Reducing the number of function evaluations to estimate *first-order* Sobol' indices

Lluís Antoni Jiménez Rugama
Joint work with: Fred J. Hickernell (IIT),
Clémentine Prieur (Univ. Grenoble), Elise Arnaud (Univ. Grenoble),
Hervé Monod (INRA), and Laurent Gilquin (Univ. Grenoble)

Room 120, Bldg E1, Department of Applied Mathematics
Illinois Institute of Technology, Chicago, 60616 IL
Email: ljimene1@hawk.iit.edu

Tuesday 22nd March, 2016

Outline

- Introduction
- Sobol Indices'
- Replicated Method

Outline

- Introduction—The ANalysis Of VAriance.
- Sobol Indices'
- Replicated Method

ANOVA

For $f \in L^2\left([0,1]^d\right)$, and $\mathcal{D} = \{1, \dots, d\}$,

$$f(\boldsymbol{x}) = \sum_{u \subseteq \mathcal{D}} f_u(\boldsymbol{x}) \,, \qquad f_\varnothing = \mu \,,$$

where,

$$f_u(\boldsymbol{x}) = \int_{[0,1]^{d-|u|}} f(\boldsymbol{x}) d\boldsymbol{x}_{-u} - \sum_{v \subset u} f_v(\boldsymbol{x}) \,.$$

Variance Decomposition

Under the previous definitions,

$$\sigma_{\varnothing}^2 = 0\,, \qquad \sigma_u^2 = \int_{[0,1]^d} f_u(\boldsymbol{x})^2 d\boldsymbol{x}\,, \qquad \sigma^2 = \int_{[0,1]^d} \left(f(\boldsymbol{x}) - \mu\right)^2 d\boldsymbol{x}\,.$$

The ANOVA identity is,

$$\sigma^2 = \sum_{u \subseteq \mathcal{D}} \sigma_u^2\,.$$

Outline

- Introduction
- Sobol Indices'—Measuring the importance of the inputs.
- Replicated Method

Sobol' Indices

We consider $f(\boldsymbol{x})$ as a random variable with $\boldsymbol{x} \sim U([0,1)^d)$.

Sobol' introduced the *global sensitivity* indices which measure the variance explained by subsets of dimensions:

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2 \,, \quad \text{and} \quad \overline{\tau}_u^2 = \sum_{v \cap u \neq \varnothing} \sigma_v^2 \,, \qquad u, v \in \mathcal{D} \,.$$

We have the following properties,

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

Normalized *First-Order* Sobol' Indices

In this particular case, we consider $|u| = 1$ and want to estimate $\underline{\tau}_u^2/\sigma^2 = \sigma_u^2/\sigma^2$. For this purpose, given $\boldsymbol{x}, \boldsymbol{x}' \in [0,1]^d$, we define the following point,

$$(\boldsymbol{x}_u : \boldsymbol{x}'_{-u}) := (x'_1, \dots, x'_{u-1}, x_u, x'_{u+1}, \dots, x'_d) \in [0,1]^d.$$

Thus, one can use the following integral form to build an estimator:

$$\frac{\sigma_u^2}{\sigma^2} = \frac{\int_{[0,1)^{2d-1}} \big(\overbrace{f(\boldsymbol{x})}^{g(\boldsymbol{x},\boldsymbol{x}'):=} - \mu\big)\big(\overbrace{f(\boldsymbol{x}_u : \boldsymbol{x}'_{-u})}^{g_u(\boldsymbol{x},\boldsymbol{x}'):=} - \mu\big) d\boldsymbol{x} d\boldsymbol{x}'_{-u}}{\int_{[0,1]^d} f(\boldsymbol{x})^2 d\boldsymbol{x} - \mu^2} = H(g, g_u)\,.$$

Numerical Integration Problem

We will focus on reducing the number of function evaluations, and to estimate σ_u^2/σ^2, only g and g_u are evaluated.

Computing all the indices one by one, if one requires n points for each estimation, the total number of function evaluations of g and g_u are

$$2dn\,,$$

However, if all indices are computed together, g only needs to be evaluated once. Therefore, the number of function evaluations becomes

$$(1+d)n\,,$$

Finally, under a special set of quasi-Monte Carlo sequences, this number is decreased to

$$2n\,.$$

Outline

- Introduction
- Sobol Indices'
- Replicated Method—Reducing the number of function evaluations to compute *first-order* indices.

Replicated points

Functions g and g_u only share input dimension u:

$$g(\boldsymbol{x},\boldsymbol{x}') = f(x_1,\ldots,x_{u-1},x_u,x_{u+1},\ldots,x_d)\,,$$
$$g_u(\boldsymbol{x},\boldsymbol{x}') = f(x'_1,\ldots,x'_{u-1},x_u,x'_{u+1},\ldots,x'_d)\,.$$

Hence, we can construct our points $\boldsymbol{x}'_i$ as follows,

$$\begin{pmatrix} x_{0,1} & \cdots & x_{0,d} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,d} \\ \vdots & & \vdots \end{pmatrix}, \qquad \begin{pmatrix} x'_{0,1} & \cdots & x'_{0,d} \\ \vdots & \ddots & \vdots \\ x'_{n,1} & \cdots & x'_{n,d} \\ \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} x_{\pi_1(0),1} & \cdots & x_{\pi_d(0),d} \\ \vdots & \ddots & \vdots \\ x_{\pi_1(n),1} & \cdots & x_{\pi_d(n),d} \\ \vdots & & \vdots \end{pmatrix}.$$

The Right Function Values

Given the right order of points:

$$\begin{pmatrix} \boldsymbol{x'}_{\pi_u^{-1}(0)} \\ \vdots \\ \boldsymbol{x'}_{\pi_u^{-1}(n)} \\ \vdots \end{pmatrix} = \begin{pmatrix} x'_{\pi_u^{-1}(0),1} & \cdots & x_{0,u} & \cdots & x'_{\pi_u^{-1}(0),d} \\ \vdots & & \vdots & & \vdots \\ x'_{\pi_u^{-1}(n),1} & \cdots & x_{n,u} & \cdots & x'_{\pi_u^{-1}(n),d} \\ \vdots & & \vdots & & \vdots \end{pmatrix}.$$

The Right Function Values

We only need to evaluate $g_u(\boldsymbol{x}, \boldsymbol{x}')$ once:

$$\begin{pmatrix} f(\boldsymbol{x}'_0) \\ \vdots \\ f(\boldsymbol{x}'_n) \\ \vdots \end{pmatrix} = \begin{pmatrix} y_0 \\ \vdots \\ y_n \\ \vdots \end{pmatrix} \Longrightarrow \begin{pmatrix} g_u(\boldsymbol{x}_0, \boldsymbol{x}'_0) \\ \vdots \\ g_u(\boldsymbol{x}_n, \boldsymbol{x}'_n) \\ \vdots \end{pmatrix} = \begin{pmatrix} y_{\pi_u^{-1}(0)} \\ \vdots \\ y_{\pi_u^{-1}(n)} \\ \vdots \end{pmatrix}$$

Quasi-Monte Carlo Sequences

For $i \in \mathbb{N}_0$ and $x \in [0,1)$,

$$i = \sum_{k \geqslant 0} i_k b^k \,, \qquad x = \sum_{k \geqslant 1} x_k b^{-k} \,, \qquad i_k, x_k \in \mathbb{F}_b.$$

Each dimension j of a digital net is constructed according to:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \vdots \end{pmatrix} = C_j \begin{pmatrix} i_0 \\ \vdots \\ i_{m-1} \\ \vdots \end{pmatrix}, \qquad C_j \in M^{\infty \times \infty}(\mathbb{F}_b)\,.$$

Quasi-Monte Carlo Sequences

For the Sobol' sequences, we can for instance, construct our permutations through an upper triangular matrix $U \in M^{\infty\times\infty}(\mathbb{F}_b)$ applied as follows,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \vdots \end{pmatrix} = C_j U \begin{pmatrix} i_0 \\ \vdots \\ i_{m-1} \\ \vdots \end{pmatrix}$$

References I

Hickernell, F. J. and Ll. A. Jiménez Rugama. 2015+. *Reliable adaptive cubature using digital sequences*, Monte Carlo and quasi-Monte Carlo methods 2014. to appear, arXiv:1410.8615 [math.NA].

I.M., Sobol. 2001. *Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates*, Mathematics and Computers in Simulation (MATCOM) **55**, no. 1, 271–280.

Niederreiter, H. 1992. *Random number generation and quasi-Monte Carlo methods*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.

Owen, A. B. 1998. *Scrambling Sobol' and Niederreiter-Xing points*, J. Complexity **14**, 466–489.

Owen, Art B. 2013. *Variance components and generalized Sobol' indices*, SIAM/ASA Journal on Uncertainty Quantification **1**, no. 1, 19–41.