Uncertainty propagation and estimation of a quantile

Labopin-Richard Tatiana, Ph-D student

Institut de mathématiques de Toulouse, Université Paul Sabatier, Toulouse.

Problem

- When a pregnant woman phones, her fetus absorbs electromagnetic waves.
  - ↪ How much ?
  - ↪ Is that in higher quantity that the safe limit ?
- This quantity depends on several parameters (position, distance, ...). We are able to measure the rate of electromagnetic waves absorbed when this parameters are fixed. Problem : **One measure is very expensive**.
- We aim to estimate a quantile of 5% of the rate of electromagnetic waves absorption distribution with **only 100 measures**.

Modelisation

We introduce the following model :

$$\mathbf{Y} = \mathbf{f}(\mathbf{X})$$

with $\mathbf{X} \sim \mathcal{U}\left([\mathbf{0},\mathbf{1}]^{\mathbf{d}}\right)$ is the vector of parameters and **f is a very expensive-to-evaluate function**.

Goal : **Find a good estimation of a quantile of 5% of Y with only 100 evaluations of f**.

Mathematic toolbox : Krigeage

Classical hypothese : **f is a realisation of a centered gaussian path with fixed covariance function k**.

Property : krigeage formula

Let be $Y \sim PG(0,k)$. Let $Y_n = (y_1, \dots, y_n)$ be the vector of evaluations in $X_n = (x_1, \dots, x_n)$. Then the law of $Y(x)$ given the σ-field $\mathcal{F}_n := \sigma(X_n, f(X_n))$ for $x \in \mathcal{X}$ follows the gaussian distribution with parameters :

$$m_n(x) = k_n(x)^T K_n^{-1} y_n$$
$$c_n(x,x') = k(x,x') - k_n(x)^T K_n^{-1} k_n(x')$$

with $k_n(x) = [k(x_1,x), \dots, k(x_n,x)]'$ and $K_n = [k(x_i,x_j)]_{1 \le i,j \le n}$. We now denote $s_n^2(x) = c_n(x,x)$.

Estimator : In this model we have considered the two following estimators (denoting $\hat{q}_n^*$ the empirical quantile) : $\hat{\mathbf{q}}^{(\mathbf{n})} = \hat{\mathbf{q}}_{\mathbf{n}}^*(\mathbf{m_n})$ and $\tilde{\mathbf{q}}^{(\mathbf{n})} = \mathbb{E}_{\mathbf{n}}[\hat{\mathbf{q}}_{\mathbf{n}}^*(\mathbf{Y})]$.

Warning : Naive implementation leads us to **inverse a big matrix $\Longrightarrow$ too expensive.**

Active learning : We want strategies in which we choose points to evaluate **one by one.**

Property : 1-step update formula (see [3])

$$m_{n+1}(x) = m_n(x) + \frac{c_n(x_{n+1},x)}{s_n^2(x_{n+1})}(Y(x_{n+1}) - m_n(x_{n+1}))$$
$$s_{n+1}^2(x) = s_n^2(x) - \frac{c_n^2(x_{n+1},x)}{s_n^2(x_{n+1})}$$
$$c_{n+1}(x,y) = c_n(x,y) - \frac{c_n(x_{n+1},x)c_n(x_{n+1},y)}{s_n^2(x_{n+1})}$$

Active learning

We work with the following class of strategies :

1) We have at hand a **grid of $[\mathbf{0},\mathbf{1}]^{\mathbf{d}}$** named $\mathcal{X}$ of size N.

2) Initialization : We evaluate the function f on a **sub-grid $\mathcal{X}_\mathbf{0}$**. We compute the estimator $\hat{q}^{(0)}$ of step 0 and we define the σ-field :

$$\mathcal{F}_\mathbf{0} = \sigma\left(\mathcal{X}_\mathbf{0}, \mathbf{f}(\mathcal{X}_\mathbf{0}), \hat{\mathbf{q}}^{(\mathbf{0})}\right).$$

3) At each step n, **using only $\mathcal{F}_{n-1}$**, we choose the new point x_n where f will be evaluated. We deduce of $f(x_n)$ the new estimator $\hat{q}^{(n)}$ and we update the σ-field :

$$\mathcal{F}_\mathbf{n} = \sigma\left(\mathcal{F}_{\mathbf{n-1}}, \mathbf{x_n}, \mathbf{f}(\mathbf{x_n}), \hat{\mathbf{q}}^{(\mathbf{n})}\right).$$

A first idea : the Stepwise Uncertainty Reduction strategy (see [1])

The idea is to minimize at each step the variance of the estimator. In fact, **at step n, for each point x of the grid, we evaluate what would be the variance of $\hat{q}^{(n)}$ if we had chosen x for x_n**. Unfortunately, we need for this the value of $f(x)$. So we replace it by **the information we have at step n : the law of $f(x)$ given $\mathcal{F}_{n-1}$**. Finally we choose $x_n^* = \min_{x \in \mathcal{X}} V_{n-1}(x)$ with

$$\mathbf{V_{n-1}(x)} = \int \mathbf{Var}(\hat{\mathbf{q}}^{(\mathbf{n})}(\mathbf{y})|\mathcal{F}_{\mathbf{n-1}}^{(\mathbf{x},\mathbf{y})})\phi_{(\mathbf{m_{n-1}(x)},\mathbf{s_{n-1}^2(x)})}(\mathbf{y})\mathbf{dy}$$

where $\phi_{(m_{n-1}(x), s_{n-1}^2(x))}$ is the gaussian density.

Disadvantage : V_n has no analytic form $\Longrightarrow$ **high computational complexity**. In fact the complexity is in N^3.

A new idea from Bayesian optimization

- In [2] and [4], the same class of strategies are used to minimize functions. **At each step n, we choose the point x_n^* for which if $x_n = x_n^*$, we have the tiniest probability to be mistaken, considering that $y_{\min} = y_{\min}^{(n)}$**. Then we choose $x_n^* = \mathrm{argmin}_{x \in \mathcal{X}} \Gamma_{n-1}(x)$ with

$$\mathbf{\Gamma_{n-1}(x)} = \int_{\mathbb{X}} \mathbb{P}_{\mathbf{n-1}}\left(\mathbf{f}(\mathbf{z}) \le \mathbf{y}_{\mathbf{min}}^{(\mathbf{n})}(\mathbf{y})|\mathbf{y} = \mathbf{f}(\mathbf{x})\right)\mathbf{dz}$$

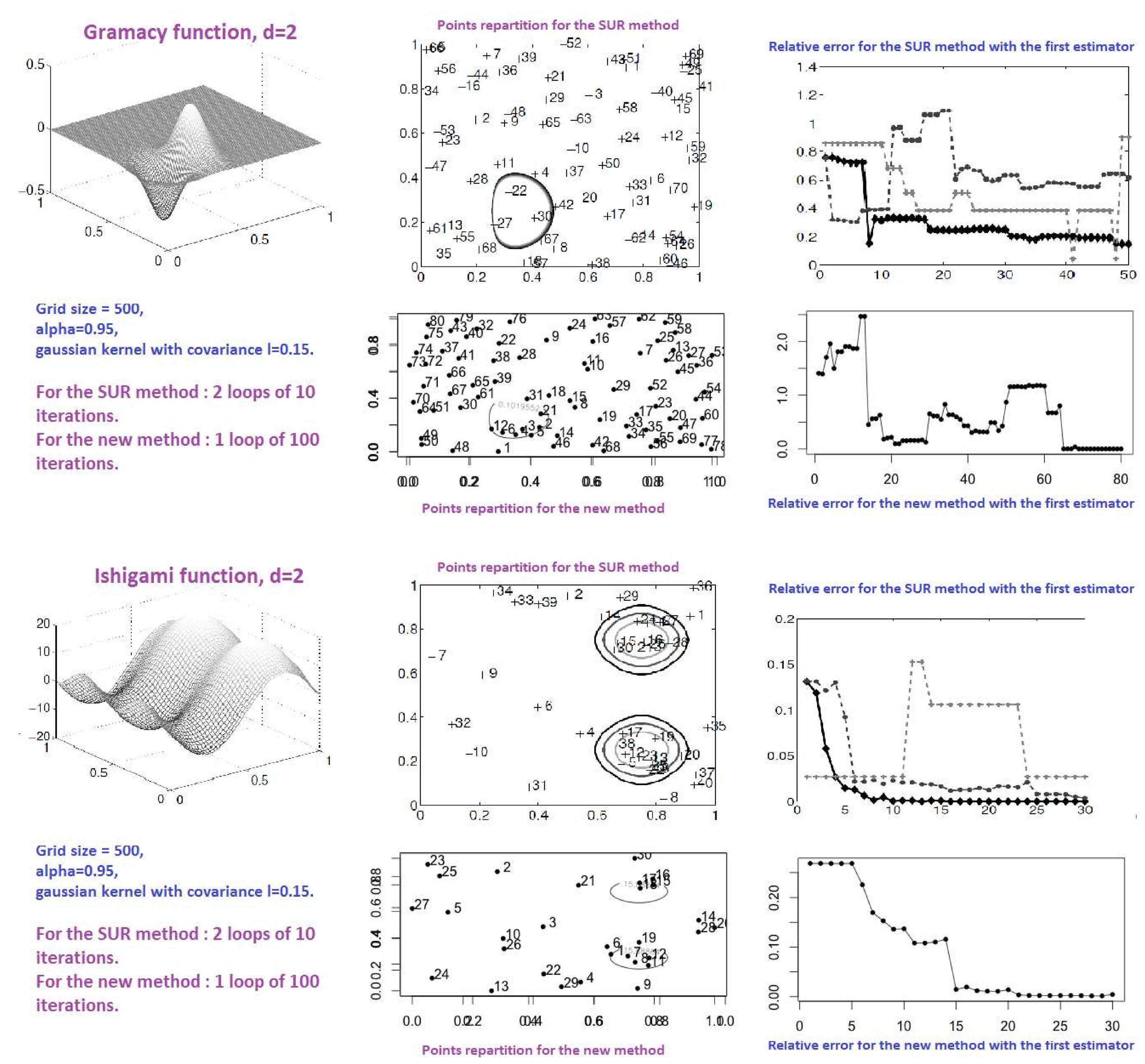
  The advantage is that we **don't need to evaluate $y = f(x)$** to compute the criterium in practice. Contrary to the SUR strategy, we can obtain here an **analytic expression of $V_{n-1}(x)$ depending only on x and $\mathcal{F}_{n-1}$** thanks to the 1-step update formula (and that's why V is indexed by $n-1$ instead of n despite we use it in step n).
- In our case, the probability to be mistaken is not the same, but we can use the same kind of probability. At each step n, we choose the point $x_n^* = \mathrm{argmin}_{x \in \mathcal{X}} \Gamma_{n-1}(x)$ where **Γ_{n-1} measures the mean of the gap in absolute value between $(1-\alpha)$ and the probability that f is superior than $\hat{q}^{(n)}$**.

$$\mathbf{\Gamma_n(x)} = \int_{\mathbb{X}} \left|\mathbb{P}_{\mathbf{n-1}}\left(\mathbf{f}(\mathbf{z}) \ge \mathbf{q}^{(\mathbf{n})}(\mathbf{y})|\mathbf{y} = \mathbf{f}(\mathbf{x})\right) - (\mathbf{1} - \alpha)\right|\mathbf{dz}$$

- In fact this criteria can be written in this way :

$$\mathbf{\Gamma_n(x)} = \int_{\mathbb{X}} \left|\mathbf{\Phi_1^n(x,z)} + \mathbf{\Phi_2^n(x,z)} + \mathbf{\Phi_3^n(z,x)} + \mathbf{\Phi_4^n(z,x)} - (\mathbf{1} - \alpha)\right|\mathbf{dz}$$

  where Φ_i^n are gaussian cumulative distribution functions whose parameters **depends only on x and $\mathcal{F}_{n-1}$ thanks to the 1-step update formula**.
- Advantage : **The complexity is now in N**.



Conclusion and Perspectives

- We found a new method **as precise as** the SUR method in dimension 2.
- But the new method is less expensive :
  On my laptop :
  SUR method : Gramacy $(20+50)$: 5h30, Ishigami $(10+30)$: 4h00.
  New method : Gramacy $(20+50)$ 1h30 (20+80 : 2H15), Ishigami $(10+30)$: 40min.
- Perspective 1 : Hope for **superior dimension (the goal is $d = 5$)** $\longrightarrow$ increase of the grid size and of the number of iterations.
- Perspective 2 : Think about a **more efficient way to find the minimum**.

Bibliography

[1] M. Jala, C. Levy-Leduc, E.Moulines, E.Conil and J. Wiart, Sequential design of computer experiments for parameter estimation withapplication to numerical dosimetry, august 2012.
[2] J. Bect, D. Ginsbourger, L. Li, V. Picheny and E. Vazquez, Sequential design of computer experiments for the estimation of a probability of failure, Springer Science, April 2011.
[3] C. Chevalier, D. Ginsbourger, Corrected kringing update formula for batch-sequential data assimilation, ArXiv : 1203.6452v1, Mars 2012.
[4] V. Picheny, Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction, preprint : http ://arxiv.org/pdf/1310.0732.pdf