Sensitivity Analysis - ModIA - 2021
ANOVA decomposition - Exercises

O. Roustant

In the following, we assume that $X_1, \dots, X_d$ are *independent* random variables with probability measures $\nu_1, \dots, \nu_d$. We denote: $X = (X_1, \dots, X_d)$, $\nu = \nu_1 \otimes \dots \otimes \nu_d$ the probability measure of X and $\Delta = \Delta_1 \times \dots \times \Delta_d$, the integration domain.

1 ANOVA decomposition in dimension 2

Here $d = 2$. Let f be in $L^2(\nu)$. We want to prove that there exists a unique decomposition

$$f(X_1, X_2) = f_0 + f_1(X_1) + f_2(X_2) + f_{1,2}(X_1, X_2)$$

such that $\mathbb{E}(f_I(X_I)|X_J) = 0$ for all $J \subsetneq I$, i.e. $\mathbb{E}(f_i(X_i)) = \mathbb{E}(f_{1,2}(X_1, X_2)|X_i) = 0$ for $i = 1, 2$.

1. Prove that necessarily, we must have:
   - $f_0 = \mathbb{E}(f(X))$
   - $f_i(X_i) = \mathbb{E}(f(X)|X_i) - f_0$ (for $i = 1, 2$)
   - $f_{1,2}(X_1, X_2) = \mathbb{E}(f(X)|X_1, X_2) - f_1(X_1) - f_2(X_2) - f_0$

   Conversely, check that these terms give the ANOVA decomposition.
2. Prove that the recursion formula for $f_{1,2}$ can be rewritten as a sum of conditional expectations with alternate signs:
   $$f_{1,2}(X_1, X_2) = \mathbb{E}(f(X)|X_1, X_2) - \mathbb{E}(f(X)|X_1) - \mathbb{E}(f(X)|X_2) + \mathbb{E}(f(X))$$
3. Prove that all the terms are orthogonal: $\mathbb{E}[f_I(X_I)f_{I'}(X_{I'})] = 0$ if $I \neq I'$.
4. Consider $f(X_1, X_2) = X_1$, and let ν_1 be such that X_1 is centered. Observe that we have the two possible decompositions:
   $$f(X_1, X_2) = 0 + X_1 + 0 + 0 = 0 + 0 + 0 + X_1$$
   What's wrong? What is the intuition of the condition "$\mathbb{E}(f_I(X_I)|X_J) = 0$ for all $J \subsetneq I$"?

2 Additive functions - 1st order polynomials & SRCs

Consider an additive function:

$$f(x) = \beta_0 + g_1(x_1) + \dots + g_d(x_d)$$

where the $g_i(X_i)$'s are *centered* (with respect to the measure ν_i) and square-integrable.

1. What should be the ANOVA decomposition of f? Prove it and compute all Sobol indices.
2. Deduce from 1 the ANOVA decomposition of a first order polynomial:
   $$f(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d$$
   and all Sobol indices. The results can be expressed in function of $m_i^{(1)} = E(X_i)$.

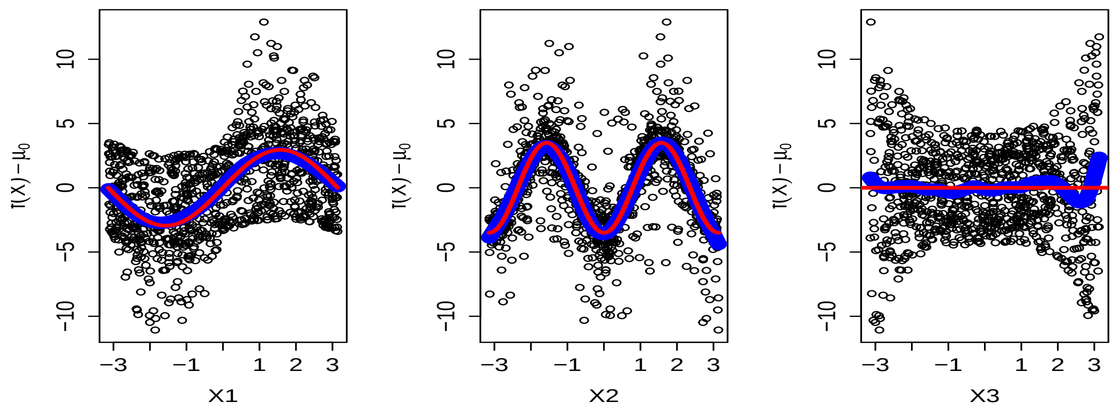

Figure 1: Main effects of Ishigami function: Theoretical (straight line) and estimated (bold line)

3 A 3-dimensional test-function

The Ishigami function is defined over $\Delta := [-\pi; \pi]^3$ by:

$$f(x) = \sin(x_1) + A\sin^2(x_2) + Bx_3^4\sin(x_1)$$

with A = 7 and B = 0.1.

1. Compute the ANOVA decomposition and Sobol indices when μ is uniform over Δ (use $a = 1/2, b = \pi^4/5$.)
2. (Computer lab) Estimate the main effects by using simulations of the inputs. Plot them on the same figure and compare with the theoretical ones. Same question with the 2-dimensional projection over (x_1, x_3) and the second order interaction $f_{1,3}(x_1, x_3)$.

4 Polynomial chaos

Polynomial chaos are defined as a tensor basis of orthonormal polynomials. It is very famous in sensitivity analysis since, once the function of interest has been decomposed on that basis, the Sobol indices are directly obtained as sums of squared coefficients. Now let us go into details.

Let f be in $L^2(\nu)$ with $\nu = \otimes_{i=1}^d \nu_i$. For each probability distribution ν_i $(i = 1, \dots, d)$, denote:

$$P_{i,0}(x_i) = 1, \quad P_{i,1}(x_i), \quad \dots \quad P_{i,\ell}(x_i), \quad \dots$$

a set of orthonormal polynomials in $L^2(\nu_i)$ (of degree $0, 1, 2, \dots, \ell, \dots$). Then the *polynomial chaos* indexed by the multi-index $\underline{\ell} = (\ell_1, \dots, \ell_d)$ is the tensor:

$$P_{\underline{\ell}}(x) = \prod_{i=1}^{d} P_{i,\ell_i}(x_i).$$

We denote by $\mathcal{I} = \mathbb{N}^d$ the set of all multi-indices.

1. For two multi-indices $\underline{\ell}, \underline{\ell'}$, compute $\mathbb{E}(P_{\underline{\ell}}(X)P_{\underline{\ell'}}(X))$. Deduce that polynomial chaos are orthonormal in $L^2(\nu)$. We admit that they form a Hilbert basis of $L^2(\nu)$.

2. Deduce that
$$f(x) = \sum_{\underline{\ell} \in \mathcal{I}} c_{\underline{\ell}} P_{\underline{\ell}}(x)$$
with $c_{\underline{\ell}} = \langle f, P_{\underline{\ell}} \rangle$. Express the total variance D in function of the $c_{\underline{\ell}}$.

3. Compute $\mathbb{E}(P_{\underline{\ell}}(X)|X_1)$.

4. Deduce that the first main effect of f is obtained by choosing the tensors that involve *only* X_1, defined by the subset $\mathcal{I}_1 = \{\underline{\ell} \in \mathcal{I} \text{ s.t. } \ell_1 \geq 1, \ell_2 = \ldots = \ell_d = 0\}$. Compute the corresponding Sobol index S_1 in function of the $c_{\underline{\ell}}$.

5. Similarly, compute $\mathbb{E}(P_{\underline{\ell}}(X)|X_{-1}) = \mathbb{E}(P_{\underline{\ell}}(X)|X_2, \ldots, X_d)$.
Show that the first total effect of f is obtained by choosing the tensors that involve *at least* X_1, defined by $\mathcal{I}_1^{\text{tot}} = \{\underline{\ell} \in \mathcal{I} \text{ s.t. } \ell_1 \geq 1\}$. Give the expression of the total Sobol index S_1^{tot}.

5 Computer lab

The aim of the lab is to perform a global SA of a complex function by using a Kriging metamodel.

5.1 On a toy highly non-linear function

We consider the Ishigami function of Exercise 3, with the uniform distribution on $[-\pi, \pi]^3$.

1. In this question, we have an unlimited budget, say $n = 1000$. Estimate the main effects by simulation. Plot them on the same figure and compare with the theoretical ones. Estimate the Sobol indices of Ishigami, with the R package sensitivity (function `fast99`).

2. Now we assume that the budget is limited to $n = 90$ function evaluations.

   (a) Check that the direct approach of question 1 is not working anymore.

   (b) Build a maximin Latin hypercube design $\mathbf{X}$ with DiceDesign (function `maximinESE_LHS`). Visualize the space-filling properties. Compute the response $\mathbf{y}$ at $\mathbf{X}$, and construct a Kriging model with DiceKriging. Look at the validation results.

   (c) Perform a sensitivity analysis on the Kriging metamodel. For that goal, we provide the following wrapper (where m is the name of the metamodel object):
```
kriging.mean <- function(Xnew, m) {
  predict(m, Xnew, "UK", se.compute = FALSE, checkNames = FALSE)$mean
}
```

   (d) Comment the results obtained with the metamodel, built with a small budget, to the results of question 1.

5.2 Performance of a machine learning algorithm

We consider the k-fold cross validation error of the SVM algorithm considered in the previous computer lab. It depends on three parameters: the kernel parameter gamma, and two SVM parameters: cost and epsilon. Which parameter is the most influential on the performance? Can you hierarchize their influence? Visualize the main effects.

Bonus. The Ishigami function has a block-additive structure. Show, by running the example of `kmAdditive` from the R package fanovaGraph, that the metamodel is much more accurate with a kernel mimicking that structure, of the form $k(x, x') = k_2(x_2, x_2') + k_{1,3}((x_1, x_3), (x_1', x_3'))$. Adapt this idea to the case of the k-fold cross validation error.