Bayesian approach of the parameter inverse problem under uncertainties

Victor Trappler

Directeurs de Thèse : Arthur VIDARD (Inria)
Élise ARNAUD (UGA)
Laurent DEBREU (Inria)

22 janvier 2020

Table des matières

(Joint) Posterior formulation

Priors

$$
\begin{aligned}
K &\sim \mathcal{U}(\mathbb{K}), \quad p(k) \\
U &\sim \mathcal{U}(\mathbb{U}), \quad p(u)
\end{aligned}
$$

Likelihood model

$$\begin{aligned} p(y \mid k, u, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} SS(k,u)\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} \|\mathcal{M}(k,u) - y\|_\Sigma^2\right] \end{aligned}$$

Now to Bayes' theorem

$$p(k, u \mid y, \sigma^2) = \frac{p(y \mid k, u, \sigma^2) p(k,u)}{\iint_{\mathbb{K}\times\mathbb{U}} p(y \mid k, u, \sigma^2) p(k,u) \,\mathrm{d}(k,u)}$$

Let us assume an hyperprior for σ^2 : $p(\sigma^2)$

GP, RR-based family of estimators

Random processes

Let us assume that we have a map f from a p dimensional space to $\mathbb{R}$:

$$\begin{array}{rccc} f : & \mathbb{X} \subset \mathbb{R}^p & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & f(x) \end{array} \tag{1}$$

This function is assumed to have been evaluated on a design of n points, $\mathcal{X} \subset \mathbb{X}^n$. We wish to have a probabilistic modelling of this function We introduce random processes as way to have a prior distribution on function This uncertainty on f is modelled as a random process :

$$\begin{array}{rcl} Z : \mathbb{X} \times \Omega & \longrightarrow & \mathbb{R} \\ (x, \omega) & \longmapsto & Z(x,\omega) \end{array} \tag{2}$$

The ω variable will be omitted next.

Linear Estimation

A linear estimation $\hat{Z}$ of f at an unobserved point $x \notin \mathcal{X}$ can be written as

$$\hat{Z}(x) = \begin{bmatrix} w_1 \dots w_n \end{bmatrix} \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} = \mathbf{W}^T f(\mathcal{X}) = \sum_{i=1}^n w_i(x) f(x_i) \tag{3}$$

Using those kriging weights $\mathbf{W}$, a few additional conditions must be added :

— Non-biased estimation : $\mathbb{E}[\hat{Z}(x) - Z(x)] = 0$

— Minimal variance : $\min \mathbb{E}[(\hat{Z}(x) - Z(x))^2]$

Translating using Eq.(3) :

$$\mathbb{E}[\hat{Z}(x)-Z(x)]=0 \iff m(\sum_{i=1}^{n} w_i(x)-1)=0 \iff \sum_{i=1}^{n} w_i(x)=1 \iff \mathbf{1}^T\mathbf{W}=1 \quad (4)$$

For the minimum of variance, we introduce the augmented vector $\mathbf{Z}_n(x)=[Z(x_1),\dots Z(x_n),Z(x)]$, and the variance can be expressed as :

$$\mathbb{E}[(\hat{Z}(x)-Z(x))^2]=\mathbb{V}\text{ar}\left[[\mathbf{W},-1]^T\mathbf{Z}_n(x)\right] \quad (5)$$

GP of the penalized cost function Δ_α

GP processes

Let $\Delta_\alpha(\mathbf{k},\mathbf{u})=J(\mathbf{k},\mathbf{u})-\alpha J^*(\mathbf{u})$. Furthermore, we assume that we constructed a GP on J on the joint space $\mathbb{K}\times\mathbb{U}$, based on a design of n points $\mathcal{X}=\left\{(\mathbf{k}^{(1)},\mathbf{u}^{(1)}),\dots,(\mathbf{k}^{(n)},\mathbf{u}^{(n)})\right\}$, denoted as $(\mathbf{k},\mathbf{u})\mapsto Y(\mathbf{k},\mathbf{u})$.

As a GP, Y is described by its mean function m_Y and its covariance function $C(\cdot,\cdot)$, while $\sigma_Y^2(\mathbf{k},\mathbf{u})=C((\mathbf{k},\mathbf{u}),(\mathbf{k},\mathbf{u}))$

$$Y(\mathbf{k},\mathbf{u})\sim\mathcal{N}\left(m_Y(\mathbf{k},\mathbf{u}),\sigma_Y^2(\mathbf{k},\mathbf{u})\right) \quad (6)$$

Let us consider now the conditional minimiser :

$$J^*(\mathbf{u})=J(\mathbf{k}^*(\mathbf{u}),\mathbf{u})=\min_{\mathbf{k}\in\mathbb{K}} J(\mathbf{k},\mathbf{u}) \quad (7)$$

Analogous to J and J^*, we define Y^* as

$$Y^*(\mathbf{u})\sim\mathcal{N}\left(m_Y^*(\mathbf{u}),\sigma_Y^{2,*}(\mathbf{u})\right) \quad (8)$$

where

$$m_Y^*(\mathbf{u})=\min_{\mathbf{k}\in\mathbb{K}} m_Y(\mathbf{k},\mathbf{u}) \quad (9)$$

The surrogate conditional minimiser is used in Ginsbourger profiles etc. The α-relaxed difference Δ_α modelled as a GP can then be written as

Considering the joint distribution of $Y(\mathbf{k},\mathbf{u})$ and $Y^*(\mathbf{u})=Y(\mathbf{k}^*(\mathbf{u}),\mathbf{u})$, we have

$$\begin{bmatrix}Y(\mathbf{k},\mathbf{u})\\ Y^*(\mathbf{u})\end{bmatrix}\sim\mathcal{N}\left(\begin{bmatrix}m_Y(\mathbf{k},\mathbf{u})\\ m_Y^*(\mathbf{u})\end{bmatrix};\begin{bmatrix}C\left((\mathbf{k},\mathbf{u}),(\mathbf{k},\mathbf{u})\right) & C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right)\\ C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) & C\left((\mathbf{k}^*(\mathbf{u}),\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right)\end{bmatrix}\right) \quad (10)$$

By multiplying by the matrix $[1 \quad -\alpha]$ yields

$$\Delta_\alpha(\mathbf{k},\mathbf{u})\sim\mathcal{N}\left(m_\Delta(\mathbf{k},\mathbf{u});\sigma_\Delta^2(\mathbf{k},\mathbf{u})\right) \quad (11)$$

$$m_\Delta(\mathbf{k},\mathbf{u})=m_Y(\mathbf{k},\mathbf{u})-\alpha m_Y^*(\mathbf{u}) \quad (12)$$

$$\sigma_\Delta^2(\mathbf{k},\mathbf{u})=\sigma_Y^2(\mathbf{k},\mathbf{u})+\alpha^2\sigma_{Y^*}^2(\mathbf{k},\mathbf{u})-2\alpha C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) \quad (13)$$

Assuming that $C((\mathbf{k},\mathbf{u}),(\mathbf{k}',\mathbf{u}')) = s\prod_{i\in\mathcal{I}_\mathbf{k}} \rho_{\theta_i}(\|k_i - k_i'\|)\prod_{j\in\mathcal{I}_\mathbf{u}}\rho_{\theta_j}(u_j - u_j'\|)$

$$C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) = s\prod_{i\in\mathcal{I}_\mathbf{k}} \rho_{\theta_i}(\|k_i - k_i^*(\mathbf{u})\|)\prod_{j\in\mathcal{I}_\mathbf{u}}\rho_{\theta_j}(0) \tag{14}$$

$$= s\prod_{i\in\mathcal{I}_\mathbf{k}} \rho_{\theta_i}(\|k_i - k_i^*(\mathbf{u})\|) \tag{15}$$

Approximation of the objective probability using GP

We are going now to use a different notation for the probabilities, taken with respect to the GP : $\mathcal{P}$, to represent the uncertainty encompassed by the GP.

Defined somewhere else, we have

$$\Gamma_\alpha(\mathbf{k}) = \mathbb{P}_\mathbf{U}\left[J(\mathbf{k},\mathbf{U}) \leq \alpha J^*(\mathbf{U})\right] \tag{16}$$

$$= \mathbb{E}_\mathbf{U}\left[\mathbb{1}_{J(\mathbf{k},\mathbf{U})\leq\alpha J^*(\mathbf{U})}\right] \tag{17}$$

This classification problem can be approached with a plug-in approach, or a probablistic one :

$$\mathbb{1}_{J(\mathbf{k},\mathbf{u})\leq\alpha J^*(\mathbf{u})} \approx \mathbb{1}_{m_Y(\mathbf{k},\mathbf{u})\leq\alpha m_Y^*(\mathbf{u})} \tag{18}$$

$$\mathbb{1}_{J(\mathbf{k},\mathbf{u})\leq\alpha J^*(\mathbf{u})} \approx \mathcal{P}\left[\Delta_\alpha(\mathbf{k},\mathbf{u})\leq 0\right] = \pi(\mathbf{k},\mathbf{u}) \tag{19}$$

Using the GPs, for a given $\mathbf{k}$, α and $\mathbf{u}$, the probability for our meta model to verify the inequality is given by Based on those two approximation, the approximated probability Γ is

$$\hat{\Gamma}_{\alpha,n}(\mathbf{k}) = \mathbb{P}_U\left[m_Y(\mathbf{k},\mathbf{u}) \leq \alpha m_Y^*(\mathbf{u})\right] \qquad \text{(plug-in)}$$
$$\hat{\Gamma}_{\alpha,n}(\mathbf{k}) = \mathbb{E}_U\left[\mathcal{P}\left[\Delta_\alpha(\mathbf{k},\mathbf{u})\leq 0\right]\right] \qquad \text{(Probabilistic approx)} \tag{20}$$

The probability of coverage for the set $\{Y - \alpha Y^*\}$ is π_α, and can be computed using the CDF of the standard normal distribution Φ :

$$\pi_\alpha(\mathbf{k},\mathbf{u}) = \Phi\left(-\frac{m_{\Delta_\alpha}(\mathbf{k},\mathbf{u})}{\sigma_{\Delta_\alpha}(\mathbf{k},\mathbf{u})}\right) \tag{21}$$

Finally, averaging over $\mathbf{u}$ yields

$$\hat{\Gamma}(\mathbf{k}) = \int_\mathbb{U} \pi_\alpha(\mathbf{k},\mathbf{u})p(\mathbf{u})\,\mathrm{d}\mathbf{u} \tag{22}$$

Sources, quantification of uncertainties, and SUR strategy ?

Formally, for a given point $(\mathbf{k},\mathbf{u})$, the event “the point is α-acceptable” has probability $\pi(\mathbf{k},\mathbf{u})$ and variance $\pi(\mathbf{k},\mathbf{u})(1-\pi(\mathbf{k},\mathbf{u}))$. Obviously, the points with the highest uncertainty have the highest variance, so have a coverage probability π around 0.5.