Gaussian process regression for Sensitivity analysis

GPSS Workshop on UQ, Sheffield, September 2016

Nicolas Durrande, Mines St-Étienne, durrande@emse.fr

We assume we are interested in a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with $d \geq 2$.

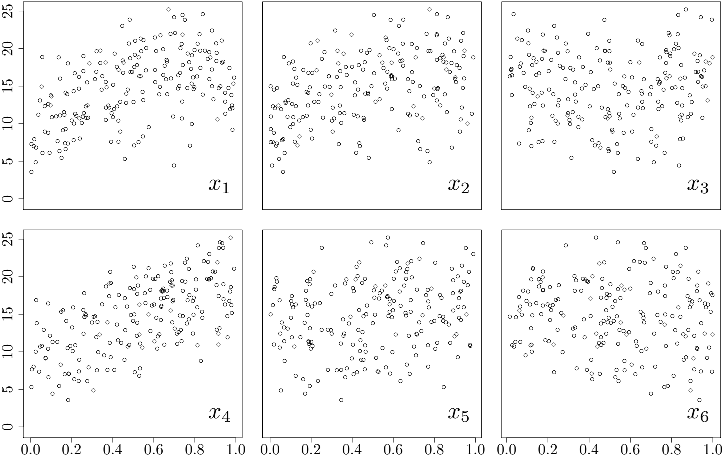
We want to get some understanding on the “structure” of f :

- What is the effect of each input variables on the output ?
- Do some variables have more influence than other ?
- Do some variables interact together ?

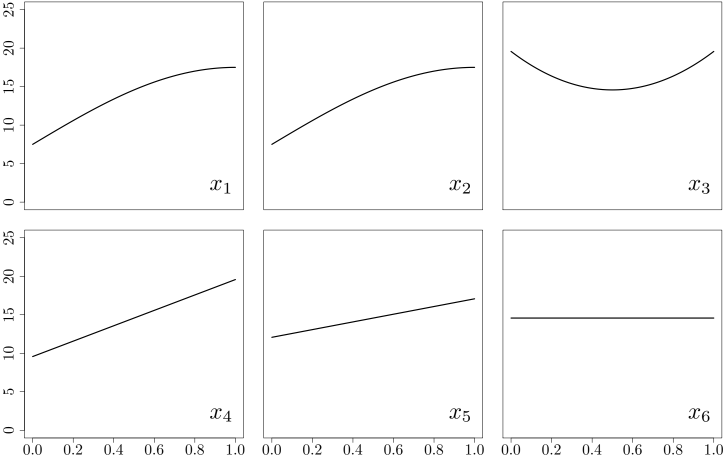
The talk will be illustrated on the following test function :

$$\begin{aligned} f : [0,\ 1]^6 &\rightarrow \mathbb{R} \\ x &\mapsto 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10 x_4 + 5 x_5 \end{aligned}$$

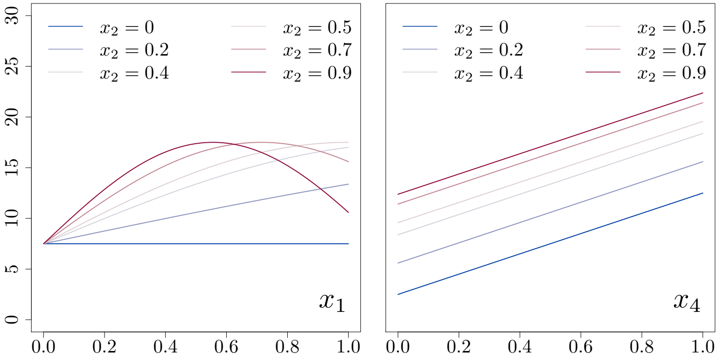
First thing one can do is to plot the output versus each input.
For 100 samples uniformly distributed over $[0,1]^6$ we get :



In a similar fashion, we can fix all variables except one. In graph bellow, all non plotted variables are set to 0.5.



In order to get an insight on the interaction between variables, we can look at the influence of changing the reference value.



One common tool for analysing the structure of f is to look at its FANOVA representation :

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i<j} f_{i,j}(x_i, x_j) + \dots + f_{1,\dots,d}(\mathbf{x})$$

This decomposition is such that :

- f_0 accounts for the constant term
  - $\Rightarrow$ all f_I are zero mean ($I \neq 0$)
- f_1 accounts for all signal that can be explained just by x_1
  - $\Rightarrow \displaystyle\int f_I(x) dx_{-1} = 0$ for all $I \notin \{0, 1\}$
  - $\Rightarrow \displaystyle\int f_I(x) dx_1 = 0$ for all $I \supset 1$

In other words, this decomposition is such that all terms are orthogonal in L^2.

The expressions of the f_I are :

$$f_0 = \int f(\mathbf{x}) \mathrm{d}\mathbf{x}$$

$$f_i(x_i) = \int f(\mathbf{x}) \mathrm{d}\mathbf{x}_{-i} - f_0$$

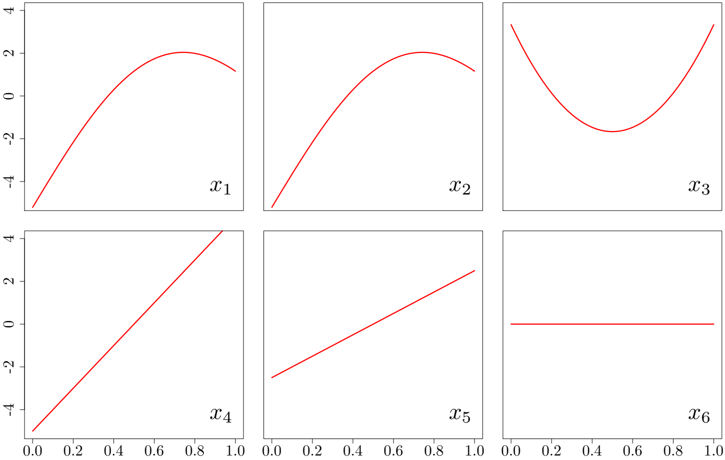
$$f_{i,j}(x_i, x_j) = \int f(\mathbf{x}) \mathrm{d}\mathbf{x}_{-ij} - f_i(x_i) - f_j(x_j) + f_0$$

It can also be interesting to look at the total effect of some inputs :

$$\tilde{f}_1(x_1) = \int f(\mathbf{x}) \mathrm{d}\mathbf{x}_{-1}$$

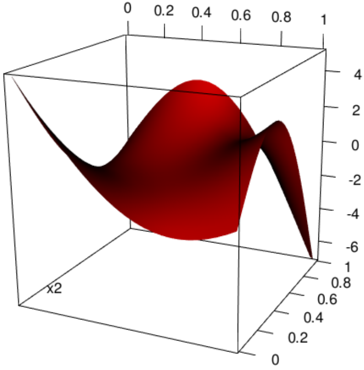
$$\tilde{f}_{1,2}(x_1, x_2) = \int f(\mathbf{x}) \mathrm{d}\mathbf{x}_{-\{1,2\}}$$

On the previous example we obtain :



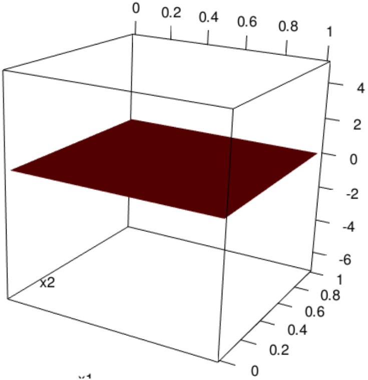
We can also look at 2nd order interactions

$f_{1,2}(x_1, x_2)$



Interaction x_1, x_2

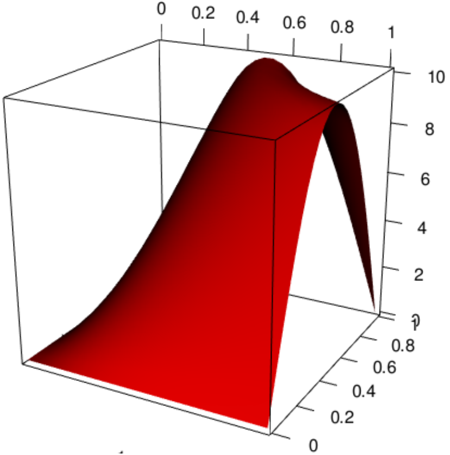
$f_{1,3}(x_1, x_3)$



Interaction x_1, x_3

The total effect of (x_1, x_2) is thus

$$\tilde{f}_{1,2}(x_1, x_2) = f_0 + f_1(x_1) + f_2(x_2) + f_{1,2}(x_1, x_2)$$



In practical application f is not analytical so the above method require numerical computations of the integrals. If there is a cost associated with the evaluation of f, surrogate models are useful.

Some models are naturally easy to interpret, for example

$$m(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

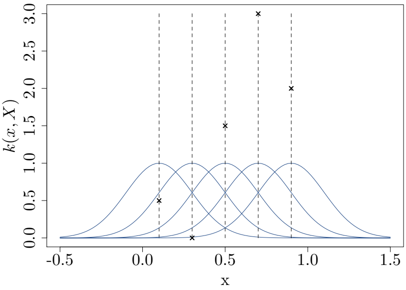
but it soon becomes more tricky.

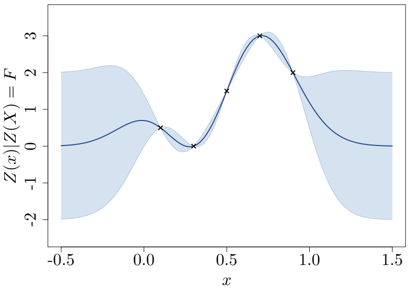
$$m(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2$$

In GPR, the mean can be seen either as a linear combination of

- the observations : $m(x) = \alpha^t F$
- the kernel evaluated at X : $m(x) = k(x, X)\beta$

For example, we have for a squared exponential kernel





The basis function have a local influence which makes the interpretation difficult.

The principle of polynomial chaos is to project f onto a basis of orthonormal polynomials.

One dimension

For $x \in \mathbb{R}$, the h_i are of order i. Starting from the constant function $h_0 = 1$, the following ones can be obtain using Gram-Schmidt orthogonalisation.

$$h_0(x) = \frac{1}{||1||},\ h_1(x) = \frac{x - \langle x, h_0 \rangle h_0}{||x - \langle x, h_0 \rangle h_0||},\ h_2(x) = \frac{x^2 - \langle x, h_0 \rangle h_0 - \langle x, h_1 \rangle h_1}{||x^2 - \langle x, h_0 \rangle h_0 - \langle x, h_1 \rangle h_1||}$$

d-dimension

In $\mathbb{R}^d$, the basis is obtained by a tensor product of one dimensional basis. For example, if $d = 2$:

$$\begin{aligned} h_{00}(x) &= 1 \times 1 \\ h_{10}(x) &= h_1(x_1) \times 1 \\ h_{01}(x) &= 1 \times h_1(x_2) \end{aligned} \qquad \begin{aligned} h_{11}(x) &= h_1(x_1) \times h_1(x_2) \\ h_{20}(x) &= h_2(x_1) \times 1 \\ \vdots &= \vdots \end{aligned}$$

The orthonormal basis H depends on the measure over the input space D.

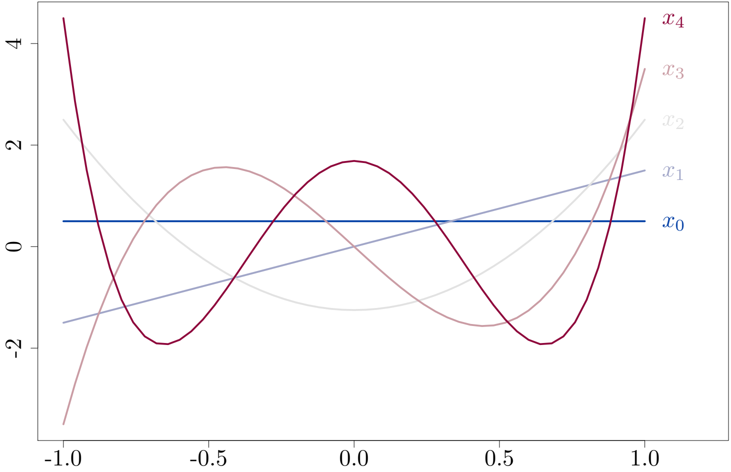
A uniform measure over $D = [-1, 1]$ gives the **Legendre basis** :

$$
\begin{aligned}
h_0(x) &= 1/2 & h_3(x) &= 7/4\ (5x^3 - 3x) \\
h_1(x) &= 3/2\ x & h_4(x) &= 9/16\ (35x^4 - 30x^2 + 3) \\
h_2(x) &= 5/4\ (3x^2 - 1) & \vdots &= \vdots
\end{aligned}
$$

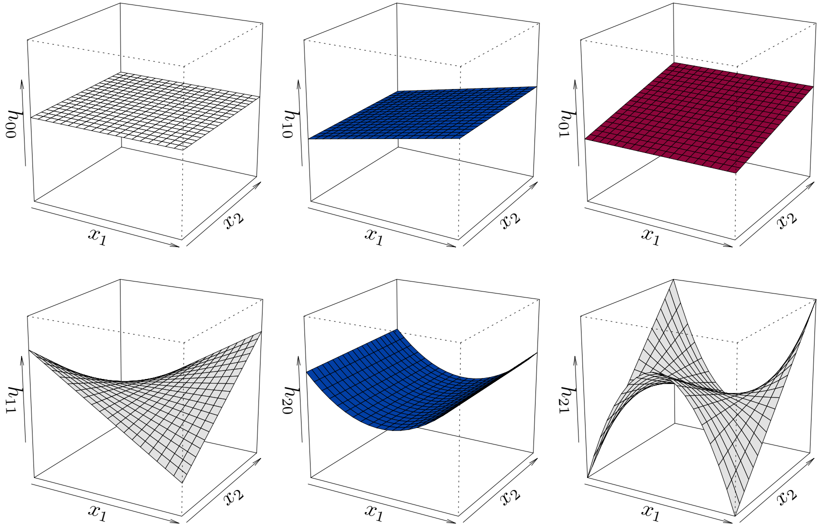
A standard Gaussian measure over $\mathbb{R}$ gives the **Hermite basis** :

$$
\begin{aligned}
h_0(x) &= 1/\sqrt{2\pi} & h_3(x) &= 1/(6\sqrt{2\pi})\ (x^3 - 3x) \\
h_1(x) &= 1/\sqrt{2\pi}\ x & h_4(x) &= 1/(24\sqrt{2\pi})\ (x^4 - 6x^2 + 3) \\
h_2(x) &= 1/(2\sqrt{2\pi})\ (x^2 - 1) & \vdots &= \vdots
\end{aligned}
$$

Legendre basis in 1D



Legendre basis in 2D



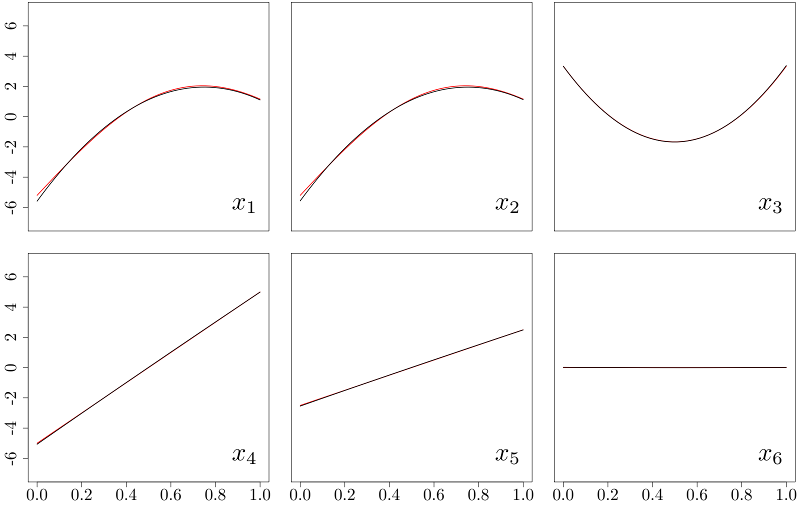
If we consider linear regression model based on polynomial chaos basis functions

$$m(x) = \sum_{I \subset \{0,\dots p\}} \beta h_I(x)$$

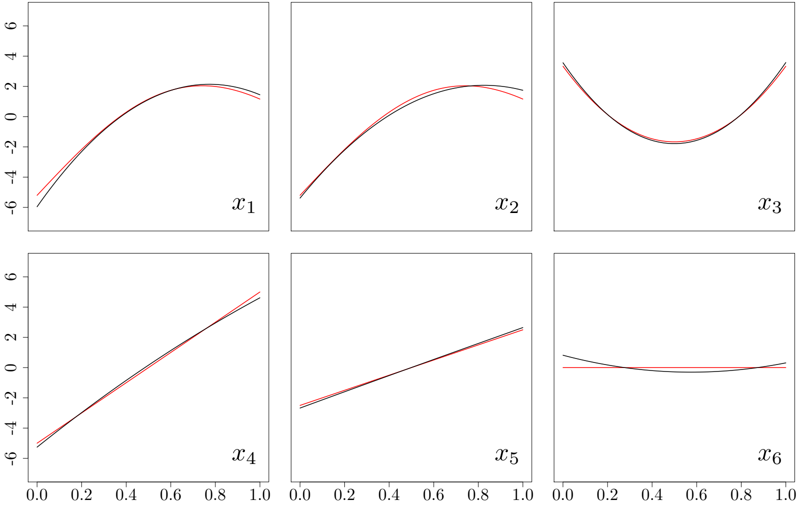
the FANOVA representation of m is straightforward. For example in 2D :

$$\begin{aligned}
m_0 &= \iint m(x) dx_1 dx_2 = \int \beta_{00} h_{00}(x) dx_1 dx_2 = \beta_{00} \\
m_1(x_1) &= \int m(x) dx_2 - m_0 \\
&= \int \beta_{00} h_{00}(x) + \beta_{10} h_{10}(x) + \beta_{20} h_{20}(x) dx_2 - m_0 \\
&= \beta_{10} h_{10}(x_1) + \beta_{20} h_{20}(x_1) \\
m_{1,2}(x) &= \dots = \beta_{11} h_{11}(x) + \beta_{12} h_{12}(x) + \beta_{21} h_{21}(x) + \beta_{22} h_{22}(x)
\end{aligned}$$

We obtain on the motivating example :



Same figure without cheating :



A first idea is to consider ANOVA kernels [Stitson 97] :

$$k(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d} (1 + k_i(x_i, y_i))$$
$$= 1 + \underbrace{\sum_{i=1}^{d} k_i(x_i, y_i)}_{additive\ \ part} + \underbrace{\sum_{i<j} k_i(x_i, y_i) k_j(x_j, y_j)}_{2^{nd}\ \ order\ \ interactions} + \dots + \underbrace{\prod_{i=1}^{d} k_i(x_i, y_i)}_{full\ \ interaction}$$

The associated GP is

$$Z(\mathbf{x}) = \underbrace{Z_0}_{cst} + \underbrace{\sum_{i=1}^{d} Z_i(x_i)}_{additive\ \ part} + \underbrace{\sum_{i<j} Z_{i,j}(x_i, x_j)}_{2^{nd}\ \ order\ \ interactions} + \dots + \underbrace{Z_{1 \dots d}(\mathbf{x})}_{full\ \ interaction}$$

However, the Z_I do not satisfy $\int Z_I(x) \mathrm{d}x_i = 0$.

If we build a GPR model based on this kernel, we obtain :

$$m(x) = k(x, X)k(X, X)^{-1}F$$

$$\begin{aligned} m(x) &= \left(1 + \sum_{i=1}^{d} k_i(x_i, y_i) + \sum_{i<j} k_i(x_i, y_i)k_j(x_j, y_j)\right) k(X, X)^{-1}F \\ &= 1^t k(X, X)^{-1}F + \underbrace{\sum_{i=1}^{d} k(x_i, X_i)k(X, X)^{-1}F}_{m_i(x_i)} \\ &\quad + \underbrace{\sum_{i<j} k_i(x_i, X_i)k(x_j, X_j)k(X, X)^{-1}F}_{m_{i,j}(x_i,x_j)} \\ &\quad + \dots \end{aligned}$$

As previously, the m_I do not satisfy $\int m_I(x)\mathrm{d}x_i = 0$.

samples with zero integrals

We are interested in building a GP such that the integral of the samples are exactly zero...

Let's consider the associated conditional GP :

$$Z_0 \overset{law}{=} Z | \int Z(s)ds{=}0$$

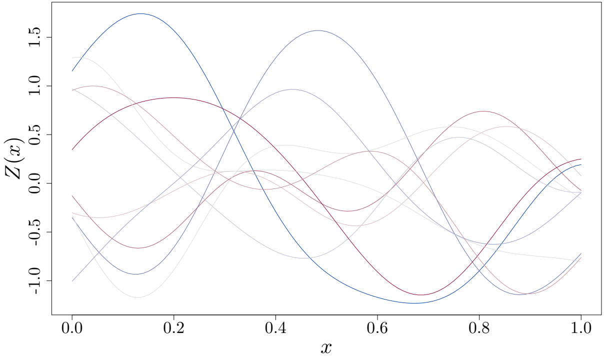
Let $\mu_0(x) = \mathrm{E}\left(Z(x) | \int Z(s)ds{=}0\right)$ denote the conditional expectation and $k_0(x,x') = \mathrm{cov}\left(Z(x), Z(x') | \int Z(s)ds{=}0\right)$

$$\mu_0(x) = \int k(x,s)ds \left(\iint k(s,t)dsdt\right)^{-1} 0$$

$$k_0(x,y) = k(x,x') - \int k(x,s)ds \left(\iint k(s,t)dsdt\right)^{-1} \int k(x,s)ds$$

Samples from Z_0 have the required property

$$\mu_0(x) = 0 \qquad k_0(x,y) = k(x,y) - \frac{\int k(x,s)\mathrm{d}s \int k(y,s)\mathrm{d}s}{\iint k(s,t)\mathrm{d}s\mathrm{d}t}$$



These 1-dimensional kernels are of great importance to create ANOVA kernels dedicated to sensitivity analysis :

$$k_{SA}(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{d}(1 + k_0(x_i, y_i))$$

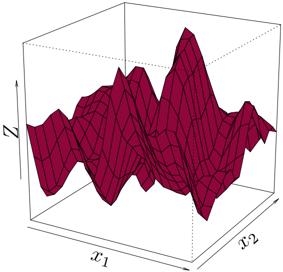
$$= 1 + \underbrace{\sum_{i=1}^{d} k(x_i, y_i)}_{additive\ \ part} + \underbrace{\sum_{i<j} k(x_i, y_i)k(x_j, y_j)}_{2^{nd}\ \ order\ \ interactions} + \cdots + \underbrace{\prod_{i=1}^{d} k(x_i, y_i)}_{full\ \ interaction}$$

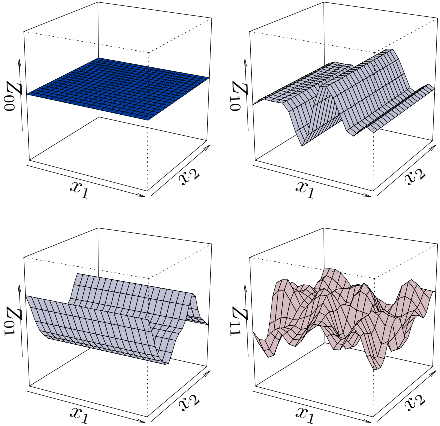
The associated GP naturally writes

$$Z_{SA}(\mathbf{x}) = \underbrace{Z_0}_{cst} + \underbrace{\sum_{i=1}^{d} Z_i(x_i)}_{additive\ \ part} + \underbrace{\sum_{i<j} Z_{i,j}(x_i, x_j)}_{2^{nd}\ \ order\ \ interactions} + \cdots + \underbrace{Z_{1\dots d}(\mathbf{x})}_{full\ \ interaction}$$

Now, the Z_I **do** satisfy $\int Z_I(x)\mathrm{d}x_i = 0$.

We get the following decomposition of samples





Furthermore, the GPR model inherits this properties

2D example

$$\begin{aligned} k(\mathbf{x}, \mathbf{y}) &= \prod_{i=1}^{2} (1 + k_0(x_i, y_i)) \\ &= 1 + k_0(x_1, y_1) + k_0(x_2, y_2) + k_0(x_1, y_1) k_0(x_2, y_2) \end{aligned}$$

The mean writes

$$\begin{aligned} m(\mathbf{x}) &= (1 + k_0(x_1, X_1) + k_0(x_2, X_2) + k_0(x_1, X_1) k_0(x_2, X_2))^t k(X, X)^{-1} F \\ &= m_0 + m_1(x_1) + m_2(x_2) + m_{12}(\mathbf{x}) \end{aligned}$$

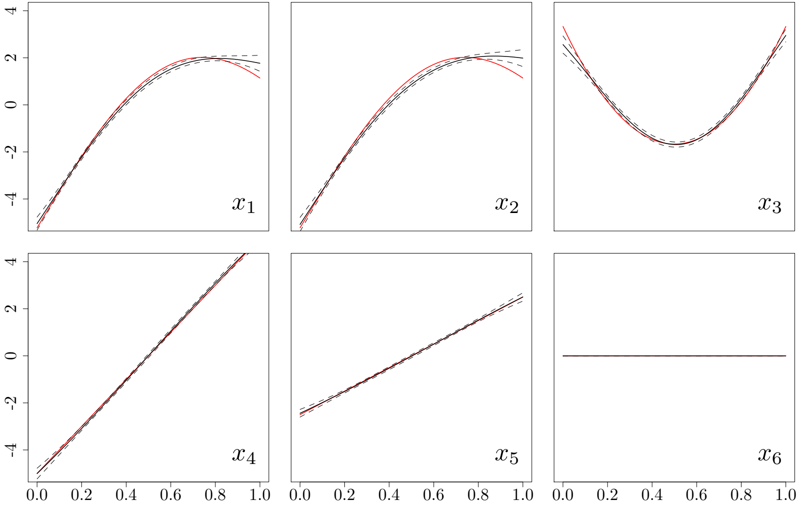
These terms correspond to the FANOVA representation of m.

The sub-models are conditional expectations :

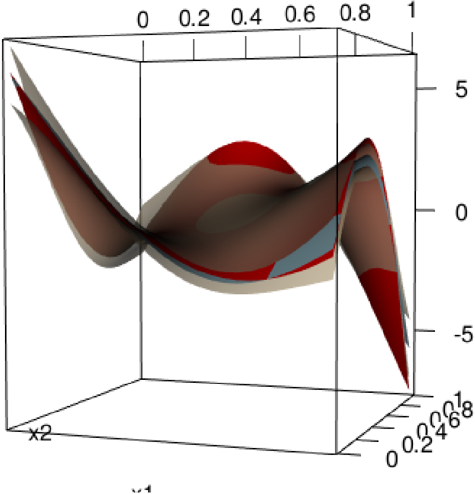
$$m_I(x) = \mathrm{E}\left(Z_I(x) | Z(X){=}F\right)$$

We can thus associate a predictive covariance to each sub-model !

We obtain on the motivating example :



We obtain on the motivating example :



The principle of sensitivity analysis is to quantify how much each input or group of inputs has an influence on the output :

- local sensitivity analysis
- global sensitivity analysis

Probabilistic framework has proven to be very interesting :

If we introduce randomness in the inputs,
how random is the output ?

Hereafter, we focus on variance based global sensitivity analysis

Let X be the random vector representing our uncertainty on the inputs. We assume its probability distribution factorises (i.e. the X_i are independents) :

$$\mu(x) = \mu(x_1) \times \mu(x_2) \times \cdots \times \mu(x_n)$$

This factorization of μ allows to use it in the FANOVA

representation of f :

$$f(x) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i<j} f_{i,j}(x_i, x_j) + \cdots + f_{1,\dots,d}(x)$$

in this expression, the f_I are orthogonal for μ.

We know plug X into this expression :

$$f(X) = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{i<j} f_{i,j}(X_i, X_j) + \cdots + f_{1,\dots,d}(X)$$

and we get interesting results...

The $f_I(X_I)$ are centred and independent

$$\mathrm{E}\left(f_I(X_I)\right) = \int f_I(x_I) d\mu(x) = 0 \qquad \text{(for } I \neq 0\text{)}$$

$$\mathrm{cov}\left(f_I(X_I), f_J(X_J)\right) = \mathrm{E}\left(f_I(X_I) f_J(X_J)\right) = \int f_I(x_I) f_J(x_J) d\mu(x) = 0$$

As a consequence, we get

$$\mathrm{var}\left(f(X)\right) = \sum_{i=1}^{d} \mathrm{var}\left(f_i(X_i)\right) + \sum_{i<j} \mathrm{var}\left(f_{i,j}(X_i, X_j)\right) + \dots + \mathrm{var}\left(f_{1,\dots,d}(X)\right)$$

The Sobol indices are defined as

$$S_I = \frac{\mathrm{var}\left(f_I(X_I)\right)}{\mathrm{var}\left(f(X)\right)}$$

These indices are in $[0, 1]$, and their sum is 1.

In practice, these indices can be computed using Monte Carlo methods.

$\Rightarrow$ This requires lots of observations

Another approach is to use surrogate models. If the model is well chosen, computational cost is almost free !

- Polynomial Chaos
- GPR with k_{sa} kernels

Polynomial Chaos

In the case of polynomial Chaos, Sobol indices are given by the squares of the β coefficients

$$D_i = Var_X[E_X(H(X)\beta|X_i)] = Var_X[H_i(X_i)\beta_i] = \beta_i^2$$
$$S_i = \frac{D_i}{\sum_k D_k}$$

For mode details, see the work from Bruno Sudret

GPR with k_{sa} kernel

The sensitivity indices can be obtained analytically :

$$S_I = \frac{\mathrm{var}\left(m_I(X_I)\right)}{\mathrm{var}\left(m(X)\right)} = \frac{F^T K^{-1} \left(\bigodot_{i \in I} \Gamma_i\right) K^{-1} F}{F^T K^{-1} \left(\bigodot_{i=1}^{d} \left(1_{n \times n} + \Gamma_i\right) - 1_{n \times n}\right) K^{-1} F}$$

where Γ_i is the matrix $\Gamma_i = \int_{D_i} k_i^0(s_i) k_i^0(s_i)^T \mathrm{d}s_i$, and $\odot$ is an entry-wise product.

The computation of Sobol indices on the mean gives :

	S_1	S_2	S_3	S_4	S_5	S_6	S_{12}
model 1	0.20	0.20	0.09	0.35	0.09	0.00	0.07
model 2	0.20	0.20	0.08	0.37	0.09	0.00	0.05
truth	0.20	0.20	0.09	0.35	0.09	0.00	0.07

For this test-function 50 observations are enough !

Sensitivity analysis

- There are interesting tools to get an insight of what's happening inside high dimensional functions
- The effective dimensionality can be much smaller
- Monte Carlo or model based approach

Some modelling tips

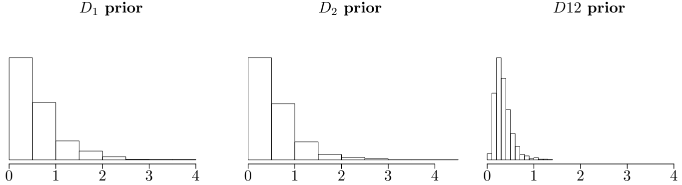
- What is the purpose of the model ?
- GPR models are not necessarily black-box...
- It is possible to include fancy observations in GPR (integrals, derivatives, ...)

This talk unfortunately focused on sensitivity analysis on the mean... the proper way is to perform SA on the conditional sample paths. See work from Marrel, Iooss et Al, SAMO 2007.

Using appropriate kernels, the computation of Sobol indices on the samples gives :

$$D_i = Var_X[E_X(Z(X)|X_i)] = Var_X[Z_i(X_i)]$$

We can easily sample from this distribution



Similarly, we can sample from the posterior to get an uncertainty measure on the indices.