Global sensitivity analysis of models with dependent and independent inputs

Sergei Kucherenko,
Imperial College London
s.kucherenko@imperial.ac.uk

Oleksiy Klymenko[b], Nilay Shah[a]
[a] *Imperial College London, UK*
[b] *University of Surrey, UK*

Outline

1. Models with independent inputs
2. Sobol sensitivity indices (SI) for models with dependent inputs
3. Dependences in a form of pair wise correlations
4. Constrained Global Sensitivity Analysis (cGSA)
5. Acceptance-rejection method
6. Rosenblatt transformation
7. Test cases

Sobol'-ANOVA decomposition and Sensitivity Indices

$$x \in \Omega \longrightarrow \boxed{\text{Model, } f(x)} \longrightarrow Y$$

Consider a model

x is a vector of input variables

f(x) is square integrable

$$Y = f(x)$$
$$x = (x_1, x_2, \ldots, x_n) \in H^n$$
$$0 \le x_i \le 1$$

Sobol' - ANOVA decomposition is unique if **variables are independent**

$$Y = f(x) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n),$$

$$\int_0^1 f_{i_1 \ldots i_s}(x_{i_1}, , \ldots, x_{i_s}) dx_{i_k} = 0, \ \forall k, \ 1 \le k \le s, \rightarrow \int_0^1 f_{i_1 \ldots i_s} f_{i_1 \ldots i_l} dx_{i_k} dx_{i_l} = 0, \ \forall i_k \ne i_l$$

Variance decomposition:

$$D = \sum_i D_i + \sum_{i,j} D_{ij} + K \, D_{1,2,\ldots,n}$$

Sobol' SI:

$$1 = \sum_{i=1}^{n} S_i + \sum_{i<j} S_{ij} + \sum_{i<j<l} S_{ijl} + \ldots + S_{1,2,\ldots,n}$$

Sobol' SI in the case of dependent inputs

Consider two subsets of variables: $y=\left(x_{i_1},\dots,x_{i_s}\right),\ 1\le s<n,\quad z=\left(x_{i_{s+1}},\dots,x_{i_n}\right)$

so that $x=(y,z),\ x\sim p(x_1,\dots,x_n)$ - joint PDF

General variance decomposition (known in statistics):

$$D=D_y[E_z(f(y,\overline{z})\,|\,y)]+E_y[D_z(f(y,\overline{z})\,|\,y)]$$

First-order (main) effect index:

$$S_y=\frac{D_y[E_z(f(y,\overline{z}))]}{D}$$

Total effect index of subset y:

$$S_y^T=\frac{E_z[D_y(f(\overline{y},z))]}{D}=\frac{D-D_z\left[E_y\left(f(\overline{y},z)\right)\right]}{D}$$

Sobol' Sensitivity Indices: integral representation

First-order (main) effect index of subset of variables $y = \left(x_{i_1}, ..., x_{i_s}\right), 1 \le s < n$ [$z = \left(x_{i_{s+1}}, ..., x_{i_n}\right)$ so that $x = (y, z)$]

$$S_y = \frac{D_y[E_z(f(y,\overline{z}))]}{D} = \frac{1}{D}\left[\int_{H^s} p(y)dy\left[\int_{H^{n-s}} f(y,\overline{z})p(y,\overline{z}\,|\,y)d\overline{z}\right]^2 - f_0^2\right]$$

$$S_y = \frac{1}{D}\left[\int_{H^n} f(y',z')p(y',z')dy'dz'\left[\int_{H^{n-s}} f(y',\hat{z})p(y',\hat{z}\,|\,y')d\hat{z} - \int_{H^n} f(y,z)p(y,z)dydz\right]\right]$$

Total effect index of subset y

$$S_y^T = 1 - \frac{1}{D}\left[\int_{H^{n-s}} p(z)dz\left[\int_{H^s} f(\overline{y},z)p(\overline{y},z\,|\,z)d\overline{y}\right]^2 - f_0^2\right]$$

$$S_y^T = \frac{1}{2D}\int_{H^{n+s}} [f(y,z) - f(\overline{y}',z)]^2 p(y,z)p(\overline{y}',z\,|\,z)dyd\overline{y}dz$$

Requires sampling from multivariate probability distributions ($p(y)$, $p(y,\overline{z}\,|\,y)$, etc)

"Modified Sobol' formulas " (S. Kucherenko et al. Comput. Phys. Commun. 2012)

Models with pairwise correlated variables. Copula. Uniform distributions

$u = u_1,...,u_n,\ u_i \in [0,1],\ i = 1,...,n,\ \Sigma_u$ - correlation matrix .

Gaussian copula function:

$C(u_1,...,u_n;\Sigma_u) = F_n(F^{-1}(u_1),...,F^{-1}(u_n);\Sigma)$.

$F_n(\xi)$ - n-variate cumulative normal distribution function (NDF)

$F\ (\xi_i) -$ univariate NDF.

F^{-1} - inverse NDF

$\bar{u}$(independent uniform) $\to\ \bar{\xi}$(independent normal)

$\to\ \xi$(dependent normal, Σ) $\to\ u$(dependent uniform, Σ_u)

(It requires mapping $\Sigma_u \to \Sigma$)

$$u = T(\bar{u})$$

We can also use the inverse transformation

$$\bar{u} = T^{-1}(u)$$

Sobol' Sensitivity Indices. Models with pairwise correlated variables.

Main effect SI:

$$S_y = \frac{1}{D}\left[\int_{R^s} \Phi_s(y)dy \left[\int_{R^{n-s}} f(\bar{G}_s^{-1}(\bar{F}_s(y)), \bar{G}_{n-s}^{-1}(\bar{F}_{n-s}(z)))\Phi_{n-s}(y,z \mid y)dz\right.\right.$$

$$\left.\left.\int_{R^{n-s}} f(\bar{G}_s^{-1}(\bar{F}_s(y)), \bar{G}_{n-s}^{-1}(\bar{F}_{n-s}(\bar{z}')))\Phi_{n-s}(y,\bar{z}' \mid y)d\bar{z}'\right] - f_0^2\right],$$

Total order effect SI:

$$S_y^T = \frac{1}{2D}\int_{R^{n+s}} [f(\bar{G}_s^{-1}(\bar{F}_s(y)), \bar{F}_{n-s}(z)) - f(\bar{G}_s^{-1}(\bar{F}_s(\bar{y}')), \bar{G}_{n-s}^{-1}(\bar{F}_{n-s}(z)))]^2 \cdot$$

$$\cdot \Phi_{n-s}(z)\Phi_s(y,z \mid z)\Phi_s(\bar{y}', z \mid z)dyd\bar{y}'dz.$$

Here

$$\bar{G}_s^{-1}(\bar{F}_s(y)) = (G_1^{-1}(F(x_1)), \ldots, G_s^{-1}(F(x_s)),$$

$$\Phi_n(x) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

Imperial College London

Correlated variables. Test example: Gaussian Hyperplane

- *Model:*

Inputs are Gaussians

$$Y = f\left(X_1, X_2, X_3\right) = X_1 + X_2 + X_3$$

- *Correlation matrix:*

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho\sigma \\ 0 & \rho\sigma & \sigma^2 \end{pmatrix}$$

- *Sensitivity indices:*

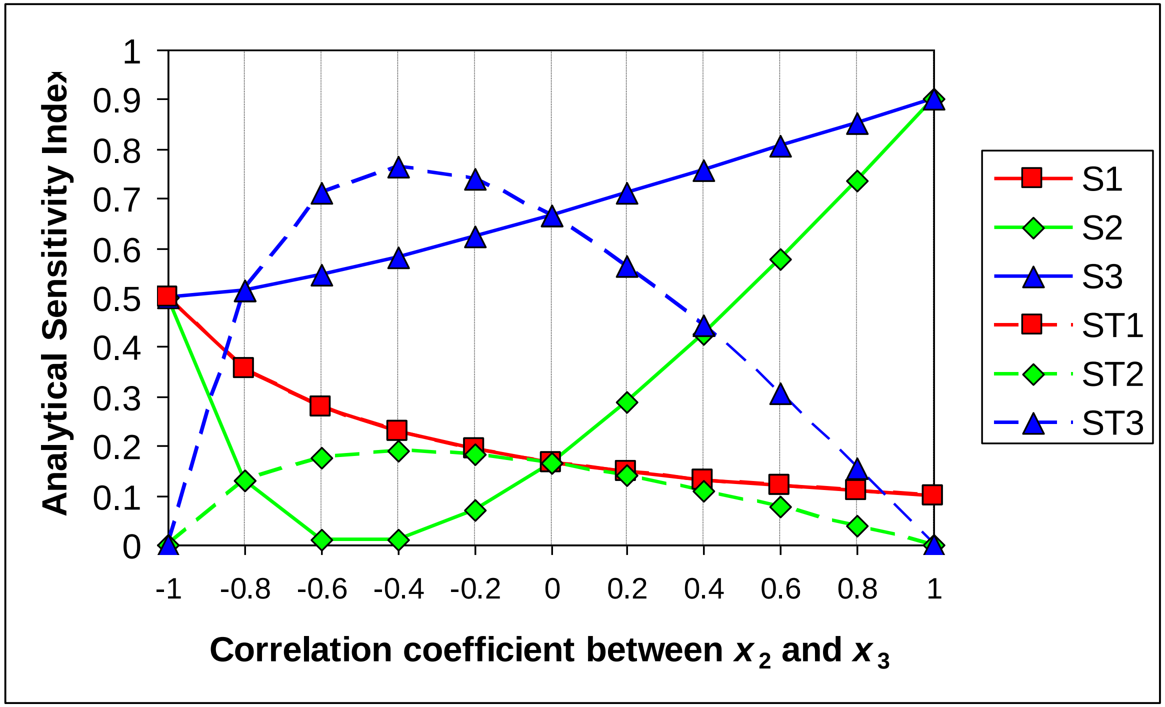
$$S_1 = \frac{1}{2+\sigma^2+2\rho\sigma}, S_1^T = \frac{1}{2+\sigma^2+2\rho\sigma};$$

$$S_2 = \frac{\left(1+\rho\sigma\right)^2}{2+\sigma^2+2\rho\sigma}, S_2^T = \frac{1-\rho^2}{2+\sigma^2+2\rho\sigma};$$

$$S_3 = \frac{\left(\sigma+\rho\right)^2}{2+\sigma^2+2\rho\sigma}, S_3^T = \frac{\sigma^2\left(1-\rho^2\right)}{2+\sigma^2+2\rho\sigma}.$$

Evolution of the first and total order indices at different values of correlation ratio ρ

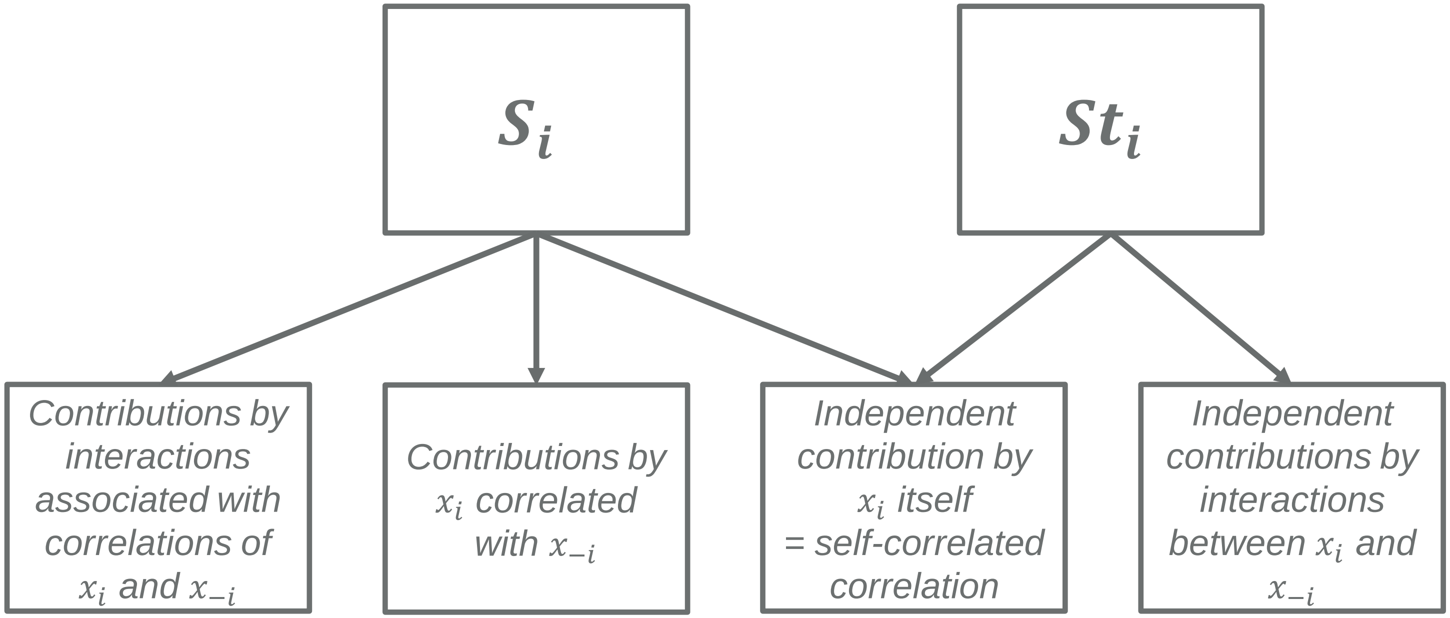
Quasi MC sample size $N = 2^{13}$, σ=2.0



$$S_i^T \le S_i \ , \ i = 2,3 \ \text{ if } \rho \ge 0 \text{ or } \rho \le -\frac{2\sigma}{\sigma^2 + 1}$$

$$S_2^T \to 0, S_3^T \to 0 \text{ if } |\rho| \to 1$$

Interpretation



Hao W , Lu Z , Li L . A new interpretation and validation of variance based importance measures for models with correlated inputs. Comput Phys Commun 2013;184:1401–13 :

Sy - the total correlated contribution

STy - the total uncorrelated contribution

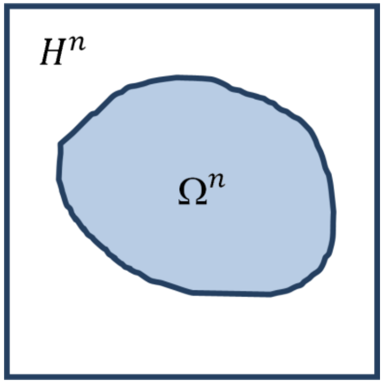
Mara TA , Tarantola S , Annoni P . Non-parametric methods for global sensitivity analysis of model output with dependent inputs. Environ Modell Soft 2015;72:173–83:

Sy - full first-order sensitivity index

STy - independent total sensitivity index

Imperial College London

Constrained Global Sensitivity Analysis (cGSA)



Problem setting: $f(x), \quad x \in \Omega^n \subset H^n$

Joint PDF of inputs: $p(x)$ in $H^n \supset \Omega^n$

or $p^{\Omega}(x)$ in Ω^n

Domain $\Omega^n \subset H^n$ may be defined by a number of constraints:

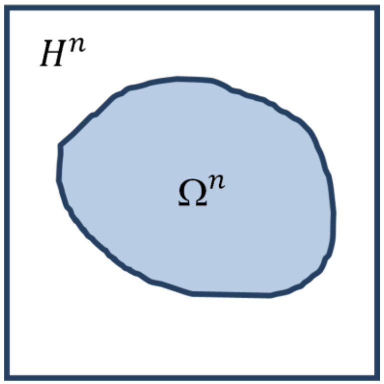
$$\Omega^n = \{x : g_m(x) \geq 0,\ m = 1, \ldots, M\}$$

Constraint types:

- geometrical, physical, chemical, biological, economical, etc.
- ‘input’ (explicit) or ‘output’ (implicit) constraints:

$$f(x) \geq f_{\min} \quad \Rightarrow \quad g(x) = f(x) - f_{\min} \geq 0$$

Acceptance-rejection method



Recall Sobol' SI's:

$$S_y = \frac{1}{D}\left[\int_{H^s} p(y)dy\left[\int_{H^{n-s}} f(y,\bar{z})\,p(y,\bar{z}\mid y)d\bar{z}\right]^2 - f_0^2\right]$$

$$S_y^T = 1-\frac{1}{D}\left[\int_{H^{n-s}} p(z)dz\left[\int_{H^s} f(\bar{y},z)\,p(\bar{y},z\mid z)d\bar{y}\right]^2 - f_0^2\right]$$

How to sample PDFs (marginal, conditional, …) in non-rectangular domains?

$$p^\Omega(y,z) = \frac{p(y,z)I^\Omega(y,z)}{\int_{\Omega^n} p(y,z)dydz} = \frac{p(y,z)I^\Omega(y,z)}{\bar{I}}$$

$$p^\Omega(y) = \int_{\Omega^n} p^\Omega(y,z)dz = \frac{1}{\bar{I}}\int_{H^{n-s}} p(y,z)I^\Omega(y,z)dz$$

$$p^\Omega(y,\bar{z}\mid y) = \frac{p^\Omega(y,z)}{p^\Omega(y)} = \frac{p(y,z)I^\Omega(y,z)}{\int_{H^{n-s}} p(y,z)I^\Omega(y,z)dz}$$

Set indicator:

$$I^\Omega(y,z) = \begin{cases} 1, & (y,z)\in\Omega^n \\ 0, & (y,z)\notin\Omega^n \end{cases}$$

Scaling factor:

$$\bar{I} = \int_{H^n} p(y,z)I^\Omega(y,z)dydz$$

Acceptance-rejection method

Explicit integral formulas for function mean and variance in Ω :

$$f_0 = \int_{\Omega^n} f(y,z)p^{\Omega}(y,z)dydz = \frac{1}{\bar{I}}\int_{H^n} f(y,z)p(y,z)I^{\Omega}(y,z)dydz$$

$$D = \int_{\Omega^n} f^2(y,z)p^{\Omega}(y,z)dydz - f_0^2 = \frac{1}{\bar{I}}\int_{H^n} f^2(y,z)p(y,z)I^{\Omega}(y,z)dydz - f_0^2$$

…and first-order and total SI in Ω:

$$S_y = \frac{1}{D}\left[\int_{H^s} \frac{\left[\int_{H^{n-s}} f(y,z)p^{\Omega}(y,z)dz\right]^2}{p^{\Omega}(y)}dy - f_0^2\right]$$

$$S_y^T = 1 - \frac{1}{D}\left(\int_{H^{n-s}} \frac{\left[\int_{H^s} f(y,z)p^{\Omega}(y,z)dy\right]^2}{p^{\Omega}(z)}dz - f_0^2\right)$$

Imperial College London

Acceptance-rejection method

Modified formulas:

$$S_y = \frac{1}{D}\left[\int_{H^n} f(y',z')p^{\Omega}(y',z')dy'dz'\left[\int_{H^{n-s}} \frac{f(y',z)}{p^{\Omega}(y')} p^{\Omega}(y',z)dz - \int_{H^n} f(y,z)p^{\Omega}(y,z)dydz\right]\right]$$

$$S_y^T = \frac{1}{2D}\int_{H^n}\int_{H^s} [f(y,z) - f(y',z)]^2 p^{\Omega}(y,z)\frac{p^{\Omega}(y',z)}{p^{\Omega}(z)} dydy'dz$$

Monte Carlo estimators

MC estimators of function mean and total variance:

$$f_0 \approx \frac{1}{\bar{I}\,N}\sum_{l=1}^{N} f(y_l, z_l) I^{\Omega}(y_l, z_l) \qquad D \approx \frac{1}{\bar{I}\,N}\sum_{l=1}^{N}\left[f(y_l, z_l) - f_0\right]^2 I^{\Omega}(y_l, z_l)$$

Scaling factor: $$\bar{I} \approx \frac{1}{N}\sum_{l=1}^{N} I^{\Omega}(y_l, z_l)$$

Double loop reordering (DLR) formula for first-order indices:

$$S_y \approx \frac{1}{D}\left[\frac{1}{\bar{I}\,N_y}\sum_{j=1}^{N_y}\frac{F^2(y_j^A)}{p^{\Omega}(y_j^A)} - f_0^2\right]$$

where

$$F(y_j^A) \approx \frac{1}{\bar{I}\,N_z}\sum_{k=1}^{N_z} f(y_{j_k}, z_{j_k}) I^{\Omega}(y_{j_k}, z_{j_k}) \qquad p^{\Omega}(y_j^A) \approx \frac{1}{\bar{I}\,N_z}\sum_{k=1}^{N_z} I^{\Omega}(y_{j_k}, z_{j_k})$$

Sample is subdivided into $N_y \approx \sqrt{N}$ 'bins'

Total number of sample points: $N_{\text{CPU}} = N = N_y N_z$

Imperial College London

Double loop reordering approach (DLR)

$$E_z(f(x_1,\overline{z})\,|\,x_1)$$

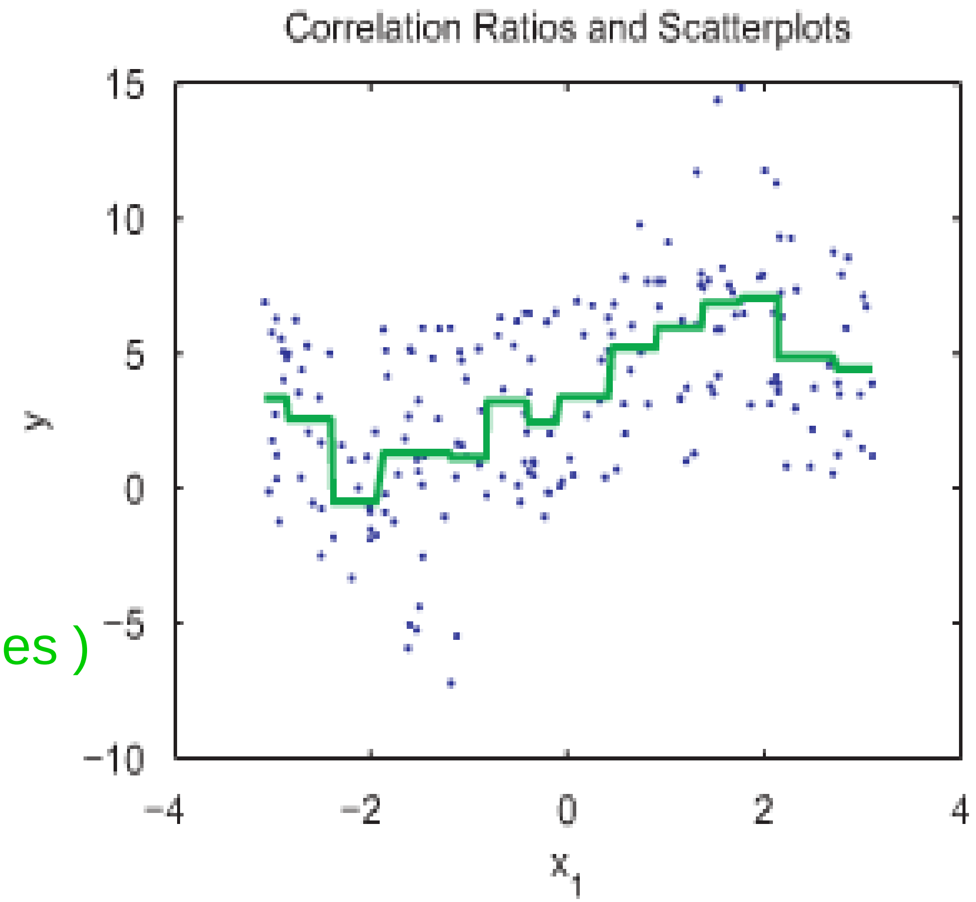
$$S_{x_1} = \frac{D_{x_1}[E_z(f(x_1,\overline{z})\,|\,x_1)]}{D}$$

Sort $(x_1, Y = f)$

Divide space into M bins

Compute local mean values (green lines)

Estimate variance D_{x1}



Imperial College London

Monte Carlo estimators

Modified formulas for first-order and total indices:

$$S_y = \frac{1}{\bar{I}^2 DN} \sum_{l=1}^{N} \left(f(y_l', z_l') I(y_l', z_l') \left(\bar{I} \frac{f(y_l', z_l) I(y_l', z_l)}{p^{\Omega}(y_l')} - f(y_l, z_l) I(y_l, z_l) \right) \right)$$

$$S_y^T = \frac{1}{2 \bar{I} DN} \sum_{l=1}^{N} \left(f(y_l, z_l) I(y_l, z_l) - f(y_l', z_l) I(y_l', z_l) \right)^2 \frac{1}{p^{\Omega}(z_l)}$$

Total number of sampled points: $N_{\mathrm{CPU}} = N(n+2)$

Rosenblatt transformation

Let $\gamma_1,\dots,\gamma_n$ be independent random numbers uniformly distributed on [0,1]. The set of random values $\{\xi_1,\dots,\xi_n\}$ defined on Ω^n obtained from

$$F_1(\xi_1)=\gamma_1,$$
$$F_2(\xi_2\mid\xi_1)=\gamma_2,$$
$$\dots\dots\dots$$
$$F_n(\xi_n\mid\xi_1,\dots,\xi_{n-1})=\gamma_n$$

has the pdf $p^{\Omega}(\xi_1,\dots,\xi_n)$. Here, $F_1(\xi_1),F_2(\xi_2\mid\xi_1),\dots,F_n(\xi_n\mid\xi_1,\dots,\xi_{n-1})$ are cumulative distributions corresponding to $\{\xi_1,\dots,\xi_n\}$

Test case 1. 2D g-function, linear constraint

Function:

$$g = \prod_{i=1}^{n} \frac{|4x_i - 2| + a_i}{1 + a_i}$$

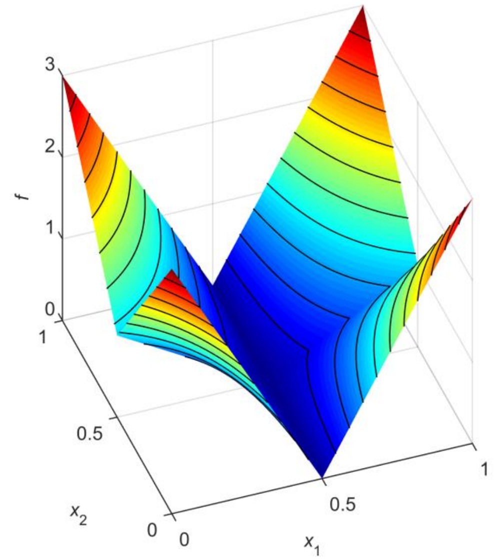
$$n = 2, \quad a_1 = 0, \quad a_2 = 1$$

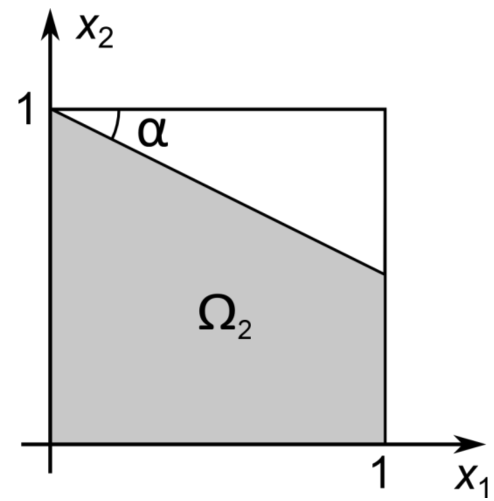
Joint PDF:

$$p(x_1, x_2) = 1$$

Permissible domain:

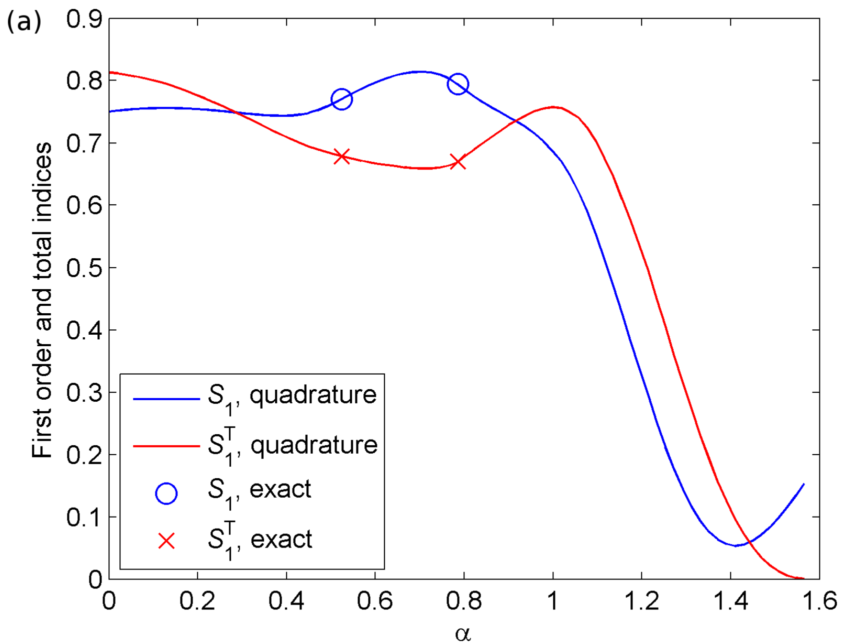
$$\Omega: \quad x_2 \le 1 - x_1 \tan\alpha$$
$$0 \le \alpha < \pi / 2$$

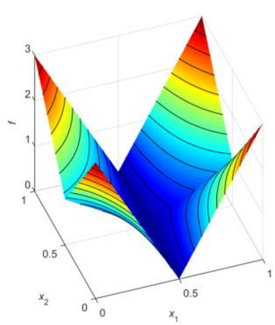




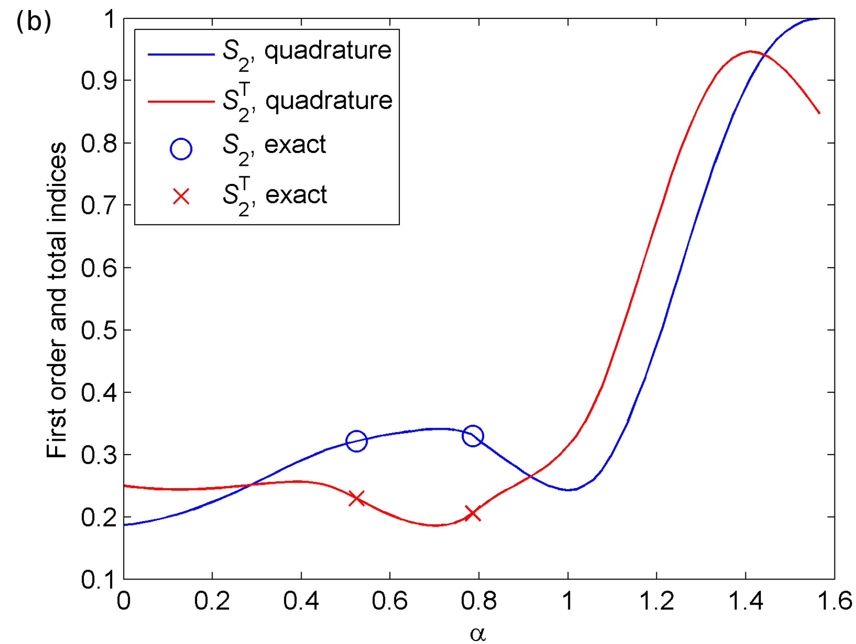
Test case 1. 2D g-function, Sobol SI's versus alfa

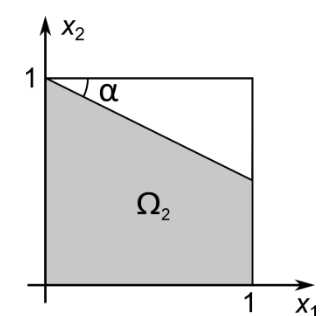
S1, S1_Total





S2, S2_Total



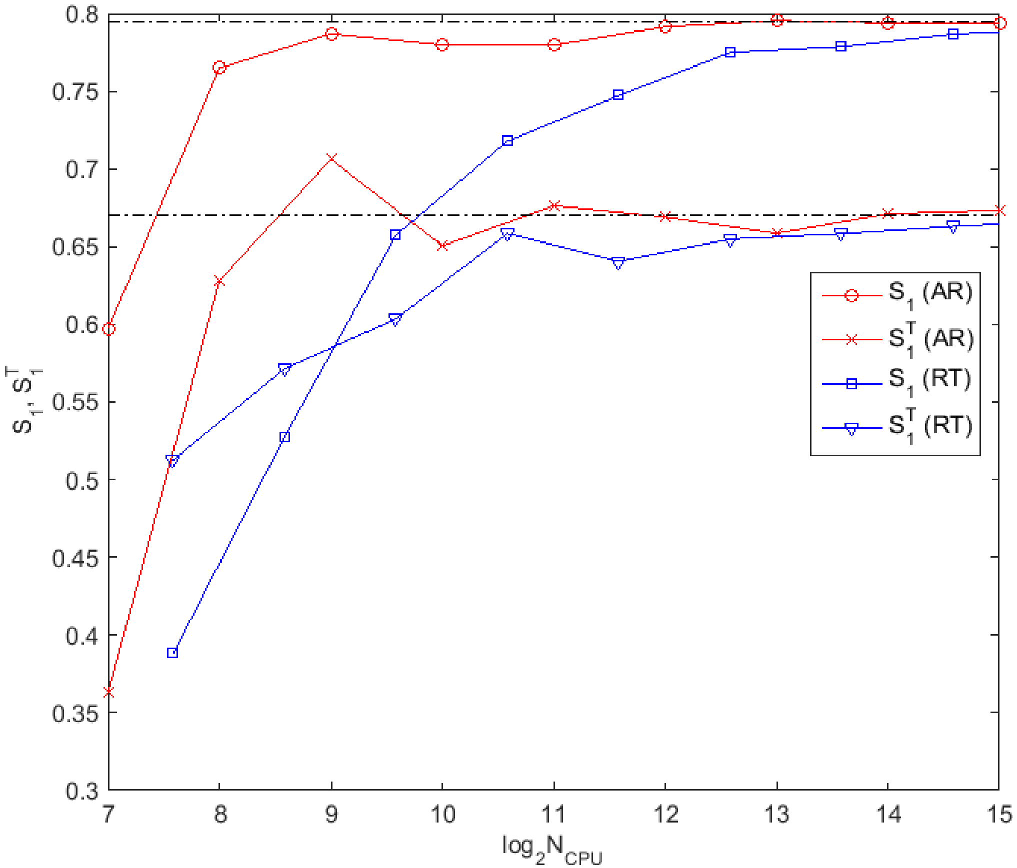


Imperial College London

Test case 1. 2D g-function, two approaches

Sampling using Rosenblatt transformation

$$x_1 = \sqrt{\gamma_1},$$
$$x_2 = x_1(\gamma_2 - 1) + 1.$$



S1, S1_Total Convergence of numerical estimates for and S1, S1_Total for the original formulas using Rosenblatt transformation (RT) with those based on the acceptance-rejection (AR) approach, alfa=PI/4

Imperial College London

Test case 2. 2D g-function, parabolic constraint

Function:

$$g = \prod_{i=1}^{n} \frac{|4x_i - 2| + a_i}{1 + a_i}$$

$$n = 2, \quad a_1 = 0, \quad a_2 = 1$$

Joint PDF:

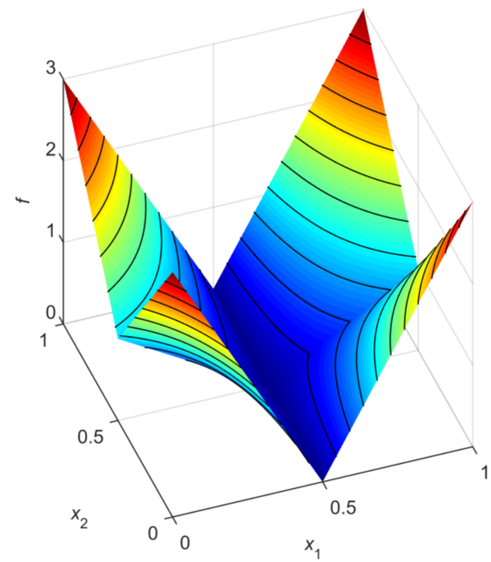
$$p(x_1, x_2) = 1$$

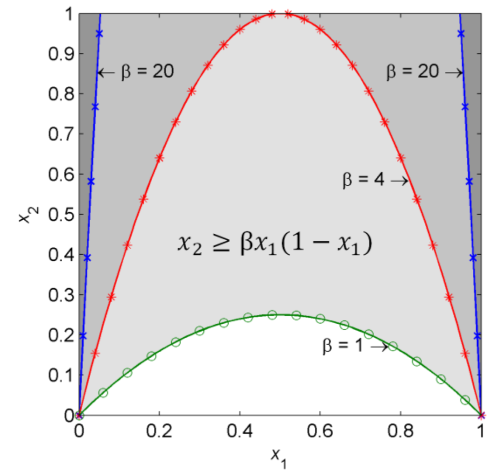
Permissible domain:

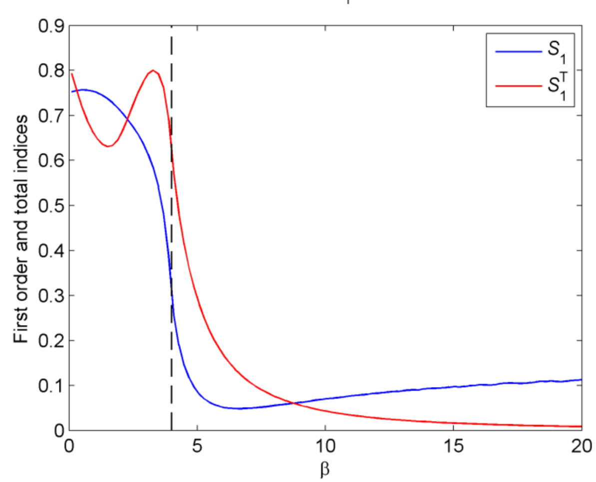
$$\Omega: \quad x_2 \geq \beta x_1 (1 - x_1)$$

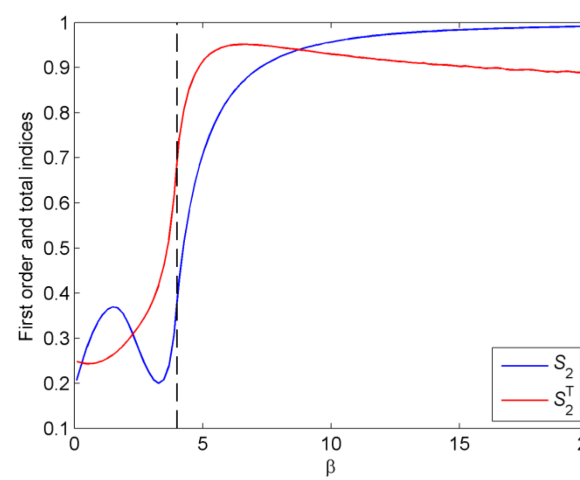
$$\beta > 0$$

Domain is disconnected for $\beta > 4$!









Test case 3. 10D g-function, minimum threshold constraint

Function:

$$g = \prod_{i=1}^{n} \frac{|4x_i - 2| + a_i}{1 + a_i}$$

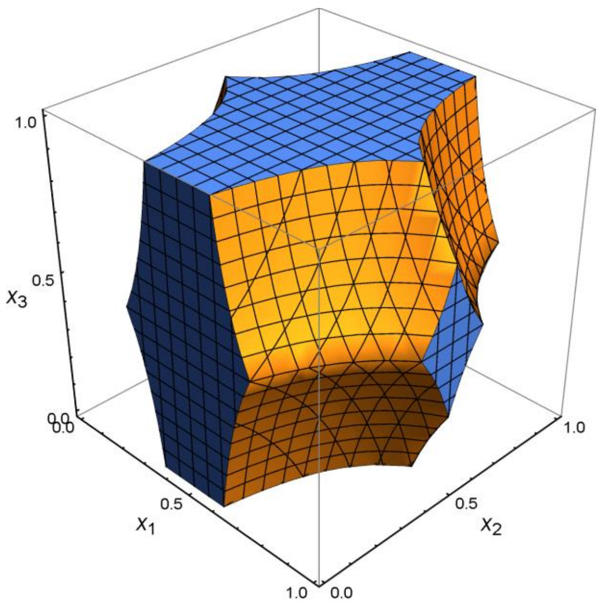
$$n = 10, \quad a_i = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]$$

Joint PDF:

$$p(x_1, x_2) = 1$$

Permissible domain:

$$g(x_1, \ldots, x_{10}) \geq 1$$



Projection of the 10-dimentional feasible domain onto the first three dimensions.

i

Imperial College London

Test case 3. 10D g-function, values of Sobol' indices

	S_i (GSA)	S_i (cGSA)	S_i^T (GSA)	S_i^T (cGSA)
1	0.56066	0.68667	0.67050	0.95605
2	0.14017	0.00482	0.20631	0.19810
3	0.06230	0.00217	0.09579	0.10126
4	0.03504	0.00137	0.05473	0.06135
5	0.02243	0.00109	0.03529	0.04120
6	0.01557	0.00091	0.02461	0.02958
7	0.01144	0.00089	0.01812	0.02248
8	0.00876	0.00114	0.01390	0.01747
9	0.00692	0.00184	0.01100	0.01402
10	0.00561	0.00103	0.00891	0.01149

Both the main and total effects of x_1 (the most influential input) increase upon the introduction of the constraint, other main indices decrease. However, their total indices exceed those for the unconstrained case owing to the presence of the structural dependences.

Imperial College London

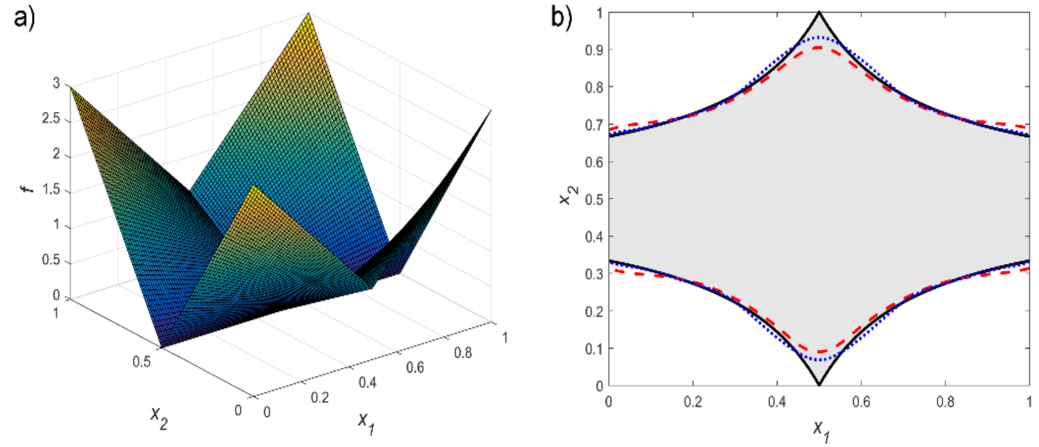
Test case 4. 2D g-function, Metamodelling approach

$$g = \prod_{i=1}^{n} \frac{|4x_i - 2| + a_i}{1 + a_i}$$

$$n = 2, \ \ a_1 = 0, \ \ a_2 = 1$$

Permissible domain:

$$g(x_1, x_2) \le 1$$



Boundary of the feasible domain: exact (solid line), 512-sample HDMR (dashed line), 1024-samlpe HDMR (dotted line).

Function&constraint	S_1	S_2	S_1^T	S_2^T
Exact	0.8086	0.0013	0.9978	0.1913
RS-HDMR(512)	0.7572	0.0179	0.9836	0.2426
RS-HDMR(1024)	0.8246	0.0061	0.9929	0.1750

Values of Sobol' indices for exact model&constraints are close to those build using metamodels. This approach can significantly reduce CPU time for complex (expensive) models.

SobolGSA: www3.imperial.ac.uk/centreforprocesssystemsengineering/software1/sobolhdmr

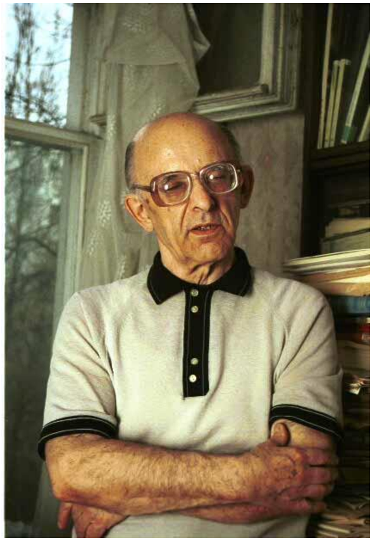
Imperial College London

Conclusions

- A large new class of models with inequality constraints can be analysed with cGSA
- Suggested acceptance-rejection formulas only require the knowledge of joint PDF and domain indicator function
- Approach based on building metamodels for model&constraints can significantly reduce CPU time for complex (expensive) models.
- Further work is required on the interpretation of cGSA results

Acknowledgments

Prof. Sobol'



Imperial College London

References

Sobol' I., Kucherenko S. On global sensitivity analysis of quasi-Monte Carlo algorithms. *Monte Carlo Methods and Simulation*, 11, 1, 1-9, 2005.

Sobol' I., Kucherenko S. Global Sensitivity Indices for Nonlinear Mathematical Models. Review, *Wilmott*, 56-61, 1, 2005.

Kucherenko S., Shah N. The Importance of being Global.Application of Global Sensitivity Analysis in Monte Carlo option Pricing *Wilmott*, 82-91, July 2007.

Sobol, I.M., S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz. Estimating the Approximation Error when fixing Unessential Factors in Global Sensitivity Analysis, Reliability Engineering & System Safety, 92(7): 957-960, 2007.

Kucherenko S., Rodriguez-Fernandez M., Pantelides C., Shah N.. Monte Carlo evaluation of derivative based global sensitivity measures. *Reliability Engineering & System Safety* ,Vol. 94, Issue 7, p 1135-1148, 2009

Gatelli D., Kucherenko S., Ratto M., Tarantola S. Calculating first order sensitivity measures: a benchmark of some recent methodologies. *Reliability Engineering and System Safety* 94 (7), p 1212-1219, 2009

Sobol' I.M., Kucherenko S., Derivative based Global Sensitivity Measures and their link with global sensitivity indices, *Mathematics and Computers in Simulation*, V 79, Issue 10, pp. 3009-3017, June 2009.

Kucherenko S., Feil B., Shah N., Mauntz W. The identification of model effective dimensions using global sensitivity analysis Reliability Engineering and System Safety 96 (2011) 440–449

Zuniga M, Kucherenko S, Shah N, Metamodelling with independent and dependent inputs, *Computer Physics Communications*, Volume 184, Issue 6, p. 1570-1580, 2013

Kucherenko S, Tarantola S., Annoni P. , Estimation of global sensitivity indices for models with dependent variables, *Computer Physics Communications, V. 183 (2012) 937–946*

Kucherenko S., Klymenko O.V. , Shah N. , Sobol' indices for problems defined in non-rectangular domains, Reliability Engineering and System Safety 167 (2017) 218–231