Notes

Victor Trappler

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

7 février 2020

Table des matières

(Joint) Posterior formulation

Priors

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

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 $\sigma^2 : p(\sigma^2)$

GP, RR-based family of estimators

Random processes

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

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

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, in order to obtain the Best Linear Unbiased Estimator :

- — 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) :

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

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 :

$$\mathbb{E}[(\hat{Z}(x) - Z(x))^2] = \mathrm{Cov}\left[[\mathbf{W}^T, -1] \cdot \mathbf{Z}_n(x)\right] \tag{5}$$

$$= [\mathbf{W}^T, -1]\, \mathrm{Cov}\left[\mathbf{Z}_n(x)\right] [\mathbf{W}^T, -1]^T \tag{6}$$

In addition, we have

$$\mathrm{Cov}\left[\mathbf{Z}_n(x)\right] = \begin{bmatrix} \mathrm{Cov}\left[\left[Z(x_1) \dots Z(x_n)\right]^T\right] & \mathrm{Cov}\left[\left[Z(x_1) \dots Z(x_n)\right]^T, Z(x)\right] \\ \mathrm{Cov}\left[\left[Z(x_1) \dots Z(x_n)\right]^T, Z(x)\right]^T & \mathbb{V}\mathrm{ar}\left[Z(x)\right] \end{bmatrix} \tag{7}$$

Once expanded, the kriging weights solve then the following optimisation problem :

$$\min_{\mathbf{W}} \mathbf{W}^T \operatorname{Cov}\left[Z(x_1)\dots Z(x_n)\right]\mathbf{W} \tag{8}$$
$$- \operatorname{Cov}\left[\left[Z(x_1)\dots Z(x_n)\right]^T, Z(x)\right]^T \mathbf{W} \tag{9}$$
$$- \mathbf{W}^T \operatorname{Cov}\left[\left[Z(x_1)\dots Z(x_n)\right]^T, Z(x)\right] \tag{10}$$
$$+ \mathbb{V}\mathrm{ar}\left[Z(x)\right] \tag{11}$$
$$\text{s.t.} \mathbf{W}^T\mathbf{1} = \mathbf{1} \tag{12}$$

This leads to

$$\begin{bmatrix}\mathbf{W}\\ m\end{bmatrix} = \begin{bmatrix}\operatorname{Cov}\left[Z(x_1)\dots Z(x_n)\right] & \mathbf{1}\\ \mathbf{1}^T & 0\end{bmatrix}^{-1}\begin{bmatrix}\operatorname{Cov}\left[\left[Z(x_1)\dots Z(x_n)\right]^T, Z(x)\right]^T\\ 1\end{bmatrix} \tag{13}$$

$$= \begin{bmatrix} C(x_1,x_1) & \cdots & C(x_1,x_n) & 1\\ C(x_2,x_1) & \cdots & C(x_2,x_n) & 1\\ \vdots & \ddots & \vdots & \vdots\\ C(x_n,x_1) & \cdots & C(x_n,x_n) & 1\\ 1 & \cdots & 1 & 0\end{bmatrix}^{-1}\begin{bmatrix}C(x_1,x)\\ C(x_2,x)\\ \vdots\\ C(x_n,x)\\ 1\end{bmatrix} \tag{14}$$

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 $\sigma^2_Y(\mathbf{k},\mathbf{u}) = C((\mathbf{k},\mathbf{u}),(\mathbf{k},\mathbf{u}))$

$$Y(\mathbf{k},\mathbf{u}) \sim \mathcal{N}\left(m_Y(\mathbf{k},\mathbf{u}), \sigma^2_Y(\mathbf{k},\mathbf{u})\right) \tag{15}$$

Let us consider now the conditional minimiser :

$$J^*(\mathbf{u}) = J(\mathbf{k}^*(\mathbf{u}),\mathbf{u}) = \min_{\mathbf{k}\in\mathbb{K}} J(\mathbf{k},\mathbf{u}) \tag{16}$$

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

$$Y^*(\mathbf{u}) \sim \mathcal{N}\left(m^*_Y(\mathbf{u}), \sigma^{2,*}_Y(\mathbf{u})\right) \tag{17}$$

where

$$m^*_Y(\mathbf{u}) = \min_{\mathbf{k}\in\mathbb{K}} m_Y(\mathbf{k},\mathbf{u}) \tag{18}$$

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

$$\begin{bmatrix} Y(\mathbf{k},\mathbf{u}) \\ Y^*(\mathbf{u}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_Y(\mathbf{k},\mathbf{u}) \\ m_Y^*(\mathbf{u}) \end{bmatrix}; \begin{bmatrix} C\left((\mathbf{k},\mathbf{u}),(\mathbf{k},\mathbf{u})\right) & C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) \\ C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) & C\left((\mathbf{k}^*(\mathbf{u}),\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) \end{bmatrix}\right) \tag{19}$$

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

$$\Delta_\alpha(\mathbf{k},\mathbf{u}) \sim \mathcal{N}\left(m_\Delta(\mathbf{k},\mathbf{u}); \sigma^2_\Delta(\mathbf{k},\mathbf{u})\right) \tag{20}$$
$$m_\Delta(\mathbf{k},\mathbf{u}) = m_Y(\mathbf{k},\mathbf{u}) - \alpha m_Y^*(\mathbf{u}) \tag{21}$$
$$\sigma^2_\Delta(\mathbf{k},\mathbf{u}) = \sigma^2_Y(\mathbf{k},\mathbf{u}) + \alpha^2\sigma^2_{Y^*}(\mathbf{k},\mathbf{u}) - 2\alpha C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) \tag{22}$$

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

$$C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) = s\prod_{i\in\mathcal{I}_\mathbf{k}}\rho_{\theta_i}(\|k_i - k_i^*(\mathbf{u})\|)\prod_{j\in\mathcal{I}_\mathbf{u}}\rho_{\theta_j}(0) \tag{23}$$
$$= s\prod_{i\in\mathcal{I}_\mathbf{k}}\rho_{\theta_i}(\|k_i - k_i^*(\mathbf{u})\|) \tag{24}$$

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

$$\Gamma_\alpha(\mathbf{k}) = \mathbb{P}_\mathbf{U}\left[J(\mathbf{k},\mathbf{U}) \leq \alpha J^*(\mathbf{U})\right] \tag{25}$$
$$= \mathbb{E}_\mathbf{U}\left[\mathbb{1}_{J(\mathbf{k},\mathbf{U})\leq\alpha J^*(\mathbf{U})}\right] \tag{26}$$

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

$$\mathbb{1}_{J(\mathbf{k},\mathbf{u})\leq\alpha J^*(\mathbf{u})} \approx \mathbb{1}_{m_Y(\mathbf{k},\mathbf{u})\leq\alpha m_Y^*(\mathbf{u})} \tag{27}$$
$$\mathbb{1}_{J(\mathbf{k},\mathbf{u})\leq\alpha J^*(\mathbf{u})} \approx \mathcal{P}\left[\Delta_\alpha(\mathbf{k},\mathbf{u}) \leq 0\right] = \pi(\mathbf{k},\mathbf{u}) \tag{28}$$

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

$$\hat{\Gamma}_{\alpha,n}(\mathbf{k}) = \mathbb{P}_U\left[m_Y(\mathbf{k},\mathbf{u}) \leq \alpha m_Y^*(\mathbf{u})\right] \qquad \text{(plug-in)}$$
$$\hat{\Gamma}_{\alpha,n}(\mathbf{k}) = \mathbb{E}_U\left[\mathcal{P}\left[\Delta_\alpha(\mathbf{k},\mathbf{u}) \leq 0\right]\right] \qquad \text{(Probabilistic approx)} \tag{29}$$

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 :

$$\pi_\alpha(\mathbf{k},\mathbf{u}) = \Phi\left(-\frac{m_{\Delta_\alpha}(\mathbf{k},\mathbf{u})}{\sigma_{\Delta_\alpha}(\mathbf{k},\mathbf{u})}\right) \tag{30}$$

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

$$\hat{\Gamma}(\mathbf{k}) = \int_{\mathbb{U}} \pi_\alpha(\mathbf{k},\mathbf{u}) p(\mathbf{u})\,\mathrm{d}\mathbf{u} \tag{31}$$

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.

Random sets

Let us start by introducing diverse tools based around Vorob'ev expectation of closed sets (ref thèse Reda), [HST12].

Let us consider A, a random closed set, such that its realizations are subsets of $\mathbb{X}$, and p is its coverage probability, that is

$$p(x) = \mathbb{P}\left[x \in A\right], x \in \mathbb{X} \tag{32}$$

For $\eta \in [0,1]$, we define the η-level set of p,

$$Q_\eta = \{x \in \mathbb{X} \mid p(x) \geq \eta\} \tag{33}$$

It may seem trivial, but let us still note that those sets are decreasing :

$$0 \leq \eta \leq \xi \leq 1 \implies Q_\xi \subseteq Q_\eta \tag{34}$$

Using those level sets for the level $\eta = 0.05$ for instance :

$$Q_{1-\frac{\eta}{2}} \subset Q_{\frac{\eta}{2}} \tag{35}$$

Recalling the objective, it gives upper bounds and lower bounds of the confidence interval of level η on the probability for each $\mathbf{k}$:

$$\hat{\Gamma_\alpha}^U(\mathbf{k}) = \mathbb{P}_{\mathbf{U}}\left[x = (\mathbf{k},\mathbf{u}) \in Q_{1-\frac{\eta}{2}}\right] \tag{36}$$

$$\hat{\Gamma_\alpha}^L(\mathbf{k}) = \mathbb{P}_{\mathbf{U}}\left[x = (\mathbf{k},\mathbf{u}) \in Q_{\frac{\eta}{2}}\right] \tag{37}$$

In [DSB11] is introduced the Margin of uncertainty, defined as the following set difference

$$\mathbb{M}_\eta = Q_{\frac{\eta}{2}} \setminus Q_{1-\frac{\eta}{2}} \tag{38}$$

Considering the

Let μ be a Borel σ-finite measure on $\mathbb{X}$. We define Vorob'ev expectation, as the η^*-level set of A verifying

$$\forall \beta < \eta^* \quad \mu(Q_\beta) \leq \mathbb{E}[\mu(A)] \leq \mu(Q_{\eta^*}) \tag{39}$$

that is the level set of p, that has the volume of the mean of the volume of the random set A.

SUR Strategies

The main idea behind Stepwise Uncertainty Reduction is to define a criterion, say κ_n, that encapsulates the epistemic uncertainty, and to minimize this criterion, in order to select the next point :

$$x^{n+1} = \arg\max_{x\in\mathbb{X}} \kappa_n(x) \tag{40}$$

where κ_n depends on $Y \mid \mathcal{X}_n$ This approach is suitable for step by step evaluations.

Integrated Mean square criterion

[SSW89] Let us consider that we have a kriging model over $\mathbb{X}$ based on a experimental design $\mathcal{X}$, that is denoted $Y \mid \mathcal{X}$

We define the Integrated Mean Square Error (IMSE) as

$$\mathrm{IMSE}(Y \mid \mathcal{X}) = \int_{\mathbb{X}} \sigma^2_{Y|\mathcal{X}}(x)\,\mathrm{d}x \tag{41}$$

where

$$Y \mid \mathcal{X} \sim \mathcal{N}(m_{Y|\mathcal{X}}(x), \sigma^2_{Y|\mathcal{X}}(x)) \tag{42}$$

$$x^{n+1} = \arg\min_{x\in\mathbb{X}} \mathbb{E}_{y\sim Y(x)}\left[\mathrm{IMSE}\left(Y \mid \mathcal{X} \cup \{(x,y)\}\right)\right] \tag{43}$$

So we choose th point minimizing the expected integrated mean square error.

Weighted IMSE

To include a more precise objective than the enrichment of the design, one can add a weight function to the integral, giving the $W-\mathrm{IMSE}$:

$$w-\mathrm{IMSE}(Y \mid \mathcal{X}) = \int_{\mathbb{X}} \sigma^2_{Y|\mathcal{X}}(x) w(x)\,\mathrm{d}x \tag{44}$$

In order to increase the accuracy of the surrogate model around some region of interest, the $w-\mathrm{IMSE}$ can be transformed into

$$w-\mathrm{IMSE}(Y \mid \mathcal{X}) = \int_{\mathbb{X}} \sigma^2_{Y|\mathcal{X}}(x)\mathcal{P}\left[x \in \mathbb{M}_\eta\right]\,\mathrm{d}x \tag{45}$$

where $\mathcal{M}_\eta$ is the η-margin of uncertainty.

UB-LB for $(p, \alpha_p, \mathbf{k}_p)$

Let us assume that we have set a probability $p \in [0, 1]$. Let us recall that the triplet $(p, \alpha_p, \mathbf{k}_p)$ verifies

$$\max_{\mathbf{k}} \Gamma_{\alpha_p}(\mathbf{k}) = \Gamma_{\alpha_p}(\mathbf{k}_p) = \mathbb{P}_{\mathbf{U}}\left[J(\mathbf{k}_p, \mathbf{U}) \leq \alpha_p J^*(\mathbf{U}) \mid \mathbf{U} = \mathbf{u}\right] = p \tag{46}$$

Let us say that $\bar{\Gamma}$ is the η-upper-bound, while $\underline{\Gamma}$ is the η-lower bounds, so

$$\mathcal{P}\left[\underline{\Gamma}(\mathbf{k}) \leq \Gamma_n(\mathbf{k}) \leq \bar{\Gamma}(\mathbf{k})\right] = \eta \tag{47}$$

- If $\underline{\Gamma}(\mathbf{k}) > p$, we are too permissive, so we should decrease α
  - by how much ?
- If $\bar{\Gamma}(\mathbf{k}) < p$, we are too conservative, so we should increase α
  - by how much again ?
- If $\underline{\Gamma}(\mathbf{k}) < p < \bar{\Gamma}(\mathbf{k})$, reduce uncertainty on $\mathbf{k}_p$

Changing the value of α does not require any further evaluation of the objective function, so can be increased until $\max \hat{\Gamma} = p$? by dichotomy for instance. This $\hat{\mathbf{k}}_p$ is then the candidate.

Criterion : stepwise reduction of the variance of the estimation of $\hat{\Gamma}(\hat{\mathbf{k}}_p) = \max_{\mathbf{k}} \hat{\Gamma}(\hat{\mathbf{k}})$
For a fixed $p \in (0, 1]$, and an initial design $\mathcal{X}$. Set an initial value for $\alpha \geq 1$.

- Define Δ_α, using $Y \mid \mathcal{X}$
- Update α such that $\max \hat{\Gamma}_{\alpha,n} = p$
- Compute measure of uncertainty that we want to reduce :
  - $\bar{\Gamma}_{\alpha,n}(\mathbf{k}) - \underline{\Gamma}_{\alpha,n}(\mathbf{k})$
  - $\pi_\alpha(\mathbf{k}, \mathbf{u})(1 - \pi_\alpha(\mathbf{k}, \mathbf{u}))$

Références

[DSB11] V. Dubourg, B. Sudret, and J.-M. Bourinet. Reliability-based design optimization using kriging surrogates and subset simulation. *Structural and Multidisciplinary Optimization*, 44(5) :673–690, November 2011.

[HST12] Philippe Heinrich, Radu S. Stoica, and Viet Chi Tran. Level sets estimation and Vorob'ev expectation of random compact sets. *Spatial Statistics*, 2 :47–61, December 2012.

[SSW89] Jerome Sacks, Susannah B. Schiller, and William J. Welch. Designs for Computer Experiments. *Technometrics*, 31(1) :41–47, February 1989.