Design of Experiments and Sensitivity analysis

Course and practical application

ExModelo Summer School

OpenMOLE

June 24, 2019

- Interactive model exploration by hand and the need for preliminary experiments
- The Design of Experiments (DOE) as the definition of computational experiments to extract information from the simulation model
- Example: NetLogo behavior space: basic grid DOE
- Sensitivity analysis as an advanced DOE

Remark 1: *terminology strongly depends on disciplines and practices*

Remark 2: *most are generally **preliminary experiments** to prepare more elaborated, question-related, experiments*

Outline

1. Basic experiments
2. High-dimensional samplings
3. Sensitivity analysis
4. Application in OpenMOLE
5. Practical application

Basic experiments

Provide explicitly sampling points on which the model (or its replication task) will be run: notion of ***direct sampling*** *in OpenMOLE (corresponds to DOE in the literature)*

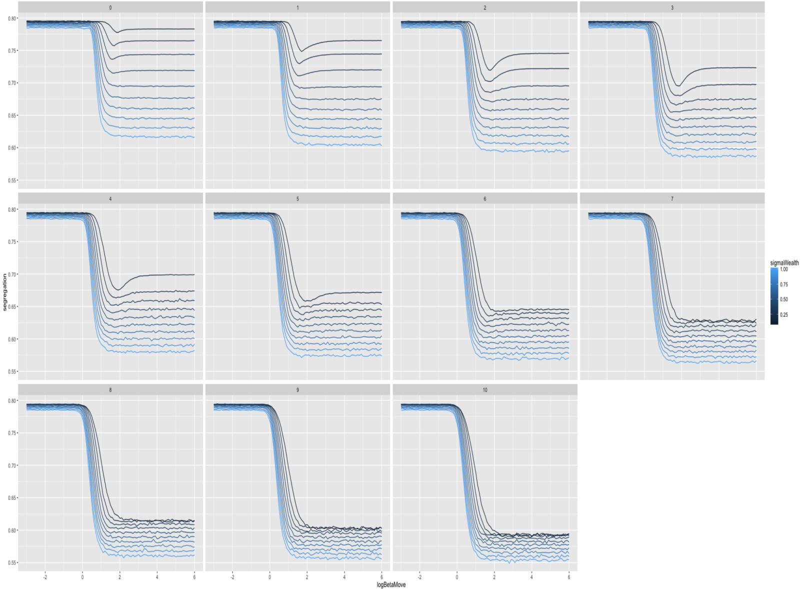
- full samplings
- elaborated sampling for high dimensions given a low computational budget (**the curse of dimensionality**)

One factor at a time

Cheapest and intuitive DOE: *all factors have nominal values and a discrete variation set, in which each is varied while others remaining fixed*

- when model is slow - or computational budget highly limited
- does not capture interaction between parameters, and highly dependent on nominal values
- seen as a bad practice **BUT** useful for models taking significant time, and prone to thematic interpretation

Example where One-At-a-Time fails



Grid sampling

Brute force DOE: *ensemble product of discrete variation ranges for factors (usually a regular grid but not necessarily)*

- quickly limited by the curse of dimensionality - in practice still powerful with a quick model and a low number of parameters
- naive approach, i.e. only DOE for many "simulation-newcomers" such as economics or some parts of physics

High-dimensional samplings

Computational limitations $\Longrightarrow$ *need specific methods to efficiently sample the parameter space*

Different methods for improving sampling in numerical experiments given limited computational resources have been proposed, as for example:

- Sobol sequences (quicker convergence of integral estimation)
- Latin Hypercube Sampling
- Orthogonal sampling

Minimizing discrepancy for a point cloud: intuitively being spread evenly across the definition space

L2-discrepancy given for normalized data points $\mathbf{X} = (x_{ij}) \in [0,1]^d$ by

$$\left\| \mathbf{t} \in [0,1]^d \mapsto \frac{1}{n} \sum_i \mathbb{1}_{\prod_j x_{ij} < t_j} - \prod_j t_j \right\|_2$$

Latin Hypercube Sampling

X				
	X			
				X
			X	
		X		

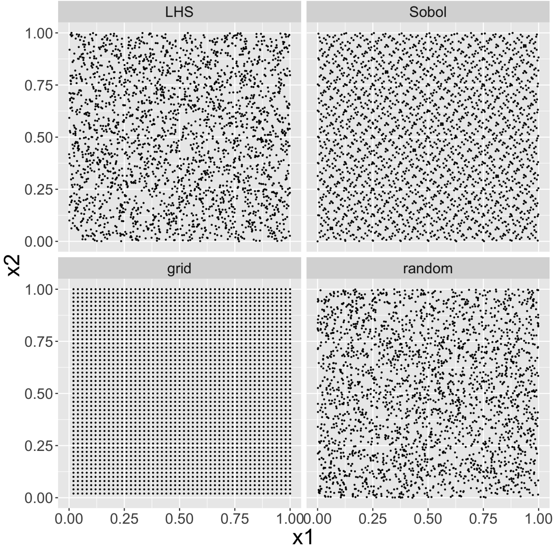
Latin cube: one point in each row and column; hypercube generalization in any dimension

Sobol sequence

Sobol sequences are a case of quasi-random sequences with low discrepancy (also Halton sequences e.g.)

- Estimate integral in $1/N$ instead of $1/\sqrt{N}$ with random sampling
- Constructed recursively (using bit representations)

For $N = 2500$ samples in 2 dimensions



Aim of sensitivity analysis methods *How to summarize model sensitivity and isolate principal factors ?*

- Most methods are *global*, i.e. provide an aggregate of factor effect on the full parameter space
- Advanced methods, still useful for preliminary experiments e.g. to discard factors from further experiments
- Examples: Morris and Saltelli methods

Idea: *Sample trajectories in the parameter space in a One-At-a-Time manner. Screening method isolating **elementary effects***

- isolate local effects of factors
- more efficient than point sampling to get individual effects
- useful as a first experiment to understand the relative influence of factors

Introduced by [Morris, 1991], improved by [Saltelli et al., 2004], [Campolor
propose to extend the method with Sobol sequences

Estimation of relative and conditional variances

$$ST_i = \frac{E_{\mathbf{X}_{\sim i}}\left[Var(Y|\mathbf{X} \sim i)\right]}{Var(Y)}$$

Syntax of the direct sampling:

```
val explo = DirectSampling(
    evaluation = model,
    sampling = ...
)
```

One-factor sampling:

```
sampling = OneFactorSampling(
    (x1 in (0.0 to 1.0 by 0.2)) nominal 0.5,
    (x2 in (0.0 to 1.0 by 0.2)) nominal 0.5
)
```

Example of samplings

- Grid sampling

```
sampling =
     (x1 in (0.0 to 1.0 by 0.5)) x
     (x2 in (0.0 to 1.0 by 0.5))
```

- LHS Sampling

```
sampling = LHS(
    100,
    x1 in (0.0,1.0),
    x2 in (0.0,1.0)
)
```

- Sobol sampling

```
sampling = SobolSampling(
     100,      x1 in (0.0,1.0),      x2 in
(0.0,1.0) )
```

Saltelli is a method in itself

```
val sen = SensitivitySaltelli(
     evaluation = (model on env by 1000),
     samples = 100000,
     inputs = Seq(
          humanFollowProbability in (0.0,1.0),
          humanInformedRatio in (0.0,1.0),
          humanInformProbability in (0.0,1.0)),
     outputs = Seq(totalZombified,halfZombified),
)
```

Morris is also a method

```
 val morrisHook =
val sen = SensitivityMorris(
     evaluation = model on env hook morrisHook,
      inputs = Seq(
           humanFollowProbability in (0.0,1.0),
           humanInformedRatio in (0.0,1.0),
           humanInformProbability in (0.0,1.0)),
      outputs = Seq(totalZombified,halfZombified),
      repetitions = 100,
      levels = 5
)
```

Your turn to run some direct samplings and/or sensitivity analysis

- given the described zombie model, what first experiment beyond stochasticity would be relevant ?
- write a script
- explore results (using e.g. the OpenMOLE GUI plots)

References I

Campolongo, F., Saltelli, A., and Cariboni, J. (2011).
From screening to quantitative sensitivity analysis. a unified approach.
Computer Physics Communications, 182(4):978–988.

Morris, M. D. (1991).
Factorial sampling plans for preliminary computational experiments.
Technometrics, 33(2):161–174.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004).
Sensitivity analysis in practice: a guide to assessing scientific models.
Chichester, England.