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 on the function: as a consequence, this must correspond 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 Db 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 $\mathbb{E}\big(X_2|X_1\big)$ is always null but the conditional variance $\mathrm{var}\big(X_2|X_1\big)$ 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:

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

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

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 for *model E*.

Exercise 5: Kernel design

Q1. We already know from the lecture notes that k_0, k_1 and k_3 are valid covariance functions, and k_2 is also a valid kernel since it is the composition of a linear kernel $\sigma_2^2 xy$ with the function $x \mapsto x^2$. Since k is a sum of covariance functions, it is also symmetric and positive semi-definite.

Q2. The following *R* code will return the covariance matrices for k_2

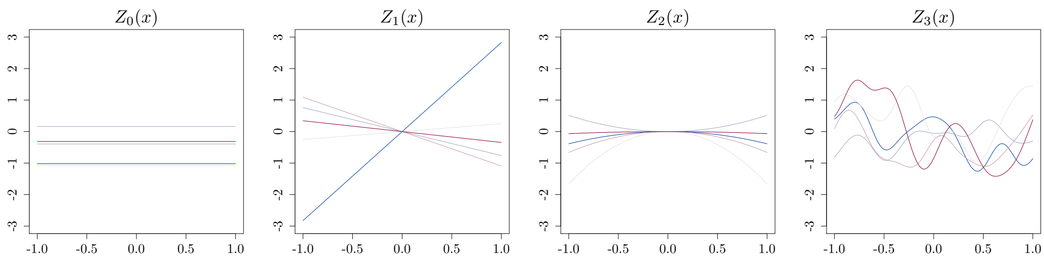
```
k2 <- function(a,b,s2=1){
    s2*outer(a^2,b^2)
}
```

Note the default value of 1 for the parameter $s2$, which allows to call the function without specifying any value for this parameter.

Q3. The sum of 4 independent Gaussian processes is also a GP, so it is sufficient to show that Z and $Z_s = Z_0 + Z_1 + Z_2 + Z_3$ have the same mean and covariance function. This is straightforward using the linearity of the expectation and the independence of the Z_i:

$$\begin{aligned}
\mathrm{E}[Z_s(x)] &= \mathrm{E}[Z_0(x) + Z_1(x) + Z_2(x) + Z_3(x)] \\
&= \mathrm{E}[Z_0(x)] + \mathrm{E}[Z_1(x)] + \mathrm{E}[Z_2(x)] + \mathrm{E}[Z_3(x)] \\
&= 0 = \mathrm{E}[Z(x)] \\
\mathrm{cov}[Z_s(x), Z_s(y)] &= \mathrm{cov}[Z_0(x) + Z_1(x) + Z_2(x) + Z_3(x), Z_0(y) + Z_1(y) + Z_2(y) + Z_3(y)] \\
&= \mathrm{cov}[Z_0(x), Z_0(y)] + \cdots + \mathrm{cov}[Z_3(x), Z_3(y)] \qquad \text{(by independence)} \\
&= k(x, y)
\end{aligned}$$

Typical examples of sample path are shown bellow:



Q4. As detailed in the lecture notes, the conditional mean and variance are:

$$\begin{aligned}
m(x) &= \mathrm{E}[Z(x)|Z(X) = F] = k(x, X)k(X, X)^{-1}F \\
v(x) &= \mathrm{var}[Z(x)|Z(X) = F] = k(x, x) - k(x, X)k(X, X)^{-1}k(X, x).
\end{aligned}$$

If the prediction point x is located far away from one observation point X_i, then

$$\begin{aligned}
k(x, X_i) &= \sigma_0^2 + \sigma_1^2 x X_i + \sigma_2^2 x^2 X_i^2 + \sigma_3^2 \exp\left(-\frac{(x - X_i)^2}{2\theta^2}\right) \\
&\approx \sigma_0^2 + \sigma_1^2 x X_i + \sigma_2^2 x^2 X_i^2
\end{aligned}$$

As a consequence, if the prediction point is far away from all observation points, the mean predictor is a linear combination of quadratic functions (each element of $k(x, X$ is a second order polynomial in x) so it behaves itself as a quadratic function.

Q5. We can expand $k(x, X)$ in the expression of m:

$$\begin{aligned} m(x) &= k(x,X)k(X,X)^{-1}F \\ &= (k_0(x,X) + k_1(x,X) + k_2(x,X) + k_3(x,X))k(X,X)^{-1}F \\ &= \underbrace{k_0(x,X)k(X,X)^{-1}F}_{m_0} + \underbrace{k_1(x,X)k(X,X)^{-1}F}_{m_1(x)} \\ &\quad + \underbrace{k_2(x,X)k(X,X)^{-1}F}_{m_2(x)} + \underbrace{k_3(x,X)k(X,X)^{-1}F}_{m_3(x)}. \end{aligned}$$

where m_0 is a constant, m_1 is linear and m_2 is quadratic. The polynomial component of m is then

$$m_p(x) = m_0 + m_1(x) + m_2(x) = (k_0(x,X) + k_1(x,X) + k_2(x,X))k(X,X)^{-1}F.$$

We can recognize here the conditional distribution of $Z_0 + Z_1(x) + Z_2(x)$ given $Z(X) = F$. The associated conditional variance is:

$$\begin{aligned} v_p(x) &= \mathrm{var}[Z_0 + Z_1(x) + Z_2(x) | Z(X) = F] \\ &= k_0(x,x) + k_1(x,x) + k_2(x,x) \\ &\quad - (k_0(x,X) + k_1(x,X) + k_2(x,X))k(X,X)^{-1}(k_0(X,x) + k_1(X,x) + k_2(X,x)). \end{aligned}$$

Exercise 6: Optimization

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. The simple kriging equations give

$$\begin{aligned} m(x) &= k(x,X)k(X,X)^{-1}F \\ &= \begin{bmatrix} \sigma^2 e^{-1} \\ \sigma^2 e^{-1} \end{bmatrix}^T \times \begin{bmatrix} \sigma^2 & \sigma^2 e^{-2} \\ \sigma^2 e^{-2} & \sigma^2 \end{bmatrix}^{-1} \times \begin{bmatrix} -0.2 \\ 0.2 \end{bmatrix} \\ &= \frac{1}{1 - e^{-4}} \begin{bmatrix} e^{-1} \\ e^{-1} \end{bmatrix}^T \times \begin{bmatrix} 1 & -e^{-2} \\ -e^{-2} & 1 \end{bmatrix} \times \begin{bmatrix} -0.2 \\ 0.2 \end{bmatrix} \\ &= 0 \end{aligned}$$

In a similar fashion, we obtain:

$$v(x) = k(x,x) - k(x,X)k(X,X)^{-1}k(X,x) = \cdots = 0.76$$

Note that the result $m(x) = 0$ could also be obtained noticing the problem symmetry.

Q2. The expression of the expected improvement is

$$EI(x) = \sqrt{v(x)}(u(x)\Phi(u(x)) + \phi(u(x))) \qquad \text{with } u(x) = \frac{\min(F) - m(x)}{\sqrt{v(x)}}$$

In $x = 0$, we thus have $u(0) = -0.2/0.87 = -0.23$. As a consequence, we have

$$EI(0) = 0.87 \times (-0.23 \times \Phi(-0.23) + \phi(-0.23)) = 0.26$$

Q3. The expected improvement is smaller when using a Gaussian kernel since the confidence intervals of the later model are mush smaller that the initial one.