Consistency between Sobol indices and stochastic orders

Global optimization using Sobol indices

Alexandre Janon (Université Paris Saclay, labo. de math. d'Orsay)

UQSay June '19

Part 1: Consistency between Sobol indices and stochastic orders

(Co-authors: A. Cousin, V. Maume-Deschamps, I. Niang)

Context

- Model output with uncertain input parameters:

$$Y = f(X_1, \dots, X_p)$$

- $X_1, \dots, X_p$: independent random variables of known distributions, encoding parameter uncertainty
- Y: random variable, supposed square integrable
- For $i = 1, \dots, p$, first-order Sobol index of X_i:

$$S_i = \frac{\mathbf{Var}\mathbf{E}(Y|X_i)}{\mathbf{Var}Y}.$$

S_i quantifies the impact of the uncertainty on X_i on the uncertainty on Y.

Problem

- How to choose distribution of input parameters ?
- How do Sobol indices change when input distributions are changed ?
- Qualitatively speaking, how do first-order Sobol indices vary when X_i is replaced by X_i^* ?

Problem

- Stochastic ordering between two r.v. X and Y:

$$X \leq Y \Leftrightarrow X \text{ carries "less uncertainty" than } Y$$

- Intuitively one would say that

$$X_i \leq X_i^* \Rightarrow \mathbf{Var}\, Y \leq \mathbf{Var}\, Y$$

and

$$X_i \leq X_i^* \Rightarrow S_i \leq S_i^* \text{ and } S_j \geq S_j^* \ \forall j \neq i$$

- Is this always the case ?
- Under what hypotheses:
  - on distributions of X_i, X_i^* ?
  - on the f function ?
- And for what ordering between rv ?

Outline of Part 1

1. Stochastic orderings
2. Effect on variances
3. Effect on Sobol indices:
    3.1 additive case,
    3.2 multiplicative case,
    3.3 “tensor” case

Stochastic orderings

- F_X: cdf of a random variable X: $F_X(x) = P(X \leq x)$
- X, Y: two random variables
- Dispersive order:

$$X \leq_{Disp} Y \Leftrightarrow F_Y^{-1} - F_X^{-1} \text{ is non-decreasing}$$

- "Usual" stochastic order:

$$X \leq_{st} Y \Leftrightarrow \forall f \text{ bounded, non-decreasing }, \mathbf{E}(f(X)) \leq \mathbf{E}(f(Y))$$

- There are others (convex, dilation, Lorenz, excess-wealth, star,...)...
  See [M. Shaked, J. Shanthikumar, Stochastic orders (2006)].

Stochastic orderings: Properties of Dispersive order

$$X \leq_{Disp} Y \Leftrightarrow F_Y^{-1} - F_X^{-1} \text{ is non-decreasing}$$

- Measures only *dispersion*, in fact, it is a "location-free' order:

$$X \leq_{Disp} Y \to X + a \leq_{Disp} Y \; \forall a \in \mathbb{R}$$

- Usual distributions:

$$\mathcal{U}(a, b) \leq_{Disp} \mathcal{U}(c, d) \Leftrightarrow b - a \leq d - c$$

$$\mathcal{E}(\lambda) \leq_{Disp} \mathcal{E}(\mu) \Leftrightarrow \mu \leq \lambda$$

$$\mathcal{N}(m_1, \sigma_1^2) \leq_{Disp} \mathcal{N}(m_2, \sigma_2^2) \Leftrightarrow \sigma_1^2 \leq \sigma_2^2$$

- Ordering of variances:

$$X \leq_{Disp} Y \Rightarrow \mathbf{Var} X \leq \mathbf{Var} Y$$

Application to UQ

- Let $i = 1, \dots, p$.
- We define

$$Y = f(X_1, \dots, X_p),\, Y^* = f(X_1, \dots, X_{i-1}, X_i^*, X_{i+1}, \dots, X_p)$$

  with $X_i \leq_{Disp} X_i^*$.
- Do we have: $\mathbf{Var}\, Y \leq \mathbf{Var}\, Y^*$?
- $Y \leq_{Disp} Y^*$ would be sufficient.
- However, it is not true in general, even for convex non-decreasing f.
- Take $X \sim \mathcal{U}(1; 1.9)$, $X^* \sim \mathcal{U}(0; 1)$. We have

$$X \leq_{Disp} X^* \text{ but } \mathbf{Var} \exp(X) > \mathbf{Var} \exp(X^*)$$

- Disp. order alone is not sufficient !

Stochastic orderings: Properties of Usual st. order

$$X \leq_{st} Y \Leftrightarrow \forall f \text{ bounded, non-decreasing }, \mathbf{E}(f(X)) \leq \mathbf{E}(f(Y))$$

- Dispersive implies stochastic under a “location condition”: $\mathrm{supp}(X) \subset (l_X, +\infty[$, $\mathrm{supp}(Y) \subset (l_Y; +\infty[$
  If $l_X = l_Y > -\infty$ then $X \leq_{st} Y \Rightarrow X \leq_{Disp} Y$.
- If $X \leq_{Disp} Y$ and $X \leq_{st} Y$ then, for any convex non-decreasing, or concave non-increasing ϕ, then

$$\phi(X) \leq_{Disp} \phi(Y)$$

- Curvature hypotheses are necessary. For instance, if
  - $f(t) = t$ on $[0; 1]$ and 1 on $[1; 10]$,
  - $X \sim \mathcal{U}(0; 1)$, $Y \sim \mathcal{U}(0; 10)$,
  - $X \leq_{Disp} Y$, $X \leq_{st} Y$ but $\mathbf{Var} f(X) > \mathbf{Var} f(Y)$.

Application to UQ (2)

$$Y = f(X_1, \ldots, X_p), Y^* = f(X_1, \ldots, X_{i-1}, X_i^*, X_{i+1}, \ldots, X_p)$$

- Hence, if
  - $X_i \leq_{Disp} X_i^*$,
  - $X_i \leq_{st} X_i^*$,
  - f is convex non-dec, or concave non-inc in its i^{th} argument,
- then

$$\mathbf{Var}\, Y \leq \mathbf{Var}\, Y^*$$

- Can we do the same with Sobol indices ?

Additive case

- We suppose that:

$$f(X_1, \ldots, X_p) = \sum_{j=1}^{p} f_j(X_j) + \mathbf{E}(Y)$$

  with:
  - f_i convex non-decreasing, or concave non-increasing.
  - $X_i \leq_{Disp} X_i^*$,
  - $X_i \leq_{st} X_i^*$,
  - X_i and X_i^* are independent.
- Then,

$$S_i \leq S_i^*$$

  and

$$S_j \geq S_j^* \; \forall j \neq i$$

Product case

- Now assume that

$$f(X) = \prod_{j=1}^{p} g_j(X_j) + \mathbf{E}(Y)$$

with $\log g_i$ convex non-decreasing, or concave non-increasing.

- If $X_i \leq_{Disp} X_i^*$, $X_i \leq_{st} X_i^*$, and X_i and X_i^* are independent, then

$$S_i^T \leq S_i^{T*}$$

and

$$S_j^T \geq S_j^{T*} \; \forall j \neq i$$

where S_j^T and S_j^{T*}, are total Sobal indices of

$$Y = f(X_1, \ldots, X_p), Y^* = f(X_1, \ldots, X_{i-1}, X_i^*, X_{i+1}, \ldots, X_p)$$

respectively.

- Only g_i convex non-decreasing, or concave non-increasing is not sufficient.

General "tensor" case

We also have a similar theorem for

$$f(X) = \sum_{\ell} \prod_{j=1}^{p} g_j^{\ell}(X_j) + \mathbf{E}(Y)$$

under similar hypotheses, plus a (seemingly necessary) tangled and unsatisfactory inequality.

Part 2: Global optimization using Sobol indices

Context

- Our goal is to minimize a (generally nonconvex) function $f : \mathcal{D} = [-1;1]^d \to \mathbb{R}$:
$$\min_{\mathcal{D}} f$$
- f is computationally expensive to compute, we want to evaluate it a only a few number of times.
- We suppose that we have (even partial) knowledge about Sobol indices of f: for instance first-order, second-order, total...
- Can we use this knowledge to improve minimization of f ?

Context (2)

- For instance, if all interaction indices are zero, f can be minimized separately on each variable, allowing substantial gain.
- In general, there is some sparsity-of-effects principle allowing to neglect high-order interactions.
- We will propose an optimization algorithm which can take advantage of this situation.

Outline of Part 2

- Presentation of the strategy
- Implementation details
- Numerical “proof of concept” illustration

Presentation of the strategy

Assume:

- $\mathcal{D} = [-1; 1]^d$ endowed with the uniform probability distribution,
- $f(X)$ has unit variance,
- $\mathcal{F}$ is a subspace of square integrable functions $\mathcal{D} \to \mathbb{R}$.

The following strategy is inspired by the one used in [C. Malherbe, N. Vayatis, Global optimization of Lipschitz functions (2017)].

Presentation of the strategy (2)

We build a "minimizing" sequence of length n with the following:

- Initialization : choose X^1 uniformly on $\mathcal{D}$;
- Iteration: for $i = 2, \ldots, n$, repeat:
  - choose X^i uniformly on:

$$\mathcal{D}_i = \{x \in \mathcal{D} \text{ s.t. } \exists g \in \mathcal{F}_i,\ g(x) < \min_{1 \le j < i} f(X^j)\}$$

where:

$$\mathcal{F}_i = \{g \in \mathcal{F},\ \forall 1 \le j < i,\ g(X^j) = f(X^j)\}$$

$\mathcal{F}_i$ is the set of "consistent" functions, and $\mathcal{D}_i$ a set of "interesting" points to explore, as they might improve the current minimum.

- In our context, $\mathcal{F}$ is a set of square-integrable functions satisfying the “prior knowledge” on Sobol indices.
- For instance, if $d = 3$ and that we assume that there is no interaction between X_1 and X_3,

$$\mathcal{F} = \{g \in L^2([-1;1]^3) \text{ s.t. } \mathbf{Var} g(X) = 1,\ S_{2,3} = S_{1,2,3} = 0\}$$

Implementation details

To sample uniformly on

$$\mathcal{D}_i = \{x \in \mathcal{D} \text{ s.t. } \exists g \in \mathcal{F}_i,\ g(x) < \min_{1 \le j < i} f(X^j)\}$$

where:

$$\mathcal{F}_i = \{g \in \mathcal{F},\ \forall 1 \le j < i,\ g(X^j) = f(X^j)\}$$

we use a “rejection” algorithm:

1. we sample x uniformly on $\mathcal{D}$,
2. solve for
$$m(x) = \min_{\mathcal{F}_i} g(x)$$
3. if $m(x) < \min_{1 \le j < i} f(X^j)$, then $x \in \mathcal{D}_i$ and we accept it; else we sample a new x.

Implementation details (2)

- Solving
$$m(x) = \min_{\mathcal{F}_i} g(x)$$
can be made in practice by introducing a tensor orthonormal L^2 basis (in our case, of normalized Legendre polynomials) and search for coefficients c of g on this basis.
- The objective is a linear function of c, and the constraints of $\mathcal{F}_i$ are:
  - linear in c for the $g(X^j) = f(X^j)$ constraints;
  - positive semi-definite in c (sum of squares) for the Sobol indices constraints.
- This gives a succession (for different x's) of high-dimensional convex problems to solve until some x is accepted.

Numerical illustration

- Rosenbrock function on $[-5;5]^3$:

$$f(\frac{X_1}{5},\frac{X_2}{5},\frac{X_3}{5})=\frac{1}{26000}\sum_{m=1}^{2}100(X_{m+1}-X_m^2)^2+(1-X_m)^2$$

- Budget of 100 convex problems $\rightarrow$ variable number N_{eval} of evaluations of f, Legendre polynomials up to degree 4.
- Two families of constraints on the Sobol indices:
  - **Esti1T:** estimations of first-order and total Sobol indices (six indices)
  - **NoInter13:** no interaction between X_1 and X_3 (hence no third-order interaction)

Result

Constraints	N_{eval}	record minimum
unit variance	93	0.0089
unit variance, Esti1T	78	0.0052
unit variance, Esti1T, NoInter13	44	0.0006
unit variance, NoInter3	45	0.0049

- Constraints on Sobol indices actually improves optimization.
- Great reduction on number of evaluations of f by using that no interaction occurs between X_1 and X_3.
- Many improvements could be tried...