Sequential design of experiments for estimating quantiles of black-box functions

Victor Picheny (*INRA*), Tatiana Labopin-Richard (*IMT, Université Toulouse 3*)

CIRM, Lumigny

May 4th, 2018

Outline

Objective: quantile of a "black-box" output

Context: expensive to compute, complex "black-box"

$$X \in \mathbb{R}^d \longrightarrow \boxed{\text{Black-box}} \longrightarrow g$$

- g non-convex, possibly observed in noise
- no derivatives available
- $2 \leq d \leq 10$

Objective: given a distribution on X, estimate the $\alpha-$quantile:

$$q^{\alpha}(g(X)) = q^{\alpha}(Y) = F_Y^{-1}(\alpha)$$

Natural idea: simple Monte-Carlo

$$(X_i)_{i=1,\dots,n} \longrightarrow (Y_i)_{i=1,\dots,n} \longrightarrow \hat{q}_n := Y_{(\lfloor n\alpha \rfloor + 1)}$$

with X_i's taken from the law of X and $Y_{(k)}$ the k-th order statistic

To overcome budget constraints: many possibilities

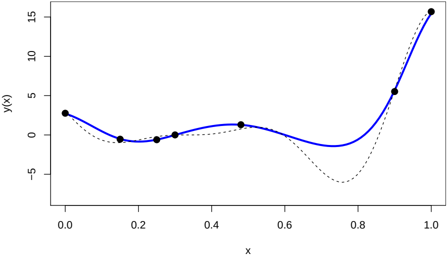
Importance / subset sampling, etc.

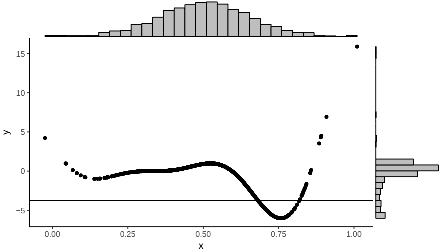
In this talk: DoE + metamodel

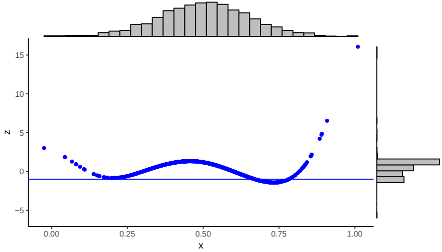
- $\mathcal{A}_n = \{(\mathbf{x}_1, g_1), (\mathbf{x}_2, g_2), \dots (\mathbf{x}_n, g_n)\}$
- Metamodel built using $\mathcal{A}_n$
- Quantile estimated using the metamodel

$\mathbf{x}_1, \dots, \mathbf{x}_n$ may <u>not</u> follow $P(X)$

A reasonable error in the metamodel can result in a large error in the quantile...







Outline

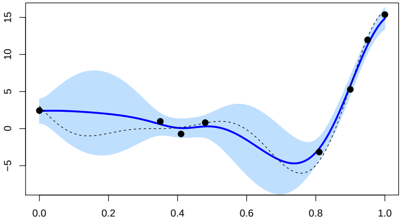
GP models

Kriging model: $G \sim \mathcal{GP}\left(m(.), k(.,.)\right)$ conditioned on $\mathcal{A}_n$

$$\begin{aligned} m_n(\mathbf{x}) &= \mathbb{E}(G(\mathbf{x})|\mathcal{A}_n) \\ &= c(\mathbf{X}_n, \mathbf{x})^T c(\mathbf{X}_n, \mathbf{X}_n)^{-1}\mathbf{g_n}, \end{aligned}$$

$$\begin{aligned} c_n(\mathbf{x}, \mathbf{x}') &= \mathrm{Cov}\left(G(\mathbf{x}), G(\mathbf{x}')|\mathcal{A}_n\right) \\ &= c(\mathbf{x}, \mathbf{x}') - c(\mathbf{X}_n, \mathbf{x})^T c(\mathbf{X}_n, \mathbf{X}_n)^{-1} c(\mathbf{X}_n, \mathbf{x}'), \end{aligned}$$

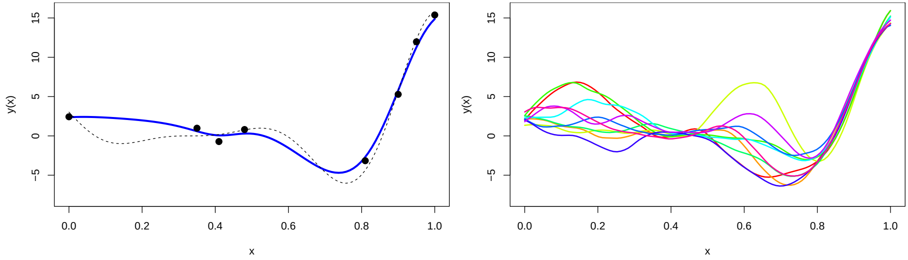
where $c(\mathbf{X}_n, \mathbf{x}) = [c(\mathbf{x}_1, \mathbf{x}), \dots, c(\mathbf{x}_n, \mathbf{x})]^T$, $c(\mathbf{X}_n, \mathbf{X}_n) = [c(\mathbf{x}_i, \mathbf{x}_j)]_{1 \le i,j \le n}$ and $\mathbf{g_n} = [g_1, \dots, g_n]$.



Two natural estimators for the quantile

$$\hat{q}_n^{(1)} = q_X\left(\mathbb{E}_G\left[G(X)|\mathcal{A}_n\right]\right) = q_X(m_n(X)),$$
$$\hat{q}_n^{(2)} = \mathbb{E}_G\left(q_X(G(X))|\mathcal{A}_n\right).$$

with q_X quantile w.r.t. the measure on X.



Except on very specific cases of c and $P(X)$: no analytical formula

$\mathbb{E}_G\left(q_X(G(X))|\mathcal{A}_n\right)$ theoretically attractive and robust but...

Double-loop Monte-Carlo: $G + X$: limits applicability

Jala, Lévy-Leduc, Moulines, Conil, Wiart (2016), Sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields, *Technometrics*

Our choice: $q_X(m_n(X)) \Rightarrow$ Monte-Carlo on X only

$$q_n = m_n(\mathbf{X}_{\text{MC}})_{(\lfloor l\alpha \rfloor + 1)}.$$

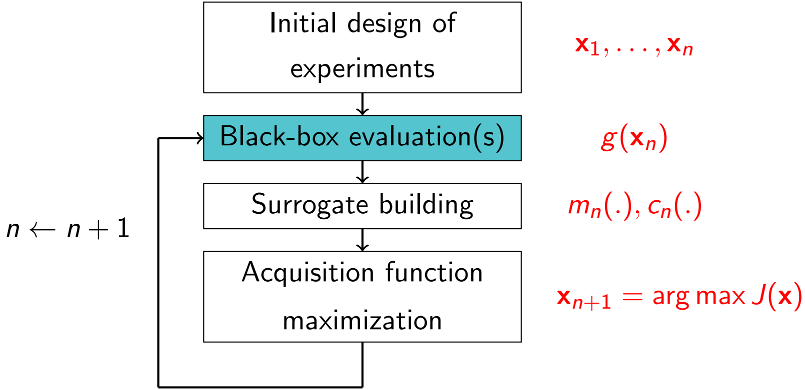
with $\mathbf{X}_{\text{MC}} = (\mathbf{x}^1_{\text{MC}}, \ldots, \mathbf{x}^l_{\text{MC}}) \sim X$.

Oakley (2004), Estimating percentiles of uncertain computer code outputs, *JRSSc*

Obviously: q_n is biased

- Generally: m_n smoother than g
- We need sequential design to make it accurate

Sequential design in a nutshell



Measuring information gain

There are shortcuts to condition GPs on new observations

$$m_{n+1}(\mathbf{x}) = \mathbb{E}\left[G(\mathbf{x}) | \{\mathcal{A}_n \cup (\mathbf{x}_{n+1}, g_{n+1})\}\right]$$

$$m_{n+1}(\mathbf{x}) =$$

$$\left[c(\mathbf{X}_n, \mathbf{x})^T c(\mathbf{x}_{n+1}, \mathbf{x})\right] \begin{bmatrix} C(\mathbf{X}_n, \mathbf{X}_n) & c(\mathbf{x}_{n+1}, \mathbf{x}) \\ c(\mathbf{x}_{n+1}, \mathbf{x})^T & c(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}_n \\ g_{n+1} \end{bmatrix}$$

After simplification:

$$m_{n+1}(\mathbf{x}) = m_n(\mathbf{x}) + \frac{c_n(\mathbf{x}_{n+1}, \mathbf{x})}{c_n(\mathbf{x}_{n+1}, \mathbf{x}_{n+1})} \left(g_{n+1} - m_n(\mathbf{x}_{n+1})\right)$$

$\Rightarrow$ The new GP mean is linear in g_{n+1}.

Conditioning on "yet-to-evaluate" observations

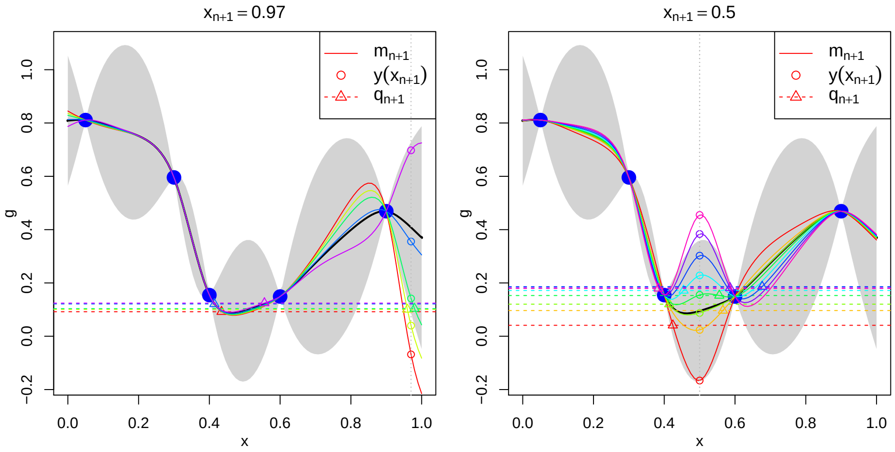
$$\begin{aligned} M_{n+1}(\mathbf{x}) &= \mathbb{E}\left[G(\mathbf{x}) \middle| \{\mathcal{A}_n \cup (\mathbf{x}_{n+1}, G_{n+1})\}\right] \\ &= m_n(\mathbf{x}) + \frac{c_n(\mathbf{x}_{n+1}, \mathbf{x})}{c_n(\mathbf{x}_{n+1}, \mathbf{x}_{n+1})} \left(G_{n+1} - m_n(\mathbf{x}_{n+1})\right) \end{aligned}$$

$\Rightarrow$ the GP mean is a random process once we have chosen $\mathbf{x}_{n+1}$ but not evaluated $g(\mathbf{x}_{n+1})$.

We can study the impact on the quantile estimator!

$q_X(M_{n+1})$ is random but entirely depends on $\{\mathbf{x}_{n+1}, G_{n+1}\}$.

Illustration: 15% quantile



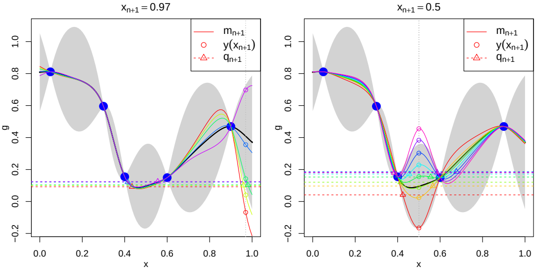
Sequential choice for $\mathbf{x}_{n+1}$

What's reasonable for G_{n+1}?

$$G_{n+1} \sim \mathcal{N}\left(m_n(\mathbf{x}_{n+1}), c_n(\mathbf{x}_{n+1}, \mathbf{x}_{n+1})\right)$$

Most influential observation = maximizer of the variance of the estimator

$$\arg \max_{\mathbf{x}_{n+1} \in \mathbb{R}^d} \mathrm{Var}_{G_{n+1}} \left[q_X(M_{n+1}) \right]$$



Outline

An update formula for the quantile

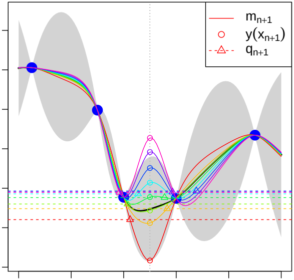
We start with

$$q_n = m_n(\mathbf{X}_{\mathrm{MC}})_{(\lfloor l\alpha \rfloor+1)} := m_n(\mathbf{x}_n^q)$$

$\mathbf{x}_n^q$= "quantile point".
Future quantile estimator is random:

$$Q_{n+1} = M_{n+1}(\mathbf{X}_{\mathrm{MC}})_{(\lfloor l\alpha \rfloor+1)} = M_{n+1}(\mathbf{x}_{n+1}^q)$$

Any $M_{n+1}(\mathbf{x})$ is Gaussian, but $\mathbf{x}_{n+1}^q$ (triangles) may change depending on G_{n+1}



An update formula for the quantile

Recall

$$M_{n+1}(\mathbf{x}) = m_n(\mathbf{x}) + \frac{c_n(\mathbf{x}_{n+1}, \mathbf{x})}{c_n(\mathbf{x}_{n+1}, \mathbf{x}_{n+1})} \left(G_{n+1} - m_n(\mathbf{x}_{n+1}) \right)$$

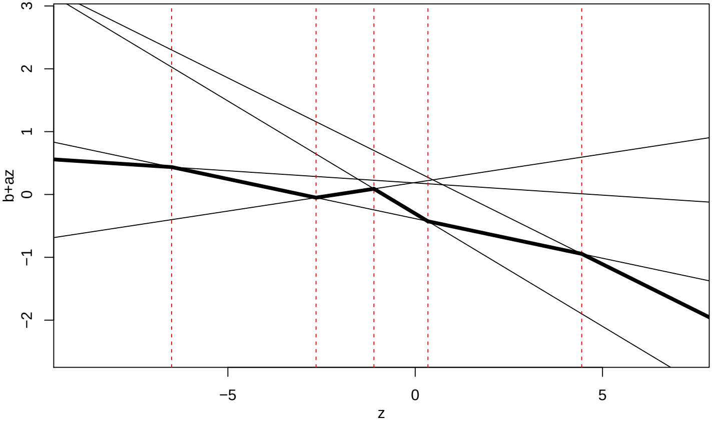
Observe now that:

$$\forall \mathbf{x}_i \in \mathbf{X}_{\mathrm{MC}}, \quad m_{n+1}(\mathbf{x}_i) = a_i + b_i z$$

- All values depend linearly on $z = G_{n+1}$
- The $(\lfloor l\alpha \rfloor + 1)$ smallest value of $m_{n+1}(\mathbf{X}_{\mathrm{MC}})$ is driven by z:

$$\begin{array}{ccc} \mathbb{R} & \rightarrow & \mathbf{X}_{\mathrm{MC}} \\ z & \rightarrow & \mathbf{x}^q_{n+1}(z) \end{array}$$

It amounts to considering a set of straight lines and looking for the $(\lfloor l\alpha \rfloor + 1)$ lowest.



- $\mathbf{x}_{n+1}^{q}(z)$ is constant over intervals of z
- j^i index of the quantile point when $z \in B_i$
  Here: $\mathbf{j} = \{2, 3, 1, 4, 3, 5\}$ (for $B_1, \ldots, B_6$)
- Fast algorithms can retrieve all j^i's and B_i's

Using the decomposition $B_1, \dots, B_p$:

The future quantile is: $Q_{n+1} = a_{j^i} + b_{j^i} Z \quad \text{if } Z \in B_i$

To compute $\mathrm{Var}(Q_{n+1})$: law of total variance

$$Var(U) = \sum_{i=1}^{p} \mathrm{Var}(U \mid E_i)\mathbb{P}(E_i) + \sum_{i=1}^{p} \mathbb{E}(U \mid E_i)^2 (1 - \mathbb{P}(E_i))\mathbb{P}(E_i)$$
$$- 2 \sum_{i=2}^{p} \sum_{j=1}^{i-1} \mathbb{E}(U \mid E_i)\mathbb{P}(E_i)\mathbb{E}(U \mid E_j)\mathbb{P}(E_j) \, .$$

Here: $(U \mid E_i) = a_{j^i} + b_{j^i} Z$ and $E_i = Z \in B_i$.

Three types of quantities

- $\mathbb{P}(I_i \leq Z \leq I_{i+1})$,
- $\mathrm{Var}\left(Z' \mid I_i < Z' < I_{i+1}\right)$,
- $\mathbb{E}\left(Z' \mid I_i < Z' < I_{i+1}\right)$.

$\Rightarrow$ Tallis formula (moments of a truncaded Gaussian).

Conditionally on $\mathcal{A}_n$ and on the choice of $\mathbf{x}_{n+1}$ (provided $s_n(\mathbf{x}_{n+1}) \neq 0$)

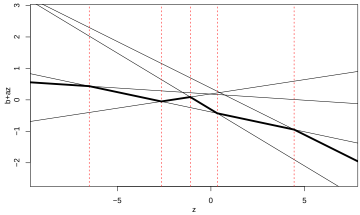
$$
\begin{aligned}
J_n^{\mathrm{Var}}(\mathbf{x}_{n+1}) = & \sum_{i=1}^{L} \left[c_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1})\right]^2 V(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) P_i \\
& + \sum_{i=1}^{L} \left[m_n(\mathbf{x}_{n+1}^q(B_i) - c_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i)\right]^2 (1 - P_i)\, P_i \\
& - 2 \sum_{i=2}^{L} \sum_{j=1}^{i-1} \left[m_n(\mathbf{x}_{n+1}^q(B_i) - c_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i)\right] P_i \\
& \qquad \left[m_n(\mathbf{x}_{n+1}^q(B_i) - c_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), I_{j+1}, I_j)\right] P_j
\end{aligned}
$$

where:

- $B_i = [I_i, I_{i+1}]$,
- $P_i = \Phi(s_n(\mathbf{x}_{n+1}) I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1}) I_i)$,
- $E(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) = \frac{1}{s_n(\mathbf{x}_{n+1})} \left(\frac{\phi(s_n(\mathbf{x}_{n+1}) I_{i+1}) - \phi(s_n(\mathbf{x}_{n+1}) I_i)}{\Phi(s_n(\mathbf{x}_{n+1}) I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1}) I_i)} \right)$, and
- $V(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) = \frac{1}{s_n(\mathbf{x}_{n+1})^2} \left[1 + \frac{s_n(\mathbf{x}_{n+1}) \phi(I_{i+1}) - s_n(\mathbf{x}_{n+1}) \phi(I_i)}{\Phi(I_{i+1}) - \Phi(I_i)} - \left(\frac{\phi(I_{i+1}) - \phi(I_i)}{\Phi(I_{i+1}) - \Phi(I_i)} \right)^2 \right]$.

Summary of this part

- We choose a large $\mathbf{X}_{\mathrm{MC}}$
- Given a potential $\mathbf{x}_{n+1}$:
  - We compute a_j and b_j ($\forall 1 \leq j \leq n_{MC}$) using GP equations
  - We decompose the variation of $G_{n+1} = Z$ in $\{B_1, \dots, B_p\}$ (using an appropriate algorithm)

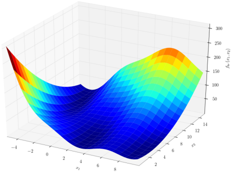


  - We compute $\mathrm{Var}\left(Q_{n+1}\right)$ using the analytical formula.
- $\mathbf{x}_{n+1} = \arg\max_{\mathbb{X}} \mathrm{Var}\left(Q_{n+1}\right)$ (using a global optimizer).

Outline

Numerical setup

2D problem - "Branin" function



$X_1, X_2 \sim \mathcal{U}[0,1]$

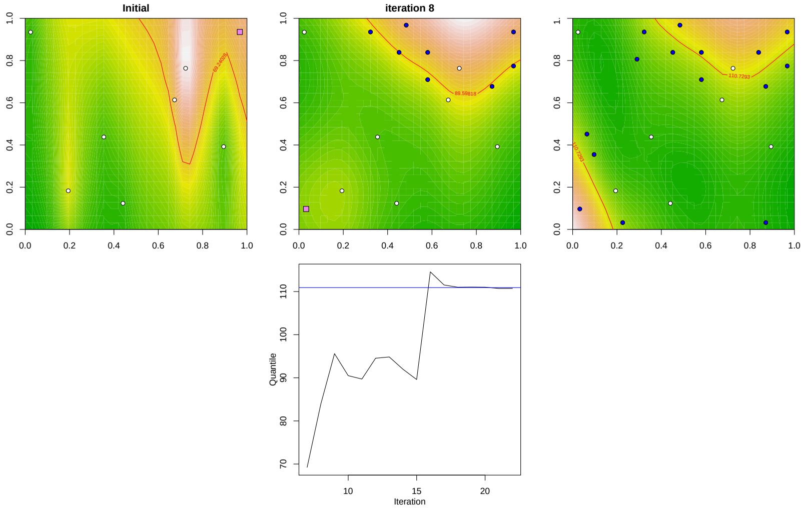
4D problem - "Hartman" function

$$g(\mathbf{x}) = \sum_{i=1}^{4} C_i \exp\left(-\sum_{j=1}^{4} a_{ji}\left(x_j - p_{ji}\right)^2\right), \ \mathbf{X} \sim \mathcal{N}\left(\frac{1}{2}, \Sigma\right)$$

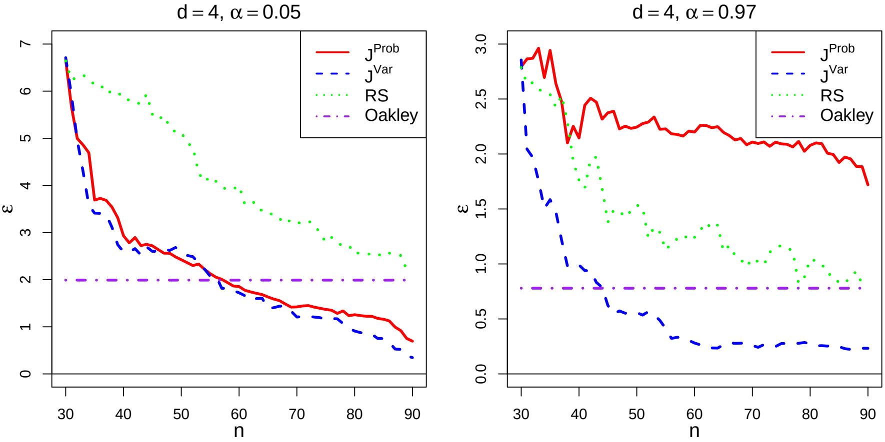
Comparison

- Random search (!): sequential space-filling design
- Oakley two-step approach (same flavour, but not sequential)

2D problem, 7 + 15 observations, 85% quantile



4D problem - 30 + 60 observations



Concluding comments

Today's trick

- Quantile estimator based on GP *mean*
- Choice of the design that maximizes the estimator variation

... opposite of Julien's talk?

Further steps

- Alternative metamodels?
- Alternative objectives? (optimization: *correlated knowledge gradient*)

Want to know more?

T. Labopin-Richard, V. Picheny (2018), Sequential design of experiments for estimating quantiles of black-box functions, *Statistica Sinica*