_p) \in V$ with $||\psi||_V \leq 1$

$$\prod_{i=1}^{p} \frac{\phi_i + \psi_i}{1 + \int_{\Omega_i} \psi_i(x_i) dx_i}$$

is a PDF on Ω. Define the Sobol' indices as an operators

$$S_k : V \to \mathbb{R} \qquad \text{and} \qquad T_k : V \to \mathbb{R}$$

Banach Space Containing PDF's

Define V_i as the Banach space of bounded functions on Ω_i with

$$||\psi||_{V_i} = \left|\left|\frac{\psi}{\phi_i}\right|\right|_{L^\infty(\Omega_i)}$$

for $i = 1, 2, \dots, p$ and $V = V_1 \times V_2 \times \cdots \times V_p$ with norm

$$||(\psi_1, \psi_2, \dots, \psi_p)||_V = \max_{1 \leq i \leq p} ||\psi_i||_{V_i}.$$

For any $(\psi_1, \psi_2, \dots, \psi_p) \in V$ with $||\psi||_V \leq 1$

$$\prod_{i=1}^{p} \frac{\phi_i + \psi_i}{1 + \int_{\Omega_i} \psi_i(x_i) dx_i}$$

is a PDF on Ω. Define the Sobol' indices as an operators

$$S_k : V \to \mathbb{R} \qquad \text{and} \qquad T_k : V \to \mathbb{R}$$

- $S_k(\phi)$ and $T_k(\phi)$ are the Sobol' indices you compute
- $S_k(\phi + \psi)$ and $T_k(\phi + \psi)$ is well defined for all $\psi \in V$, $||\psi||_V \leq 1$
- Compute the Fréchet derivative of S_k and T_k with respect to ϕ

Fréchet Differentiability

The total Sobol' index is given by

$$T_k(\eta) = \frac{G_k(\eta)}{H_k(\eta)}$$

where

$$G_k(\eta) = \frac{1}{2}\frac{1}{\int_{\Omega_k}\eta_k(y)dy}\int_{\Omega\times\Omega_k}(f(\mathbf{x})-f(\mathbf{x}'))^2\prod_{i=1}^{p}\eta_i(x_i)\eta_k(x_k')d\mathbf{x}d\mathbf{x}_k'$$

$$H_k(\eta) = \int_{\Omega} f(\mathbf{x})^2\prod_{i=1}^{p}\eta_i(x_i)d\mathbf{x} - \frac{1}{\prod_{i=1}^{p}\int_{\Omega_i}\eta_i(y)dy}\left(\int_{\Omega} f(\mathbf{x})\prod_{i=1}^{p}\eta_i(x_i)d\mathbf{x}\right)^2$$

Fréchet Differentiability

The Fréchet derivative of the total Sobol' index is given by

$$\mathcal{D} T_k(\phi)\psi = \frac{\mathcal{D} G_k(\phi)\psi}{H_k(\phi)} - T_k(\phi)\frac{\mathcal{D} H_k(\phi)\psi}{H_k(\phi)}$$

Fréchet Differentiability

The Fréchet derivative of the total Sobol' index is given by

$$\mathcal{D}T_k(\phi)\psi = \frac{\mathcal{D}G_k(\phi)\psi}{H_k(\phi)} - T_k(\phi)\frac{\mathcal{D}H_k(\phi)\psi}{H_k(\phi)}$$

where

$$\mathcal{D}G_k(\phi)\psi = \frac{1}{2}\int_{\Omega\times\Omega_k}(f(\mathbf{x})-f(\mathbf{x}'))^2\left(\frac{\psi_k(x_k')}{\phi_k(x_k')}+\sum_{i=1}^{p}\frac{\psi_i(x_i)}{\phi_i(x_i)}\right)\prod_{i=1}^{p}\phi_i(x_i)\phi_k(x_k')d\mathbf{x}d\mathbf{x}_k'$$
$$-\frac{1}{2}\int_{\Omega_k}\psi_k(y)dy\int_{\Omega\times\Omega_k}(f(\mathbf{x})-f(\mathbf{x}'))^2\prod_{i=1}^{p}\phi_i(x_i)\phi_k(x_k')d\mathbf{x}d\mathbf{x}_k'$$

Fréchet Differentiability

The Fréchet derivative of the total Sobol' index is given by

$$\mathcal{D}T_k(\phi)\psi = \frac{\mathcal{D}G_k(\phi)\psi}{H_k(\phi)} - T_k(\phi)\frac{\mathcal{D}H_k(\phi)\psi}{H_k(\phi)}$$

where

$$\begin{aligned}
\mathcal{D}G_k(\phi)\psi =& \frac{1}{2}\int_{\Omega\times\Omega_k}(f(\mathbf{x})-f(\mathbf{x}'))^2\left(\frac{\psi_k(x_k')}{\phi_k(x_k')}+\sum_{i=1}^{p}\frac{\psi_i(x_i)}{\phi_i(x_i)}\right)\prod_{i=1}^{p}\phi_i(x_i)\phi_k(x_k')d\mathbf{x}d\mathbf{x}_k' \\
&-\frac{1}{2}\int_{\Omega_k}\psi_k(y)dy\int_{\Omega\times\Omega_k}(f(\mathbf{x})-f(\mathbf{x}'))^2\prod_{i=1}^{p}\phi_i(x_i)\phi_k(x_k')d\mathbf{x}d\mathbf{x}_k' \\
\mathcal{D}H_k(\phi)\psi =& \int_{\Omega}f(\mathbf{x})^2\left(\sum_{i=1}^{p}\frac{\psi_i(x_i)}{\phi_i(x_i)}\right)\prod_{i=1}^{p}\phi_i(x_i)d\mathbf{x} \\
&+\left(\int_{\Omega}\left(\sum_{i=1}^{p}\frac{\psi_i(x_i)}{\phi_i(x_i)}\right)\prod_{i=1}^{p}\phi_i(x_i)d\mathbf{x}\right)\left(\int_{\Omega}f(\mathbf{x})\prod_{i=1}^{p}\phi_i(x_i)d\mathbf{x}\right)^2 \\
&-2\left(\int_{\Omega}f(\mathbf{x})\prod_{i=1}^{p}\phi_i(x_i)d\mathbf{x}\right)\left(\int_{\Omega}f(\mathbf{x})\left(\sum_{i=1}^{p}\frac{\psi_i(x_i)}{\phi_i(x_i)}\right)\prod_{i=1}^{p}\phi_i(x_i)d\mathbf{x}\right)
\end{aligned}$$

Optimal Perturbation of ϕ

Idea: Using the existing evaluations of f,

- Determine closed form solutions to the optimization problems

$$\psi_{S_k}^\star = \underset{\substack{\psi \in V \\ ||\psi||_V \leq 1}}{\arg\max} |\mathcal{D} S_k(\phi)\psi| \qquad \psi_{T_k}^\star = \underset{\substack{\psi \in V \\ ||\psi||_V \leq 1}}{\arg\max} |\mathcal{D} T_k(\phi)\psi|$$

  where $\mathcal{D}$ denotes the Fréchet derivative operator
- Estimate perturbed Sobol' indices for each optimal perturbation by reweighting the existing function evaluations
- Weight the perturbation of each marginal according to its Fréchet derivative
- Use sample estimator standard deviation to determine the maximum admissible perturbation size

Users Guide to Sobol' Indices with Robustness Assessment

1. Define distributions for $X_1, X_2, \dots, X_p$
2. Draw samples from these distribution
3. Evaluate f at these samples
4. Estimate the Sobol' indices using these evaluations of f
5. For each Sobol' index, use the existing evaluations of f to determine an optimal PDF perturbation and estimate the Sobol' indices corresponding to this perturbed PDF using weighted averages
   - Does not require any additional evaluations of f
   - Computational cost of step 5 is small relative to step 3
   - Robustness and corresponding optimal perturbations may be visualize alongside the Sobol' indices

An Illustrative Example

$$f(\mathbf{X}) = 2X_2 e^{-2X_1} + X_3^2.$$

- X_i is uniformly distributed on the interval $[A_i, B_i]$, $i = 1, 2, 3$
- A_i and B_i are random variables which are uniformly distributed on $[0, .1]$ and $[.9, 1]$, respectively, $i = 1, 2, 3$
- Emulates an application with uncertainty in the distributions support

Perturbed Total Sobol' Indices

$$f(\mathbf{X}) = 2X_2 e^{-2X_1} + X_3^2.$$

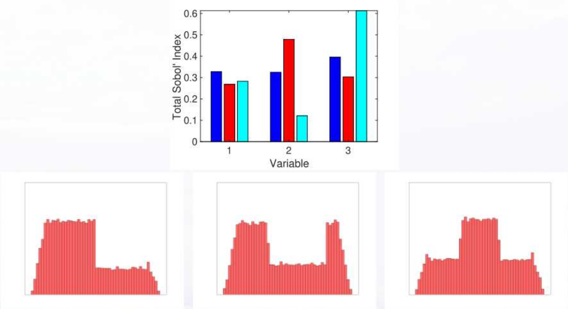

Figure: Top: Total Sobol' indices. Blue bars denote the indices computed using the nominal distribution, red bars denote the indices when the distribution is perturbed to maximize T_2, cyan bars denote the indices when the distribution is perturbed to minimize T_2. Bottom: perturbed distributions corresponding to the perturbed which maximizes T_2.

A PDE Example

$$
\begin{aligned}
-\epsilon \triangle u + v \cdot \nabla u &= s && \text{on } (0,1)^2 \\
u &= 0 && \text{on } \{0\} \times (0,1) \\
u &= 0 && \text{on } (0,1) \times \{0\} \\
\mathbf{n} \cdot \nabla u &= \nu_1 u && \text{on } \{1\} \times (0,1) \\
\mathbf{n} \cdot \nabla u &= \nu_2 u && \text{on } (0,1) \times \{1\}
\end{aligned}
$$

- $v(y_1, y_2) = (\alpha_1(y_1 + 0.5), \alpha_2(y_2 + 0.5))$
- $s(y_1, y_2) = \beta e^{-\gamma((y_1-\xi_1)^2+(y_2-\xi_2)^2)}$
- $\mathbf{X} = (\epsilon, \alpha_1, \alpha_2, \xi_1, \xi_2, \gamma, \beta, \nu_1, \nu_2)$
- assume inputs are independent uniform random variables with $\pm 30\%$ uncertainty about nominal values
- define $f(\mathbf{X}) = \int_{.5}^{.7} \int_{.5}^{.7} u(y_1, y_2) dy_1 dy_2$

The PDE Solution

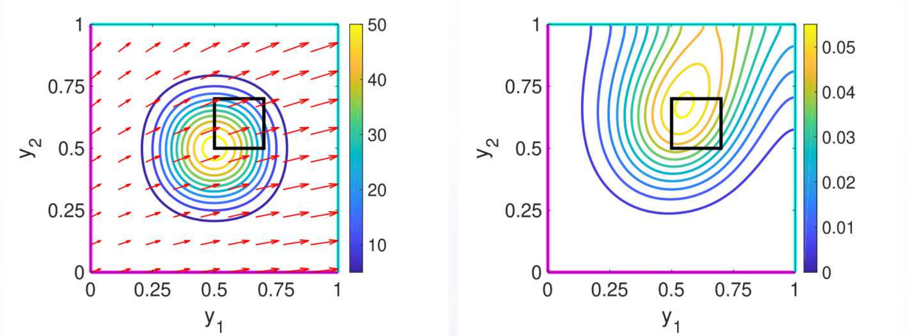

Figure: Left: contours of the average source with the average velocity field overlaid with red arrows. Right: contours of the average contaminant concentration u. The cyan and magenta lines indicate the Robin and Dirichlet boundaries, respectively. The black box encloses the region over which we integrate to compute f.

Minimum and Maximum Total Sobol' Indices

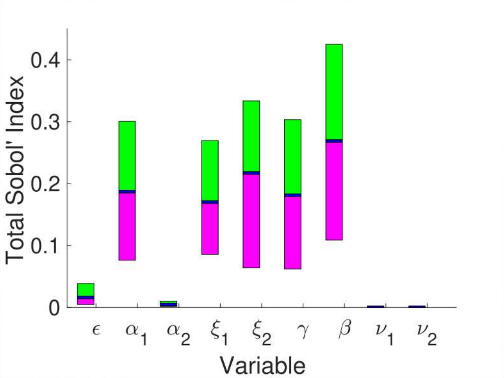

Figure: Total Sobol' indices for the advection diffusion example. The top of the green bars and bottom of the magenta bars denote the maximum and minimum perturbed total Sobol' indices, respectively; the blue area between them indicates the nominal total Sobol' indices.

A Particular Perturbation

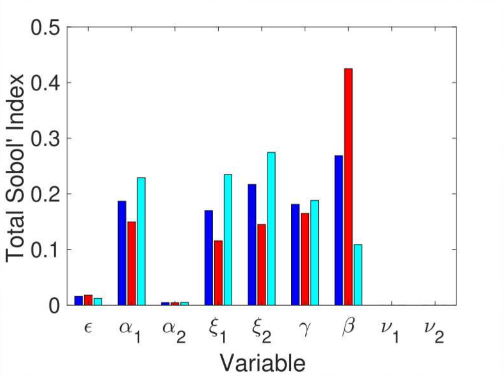

Figure: Total Sobol' indices for the advection diffusion example. Blue bars denote the indices computed using the nominal distribution, red bars denote the indices when the distribution is perturbed to maximize the total Sobol' index for β, cyan bars denote the indices when the distribution is perturbed to minimize the total Sobol' index for β.

What About Statistical Dependencies?

- The previous developments assume that the inputs are independent

What About Statistical Dependencies?

- The previous developments assume that the inputs are independent
- An analogous framework can be applied to find optimal perturbations of joint distributions instead of marginals

What About Statistical Dependencies?

- The previous developments assume that the inputs are independent
- An analogous framework can be applied to find optimal perturbations of joint distributions instead of marginals
- Perturbations of the joint distribution may be useful when you have assumed the inputs are independent but you are concerned they might not be

What About Statistical Dependencies?

- The previous developments assume that the inputs are independent
- An analogous framework can be applied to find optimal perturbations of joint distributions instead of marginals
- Perturbations of the joint distribution may be useful when you have assumed the inputs are independent but you are concerned they might not be
- Assuming independence gives computational advantages in high dimensions because the dimensions are decoupled, the method in this talk scales more efficiency to high dimensions

What About Statistical Dependencies?

- The previous developments assume that the inputs are independent
- An analogous framework can be applied to find optimal perturbations of joint distributions instead of marginals
- Perturbations of the joint distribution may be useful when you have assumed the inputs are independent but you are concerned they might not be
- Assuming independence gives computational advantages in high dimensions because the dimensions are decoupled, the method in this talk scales more efficiency to high dimensions
- A comparison of the two approaches is given in the paper (citation on final slide)

Summary

- Wrap classical theory for Sobol' indices in a function theoretic framework
- Use a derivative based approach to assess the robustness of Sobol' indices to distribution uncertainty with no additional model evaluations
- Easily visualize the resulting perturbations and perturbed indices to gain physical insights
- The approach has two important properties
  - We do not make assumptions on the parametric form of the distributions
  - We do not require any additional model evaluations

Questions?

Joseph Hart

Sandia National Laboratories

joshart@sandia.gov

J Hart and P.A. Gremaud. Robustness of the Sobol' indices to marginal distribution uncertainty. SIAM/ASA Journal on Uncertainty Quantification, 7(4) 1224-1244, 2019.

J. Hart and P.A. Gremaud. Robustness of the Sobol' indices to distributional uncertainty. International Journal for Uncertainty Quantification, 9(5) 453-469, 2019.

https://github.com/jlhart352/Sobol_Index_Robustness