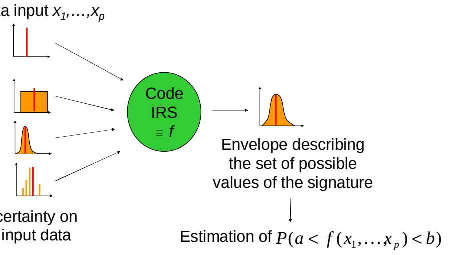
ONERA — THE FRENCH AEROSPACE LAB — retour sur innovation

Sequence Construction For Integral Estimation

S. Varet[1], S. Lefebvre[1], A. Roblin[1], G. Durand[1], S. Cohen[2]
[1]ONERA/DOTA - Chemin de la Hunière - 91761 Palaiseau Cedex
[2]IMT/LSP - 118 route de Narbonne - 31062 Toulouse Cedex 9

Context

Data input $x_1,\dots,x_p$



Dimension Reduction

Usual methods :

- Factorial design + F-test :
  - Gaussian assumption
  - **The IRS is not Gaussian**
- Functional ANOVA :
  - Independence assumption and
  - requires a lot of function evaluation
  - **The input variables are not independent**
  - **Too time consuming**

Integral estimation

$$P(a < f(X) < b) \approx \frac{1}{N}\sum_{i=1}^{N} 1\!\!1_{[a,b]^p} f(x^{(i)})$$

Monte Carlo method

$(x^{(i)})_{i=1,\dots,N}$ Uniform distribution

Convergence speed in $O(\frac{1}{\sqrt{N}})$

Too time consuming

Quasi-Monte Carlo method

$(x^{(i)})_{i=1,\dots,N}$ Low discrepancy sequence

Convergence speed in $O(\frac{(\log N)^p}{N})$

Better rate of convergence

Combination of the 2 steps

Sobol indices estimation with a 2 levels fractional factorial design

$$IS_u \approx IS_u^{frac}$$

Definition of the effective discrepancy:

Weighted sum of the projections discrepancies with the Sobol indices as weights

Gives more importance to important projections

$$D_i(x^{(1)},\dots, x^{(i)}) = \sum_u IS_u^{frac} D_{L2}(x_u^{(1)},\dots, x_u^{(i)})$$

Function dependent (IS_u^{frac}) — **Points dependent** ($D_{L2}(x_u^{(1)},\dots, x_u^{(i)})$)

Construction algorithm

```
While i <N:
  Propose a point x^(i) = (x_1^i,..., x_p^i)
  Compute the effective discrepancy D_i of (x^(1),...,x^(i))
  If D_{i-1} > k_i D_i then accept the point
  Else accept the point with probability p_i
  i = i + 1
```

The effective discrepancy

The effective discrepancy of several sequences for *f(x)*:

$f(x) : [-1; 1]^{20} \to \mathbb{R}$

$$f(x_1,\dots,x_{20}) = \sum_{i=1}^{20} \lambda_i x_i + 31x_1x_3 + 34x_2x_4 + 0{,}00001x_{18}x_{19} + 0{,}000015x_{14}x_{15}$$

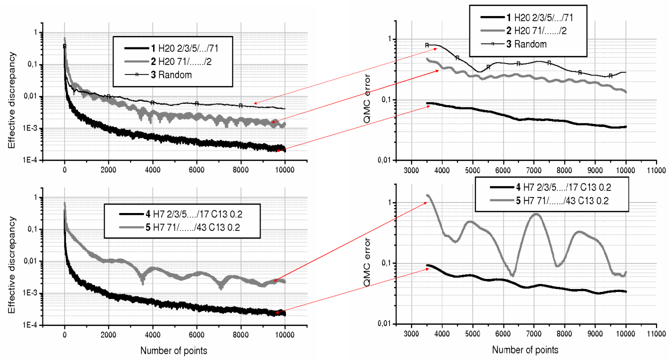
i	1	2	3	4	5	6	7	8	9	10
λ_i	30	4.10^{-5}	32	3.10^{-5}	2.10^{-5}	1	1	1	5.10^{-5}	10^{-5}

i	11	12	13	14	15	16	17	18	19	20
λ_i	6.10^{-5}	$1,5.10^{-5}$	$2,5.10^{-5}$	$3,5.10^{-5}$	$4,5.10^{-5}$	$5,5.10^{-5}$	$6,5.10^{-5}$	7.10^{-5}	$7,5.10^{-5}$	8.10^{-5}

Important factors: x_1, x_3, x_2, x_4

Moderately important factors: x_6, x_7, x_8

Important interactions: x_1x_3, x_2x_4



The effective discrepancy is correlated to the integration error

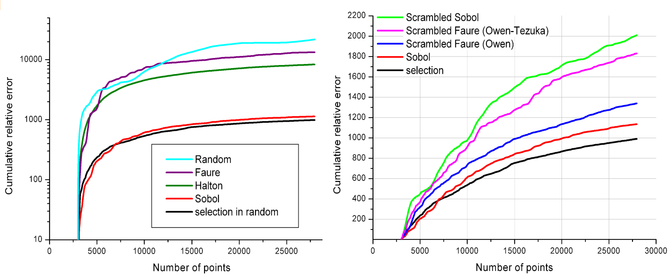
Algorithm: Test 1

$f(x) : [-1; 1]^{30} \to \mathbb{R}$ Estimation of $E(f(X)) = 1{,}3$

$$f(x_1,\dots,x_{30}) = \sum_{i=1}^{30} \lambda_i x_i + 129x_{29}x_{30} - 132x_{27}x_{28} + 133x_{29}x_{26} - 134x_{25}x_{26} + 135x_{26}x_{27} + 1{,}3$$

i	1-10	11-15	16-20	21	22	23	24	25	26	27	28	29	30
λ_i	$(-1)^{i+1} i 10^{-3}$	$(-1)^{i+1}(i-9)10^{-2}$	$(-1)^i i$	131	124	-150	-160	125	126	127	128	140	130

Algorithm parameters : $k_i = 1+1/(2i)$ and $p_i = 50/i$



Algorithm: Test 2

$g(x) : [0; 1]^6 \to \mathbb{R}$ Estimation of $E(g(X)) = 0{,}259699$

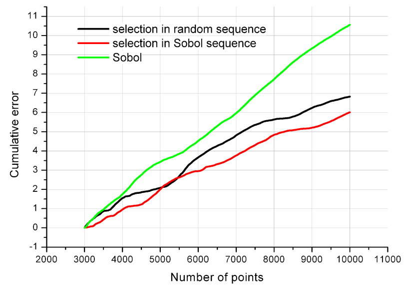
$$g(x_1,\dots,x_6) = 1\!\!1_{f(x) < -40} = 1\!\!1_{-6,5.10^{-5} - 2x_1 - 68x_2 + 2x_3 - 68x_4 + 4.10^{-5}x_5 + 3.10^{-4}x_6 + 136x_2x_4 < -40}$$

IS^{frac} of f

Algorithm parameters :

$k_i = 1+1/(2i)$

$p_i = 50/i$



References

[1] J.-J. Droesbeke, J. Fine et G. Saporta - *Plans d'expériences. Applications à l'entreprise*, Technip, 2002
[2] B. Efron et C. Stein - *The jackknife estimate of variance*, The Annals of Statistics 9 (1981), no. 3, p. 586–596
[3] E. Hlawka - *Funktionen von beschränkter variation in der theorie der gleichverteilung*, Annali di Matematica Pura ed Applicata 54 (1961), p. 325–333.
[4] W. J. Morokoff et R. E. Caflisch - *Quasi-random sequences and their discrepancies*, SIAM Journal on Scientific Computing 15 (1994), no. 6, p. 1251–1279.

Conclusions

- A new way to estimate the Sobol indices
- A new criterion to quantify the adequacy of the sequence to the function of interest
- A new algorithm to select the points adapted to the function of interest