Exam Solutions: Statistical modelling and its applications

Mines Saint-Étienne – majeure Data Science –21 Décembre 2016

Exercise 1 : The Morris method

For the graph A, the location of $X2$ indicates that this variable has no influence at all ton the function: as a consequence, this must corresonds to f_4.

The graph C is associated to a function such that $X2$ has a purely linear influence but $x1$ has a more complex behaviour. Therefore, this graph is the one of f_5

The graphs B and C show a similar pattern where both variables has a non-linear influence on the output. They must thus correspond to f_1 and f_2. It can be notices that f_1 is a symmetric function and that the two points in D are on top of each others. As a consequence, these two are associated one with each other.

Exercise 2 : Sobol-Hoeffding decomposition over the disk

Let $\mathcal{D} = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$ be the unit disk in Cartesian coordinates and $X = (X_1, X_2)$ a random point, uniformly distributed over $\mathcal{D}$. We are interested in performing sensitivity analysis on the following function:

$$f(x_1, x_2) = x_1 + \sqrt{x_1^2 + x_2^2}.$$

Q1. Given X_1, X_2 is uniformly distributed on $\left[-\sqrt{1 - X_1^2}, \sqrt{1 - X_1^2}\right]$ As a consequence, the conditional mean $\mathbb{E}(X_2|X_1)$ is always null but the conditional variance $\mathrm{var}(X_2|X_1)$ decreases with $|X_1|$. This implies that X_1 and X_2 are not independent so it is not possible to perform Sobol sensitivity analysis directly on f.

Q2. Since f_p writes as a product, we have

$$\begin{aligned} f_p(\rho, \theta) &= \rho(\cos(\theta) + 1) \\ &= (\rho - m_r + m_r)(\cos(\theta) - m_a + m_a + 1) \\ &= m_r(m_a + 1) + (\rho - m_r)(m_a + 1) + m_r(\cos(\theta) - m_a) + (\rho - m_r)(\cos(\theta) - m_a) \end{aligned}$$

By construction, this decomposition satisfies the centring and non-simplification conditions.

Exercise 3 : Polynomial Chaos [4 pts]

We consider the uniform probability measure on $[-1, 1]$ and the Legendre basis which consists in orthonormal polynomials for L^2 with increasing orders:

$$h_0(x) = 1 \qquad h_2(x) = \sqrt{5}/2(3x^2 - 1)$$
$$h_1(x) = \sqrt{3}\ x \qquad \dots$$

Q1. [1 pt] The next basis function h_3 can be obtained by Gram-Schmitt orthonormalization : staring from the function $x \to x^3$, we subtract its projections onto the previous basis functions and we normalizing it. Since x^3 is odd and, h_0 and h_1 are even, we know this functions are already orthogonal. On the other hand, we have:

$$\langle x^3, \sqrt{3}x \rangle = \int_{-1}^{1} \sqrt{3}x^4 \frac{1}{2}\, dx = \frac{\sqrt{3}}{2} \left[\frac{1}{5}x^5\right]_{-1}^{1} = \frac{\sqrt{3}}{5}$$

As a consequence, $x^3 - \frac{3}{5}x$ is orthogonal to all previous basis functions. The last task left is to normalise this function:

$$\begin{aligned}\left\|x^3 - \frac{3}{5}x\right\|^2 &= \int_{-1}^{1}\left(x^3 - \frac{3}{5}x\right)^2 \frac{1}{2}\,dx = \int_{-1}^{1}\left(x^3 - \frac{3}{5}x\right)x^3\frac{1}{2}\,dx \\ &= \frac{1}{2}\left[\frac{1}{7}x^7 - \frac{3}{25}x^5\right]_{-1}^{1} = \frac{1}{7} - \frac{3}{25} = \frac{4}{175}\end{aligned}$$

Finally, the order three basis function is $h_3(x) = \frac{\sqrt{175}}{2}\left(x^3 - \frac{3}{5}x\right)$.

Q2. By construction, we have

$$\begin{aligned}\langle h_{ij}, h_{kl}\rangle &= \int_{-1}^{1}\int_{-1}^{1} h_{ij}(x)h_{kl}(x)\frac{1}{4}\,dx_1\,dx_2 \\ &= \int_{-1}^{1} h_i(x_1)h_k(x_1)\frac{1}{2}\,dx_1 \int_{-1}^{1} h_j(x_2)h_l(x_2)\frac{1}{2}\,dx_2 \\ &= \begin{cases}1 & \text{if } i = k \text{ and } j = l \\ 0 & \text{otherwise}\end{cases}\end{aligned}$$

All the h_{ij} (except h_{00}) are orthogonal to the constant function h_{00} so they satisfy the centring conditions. Furthermore $\mathrm{E}[h_{ij}(X)|X_1] = h_i(X_1)\mathrm{E}[h_j(X_2)] = 0$ for $j \neq 0$ so they also satisfy the non simplification properties.

Q3. According to the previous questions we have

$$m = \underbrace{\beta_{00}h_{00}}_{m_0} + \underbrace{\beta_{10}h_{10} + \beta_{20}h_{20}}_{m_1(x_1)} + \underbrace{\beta_{01}h_{01} + \beta_{02}h_{02}}_{m_2(x_2)} + \underbrace{\beta_{11}h_{11} + \beta_{21}h_{21} + \beta_{12}h_{12} + \beta_{22}h_{22}}_{m_{12}(x)}.$$

The orthonormality of the h_i implies:

$$\begin{aligned}\mathrm{E}[h_i(X_1)] &= 0 && \text{for } i \neq 0 \\ \mathrm{var}[h_i(X_1)] &= 1 && \text{for } i \neq 0 \\ \mathrm{var}[h_i(X_1) + h_j(X_1)] &= \mathrm{var}[h_i(X_1)] + \mathrm{var}[h_j(X_1)] && \text{for } i \neq j\end{aligned}$$

This implies that $\mathrm{var}[m(X)] = -\beta_{0,0}^2 + \sum_{i,\ j}\beta_{i,j}^2$ and similarly for the $\mathrm{var}[m_i(X)]$. The Sobol indices are then:

$$\begin{aligned}S_1 &= \frac{\mathrm{var}[m_1(X_1)]}{\mathrm{var}[m(X)]} = \frac{\beta_{1,0}^2 + \beta_{2,0}^2}{-\beta_{0,0}^2 + \sum_{i,\ j}\beta_{i,j}^2} \\ S_2 &= \frac{\mathrm{var}[m_2(X_2)]}{\mathrm{var}[m(X)]} = \frac{\beta_{0,1}^2 + \beta_{0,2}^2}{-\beta_{0,0}^2 + \sum_{i,\ j}\beta_{i,j}^2} \\ S_{1,2} &= \frac{\mathrm{var}[m_{1,2}(X)]}{\mathrm{var}[m(X)]} = \frac{\beta_{1,1}^2 + \beta_{2,1}^2 + \beta_{1,2}^2 + \beta_{2,2}^2}{-\beta_{0,0}^2 + \sum_{i,\ j}\beta_{i,j}^2}\end{aligned}$$

Exercise 4 : Kriging models [2 pts]

The *model 1* has some observation noise but no trend so it must be *model C*

The *model 2* has a linear trend, so it corresponds to *model D*. **bonus:** Since the confidence intervals width is constant in the region far away from the observations, the prediction type is *simple kriging*.

The *model 3* has a very short length-scale, as can be seen in *model A*.

The *model 4* has an exponential covariance function, as *model E*.

Exercise 5: Kernel design [7 pts]

All the questions from this exercise may be treated independently.

We consider the following kernel for $x,\ y \in \mathbb{R}$:

$$k(x, y) = k_0(x, y) + k_1(x, y) + k_2(x, y) + k_3(x, y)$$

$$\text{where} \qquad k_0(x,y) = \sigma_0^2 \qquad\qquad k_1(x,y) = \sigma_1^2 xy$$
$$k_2(x,y) = \sigma_2^2 x^2 y^2 \qquad\qquad k_3(x,y) = \sigma_3^2 \exp\left(-\frac{(x-y)^2}{2\theta^2}\right).$$

Q1. [1 pt] Show that k is a valid covariance function. You can use the fact that the kernels in Table 1.1 from page 14 of the lecture notes are valid covariance functions without proving it.

Q2. [1 pt] Write a *R* function (you may choose another language if you prefer) that takes as inputs two vectors (say a, b) and a variance parameter $s2$ and that returns the covariance matrix with general term $k_2(a_i, b_j)$.

Q3. [2 pts] Let Z be a centred Gaussian process with covariance k. Show carefully that Z has the same distribution as the sum of 4 independent Gaussian processes:

$$Z_0 \sim \mathcal{N}(0, k_0)\,,\ Z_1 \sim \mathcal{N}(0, k_1)\,,\ Z_2 \sim \mathcal{N}(0, k_2) \text{ and } Z_3 \sim \mathcal{N}(0, k_3).$$

Represent graphically (for $x \in [-1, 1]$) a typical sample from each of this processes for the following parameters values: $\sigma_i^2 = 1$ and $\theta = 0.2$.

Q4. [1 pt] let (X, Y) be a set of n observation points. What is the expression of the conditional mean and variance of $Z(x)$ given $Z(X) = F$. How does the mean function behaves when the prediction point x is far away from the observations X?

Q5. [2 pts] Show that the conditional mean writes as a sum of conditional mean functions : $m(x) = m_0 + \cdots + m_3(x)$. According to this, what is the polynomial content m_p of the mean function m? What is the conditional variance that can be associated to m_p?

bonus: Would we obtain the exact same model as m if we were to construct a universal kriging model with kernel k_3 and a trend of the form $\beta_0 + \beta_1 x + \beta_2 x^2$?

Exercise 6: Optimization [3 pts]

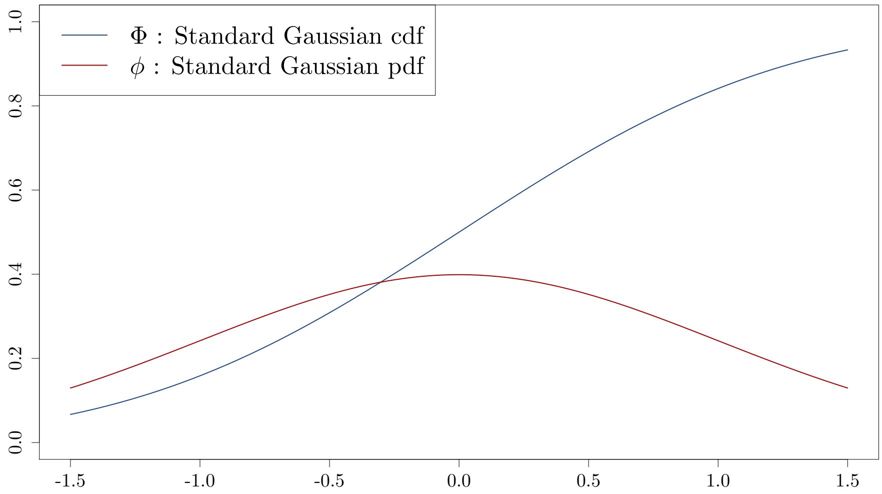
Consider a kriging model based on the design of experiments $X = (-0.2, 0.2)$ and the vector of observations $Y = (-0.2, 0.2)$. We will consider fixed values for the variance and length-scale parameters: $\sigma^2 = 1$ and $\theta = 0.2$.

Q1. [1 pt] Use simple kriging equations with exponential covariance kernel to get the prediction at $x = 0$. We recall the inversion formula for a 2×2 matrix:

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix}^{-1} = \frac{1}{a^2 - b^2} \times \begin{bmatrix} a & -b \\ -b & a \end{bmatrix} \qquad (1)$$

Q2. [1 pt] Evaluate the Expected Improvement criterion at $x = 0$. The figure bellow may be helpful.

Q3. [1 pt] The same evaluation for a model based on a Gaussian kernel gives $EI(0) = 0.15$. Compare with the previous result and analyse.



Φ : Standard Gaussian cdf
φ : Standard Gaussian pdf
0.0
0.2
0.4
0.6
0.8
1.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5