Automatic estimation of Sobol' indices based on quasi-Monte Carlo methods

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

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

Tuesday 2^{nd} August, 2016

Outline

- Introduction
- Sobol' Indices
- Quasi-Monte Carlo Methods
- Numerical example

Outline

- Introduction—The ANalysis Of VAriance.
- Sobol' Indices
- Quasi-Monte Carlo Methods
- Numerical example

ANOVA

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

$$f(\boldsymbol{x}) = \sum_{u \subseteq 1:d} f_u(\boldsymbol{x})\,, \qquad f_\varnothing = \mu\,,$$

where,

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

- $|u|$ the cardinality of u.
- $-u := u^c$.

Variance Decomposition

Under the previous definitions,

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

The ANOVA identity is,

$$\sigma^2 = \sum_{u \subseteq 1:d} \sigma_u^2\,.$$

Outline

- Introduction
- Sobol' Indices—Measuring the importance of each input.
- Quasi-Monte Carlo Methods
- Numerical example

Sobol' Indices for Uncertainty Quantification

We consider the random variable $Y = f(\boldsymbol{X})$ with $\boldsymbol{X} \sim U\left([0,1)^d\right)$.

Sobol' introduced the *global sensitivity* indices which measure the variance explained by any dimension subset $u \subseteq 1:d$:

$$\underline{\tau}_u^2 = \sum_{\substack{v \subseteq u \\ v \subseteq 1:d}} \sigma_v^2 \,, \qquad \text{and} \qquad \overline{\tau}_u^2 = \sum_{\substack{v \cap u \neq \varnothing \\ v \subseteq 1:d}} \sigma_v^2 \,.$$

We have the following properties,

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

Estimation of the Integral

Indices $\underline{\tau}_u^2$ and $\overline{\tau}_u^2$ can be estimated with integrals:

$$\underline{\tau}_u^2 = \int_{[0,1]^{2d-1}} f(\boldsymbol{x}) f(\boldsymbol{x}_u : \boldsymbol{x'}_{-u}) d\boldsymbol{x} d\boldsymbol{x'} - \left(\int_{[0,1]^d} f(\boldsymbol{x}) d\boldsymbol{x} \right)^2,$$

$$\overline{\tau}_u^2 = \frac{1}{2} \int_{[0,1]^{d+1}} \left(f(\boldsymbol{x'}) - f(\boldsymbol{x}_u : \boldsymbol{x'}_{-u}) \right)^2 d\boldsymbol{x} d\boldsymbol{x'}.$$

where the point $\boldsymbol{z} := (\boldsymbol{x}_u : \boldsymbol{x'}_{-u})$ has coordinates $z_j = x_j$ if $j \in u$, and $z_j = x'_j$ if $j \notin u$.

Outline

- Introduction
- Sobol' Indices
- Quasi-Monte Carlo Methods—How can we estimate integrals in high dimensions?
- Numerical example

Numerical Integration Problem

Given ε_a and $\boldsymbol{x} \mapsto f(\boldsymbol{x})$, we want $\hat{I}$ such that

$$\left| \int_{[0,1)^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \hat{I}\left(\boldsymbol{x} \mapsto f(\boldsymbol{x}), \varepsilon_a\right) \right| \leqslant \varepsilon_a,$$

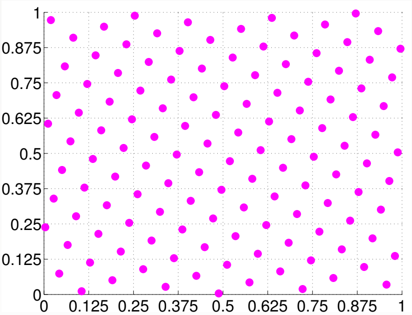
where

$$\hat{I}\left(\boldsymbol{x} \mapsto f(\boldsymbol{x}), \varepsilon_a\right) = \frac{1}{2^m} \sum_{i=0}^{2^m-1} f(\boldsymbol{z}_i \oplus \boldsymbol{\Delta}),$$

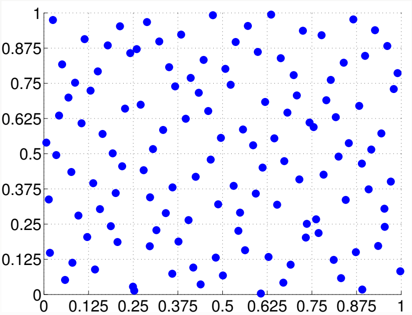
for some automatic and adaptive choice of m, $\{\boldsymbol{z}_i\}_{i=0}^{\infty} \in \left\{\begin{matrix}\text{Lattice} \\ \text{Digital}\end{matrix}\right\}$ sequence, and

$$\mathrm{cost}\left(\hat{I}\left(\boldsymbol{x} \mapsto f(\boldsymbol{x}), \varepsilon_a\right)\right) = \mathcal{O}\left((m + \$(f))2^m\right)$$

Examples of Sequences



Shifted rank-1 lattice sequence with generating vector $(1, 47)$.



Digitally shifted scrambled Sobol' sequence.

Adaptive Algorithm

The idea behind the results in Jiménez Rugama and Hickernell (2016+) and Hickernell and Jiménez Rugama (2016+) is that for all $f \in \mathcal{C}$,

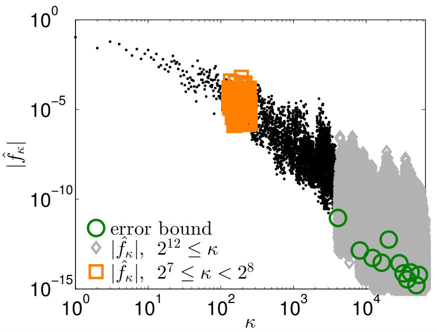
$$\left| \int_{[0,1)^d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \hat{I}\left(\boldsymbol{x} \mapsto f(\boldsymbol{x}), \varepsilon_a\right) \right| \leqslant a(r,m) \sum_{\kappa=\lfloor 2^{m-r-1} \rfloor}^{2^{m-r}-1} \left| \tilde{f}_{m,\kappa} \right|.$$

- $\tilde{f}_{m,\kappa}$ = discrete Fourier $\left\{ \begin{matrix} \text{Exponential} \\ \text{Walsh} \end{matrix} \right\}$ coefficients of f.
- $a(r,m)$ = inflation factor that depends on $\mathcal{C}$.

QMC Error Bounds Depend Only on Some Fourier Coefficients for $f \in \mathcal{C}$

In the quasi-Monte Carlo case,

$$\left| I(f) - \frac{1}{2^m} \sum_{i=0}^{2^m-1} f(\boldsymbol{z}_i \oplus \boldsymbol{\Delta}) \right| \leqslant \sum \bigcirc \leqslant a(r,m) \sum_{\kappa=\lfloor 2^{m-r-1} \rfloor}^{2^{m-r}-1} \left| \tilde{f}_{m,\kappa} \right|$$



$$\mathcal{C} = \left\{ \sum \square \text{ bounds } \sum \lozenge \atop \sum \lozenge \text{ bounds } \sum \bigcirc \right\}$$

Outline

- Introduction
- Sobol' Indices
- Quasi-Monte Carlo Methods
- Numerical example—The Bratley et al. (1992) function

Bratley et al. (1992)

$$f(\boldsymbol{X}) = \sum_{i=1}^{6} (-1)^i \prod_{j=1}^{i} \boldsymbol{X}_j = -\boldsymbol{X}_1 + \boldsymbol{X}_1\boldsymbol{X}_2 - \boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3 + \ldots$$

Values	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
S_j	65.3%	17.9%	3.7%	1.3%	0.1%	0.1%
S_j^{tot}	74.0%	26.6%	7.7%	3.3%	0.6%	0.7%

n	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
S_j	65536	32768	16384	8192	2048	1024
S_j^{tot}	32768	16384	8192	4096	1024	1024

Computation time was 0.37 seconds for $\varepsilon_a = 10^{-3}$.

Conclusions and Future Work

- The Sobol' indices algorithm accepts user input relative error tolerances.
- The estimation can be performed over the normalized indices.
- First-order indices can also be estimated using the replicated method because quasi-Monte Carlo methods are compatible with the replicated method.

References I

Bratley, P., B. L. Fox, and H. Niederreiter. 1992. *Implementation and tests of low-discrepancy sequences*, ACM Trans. Model. Comput. Simul. **2**, 195–213.

Cools, R. and D. Nuyens (eds.) 2016+. *Monte Carlo and quasi-Monte Carlo methods 2014*, Springer-Verlag, Berlin.

Hickernell, F. J. and Ll. A. Jiménez Rugama. 2016+. *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.

Jiménez Rugama, Ll. A. and F. J. Hickernell. 2016+. *Adaptive multidimensional integration based on rank-1 lattices*, Monte Carlo and quasi-Monte Carlo methods 2014. to appear, arXiv:1411.1966.

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

References II

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

Inside and Outside $\mathcal{C}$