Computer experiments Exam

INSA Toulouse – ModIA – 2023 October

1) Authorized documents: a single handwritten A4 sheet (recto-verso).
2) Electronic devices (mobile phone, calculator, laptop, etc.) are not allowed.
3) All answers must be justified.

1 GP engineering (10 pts)

1. (3 pts) Let $X = (X(t))_{t\in\mathbb{R}}$ be a centered Gaussian Process (GP) on $\mathbb{R}$ with kernel k. We denote by $X'(t), X''(t)$ the derivatives of $X(t)$ at t and $\int_{\mathbb{R}} X(t)dt$ the mean over the real line. We assume that the mathematical conditions are satisfied to ensure the validity of these objects, as well as the next computations.

   (a) What is the kernel of the process $X'(t) + X''(t)$?

   (b) To which covariance corresponds $\frac{\partial^2 k(s,t)}{\partial s^2}$?

   (c) What is the covariance between $X'(s)$ and $\int_{\mathbb{R}} X(t)dt$?

   (a) Using the linearity of derivation:

   $$k_Y(s,t) = \mathrm{cov}(X'(s) + X''(s), X'(t) + X''(t)) = \frac{\partial^2 k}{\partial s \partial t}(s,t) + \frac{\partial^3 k}{\partial s \partial t^2}(s,t) + \frac{\partial^3 k}{\partial s^2 \partial t}(s,t) + \frac{\partial^4 k}{\partial s^2 \partial t^2}(s,t)$$

   (b) It is $\mathrm{cov}\left(X''(s), X(t)\right)$.

   (c) Using the linearity of derivative / integration and the bilinearity of covariance, we have:

   $$\mathrm{cov}\left(X'(s), \int_{\mathbb{R}} X(t)dt\right) = \int_{\mathbb{R}} \frac{\partial k}{\partial s}(s,t)dt$$

2. (3 pts) Let $Y_0(x_1, x_2)$ be a 2-dimensional GP defined on the unit square $\chi = [0,1]^2$ and denote by Δ the diagonal: $x_1 = x_2$. We aim at defining a GP whose sample paths are symmetric with respect to Δ. Let s be the symmetry with respect to Δ. We propose to consider for all $x = (x_1, x_2) \in \chi$:

   $$Y(x) = Y_0(x) + Y_0(s(x))$$

   (a) Give the expression of $s(x)$ in function of x_1 and x_2 (make a picture of χ). Deduce that the sample paths of Y are indeed symmetric with respect to Δ.
   We have $s(x_1, x_2) = (x_2, x_1)$. Then a direct computation gives $Y(s(x)) = Y(x)$, which means that the sample paths of Y are symmetric with respect to Δ.

   (b) Let $C = \alpha_1 Y(x^1) + \cdots + \alpha_n Y(x^n)$ a linear combination extracted from Y at points $x^1, \ldots, x^n \in \chi$. Prove that C is normally distributed. What does it mean for Y?
   We have

   $$C = \alpha_1 Y_0(x^1) + \cdots + \alpha_n Y_0(x^n) + \alpha_1 Y_0(s(x^1)) + \cdots + \alpha_n Y_0(s(x^n))$$

   Thus C is a linear combination extracted from Y_0 at points $x^1, \ldots, x^n, s(x^1), \ldots, s(x^n)$. As Y_0 is a GP, C is normally distributed. As it is true for all linear combination, Y is a GP.

(c) Let k_0 be the kernel of Y_0, and k the kernel of Y. Express $k_Y(x, x')$ in function of k_0, for all $x, x' \in \chi$.
Using the bilinearity of covariance, we have

$$\begin{aligned} k_Y(x, x') &= \mathrm{Cov}(Y_0(x) + Y_0(s(x)), Y_0(x') + Y_0(s(x'))) \\ &= k_0(x, x') + k_0(s(x), x') + k_0(x, s(x')) + k_0(s(x), s(x')) \end{aligned}$$

3. (4 pts) Let us consider the subverse function of the computer lab

$$S = \left(\frac{Q}{BK_s\sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6} + Z_v - H_d - C_b$$

In the lab, we have used a centered GP metamodel with a tensor-product kernel k_d in dimension $d = 8$. Assume now that you have access to the expression of S. Contrarily to the lab, where you have played with the trend of the GP, here we consider only GPs with a constant mean. What new GP Ycan you propose that will mimick the structure of the function S ? You may define this GP in function of four other "small" GPs $Y_1, ..., Y_4$ acting on subsets of variables. Denote by d_i the dimension of Y_i, and choose k_{d_i} as a kernel of Y_i. Compute the kernel of Y in function of k_{d_i} $(i = 1, ..., 4)$. How can you adapt this approach such that the small GPs are defined on *disjoint* subsets of variables ?
We can see that S has an additive form, with respect to $x_1 = (Q, B, K_s, Z_m, Z_v, L), x_2 = Z_v, x_3 = H_d, x_4 = C_b$. Indeed,

$$S = g_1(x_1) + g_2(x_2) + g_3(x_3) + g_4(x_4)$$

Hence, an appropriate GP should have the same form. Denoting $x = (x_1, x_2, x_3, x_4)$, we define Y by

$$Y(x) = Y_1(x_1) + Y_2(x_2) + Y_3(x_3) + Y_4(x_4)$$

where Y_1, Y_2, Y_3, Y_4 are independent GPs in dimensions $6, 1, 1, 1$ respectively. We know that a sum of independent GPs is a GP, thus Y is a GP. Its kernel can be computed using the bilinearity of the covariance function,

$$k_Y(x, x') = k_6(x_1, x'_1) + k_1(x_2, x'_2) + k_1(x_3, x'_3) + k_1(x_4, x'_4)$$

If we want to have disjoint subsets of variables, we may discard x_2 as it is already contained in x_1. In particular, S has also the form

$$S = h_1(x_1) + g_3(x_3) + g_4(x_4)$$

with $h_1(x_1) = g_1(x_1) + g_2(x_2)$. Thus Y can be defined as a sum of 3 independent GPs in dimensions $6, 1, 1$.

2 Sensitivity analysis and design of experiments (10 pts)

We consider three independent random variables X_1, X_2, X_3, following the uniform distribution on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. We want to perform a global sensitivity analysis of the 3-dimensional function

$$f(X) = X_1 + aX_2^2 + bX_3^2X_1,$$

where $X = (X_1, X_2, X_3)$ and a, b are given non-zero real numbers.
Recall that for all $i = 1, 2, 3$, we have $\mathbb{E}(X_i) = 0$ and $\mathrm{Var}(X_i) = \frac{1}{12}$. We denote $v_2 = \mathrm{Var}(X_i^2)$.

1. (1 pt) A naive idea is that the Sobol-Hoeffding (or ANOVA) decomposition is simply given by $f_0 = 0$, $f_1(X_1) = X_1$, $f_2(X_2) = aX_2^2$ and $f_{1,3}(X_1, X_3) = bX_3^2 X_1$. By considering for instance $f_2(X_2)$, explain why it cannot be true.
In an ANOVA decomposition, all terms must be centered, which is not the case of $f_2(X_2)$.

2. (3 pts) Compute the global mean $f_0 = \mathbb{E}[f(X)]$ and all the main effects $f_i(X_i) = \mathbb{E}[f(X)|X_i] - f_0$ ($i = 1, 2, 3$). Compute the unnormalized Sobol indices $D_i = \mathrm{Var}(f_i(X_i))$, $i = 1, 2, 3$ in function of v_2.
By linearity of the expectation, and independence of X_1 and X_3, we first have:
$$f_0 = \mathbb{E}(X_1) + a\mathbb{E}(X_2^2) + b\mathbb{E}(X_3^2)\mathbb{E}(X_1) = \frac{a}{12}.$$
Then using the properties of conditional expectation, and independence of X_1, X_2, X_3:
$$\begin{aligned} f_1(X_1) &= X_1 + a\mathbb{E}(X_2^2) + bX_1\mathbb{E}(X_3^2) - f_0 = \left(1 + \frac{b}{12}\right) X_1 \\ f_2(X_2) &= \mathbb{E}(X_1) + aX_2^2 + b\mathbb{E}(X_3^2)\mathbb{E}(X_1) - f_0 = a\left(X_2^2 - \frac{1}{12}\right) \\ f_3(X_3) &= \mathbb{E}(X_1) + a\mathbb{E}(X_2^2) + bX_3^2\mathbb{E}(X_1) - f_0 = 0 \end{aligned}$$
We deduce immediately $D_3 = 0$, $D_1 = \left(1 + \frac{b}{12}\right)^2 \frac{1}{12}$, and $D_2 = a^2 v_2$ (since $\mathrm{Var}\left(X_2^2 - \frac{1}{12}\right) = \mathrm{Var}(X_2^2)$).

3. (2 pts) Compute $f_{1,3}(X_1, X_3) := \mathbb{E}[f(X)|X_1, X_3] - f_1(X_1) - f_3(X_3) - f_0$ and the corresponding unnormalized Sobol index $D_{1,3} = \mathrm{Var}(f_{1,3}(X_1, X_3))$.
We have
$$f_{1,3}(X_1, X_3) = X_1 + a\mathbb{E}(X_2^2) + bX_3^2 X_1 - f_1(X_1) - f_3(X_3) - f_0 = bX_1\left(X_3^2 - \frac{1}{12}\right).$$
As all terms of the ANOVA decomposition are centered, its variance is given by
$$\mathrm{Var}(f_{1,3}(X_1, X_3)) = \mathbb{E}\left[f_{1,3}^2(X_1, X_3)\right] = b^2\mathbb{E}\left[X_1^2\right]\mathbb{E}\left[\left(X_3^2 - \frac{1}{12}\right)^2\right] = b^2\frac{1}{12}v_2.$$

4. (1 pt) We admit that $f_{1,2}(X_1, X_2) = f_{2,3}(X_2, X_3) = 0$. Without further computation, write the ANOVA decomposition of f and give the expression of the global variance $D = \mathrm{Var}[f(X)]$ in function of v_2.
The ANOVA decomposition of f is thus given by $f(X) = f_0 + f_1(X_1) + f_2(X_2) + f_{1,3}(X_1, X_3)$. By orthogonality of the terms, the global variance is equal to
$$D = D_1 + D_2 + D_{1,3} = \left(1 + \frac{b}{12}\right)^2 \frac{1}{12} + \left(a^2 + \frac{b^2}{12}\right) v_2$$

5. (1,5 pts) Explain how we can estimate the main effect $f_i(X_i)$ by simulation (without using the expression of question 1). How can we visualize it as a curve?
We consider a sample $X^1, \dots, X^n$ of X drawn from the uniform distribution on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, obtained by sampling each coordinate independently. Then we can compute $Y^n = f(X^n)$. We can estimate the global mean $\widehat{f_0}$ as the average of $Y^1, \dots, Y^n$. The conditional expectation $\mathbb{E}[f(X)|X_i]$ is estimated by smoothing the points (X_i^n, Y^n), and an estimate of the main effect is deduced by removing $\widehat{f_0}$. Finally, the main effect is visualized as a curve in the scatterplot of $f(X)$ versus X_i.

6. (1,5 pts) An experimenter has a computational budget of 27 points to learn f, and in particular to estimate its main effects (see question 5). He proposes to use a grid with 3 points per dimension. Considering for instance the estimation of the main effect $f_2(X_2)$, do you think it is a good idea? Why? If not, what other design could you propose to him to improve the estimation of the main effects of f?
To estimate the main effect $f_2(X_2)$, we will have only 3 different values of X_2, compared to the 27 points!! To avoid this drawback, a maximin Latin hybercube design is more appropriate.