Statistical modelling and its applications

Exam Solutions – Mines Saint-Étienne – 5th January 2016

Exercise 1

1. Let X be the subset of input space points where we want the sample to be evaluated (typically $X = \{0,\ 0.01,\ \ldots,\ 1\}$). We will denote n its number of elements. The 4 steps for generating samples from a Gaussian process are:
   (a) Compute the vector of the process mean at X: $m = \mu(X)$.
   (b) Compute the covariance matrix of the process evaluated at X: $K = k(X, X)$ and its Cholesky decomposition: $K = CC^t$.
   (c) Generate a vector Y of n iid samples $\mathcal{N}(0, 1)$.
   (d) A sample of the process evaluated at X is then given by $Z = m + CY$.
2. The results are summarized in the following table:

process	kernel	centred	stationary
Z_1	Brownian	no	no
Z_2	squared exponential	no	no
Z_3	Matérn 3/2	yes	yes
Z_4	exponential	yes	yes

3. In this figures, Z_5 has variance 1 and length scale 0.1 and Z_6 has parameters $(\sigma^2, \theta) = (100, 10)$.
4. The main steps for computing the mean value of a function using no more that 50 observations are:
   (a) Choose one or various kernels according to the prior belief on the function to approximate.
   (b) Choose a type of Gaussian process model (simple Kriging, ordinary Kriging, Universal Kriging) according to the prior belief on f.
   (c) Estimate the model parameters by maximising the likelihood.
   (d) Validate the models using cross validation methods (both the mean value and predicted conditional covariance must be validated) and choose the best model.
   (e) An estimate of the mean value of f is given by the integrals of conditional samples.

Exercise 2

Recall that $f(x_1, x_2) = x_1 + x_2 + x_1x_2$, and X_1, X_2 are independent random variables, with $X_1 \sim \mathcal{U}[-\frac{a}{2}, \frac{a}{2}]$ and $X_2 \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$, with $a > 0$.

1. We immediately have $\mathrm{E}(X_1) = E(X_2) = 0$, and by independence $E(X_1X_2) = E(X_1)E(X_2) = 0$, proving the centring conditions. Furthermore $E(X_1X_2|X_1) = X_1E(X_2|X_1) = X_1E(X_2) = 0$, and similarly $E(X_1X_2|X_2) = 0$, proving the non-simplification conditions. By unicity, the Sobol-Hoeffding decomposition of $f(X_1, X_2)$ is then given by:

$$\mu_0 = 0, \quad \mu_1(X_1) = X_1, \quad \mu_2(X_2) = X_2, \quad \mu_{1,2}(X_1, X_2) = X_1X_2$$

2. We immediately have $D_1 = \mathrm{var}(X_1) = \frac{a^2}{12}$ and $D_2 = \mathrm{var}(X_2) = \frac{1}{12}$.
Further, using once again the independence of X_1, X_2, we get:

$$D_{1,2} = \mathrm{var}(X_1X_2) = \mathrm{E}(X_1^2X_2^2) = \mathrm{E}(X_1^2)\mathrm{E}(X_2^2) = \mathrm{var}(X_1)\mathrm{var}(X_2) = \frac{a^2}{12} \times \frac{1}{12}$$

Hence $D = D_1 + D_2 + D_{1,2} = \frac{1}{12}\left(1 + \frac{13}{12}a^2\right)$

3. We have $S_1 = \frac{a^2}{1+\frac{13}{12}a^2} = \frac{1}{\frac{13}{12}+\frac{1}{a^2}}$ which is an increasing function of a. Similarly, $S_2 = \frac{1}{1+\frac{13}{12}a^2}$ is a decreasing function of a. This is rather logical: Increasing the uncertainty of X_1 increases the importance of X_1 in the output $f(X_1, X_2)$ and reduces the importance of X_2.

We now assume that X_1, X_2 are independent random variables, with $X_1, X_2 \sim \mathcal{U}[0, 1]$.

4. We cannot have $\mu_1(X_1) = X_1$ since for instance $\mathrm{E}(X_1) = \frac{1}{2} \neq 0$.

5. The Sobol decomposition of $f(X_1, X_2)$ can be obtained with the recursion formulae, or by using a Taylor expansion of f at $(\frac{1}{2}, \frac{1}{2})$. By one of these methods, we obtain:

$$\mu_0 = \frac{5}{4}, \quad \mu_1(X_1) = \frac{3}{2}\left(X_1 - \frac{1}{2}\right), \quad \mu_2(X_2) = \frac{3}{2}\left(X_2 - \frac{1}{2}\right), \quad \mu_{1,2}(X_1, X_2) = \left(X_1 - \frac{1}{2}\right)\left(X_2 - \frac{1}{2}\right)$$

Exercise 3: ANOVA kernels

1. Since the function $(x, y) \mapsto 1$ is symmetric and positive semi-definite (*spsd*), $1 + k_i(x_i, y_i)$ is also *spsd* as a sum of *spsd* functions. ANOVA kernels then can then be seen as a product of *spsd* functions, which is a valid covariance function.

2. Knowing the regularity of f, one should consider smooth kernels such as the squared exponential (or Gaussian) kernel. Since we have no information regarding the mean of the process, one can try ordinary Kriging.

3. For the simple Kriging case, the expressions of the predicted mean and variance are:

$$m(x) = \hat{\mu} + k(x, X)k(X, X)^{-1}(F - \hat{\mu})$$
$$v(x) = k(x, x) - k(x, X)k(X, X)^{-1}k(X, x) + \frac{(1 + k(x, X)k(X, X)^{-1}\mathbf{1})^2}{\mathbf{1}^t k(X, X)^{-1}\mathbf{1}}$$

where $\hat{\mu}$ is the maximum likelihood estimation of the process mean : $\hat{\mu} = \dfrac{1^t k(X, X)^{-1}F}{1^t k(X, X)^{-1}1}$. The 95% confidence intervals are given by $m(x) \pm 2\sqrt{v(x)}$.

4. Since the first term of the mean expression can be written as an entry-wise product $k(x, X) = \prod_{i=1}^{d} \left(1 + k_i(x_i, X_{[.,i]})\right)$, the expression of m can be expended as:

$$m(x) = \prod_{i=1}^{d} \left(1^t + k_i(x_i, X_{[.,i]})\right) k(X, X)^{-1} F$$

$$m(x) = \left(1^t + \sum_{i=1}^{d} k_i(x_i, X_{[.,i]}) + \sum_{i<j}^{d} k_i(x_i, X_{[.,i]}) k_j(x_j, X_{[.,j]}) + \cdots + \prod_{i=1}^{d} k_i(x_i, X_{[.,i]})\right) k(X, X)^{-1} F$$

$$m(x) = 1^t k(X, X)^{-1} F + \sum_{i=1}^{d} k_i(x_i, X_{[.,i]}) k(X, X)^{-1} F + \cdots + \left(\prod_{i=1}^{d} k_i(x_i, X_{[.,i]})\right) k(X, X)^{-1} F$$

This corresponds to a sum of 2^d functions with increasing interaction order: $m(x) = m_0 + \sum m_i(x_i) + \cdots + \sum m_{1,\ldots,d}(x)$. However, this decomposition does not coincides with the Sobol decomposition since, for example, the non constant terms of the decomposition are not zero-mean.

5. Writing the kernel as a sum of kernels implies that the process can be seen as a sum of independent processes: $Z(x) = Z_0 + \sum Z_i(x_i) + \cdots + \sum Z_{1,\ldots,d}(x)$. For $I \subset \{1, \ldots, d\}$, the term m_I can thus be seen as $m_I(x_I) = \mathrm{E}[Z_I(x_I) | Z(X) = F]$.

6. If the mean is known, we should consider simple Kriging. Furthermore, since there is no interaction of order higher than 2, the kernel expression can be simplified to

$$k(x, y) = 1 + \sum_{i=1}^{d} k_i(x_i, y_i) + \sum_{i<j}^{d} k_i(x_i, y_i) k_j(x_j, y_j).$$

7. If the k_i satisfy $\int_0^1 k_i(s, x)\, ds = 0$ for all $x \in [0, 1]$, then all the sub-models m_I and m_J are L^2-orthogonal for $I \neq J$. As a consequence, the sub-models correspond directly to the Sobol-Hoeffding decomposition.

Bonus: Let Z be a centred process over $[0, 1]$ with kernel k such that $\left(Z, \int Z(s)\, ds\right)$ is Gaussian. We thus have

$$\mathrm{cov}\left[Z(x), \int Z(s)\, ds\right] = \int k(x, s)\, ds$$

$$\mathrm{cov}\left[\int Z(s)\, ds, \int Z(s)\, ds\right] = \iint k(s, t)\, ds\, dt.$$

The usual multivariate normal conditioning formula apply to compute the conditional covariance:

$$\mathrm{cov}\left[Z(x), Z(y) \,\middle|\, \int Z(s)\, ds = 0\right] = k(x, y) - \frac{\int k(x, s)\, ds \int k(y, s)\, ds}{\iint k(s, t)\, ds\, dt}$$

This expression can be used as univariate kernels k_i and it does satisfy $\int_0^1 k_i(s, x)\, ds = 0$.