Exam Solutions: Statistical modelling and its applications

Mines Saint-Étienne – 21 Décembre 2016

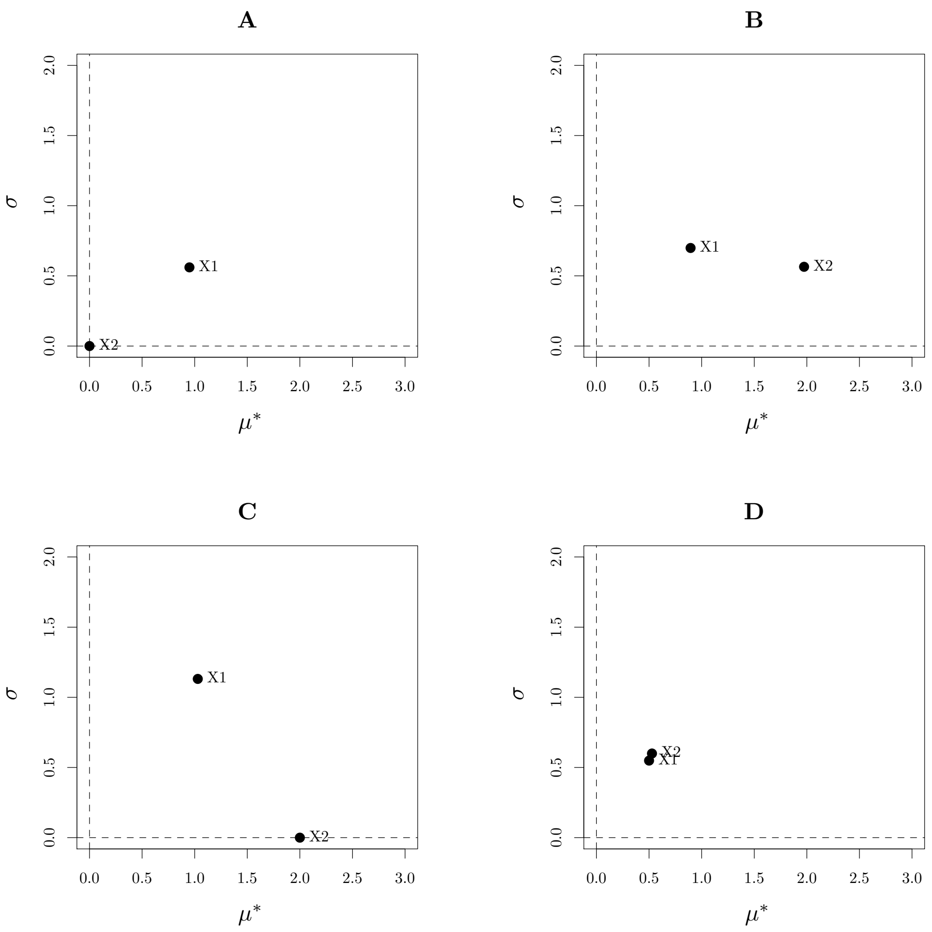
No document allowed except the Gaussian process regression handout, the slides of the optimization lecture and an A4 sheet with hand written notes. The scoring scale indicated for each question and exercise is only an indication, it may be adapted for the final grading.

Exercise 1 : The Morris method [2 pts]

Let $f_1, \ \dots\ , \ f_6$ be six functions over $[-\frac{1}{2}, +\frac{1}{2}]$ defined as:

$$f_1(x_1, x_2) = x_1^2 + x_2^2 \qquad f_2(x_1, x_2) = x_1 + 2x_2 + 2x_1x_2$$
$$f_3(x_1, x_2) = (x_2 + 1)^2 \qquad f_4(x_1, x_2) = x_1(1 + x_1)$$
$$f_5(x_1, x_2) = x_1 - 2x_2 + 2x_1^2 \qquad f_6(x_1, x_2) = 1 + 2x_1 - x_2.$$

For four of these functions, we apply the Morris method ($r = 20$ repetitions) and we obtain the following graphs for the averages μ^* and the standard deviations σ of the absolute finite differences:



Q1. [2 pts] Can you recover the functions associated to each graph? Justify briefly your answers.

Exercise 2 : Sobol-Hoeffding decomposition over the disk [2 pts]

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. [1 pt] Explain geometrically (based on a graphic for instance) that:

$$\begin{cases} \mathbb{E}\big(X_2 | X_1 = \frac{3}{4}\big) = \mathbb{E}\big(X_2 | X_1 = \frac{1}{4}\big), \\ \mathrm{var}\big(X_2 | X_1 = \frac{3}{4}\big) < \mathrm{var}\big(X_2 | X_1 = \frac{1}{4}\big). \end{cases}$$

As a consequence, is it possible to perform Sobol sensitivity analysis directly on f ?

Q2. [1 pt] We now consider the polar representation of f:

$$f_p(\rho, \theta) = \rho \cos(\theta) + \rho\,,$$

and we denote by R and A two independent random variables over $[0, 1]$ and $[0, 2\pi)$ and we use the following notations: $m_r = \mathbb{E}\big(R\big)$, $m_a = \mathbb{E}\big(\cos(A)\big)$. What is the Sobol decomposition of f_p ? The result can be expressed as a function of m_r and m_a.

bonus: We now assume that $\big(R\cos(A), R\sin(A)\big)$ is uniformly distributed on the disk. What is the distribution of A? Compute the cumulative distribution function (la fonction de répartition) of R. Deduce the associated values of m_a and m_r.

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] Detail how to compute the next basis function h_3 and give its expression.

Q2. [1 pt] We now consider the following basis of functions over $[-1, 1]^2$:

$$\begin{array}{lll} h_{00}(x) = h_0(x_1) \times h_0(x_2) & h_{11}(x) = h_1(x_1) \times h_1(x_2) & h_{02}(x) = h_0(x_1) \times h_2(x_2) \\ h_{10}(x) = h_1(x_1) \times h_0(x_2) & h_{20}(x) = h_2(x_1) \times h_0(x_2) & h_{12}(x) = h_1(x_1) \times h_2(x_2) \\ h_{01}(x) = h_0(x_1) \times h_1(x_2) & h_{21}(x) = h_2(x_1) \times h_1(x_2) & h_{22}(x) = h_2(x_1) \times h_2(x_2) \end{array}$$

Show that these basis functions are orthonormal for a uniform probability measure on $[-1, 1]^2$. Deduce that these functions are centred and satisfy the Sobol non simplification conditions.

Q3. [2 pt] Let f be a function over $[-1, 1]^2$, and let X be a set of 20 points in $[-1, 1]^2$ for which the value of f is known: $f(X) = F$. We denote by β the vector of the coefficients of a linear regression model based on these observations and the above basis:

$$m(x) = \sum_{0 \leq i,j \leq 2} \beta_{i,j} h_{i,j}(x).$$

What is the Sobol decomposition of m. Compute the Sobol indices of m, your result should be expressed as a function of the $\beta_{i,j}$.

Exercise 4 : Kriging models [2 pts]

You will find bellow the descriptions and the graphs of various models. Can you recover which graph is associated with each model? Justify briefly your answers.

Model 1

```
Trend  coeff.:

 (Intercept)      0.0000

Covar. type   : matern5_2
Covar. coeff.:

    theta(x)      0.3000

Variance: 9

Nugget effect : 0.2
```

Model 2

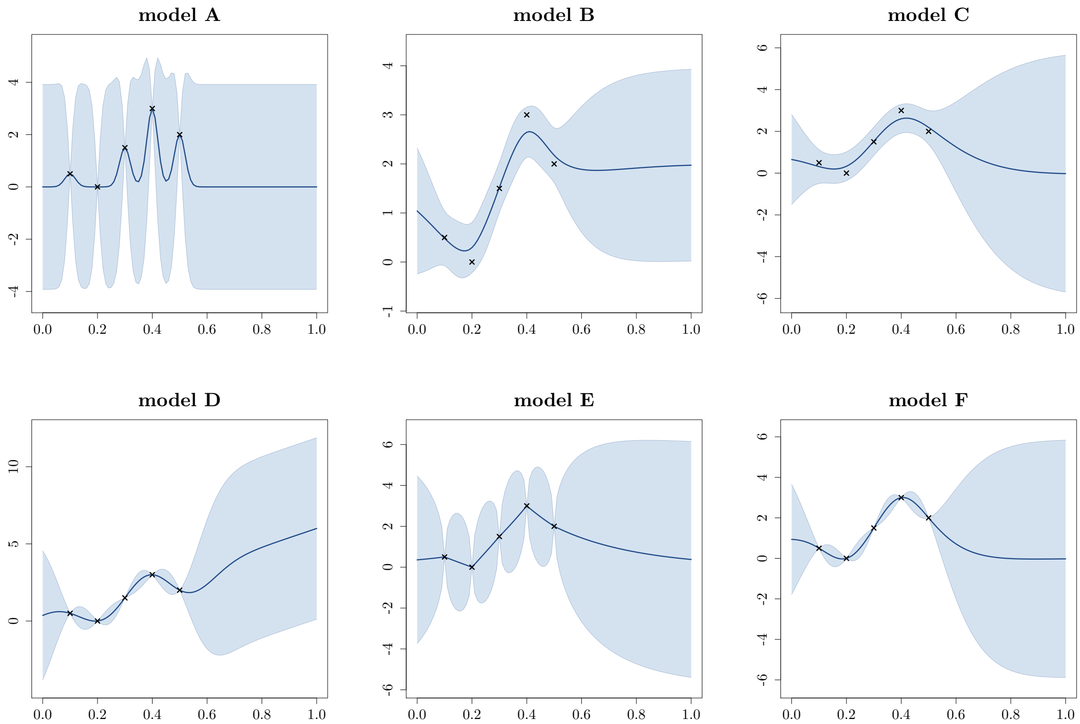
```
Trend  coeff.:

 (Intercept)      0.0000
           x      6.0000

Covar. type   : gauss
Covar. coeff.:

    theta(x)      0.1000

Variance: 9

Nugget effect : 0
```

Model 3

```
Trend  coeff.:

 (Intercept)      0.0000

Covar. type   : gauss
Covar. coeff.:

    theta(x)      0.0200

Variance: 4

Nugget effect : 0
```

Model 4

```
Trend  coeff.:

 (Intercept)      0.0000

Covar. type   : exp
Covar. coeff.:

    theta(x)      0.3000

Variance: 9

Nugget effect : 0
```



bonus: Regarding the plot of model D, can you specify if the prediction type is *simple*, *ordinary* or *universal kriging*?

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 : $m(x) = m_0 + \dots + m_3(x)$. 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:

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

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.