A class of ANOVA kernels dedicated to sensitivity analysis

N. Durrande, D. Ginsbourger, O. Roustant

SAMO 2013 — Nice

Outline:

1. ANOVA kernels
2. HDMR, FANOVA, Hoeffding-Sobol, ...
3. Kernels of zero-mean functions
4. Application to sensitivity analysis
5. Examples

Let f be the function of interest. We assume $f \in L^2(\mu)$ with:

- $D = D_1 \times \cdots \times D_d$ with $D_i \subset \mathbb{R}$.
- μ is a separable probability measure: $\mu = \mu_1 \times \cdots \times \mu_d$

Given n observations $f(\mathbf{x}^{(i)}) = y^{(i)}$, we are interested in:

- The ANOVA representation of f
- Sobol sensitivity indices

A common approach is to approximate f with a mathematical model and to compute the sensitivity indices of m.

We focus here on Gaussian process regression:

$$m(\mathbf{x}) = \mathbf{k}(\mathbf{x})^t \mathbf{K}^{-1} \mathbf{Y}$$
$$c(\mathbf{x}, \mathbf{y}) = k(\mathbf{x}, \mathbf{y}) - \mathbf{k}(\mathbf{x})^t \mathbf{K}^{-1} \mathbf{k}(\mathbf{y})$$

The choice of the kernel has a great impact on the model... Is there a specific kernel that gives directly m's FANOVA?

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

$$K(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d} (1 + k_i(x_i, y_i))$$

With such kernels, the decomposition of m can be obtained directly:

Example

In 2D we have $K = (1 + k_1) \times (1 + k_2) = 1 + k_1 + k_2 + k_1 k_2$.
The best predictor can be written as

$$\begin{aligned} m(\mathbf{x}) &= (1 + k_1(x_1) + k_2(x_2) + k_1(x_1)k_2(x_2))^t \mathrm{K}^{-1} F \\ &= \underbrace{1^t \mathrm{K}^{-1} F}_{m_0} + \underbrace{k_1(x_1)^t \mathrm{K}^{-1} F}_{m_1(x_1)} + \underbrace{k_2(x_2)^t \mathrm{K}^{-1} F}_{m_2(x_2)} + \underbrace{k_1(x_1)k_2(x_2)^t \mathrm{K}^{-1} F}_{m_{12}(\mathbf{x})} \end{aligned}$$

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

$$K(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{d}(1 + k_i(x_i, y_i))$$

With such kernels, the decomposition of m can be obtained directly:

Example

In 2D we have $K = (1 + k_1) \times (1 + k_2) = 1 + k_1 + k_2 + k_1 k_2$.
The best predictor can be written as

$$\begin{aligned} m(\mathbf{x}) &= (1 + k_1(x_1) + k_2(x_2) + k_1(x_1)k_2(x_2))^t \mathrm{K}^{-1} F \\ &= \underbrace{1^t \mathrm{K}^{-1} F}_{m_0} + \underbrace{k_1(x_1)^t \mathrm{K}^{-1} F}_{m_1(x_1)} + \underbrace{k_2(x_2)^t \mathrm{K}^{-1} F}_{m_2(x_2)} + \underbrace{k_1(x_1)k_2(x_2)^t \mathrm{K}^{-1} F}_{m_{12}(\mathbf{x})} \end{aligned}$$

However, the m_I do not satisfy $\int_{D_i} m_I(\mathbf{x}_I)\mathrm{d}x_i = 0$

By construction $L^2(\mu) = \bigotimes_{i=1}^{d} L^2(\mu_i)$. Furthermore, $L^2(\mu_i)$ can be decomposed as

$$\begin{aligned} L^2(\mu_i) &= \mathbb{1}_{D_i} + L_0^2(\mu_i) \\ f_i(x_i) &= \int_{D_i} f_i(s)\mathrm{d}\mu(s) + \left(f_i(x_i) - \int_{D_i} f_i(s)\mathrm{d}\mu(s) \right). \end{aligned}$$

We thus obtain:

$$L^2(\mu) = \bigotimes_{i=1}^{d} \left(\mathbb{1}_{D_i} + L_0^2(\mu_i) \right) = \mathbb{1}_D + \sum_{i=1}^{d} L_0^2(\mu_i) + \cdots + \prod_{i=1}^{d} L_0^2(\mu_i)$$

FANOVA is given by the projections onto these subspaces.

The best predictor m belongs to the RKHS $\mathcal{H}$ associated to K

Definition (RKHS)

A RKHS $\mathcal{H}$ with kernel K is a Hilbert space of function such that

- $K(\mathbf{x},.) \in \mathcal{H}$ for all $\mathbf{x} \in D$
- $\langle K(\mathbf{x},.), h \rangle_{\mathcal{H}} = h(\mathbf{x})$, for all $h \in \mathcal{H}$, for all $\mathbf{x} \in D$.

Wahba suggests to consider the following structure for the RKHS:

$$\mathcal{H} = \bigotimes_{i=1}^{d} \left(\mathbb{1}_{D_i} + \mathcal{H}_i^0 \right)$$

How can we build a RKHS of zero mean function?

Let $\mathcal{H}$ be a RKHS of 1-dimensional functions with kernel k.

Theorem

If the kernel satisfies $\int_D \sqrt{k(s,s)}\mathrm{d}\mu(s) < \infty$ *then* $\mathcal{H}$ *can be written*

$$\mathcal{H} = \mathcal{H}_0 \overset{\perp}{\oplus} \mathcal{H}_1$$

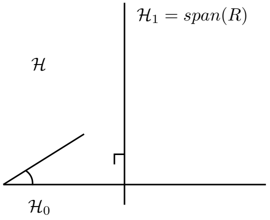
where $\mathcal{H}_0$ *is a RKHS of zero-mean functions for* μ

$\mathcal{H}_1$ *is at most 1-dimensional.*

Let R be the Riesz representer of $\int .\mathrm{d}x$:

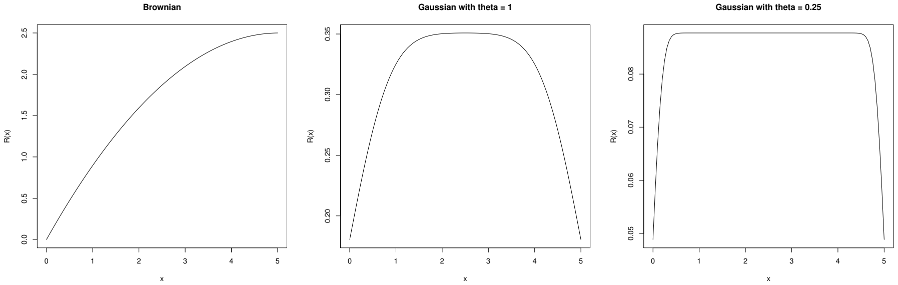
$$\int_D h(s)\mathrm{d}\mu(s) = \langle h, R \rangle_{\mathcal{H}}.$$

$\mathcal{H}_0$ corresponds to $R^{\perp}$.



The expression of $R(x)$ can be obtained easily

$$R(x) = \langle R, k(x,.) \rangle_{\mathcal{H}} = \int_D k(x,s) \mathrm{d}s$$



Finally, we have $\mathcal{H} = \mathcal{H}^1 \overset{\perp}{\oplus} \mathcal{H}^0$ with

- $\mathcal{H}^1 = span(R)$ a one dimensional RKHS
- $\mathcal{H}^0$ a RKHS of zero mean functions

The kernels of those spaces are:

$$k^1(x,y) = \frac{\langle k(x,.), R\rangle_{\mathcal{H}} R(y)}{||R||^2} = \frac{\displaystyle\int k(x,s)\mathrm{d}s \int k(y,s)\mathrm{d}s}{\displaystyle\iint k(s,t)\mathrm{d}s\mathrm{d}t}$$

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

As for the ANOVA representation in L^2, we can build a RKHS $\mathcal{H}$

$$\mathcal{H} = \bigotimes_{i=1}^{d} (\mathbb{1}_{D_i} + \mathcal{H}_i^0)$$

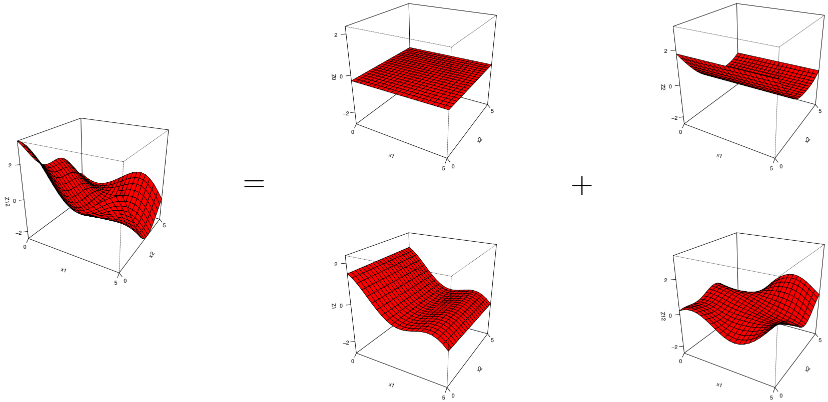
$$K(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d} (1 + k_i^0(x_i, y_i))$$

with this space, the ANOVA representation is obtained naturally

$$\begin{aligned} m(\mathbf{x}) &= (1 + k_1^0(x_1) + k_2^0(x_2) + k_1^0(x_1)k_2^0(x_2))^t K^{-1} F \\ &= m_0 + m_1(x_1) + m_2(x_2) + m_{12}(\mathbf{x}) \end{aligned}$$

here the m_I satisfy $\int_{D_i} m_I(\mathbf{x}_I) \mathrm{d}x_i = 0$.

The decomposition of the kernel gives directly a decomposition of the Gaussian process $Z(\mathbf{x}) = Z_0 + Z_1(x_1) + Z_2(x_2) + Z_{12}(\mathbf{x})$:



where the Z_I:

- satisfy the FANOVA properties
- are independent

Using K, the sensitivity indices S_I can be computed analytically:

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

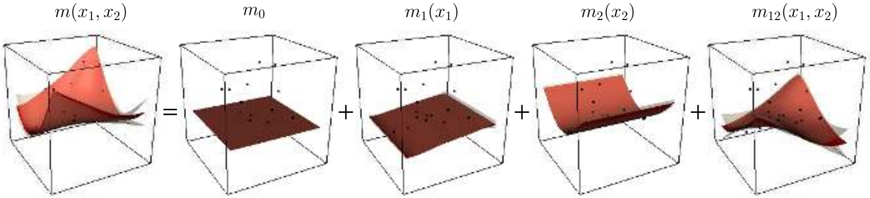
where Γ_i is the matrix $\Gamma_i = \int_{D_i} k_i^0(s_i) k_i^0(s_i)^T \mathrm{d}s_i$, $1_{n \times n}$ is the matrix of 1 and where $\odot$ is a term wise product.

Contrarily to other methods, the computation of S_I do not require to compute all S_J for $J \subset I$.

We consider a test function defined on $[-5, 5]^2$

$$f(\mathbf{x}) = x_1 + x_2^2 + x_1 x_2$$

Using the kernels described above, we obtain:



The computation of the sensitivity indices on m gives

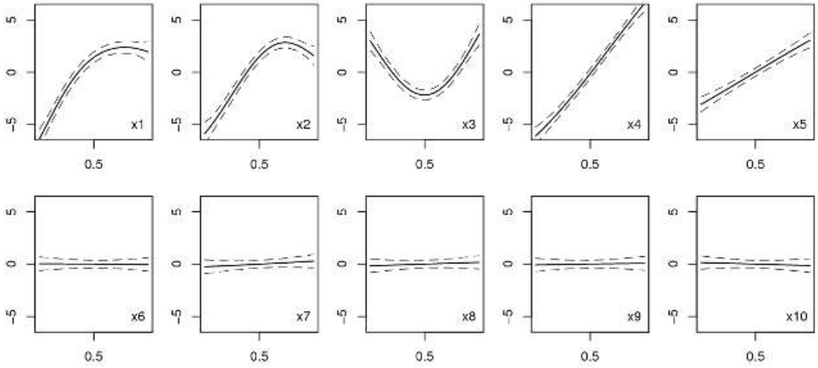
	S_1	S_2	S_{12}
model	0.23	0.48	0.29
analytical	0.25	0.5	0.25

On this example, the model gives a very good approximation.

Let us consider the random test function $f : [0,\ 1]^{10} \to \mathbb{R}$:

$$x \mapsto 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \mathcal{N}(0,1)$$

The univariate sub-models are:



Conclusion:

- kernels can be adapted to the ANOVA representation
- The Sobol sensitivity indices can be computed efficiently

Future work:

- taking the prediction variance into account
- Estimation of the kernel parameters
- RKHS orthogonal to other operators than $\int$

References:

- N. Durrande, D. Ginsbourger, O. Roustant, L. Carraro, *ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis*, Journal of Multivariate Analysis, Vol 155, 2013
- Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Publishers, 2004.
- Wahba, Wang, Gu, Klein *Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy* he Annals of Statistics, 1995