Imperial College London

Global sensitivity analysis of models with correlated inputs and models with constraints

S. Kucherenko, *O.V. Klymenko, N. Shah*

Imperial College London

s.kucherenko@imperial.ac.uk

Imperial College London

Outline

1. Global Sensitivity Analysis (GSA) and Sobol' Sensitivity Indices for case of independent inputs
2. Sobol' Sensitivity Indices for case of dependent inputs
3. Correlated inputs
4. Constrained Global Sensitivity Analysis (cGSA)
5. Conclusions

Imperial College London

ANOVA decomposition and Sensitivity Indices

Consider a model

x is a vector of input variables

f(x) is square integrable

$$Y = f(X)$$
$$X = (X_1, X_2, \ldots, X_n) \in H^n$$
$$0 \le X_i \le 1$$

ANOVA decomposition is unique if ***variables are independent***

$$Y = f(X) = f_0 + \sum_{i=1}^{n} f_i(X_i) + \sum_i \sum_{j>i} f_{ij}(X_i, X_j) + \ldots + f_{1,2,\ldots,n}(X_1, X_2, \ldots, X_n),$$

$$\int_0^1 f_{i_1 \ldots i_s}(X_{i_1}, \ldots, X_{i_s}) dX_{i_p} = 0, \ \forall p, \ 1 \le p \le s, \rightarrow \int_0^1 f_{i_1 \ldots i_s} f_{i_1 \ldots i_l} dX_{i_p} dX_{i_l} = 0, \ \forall i_p \ne i_l$$

Variance decomposition:

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

Sobol' SI:

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

Imperial College London

Sobol' Sensitivity Indices (SI)

- *Definition:* $$S_{i_1 \ldots i_s} = D_{i_1 \ldots i_s} / D$$

$$D_{i_1 \ldots i_s} = \int_0^1 f^2_{i_1 \ldots i_s}\left(x_{i_1}, \ldots, x_{is}\right) dx_{i_1}, \ldots, x_{is}$$ *- partial variances*

$$D = \int_0^1 \left(f\left(x\right) - f_0\right)^2 dx$$ *- total variance*

- *Sensitivity indices for subsets of variables:* $$x = \left(y, z\right)$$

$$D_y = \sum_{s=1}^{m} \sum_{(i_1 \langle \ldots \langle i_s) \in K} D_{i_1, \ldots, i_s}$$

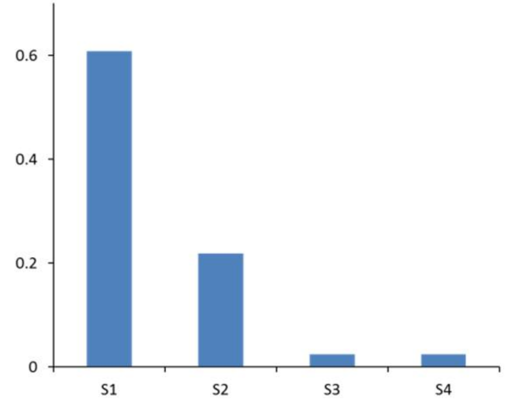
Total variance for a subset: $$\left(D_y^T\right)^2 = D - D_z$$

- *Corresponding global sensitivity indices:*

$$S_y = \sigma_y^2 / \sigma^2, \qquad S_y^T = D_y^T / D.$$

Evaluation of Sobol' Sensitivity Indices

GSA is used for:
- identification of key parameters whose uncertainty most affects the output
- decreasing problem dimensionality



Direct formulas: Sobol' 1990, 2001, 2007
Evaluation of Sobol' SI is reduced to high-dimensional integration using MC/QMC methods
Further improvements - allow a reduced number of function evaluations compared to alternatives: Saltelli 2002, 2010
S. Kucherenko 2002, 2007, 2011, 2016
A. Owen 2012, 2013

What if inputs are not independent ?
1) ANOVA decomposition is not unique;
2) Direct Sobol' formulas are not valid

Imperial College London

Sobol' SI in the case of dependent inputs

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

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

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

First-order (main) effect index

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

Total effect index of subset y

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

Imperial College London

Sobol' Sensitivity Indices

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

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

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

Total effect index of subset y

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

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

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

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

Copula. Uniform distrubutions

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

Gaussian copula function:

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

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

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

F^{-1} - inverse NDF

$\overline{u}$(independent uniform) $\rightarrow\ \overline{\xi}$(independent normal)

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

(Require mapping $\Sigma_u \rightarrow \Sigma$)

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

We can also use the inverse transformation

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

Copula. Arbitrary distributions

In a general case of arbitrary distributed r.v. X with Σ_x

and cumulative marginal distribution functions $\mathrm{G}_i(\mathrm{X}_i)$

Gaussian copula function:

$$C(G_1(X_1),\ldots,G_n(X_n);\Sigma_x) = F_n(F^{-1}(G_1(X_1)),\ldots,F^{-1}(G_1(X_n));\Sigma)$$

$\overline{u}$(independent uniform) $\rightarrow$ $\overline{\xi}$(independent normal)

$\rightarrow$ ξ(dependent normal, Σ) $\rightarrow$ X(dependent r.v, Σ_x)

(Require mapping $\Sigma_x \rightarrow \Sigma$)

$$X = T(\overline{u})$$

We can also apply the inverse transformation

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

Evaluation of Sobol' Sensitivity Indices

Main effect SI:

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

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

Total order effect SI:

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

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

Here

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

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

Correlated Inputs. Gaussian additive model

- *Model*

Inputs are Gaussians

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

- *Correlation matrix*

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

- *SI :*

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

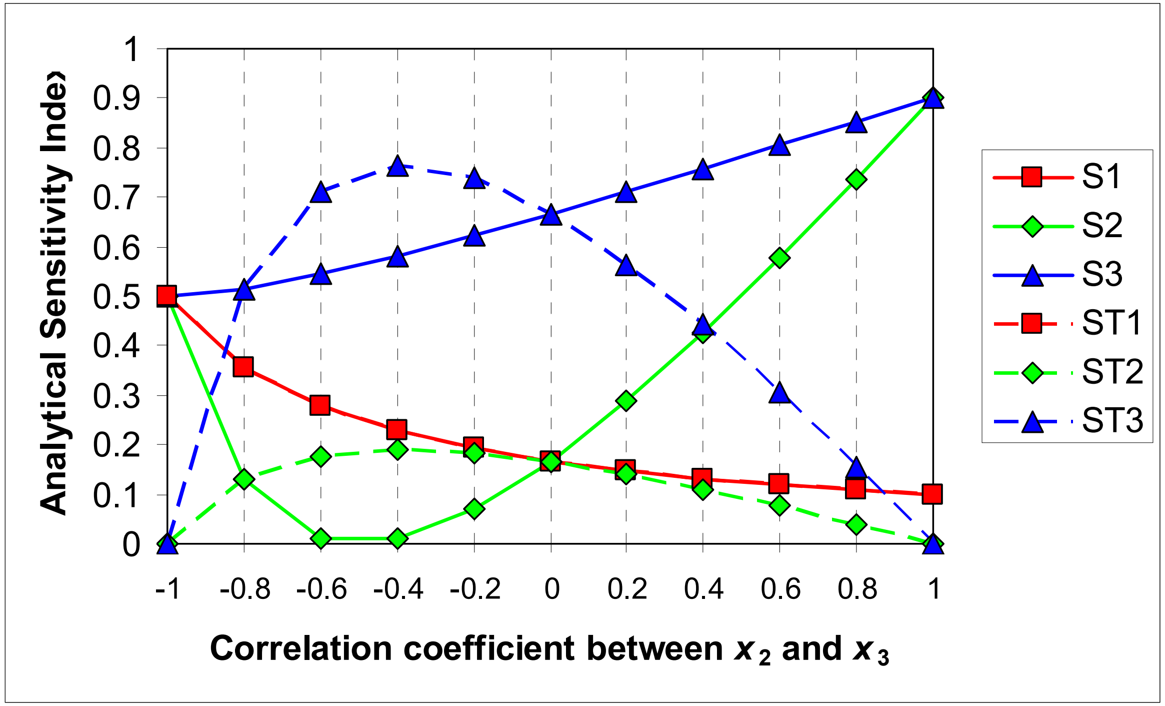
$$S_2 = \frac{\left(1+\rho\sigma\right)^2}{2+\sigma^2+2\rho\sigma}, S_2^T = \frac{1-\rho^2}{2+\sigma^2+2\rho\sigma};$$

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

Imperial College London

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

Quasi MC sample size N = 2[13]

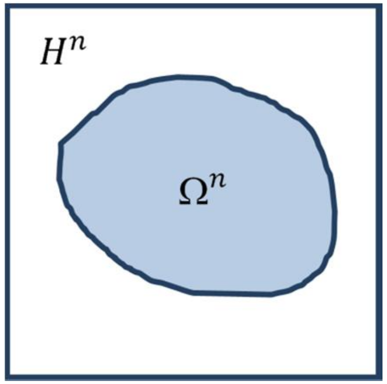


$$S_i^T \le S_i\,,\ i = 2,3 \ \text{ if } \rho \ge 0 \text{ or } \rho \le -\frac{2\sigma}{\sigma^2+1}$$

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

Constrained Global Sensitivity Analysis (cGSA)

Imperial College London



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:

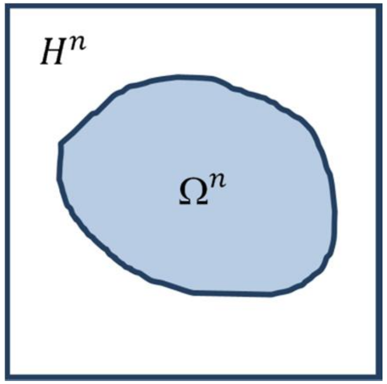
$$\Omega^n = \{x : g_m(x) \geq 0,\ m = 1, \ldots, M\}$$

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' SIs:

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

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

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

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

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

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

Set indicator:

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

Scaling factor:

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

Imperial College London

Acceptance-rejection method

Explicit integral formulas for function mean and variance in Ω :

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

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

…and first-order and total SI in Ω:

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

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

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)

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

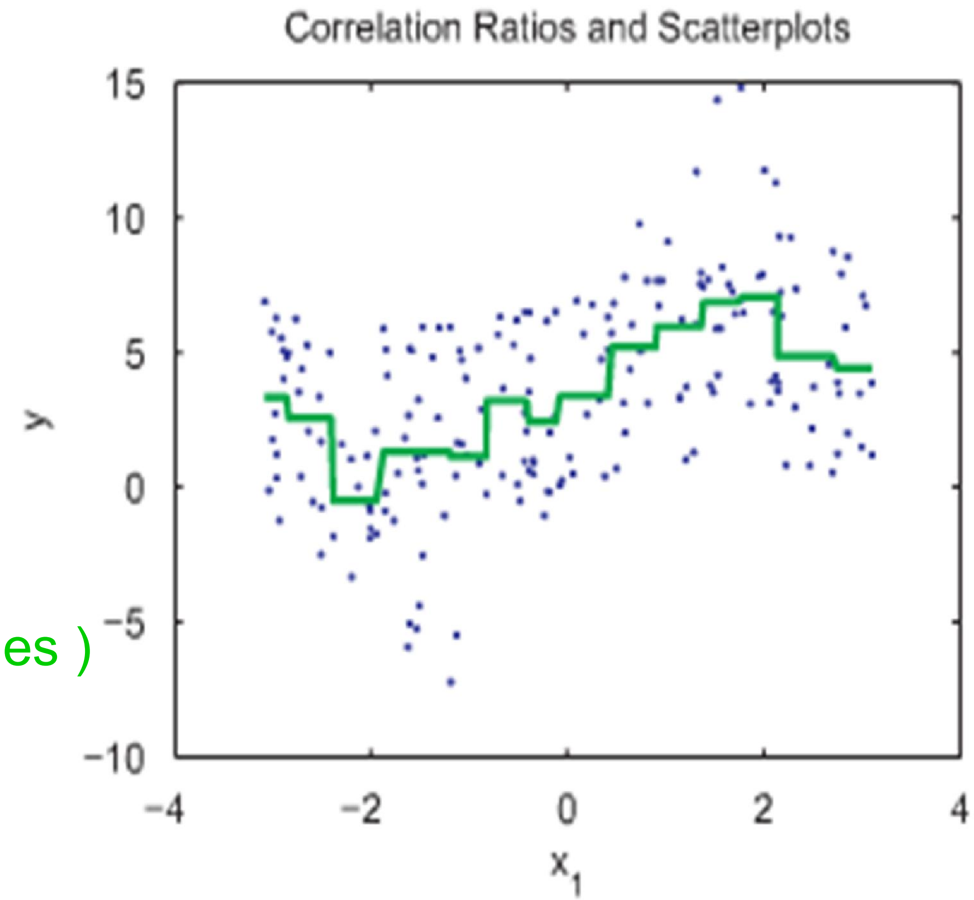
$$S_{x_1} = \frac{D_{x_1}[E_z(f(x_1,\overline{z})\,|\,x_1)]}{D}$$

Sort (x_1,Y = *f*)

Divide space into M bins

Compute local mean values (green lines)

Estimate variance D_{x1}



Monte Carlo estimators

Modified formulas for first-order and total indices:

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

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

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

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

Function:

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

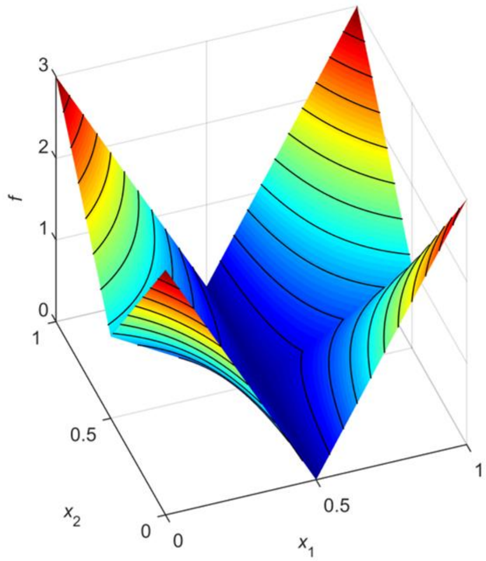
$$n = 2, \quad a_1 = 0, \quad a_2 = 1$$

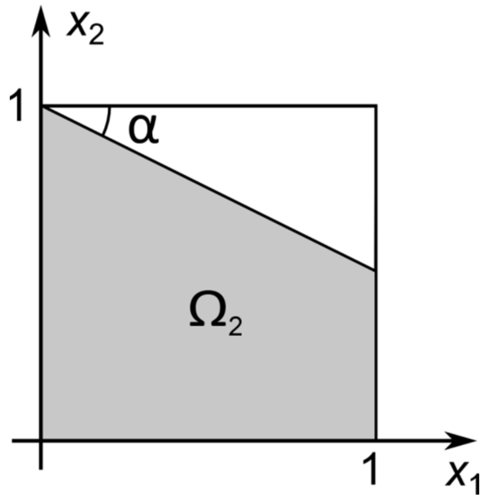
Joint PDF:

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

Permissible domain:

$$\Omega: \quad x_2 \le 1 - x_1 \tan\alpha$$
$$0 \le \alpha < \pi / 4$$

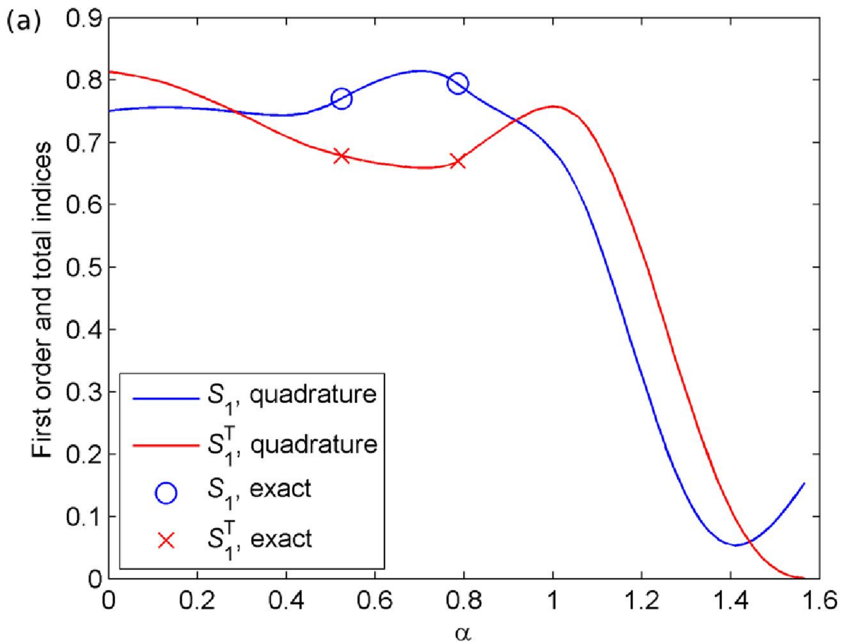


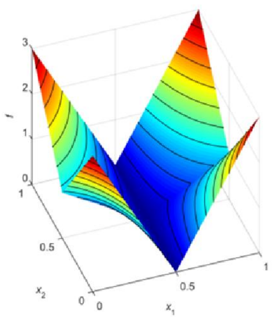


Imperial College London

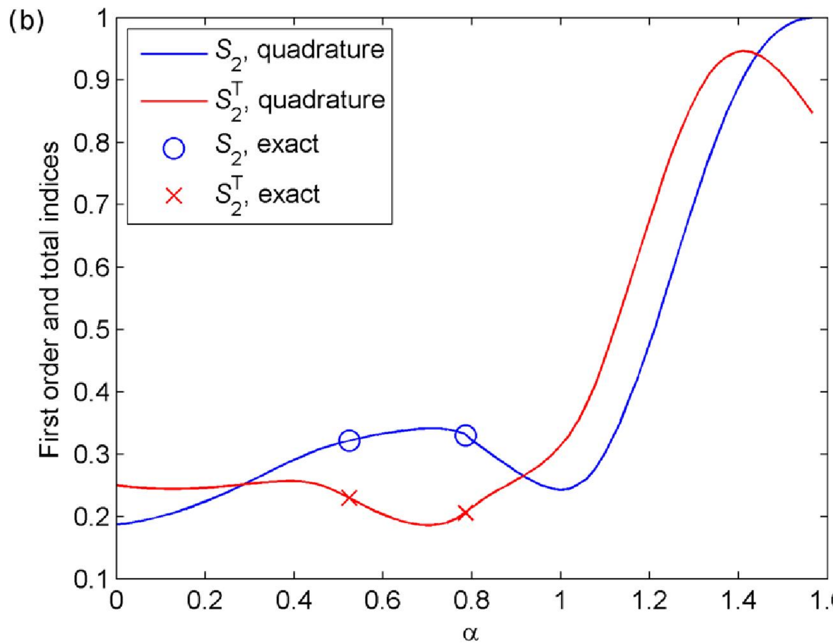
Test case 1. 2D g-function, Sobol SI's versus alfa

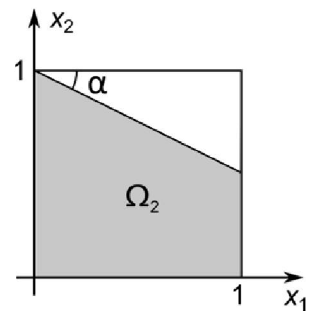
S1, S1_Total





S2, S2_Total

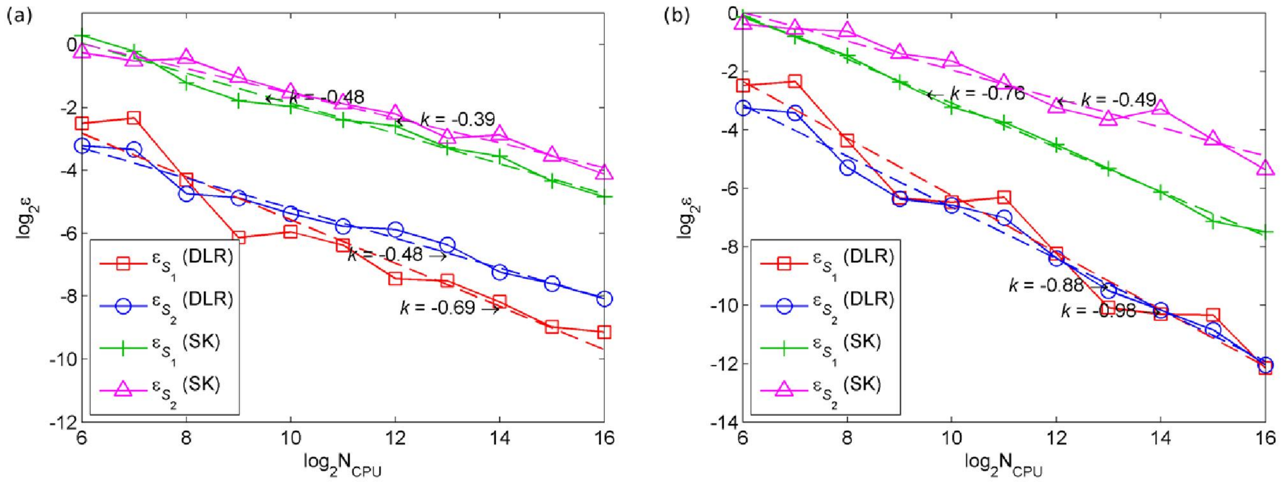




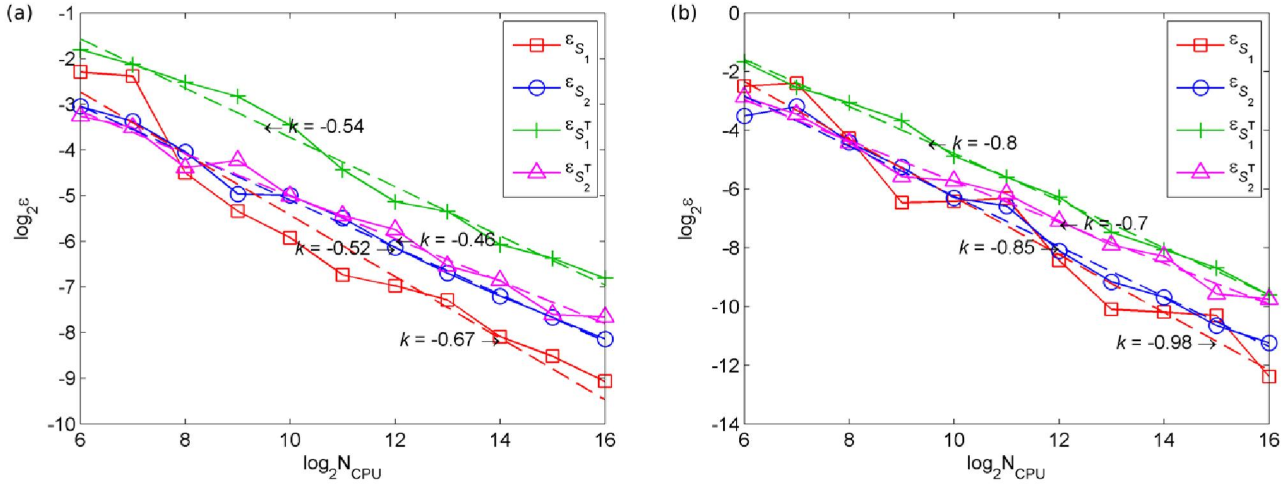
Test case 1. 2D g-function, convergence plots

$\alpha = \pi/6$

random sampling | Sobol' LDS



S_i (DLR)
S_i (modified)



S_i (DLR)
S_i^T (modified)

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

Function:

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

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

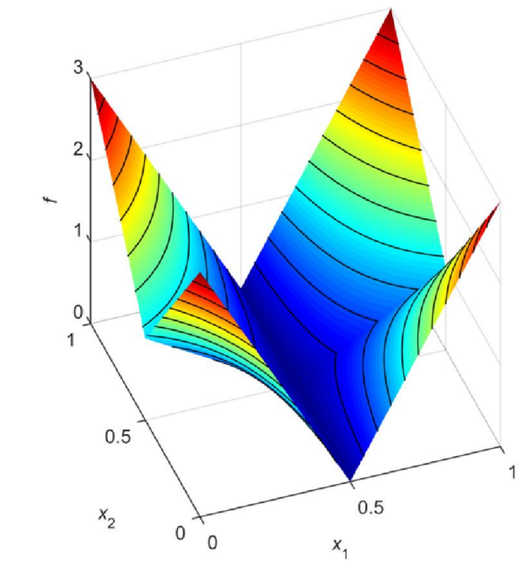
Joint PDF:

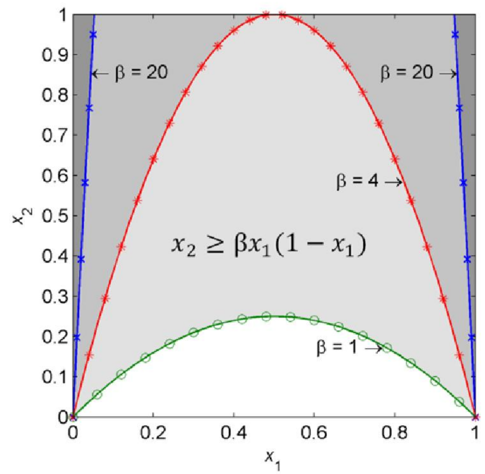
$$p(x_1, x_2) = 1$$

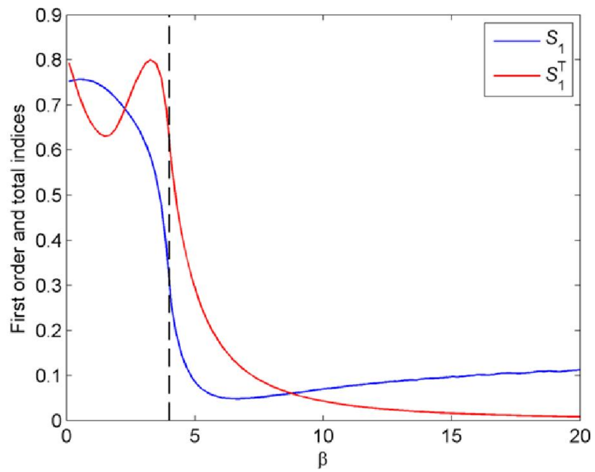
Permissible domain:

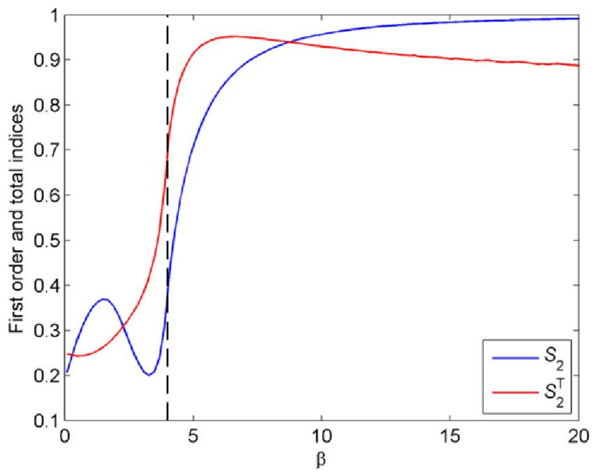
$$\Omega: \quad x_2 \geq \beta x_1 (1 - x_1)$$
$$\beta > 0$$

Domain is disconnected for $\beta > 4$!









Imperial College London

Test case 3. 4D K-function, linear constraints

Function:

$$K = -x_1 + x_1x_2 - x_1x_2x_3 + x_1x_2x_3x_4 \qquad \left(= \sum_{i=1}^{n} (-1)^i \prod_{j=1}^{i} x_j \right)$$

Permissible domains:

Two important variables constrained:

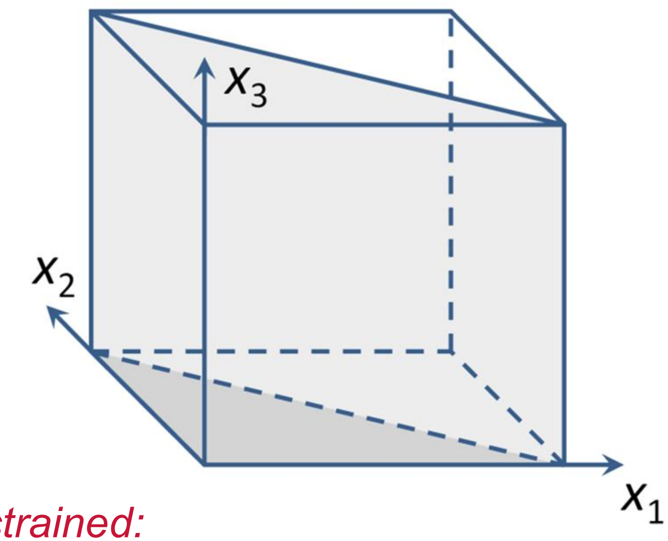
$$\Omega_1 : \qquad x_1 + x_2 \le 1$$

Two unimportant variables constrained:

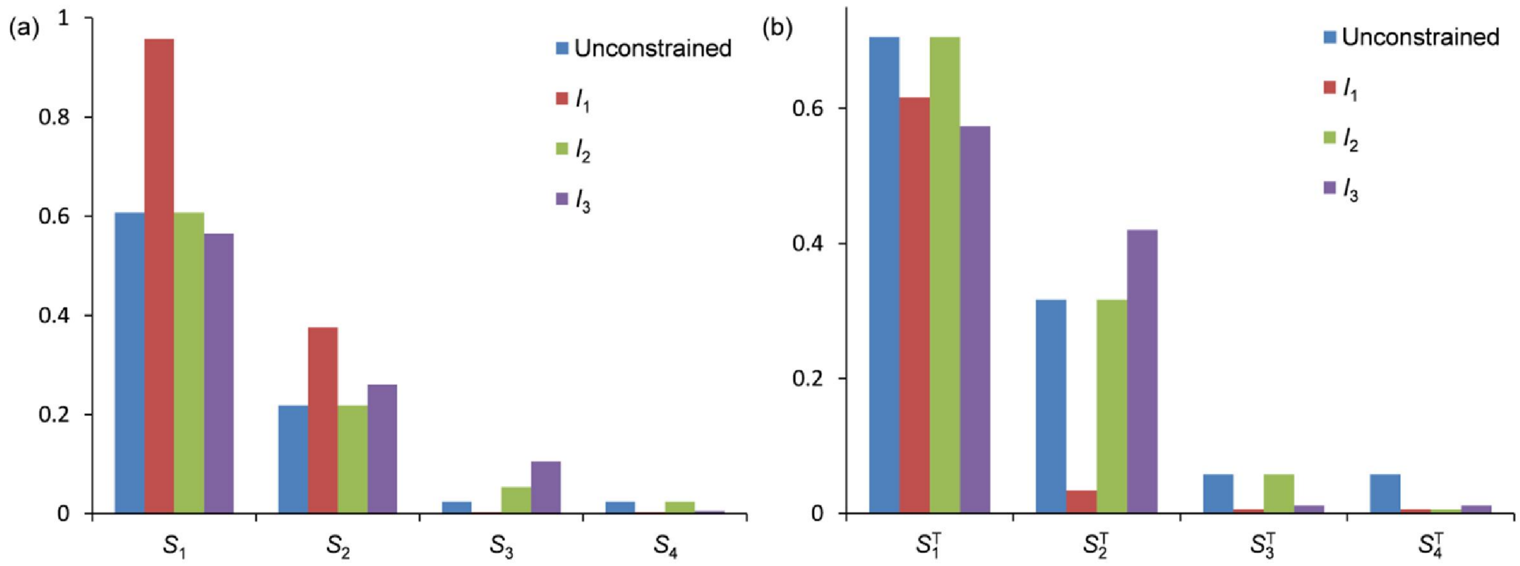
$$\Omega_2 : \qquad x_3 + x_4 \le 1$$

One important and one unimportant variables constrained:

$$\Omega_3 : \qquad x_1 + x_3 \le 1$$



Test case 3. 4D K-function, Sobol SI's



$$\Omega_1 : \quad x_1 + x_2 \le 1$$
$$\Omega_2 : \quad x_3 + x_4 \le 1$$
$$\Omega_3 : \quad x_1 + x_3 \le 1$$

Imperial College London

Conclusions

- For the case of correlated inputs a gaussian copula based approach is proposed for sampling from arbitrary multivariate probability distributions.
- 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
- Further work is required on the interpretation of cGSA results

- *References*

- *S. Kucherenko, S. Tarantola, P. Annoni. Estimation of global sensitivity indices for models with dependent variables, Computer Physics Communications, V. 183 (2012) 937–946*
- *S. Kucherenko, O.V. Klymenko, N. Shah Sobol' indices for problems defined in non-rectangular domains, Submitted to Reliability Engineering and System Safety 2016, https://arxiv.org/abs/1605.05069*