Airbus Group Innovations

Quantification des incertitudes - Partie II

Saclay, February 2014

Fabien MANGEANT, fabien.mangeant@eads.net

February 12, 2014

EADS

Outline

1. **Probability and statistics**
   - Introduction
2. **Link with computer experiments**
   - Uncertainty management
3. **Computer experiments**
   - Introduction
   - Our motivation in terms of computer experiment
4. **Reduced model techniques**
   - Introduction
   - Polynomial chaos expansion
5. **Sensitivity analysis**
6. **Formalization of the problem in a statistical framework**
   - Link with statistical learning
   - Contrast and risk function
   - Mono feature estimation by a single model approach
7. **Conclusion and perspectives**
8. **Bibliography**

Outline

1 **Probability and statistics**
- Introduction

2 **Link with computer experiments**
- Uncertainty management

3 **Computer experiments**
- Introduction
- Our motivation in terms of computer experiment

4 **Reduced model techniques**
- Introduction
- Polynomial chaos expansion

5 **Sensitivity analysis**

6 **Formalization of the problem in a statistical framework**
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 **Conclusion and perspectives**

8 **Bibliography**

Probability

Interpretation of a probability

As a matter of fact, the core mathematical concept which is used is the concept of probability. To simplify our purpose, we consider two main interpretations of a probability:

- **Frequentist/objective**: the probability is the asymptotic measure of the following ratio:

$$\mathbb{P}(A) = \frac{\text{Number of observations of the Event A}}{\text{Total number of observations}}$$

  For instance, if the modelling and simulation are accurate, it is possible to compare a sample of simulated lifetime of a modelled device with a sample of real lifetimes of the device.

- **Bayesian/subjective**: the probability is considered as a measure of risk, it is not directly linked to the frequency of an event but to the level of certainty/uncertainty expressed on the phenomenon.

It is important to note that, even if the interpretation between the two propositions is fundamentally different, the mathematical representation of both is identical.

Outline

1 Probability and statistics
- Introduction

2 **Link with computer experiments**
- Uncertainty management

3 Computer experiments
- Introduction
- Our motivation in terms of computer experiment

4 Reduced model techniques
- Introduction
- Polynomial chaos expansion

5 Sensitivity analysis

6 Formalization of the problem in a statistical framework
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 Conclusion and perspectives

8 Bibliography

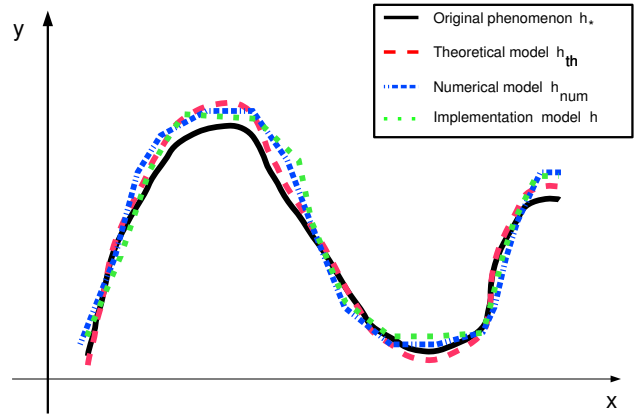
EADS

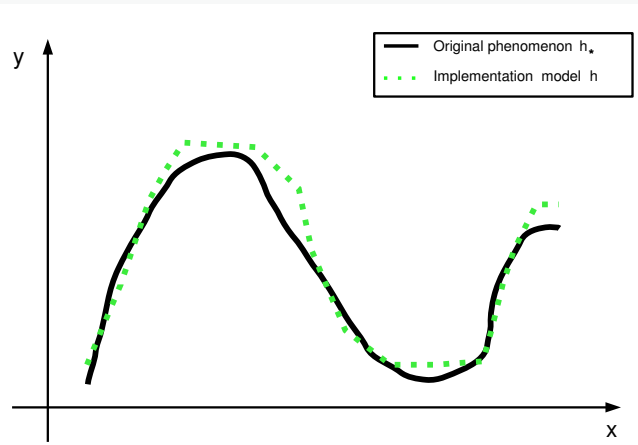
What is the nature of uncertainty in this context?

"Model" uncertainty

- **Reference model** h^*: Usually not accessible, expression of a natural or a complex technical object.
- **Theoretical model** h_{th}: Scientific expert activity (modelling activity, theoretical solution of a PDE system, ...), corresponding to the level of understanding and simplification of the problem.
- **Numerical model** h_{num}: Numerical solution of the theoretical model (effects of meshing, choice of a numerical scheme, truncature effects, ...)
- **Implementation model** h: Software implementation of the model on a given hardware architecture (computer accuracy, choice of coding rules, ...)

$$h^* \rightsquigarrow h$$





What is the nature of uncertainty in this context?

Parametric input uncertainty

- For a given numerical model $h : (\mathbf{x}, \theta) \in \mathcal{X} \times \Theta \mapsto \mathbf{y} = h(\mathbf{x}, \theta) \in \mathcal{Y}$, we consider **an uncertainty attached to the input variables X modelled by a statistical law** $\mathbb{P}^*_{\mathbf{X}}$.
- In practical contexts, it is often difficult to build $\mathbb{P}^*_{\mathbf{X}}$ due to scarsity of data, heterogeneous database, lack of information on the dependency, ... As a matter of fact, one has to work with an **approximate statistical law** $\mathbb{P}_{\mathbf{X}}$.

$$\mathbb{P}^*_{\mathbf{X}} \rightsquigarrow \mathbb{P}_{\mathbf{X}}$$

Computational budget $\mathcal{B}$

- In many situations, it is difficult to compute analytically the risk measures $\rho(h(\mathbf{X}, \theta))$. Numerical methods $\mathcal{M}(\mathcal{B}, \varepsilon, h(\mathbf{X}, \theta)$ (either stochastic or not) are required using a fixed computational budget $\mathcal{B}$ for a given accuracy ε

$$\rho\big(h(\mathbf{X}, \theta)\big) \rightsquigarrow \mathcal{M}\big(h(\mathbf{X}, \theta), \mathcal{B}, \varepsilon\big)$$

How to manage all the components of the error?

Recap of the errors

1. Building of the model: $\mathcal{N}_{\mathcal{S}}(h^*, h_{th})$
2. Numerical approximation: $\mathcal{N}_{\mathcal{N}}(h_{th}, h_{num})$
3. Hardware/Software implementation: $\mathcal{N}_{\mathcal{I}}(h_{num}, h)$
4. Model paramaters uncertainty: $\mathcal{N}_{\mathcal{Q}}(\mathbb{P}^*_{\mathbf{X}}, \mathbb{P}_{\mathbf{X}})$
5. Uncertainty propagation error: $\mathcal{N}_{\mathcal{P}}(\rho(h(\mathbf{X}, \theta)), \mathcal{M}(h(\mathbf{X}, \theta), \mathcal{B}, \varepsilon))$

Naive form of the total error

$$\Delta \leq \underbrace{\mathcal{N}_{\mathcal{S}}(h^*, h_{th})}_{\textit{Scientific Validation}} + \underbrace{\mathcal{N}_{\mathcal{N}}(h_{th}, h_{num})}_{\textit{Numerical Validation}} + \underbrace{\mathcal{N}_{\mathcal{I}}(\hat{h}, h)}_{\textit{Hardware/Software Validation}} + \underbrace{\mathcal{N}_{\mathcal{Q}}(\mathbb{P}^{\mathbf{X}}_*, \mathbb{P}^{\mathbf{X}})}_{\textit{Statistical Validation}} + \underbrace{\mathcal{N}_{\mathcal{P}}(\rho(Y), \hat{\rho}_{\mathcal{B}}(Y))}_{\textit{Propagation Validation}}$$

EADS

How to manage some components of the error?

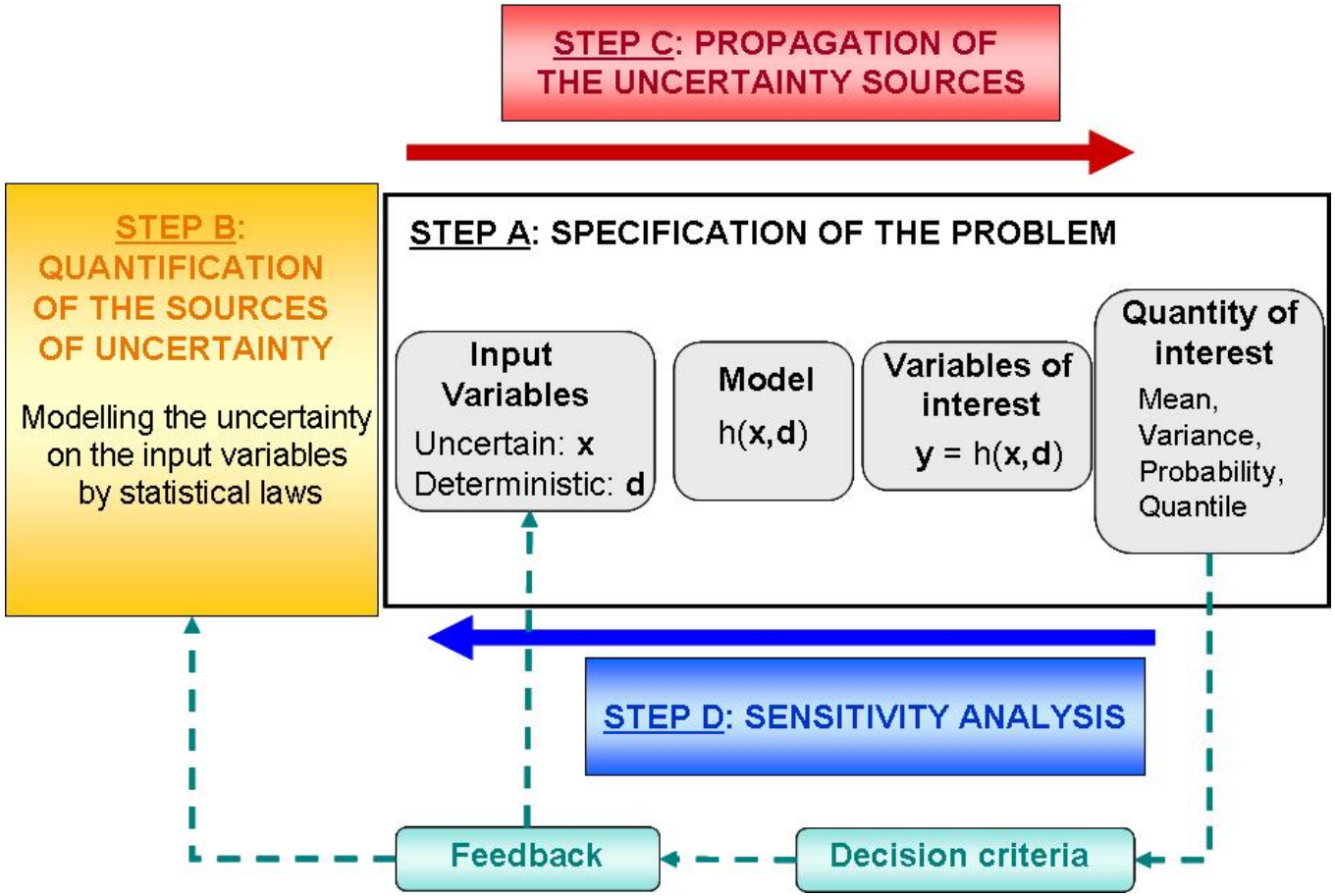


Figure : Four steps engineering process for uncertainty study (ref EsREDA [9], Open TURNS [23])

EADS

Outline

1 Probability and statistics
- Introduction

2 Link with computer experiments
- Uncertainty management

3 Computer experiments
- Introduction
- Our motivation in terms of computer experiment

4 Reduced model techniques
- Introduction
- Polynomial chaos expansion

5 Sensitivity analysis

6 Formalization of the problem in a statistical framework
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 Conclusion and perspectives

8 Bibliography

EADS

Computer experiments

Scope of computer experiments in statistics

The literature in computer experiments classically uses the computer code h as a predefined object (Owen [24], Sack [27], Santner [29]). Several objectives are classically assigned to computer experiment:

- Optimization
- Sensitivity analysis
- Approximation of some measures of risk
- Visualization
- Calibration

Computer experiments

Scope of computer experiments in numerical analysis

h is the solution of a PDE. Some coefficients of the PDE are unknown for some reasons. The methods and tools that are developed in this framework take advantage of the structure of the problem (Ghanem and Spanos [6], Ladeveze [15], LeMaitre [16], Lucor [17], Maday [18]). The same type of goals is assigned to the computer experiments:

- Optimization
- Sensitivity analysis
- Approximation of some measures of risk

EADS

Our motivation in terms of computer experiment

Goal

Let $\mathbb{Q}$ be the unknown probability measure associated to the real random variable $\mathbf{Y}^*$ defined over $(\mathbb{R}^Q, \mathcal{B}(R^Q), \mathbb{Q})$. Our goal is to predict some feature $\rho(\mathbb{Q}) \in \mathbb{F}$ of the distribution $\mathbb{Q}$ (also abusively noted $\rho(\mathbf{Y}^*)$). This feature corresponds to the measure of risk over our variable of interest $\mathbf{Y}^*$.

Examples of probabilistic measures of risk $\rho(\mathbf{Y}^*)$

Mean:	$\rho_\mu(\mathbf{Y}^*) = \mathbb{E}[\mathbf{Y}^*]$	$\in \mathbb{F} = \mathbb{R}$
Variance:	$\rho_\sigma(\mathbf{Y}^*) = \mathrm{Var}[\mathbf{Y}^*]$	$\in \mathbb{F} = \mathbb{R}_+$
Quantile:	$\rho_q(\mathbf{Y}^*) = q_r(\mathbf{Y}^*)$	$\in \mathbb{F} = \mathbb{R}_+$
Probability:	$\rho_p(\mathbf{Y}^*) = \mathbb{P}(\mathbf{Y}^* \in \mathcal{D}_P)$	$\in \mathbb{F} = [0,1]$
CDF:	$\rho_{cdf}(\mathbf{Y}^*) = \mathbb{P}(\mathbf{Y}^* \leq \mathbf{y}^*)$	$\in \mathbb{F} = \mathcal{F}_{cdf}(\mathbb{R}^Q, [0,1])$
PDF:	$\rho_{pdf}(\mathbf{Y}^*) = f_{\mathbf{Y}^*}(\mathbf{y}^*)$	$\in \mathbb{F} = \mathcal{F}_{pdf}(\mathbb{R}^Q, \mathbb{R}_+)$

Our motivation in terms of computer experiment

Properties of a numerical model h

- **Dimension**: h is classically a real function belonging to $\mathcal{F}(\mathbb{R}^P \times \mathbb{R}^T, \mathbb{R}^Q)$. Even if the dimension of $\mathbf{x}$ can be large, most of the engineering problems we are focused on only contain $P \leq 50$ and $Q \leq 5$.
- **Computational budget**: A single computation of h can be very expensive. The computational budget $\mathcal{B}$ will be represented by the number m of runs affordable to solve the problem.
- **Black box/white box**: h is either a black box (*the inner operations of the model are not accessible*), a grey box (*part of the inner operations is accessible*) or a white box (*all the operations of the model are accessible*).
- **Mathematical properties**: the basic mathematical properties (regularity, monotony, linearity or non linearity towards certain parameters) may be unknown to the engineer.
- **Domain of validity**: h should be delivered with its domain of validity $\mathcal{V}^{[\epsilon]} \subseteq \mathbb{R}^P \times \mathbb{R}^T$.

Our motivations in terms of computer experiment

Prediction of the feature $\rho(\mathbb{Q}) = \rho(\mathbf{Y}^*)$ thanks to a pre defined model $h(\mathbf{x}, \theta) = \mathbf{y}$, a statistical law $\mathbb{P}_{\mathbf{X}}$ and a numerical method $\mathcal{M}(m, \varepsilon)$

- The probability measure $\mathbb{Q}$ being unknown, it is approximated by the composition of a model h, defined over $\mathcal{X} \times \Theta$ and a statistical law $\mathbb{P}_{\mathbf{X}}$. Thus, it is possible to approximate the feature $\rho(\mathbf{Y}^*)$ by $\rho(h(\mathbf{X}, \theta))$.

$$\rho(\mathbf{Y}^*) \approx \rho(h(\mathbf{X}, \theta))$$

- As it is quite rare to compute exactly $\rho(h(\mathbf{X}, \theta))$, it is approximated by either a deterministic or a stochastic numerical method $\mathcal{M}$. $\mathcal{M}$ is characterized by its accuracy ε for a given budget of computations $\mathcal{B}$.

$$\mathcal{M}(h(\mathbf{X}, \theta), \mathcal{B}, \varepsilon) \xrightarrow{\mathcal{B} \to \infty, \varepsilon \to 0} \rho(h(\mathbf{X}, \theta))$$

- Moreover, alternative strategies using surrogate models $\tilde{h}$ are often used to obtain a better estimation of the feature $\rho(\mathbf{Y}^*)$. The final approximation with numerical method $\mathcal{M}'$ should be more accurate in a certain sense $\|\|$ than the previous one:

$$\|\mathcal{M}'\left(h(\mathbf{X}, \theta), \tilde{h}(\mathbf{X}, \eta), \mathcal{B}, \varepsilon\right) - \rho(h(\mathbf{X}, \theta)\| \leq \|\mathcal{M}(h(\mathbf{X}, \theta), \mathcal{B}, \varepsilon) - \rho(h(\mathbf{X}, \theta)\|$$

EADS

Our motivations in terms of computer experiment

Sensitivity analysis to the choice of predefined model $h(\mathbf{x}, \theta) = \mathbf{y}$ and the statistical law $\mathbb{P}_{\mathbf{X}}$ on the prediction of the feature $\rho(\mathbb{Q}) = \rho(\mathbf{Y}^*)$

- The probability measure $\mathbb{Q}$ being unknown, it is approximated by the composition of a model h, defined over $\mathcal{X} \times \Theta$ and a statistical law $\mathbb{P}_{\mathbf{X}}$. Thus, it is possible to approximate the feature $\rho(\mathbf{Y}^*)$ by $\rho(h(\mathbf{X}, \theta))$.

$$\rho(\mathbf{Y}^*) \approx \rho(h(\mathbf{X}, \theta))$$

- Influence of the group of input variables $\mathbf{X}^K$ $(K \subseteq \{1, \cdots, P\})$ on the feature of interest $\rho(h(\mathbf{X}, \theta))$:

$$\rho(h(\mathbf{X}, \theta) | \mathbf{X}^K = \mathbf{x}^K) = \rho(h(\mathbf{X}, \theta)) \; ??$$

- Influence of the statistical model $\mathbb{P}_{\mathbf{X}}$ on the feature of interest $\rho(h(\mathbf{X}, \theta))$

$$\rho(h(\mathbb{P}^1_{\mathbf{X}}, \theta)) = \rho(h(\mathbb{P}^2_{\mathbf{X}}, \theta)) \; ??$$

- Influence of the choice of model h_i among the panoply of model $\mathcal{H} = \{h_1, \cdots, h_D\}$

$$\rho(h_i(\mathbf{X}^{(i)}, \theta^{(i)})) = \rho(h_i(\mathbf{X}^{(i)}, \theta^{(i)})) \; ??$$

EADS

Outline

1 Probability and statistics
- Introduction

2 Link with computer experiments
- Uncertainty management

3 Computer experiments
- Introduction
- Our motivation in terms of computer experiment

4 Reduced model techniques
- Introduction
- Polynomial chaos expansion

5 Sensitivity analysis

6 Formalization of the problem in a statistical framework
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 Conclusion and perspectives

8 Bibliography

Strategy to decrease the computational cost

Surrogate model

Let h be the numerical model defined over $\mathcal{X} \times \Theta$ with values in $\mathcal{Y}$. $\tilde{h}$ is a surrogate model, belonging to $\mathcal{F}(\mathcal{X} \times \Theta, \mathcal{Y})$, with the following properties:

- $\tilde{h}$ is "close" from h in a domain of interest V : $\forall \mathbf{x} \in V, \|h(\mathbf{x}, \theta) - \tilde{h}(\mathbf{x})\| \leq \epsilon$ with a certain norm $\|.\|$ and a criterion ϵ.
- The computational cost of $\tilde{h}$ is much cheaper than the one of h (either in memory or CPU cost).
- Notation: $\tilde{h}(\mathbf{X}, \theta) = \tilde{\mathbf{Y}}$

Examples

Polynomial chaos, Kriging models, Radial based functions, Reduced basis, Neural networks, SVM, Taylor expansions, ...

What is a polynomial chaos expansion ?

Spectral approach

- The random vector $\mathbf{Y} = h(\mathbf{X}, \theta)$ is considered as an element of a functional space $\mathcal{F}$ with condition $\mathbb{E}\left[||\mathbf{Y}||^2\right] < \infty$.
- The goal is to build a basis of this functional space, $\mathbf{Y}$ will be represented by its coordinates in this basis.

Meaning in practice

- In other terms, the model $h(\mathbf{X}, \theta)$ is replaced by a decomposition in an adequate basis.
- As a first step, only non intrusive methods are considered.
- In a practical way, it means that the CPU cost will be paid in two steps:
  1. A cost to build the truncated polynomial chaos expansion $\tilde{h}$ (by a numerical and approximation technique).
  2. A cost to estimate the feature of interest $\rho(\mathbf{Y})$ with the help of the surrogate model $\tilde{h}$. With $M >> N$:

$$\hat{\rho}_M(\tilde{h}(\mathbf{X}, \theta)) \rightarrow \hat{\rho}_N(h(\mathbf{X}, \theta)) \rightarrow \rho(h(\mathbf{X}, \theta))(= \rho(\mathbf{Y}))$$

EADS

General approach

Representation of the model

$$\mathbf{Y} = h(\mathbf{X}, \theta) \approx \sum_{j \in \mathbb{N}} \tilde{y}_j \Psi_j(\mathbf{X})$$

- $\mathcal{B} = (\Psi_j(\mathbf{X}))_{j \in \mathbb{N}}$: basis of the polynomial chaos, which is a basis of the space $\mathcal{L}^2(\mathbb{P}_{\mathbf{X}}, \mathbb{R}^P, \mathbb{R})$
- $\tilde{y}_j \in \mathbb{R}^Q$: coordinates in the basis $\mathcal{B}$

Two algorithmic steps for implementation

1. Building of the polynomial chaos basis $\mathcal{B} = (\Psi_1, ...\Psi_R)$
2. Computation of the coefficients $\tilde{y}_j$ in the basis

EADS

Building of the PC basis

The basis is adapted to the input distribution $\mathbb{P}_X$ of the input random vector $\mathbf{X}$.

- In case of dependency, develop an iso probabilistic transformation to obtain a transformed distribution with independent components.
- When the components are independent,

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{P} f_{X_i}(x_i)$$

  - Building orthogonal polynoms towards the measure $f_{X_i}(x_i)$:

$$< \psi_k^i, \psi_l^i > \equiv \int \psi_k^i(x)\psi_l^i(x) f_{X_i}(x) dx = \delta_{kl}$$

  - Tensorial product of unidimensional polynoms:

$$\alpha \equiv \{\alpha_1, \cdots, \alpha_P\} \quad \Psi_\alpha(x) = \prod_{i=1}^{P} \psi_{\alpha_i}^i(x_i)$$

EADS

Computation of the coefficients $\tilde{y}_i$

Methods by projection

- Representation: $\mathbf{Y} = \sum_{\alpha \in \mathbb{N}^P} \tilde{y}_\alpha \Psi_\alpha(\mathbf{X})$
- $\tilde{y}_\alpha = \mathbb{E}\left[\mathbf{Y}.\Psi_\alpha(\mathbf{X})\right] = \int_{\mathcal{X}} h(\mathbf{x}, \theta)\Psi_\alpha(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})$
- Computation by numerical integration (quadrature formula) or Monte Carlo methods

Methods by regression

- Regression of the model on a truncated basis:

$$\mathbf{Y} = \sum_{j=O}^{R-1} \tilde{y}_j \Psi_j(\mathbf{x}) + \epsilon_R$$

- Classical least square minimization to obtain the coefficients for example or more elaborated regression techniques including penalization over dimension to increase the sparsity of the representation (Ridge regression, LASSO, elastic net, ...).

EADS

Outline

1 Probability and statistics
- Introduction

2 Link with computer experiments
- Uncertainty management

3 Computer experiments
- Introduction
- Our motivation in terms of computer experiment

4 Reduced model techniques
- Introduction
- Polynomial chaos expansion

5 Sensitivity analysis

6 Formalization of the problem in a statistical framework
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 Conclusion and perspectives

8 Bibliography

Sensitivity analysis

Context

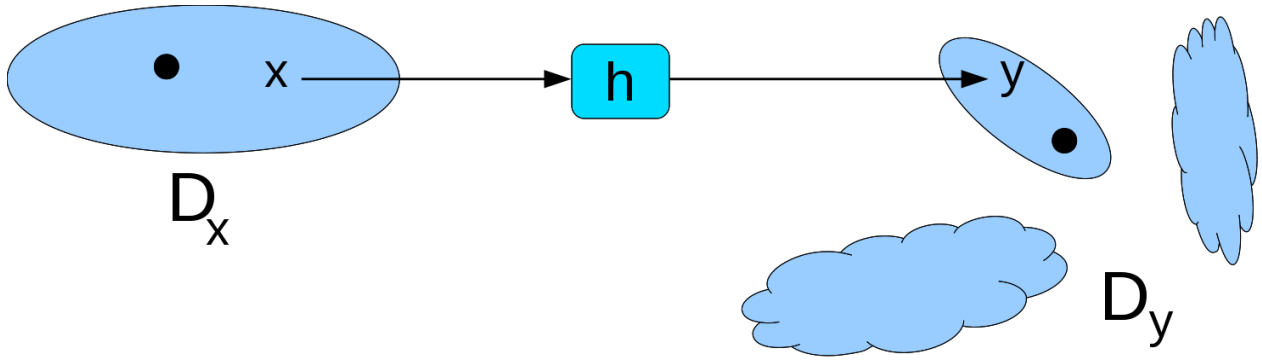
In this part, we suppose that the model $h(\mathbf{x}, \theta)$ and the statistical law $\mathbb{P}_{\mathbf{X}}$ are known. Thus, it is possible to define a random variable $\mathbf{Y} = h(\mathbf{X}, \theta)$ defined by its statistical law $\mathbb{P}_{\mathbf{Y}}$. See references Antoniadis [1], Saltelli [28], Sobol [30].

Objectives of sensitivity analysis

- **To determine the most influential input variables from X** that contribute to the uncertainty measured by the measure of risk $\rho(\mathbf{Y})$. *The uncertainties over these variables could be reduced or the model adapted to reduce the uncertainty.*
- **To determine the less influential input variables from X** that contribute to the uncertainty measured by the measure of risk $\rho(\mathbf{Y})$. *These variables could be considered deterministic for future studies.*
- **To analyse the interactions** between some input variables or some groups of input variables.
- **To check the fidelity of the model** h towards the problem studied. *An other model h could be developed.*

EADS

Several sensitivity analysis techniques



- **LOCAL** : $\bullet \rightarrow \bullet$
  The local sensivity techniques are based on the local behaviour of the model, classically linearized around a point of interest (*differentiation by direct or adjoint, ...*).
- **GLOBAL** : $D_X \rightarrow D_{\mathbf{Y}}$
  The global sensitivity techniques aim at exploring the full domain of variation D_x without any *a priori* assumption on the model.

Several sensitivity analysis techniques

Local sensitivity analysis

Local sensitivity analysis enabled to obtain good results for linearized systems. Efficient methods are developed from the differential analysis (adjoint methods, ...):

$$\mathbf{y} = h(\mathbf{x}, \theta) \Rightarrow \delta\mathbf{y} \simeq \frac{\partial h}{\partial \mathbf{x}} \delta\mathbf{x} = S\delta\mathbf{x} \Rightarrow \mathrm{Var}\left[\mathbf{Y}\right] = S.\mathrm{Var}\left[\mathbf{X}\right].S^T$$

Global sensitivity analysis

Global sensitivity analysis aims at exploring all possible variations of the parameters, without *a priori* simplification of the model.

- **Sobol indexes**:
  - First-order indexes:

$$S_i = \frac{\mathrm{Var}\left[\mathbb{E}\left[\mathbf{Y}|X_i\right]\right]}{\mathrm{Var}\left[\mathbf{Y}\right]}$$

  - Higher order indexes

$$S_{i_1,\cdots,i_k} = \frac{\mathrm{Var}\left[\mathbb{E}\left[\mathbf{Y}|X_{i_1},\cdots,X_{i_k}\right]\right]}{\mathrm{Var}\left[\mathbf{Y}\right]}$$

EADS

Outline

1 Probability and statistics
- Introduction

2 Link with computer experiments
- Uncertainty management

3 Computer experiments
- Introduction
- Our motivation in terms of computer experiment

4 Reduced model techniques
- Introduction
- Polynomial chaos expansion

5 Sensitivity analysis

6 Formalization of the problem in a statistical framework
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 Conclusion and perspectives

8 Bibliography

Link with statistical learning

Classical learning areas (see Hastie et al [12], Massart [20])

- **Unsupervised learning**: We observe $\mathbf{X}_1^*, ..., \mathbf{X}_n^*$ i.i.d $\mathbb{P}_{\mathbf{X}}^*$ (unknown) and we look for a map $g : \mathcal{X}^* \to \mathcal{Y}^*$
- **Semi-supervised learning** With $l < n$, we observe $(\mathbf{X}_i^*, \mathbf{Y}_i^*)_{i \leq l}$ + $\mathbf{X}_{l+1}^*, ..., \mathbf{X}_n^*$ and we look for a map $g : \mathcal{X}^* \to \mathcal{Y}^*$
- **Supervised/inductive learning**: We observe $(\mathbf{X}_1^*, \mathbf{Y}_1^*), ..., (\mathbf{X}_n^*, \mathbf{Y}_n^*)$ and we look for a map $g : \mathcal{X}^* \to \mathcal{Y}^*$

Our learning context

- If the $\mathbf{X}_i^*$'s are observed ?
  Data at disposal:
  $(\mathbf{X}_1^*, \mathbf{Y}_1^*), ..., (\mathbf{X}_n^*, \mathbf{Y}_n^*) + (\mathbf{X}_1, h(\mathbf{X}_1, \theta)), ..., (\mathbf{X}_m, h(\mathbf{X}_m, \theta)), \quad m >> n$
  The framework $\mathbf{Y}_1^*, ..., \mathbf{Y}_n^* + \mathbf{X}_1, ..., \mathbf{X}_m$ *may be seen between Supervised and Semi-supervised learning...*
- If the $\mathbf{X}_i^*$'s are NOT observed ?
  Data at disposal: $\mathbf{Y}_1^*, ..., \mathbf{Y}_n^* + h(\mathbf{X}_1, \theta), ..., h(\mathbf{X}_m, \theta), \quad m >> n$
  The framework $\mathbf{Y}_1^*, ..., \mathbf{Y}_n^* + \mathbf{X}_1, ..., \mathbf{X}_m$ *may be seen between Unsupervised and Semi-supervised learning...*

Definitions

Contrast

Definition: A contrast function is defined by:

$$\begin{aligned} \Psi : \quad \mathbb{F} \times \mathcal{Y} &\longrightarrow \mathbb{R} \\ (\rho, y) &\mapsto \Psi(\rho, y) \end{aligned}$$

Examples

- $\mathbb{F} = \mathbb{R}$:
  - **Mean-contrast**: $\Psi(\rho, y) = (y - \rho)^2$
- $\mathbb{F} = \{\text{Set of density function}\}$:
  - **Log-contrast**: $\Psi(\rho, y) = -\log(\rho(y))$
  - **L_2-contrast**: $\Psi(\rho, y) = \|\rho\|_2^2 - 2\rho(y)$

EADS

Definitions

Risk function

Definition: Given $(\Psi, \mathbb{F}, \mathbb{Q})$, the risk function $\mathcal{R}_\Psi$ is a real function defined as:

$$\forall \rho \in \mathbb{F}, \quad \mathcal{R}_\Psi(\rho) := \int_{\mathcal{Y}} \Psi(\rho, v)\, \mathbb{Q}(dv) = \mathbb{E}_{V \sim \mathbb{Q}}\left[\Psi(\rho, V)\right]$$

Application to our problem

- $\rho = \rho_h(\theta)$
- $\mathcal{R}_\Psi(h, \theta) = \mathbb{E}_{\mathbf{Y}^* \sim \mathbb{Q}}\left[\Psi(\rho_h(\theta), \mathbf{Y}^*)\right]$
- Some classical risk functions:
  - The **mean-squared** contrast gives a distance between means: $\mathcal{R}_\Psi(h, \theta) = \left(\mathbb{E}\left[\mathbf{Y}^*\right] - \rho_h(\theta)\right)^2 + \mathrm{Var}\left[\mathbf{Y}^*\right]$
  - The **log-contrast** gives the Kullbach-Leiber divergence between pdfs: $R_\Psi(h, \theta) = KL(f_{\mathbf{Y}^*}, \rho_h(\theta)) - \mathbb{E}\left[\log(\mathbf{Y}^*)\right]$, where $KL(g_1, g_2) = \int \log(\frac{g_1}{g_2}(y)\, g_1(y)\, dy$

EADS

Pb 1: Mono feature estimation by a single model approach

Mathematical goal

Let $\mathbb{Q}$ be the unknown probability measure associated to the real random variable $\mathbf{Y}^*$ defined over $(\mathbb{R}^Q, \mathcal{B}(R^Q), \mathbb{Q})$. Our main goal is to predict one feature $\rho(\mathbb{Q})$ of the distribution $\mathbb{Q}$.

General description of the statistical problem

We want to develop robust estimation procedures of the feature ρ based upon the availability of a reference database $(\mathbf{Y}_1^*, \cdots, \mathbf{Y}_n^*)$, a numerical model $h(\mathbf{x}, \theta)$, with $\mathbf{X}$ following $\mathbb{P}_{\mathbf{X}}$ and a computational budget $\mathcal{B}$ that can be spent either m times all at once or in an adptative way.

Pb1: Example of density prediction

Suppose that $(\mathbf{X}_1^*, \mathbf{Y}_1^*), ..., (\mathbf{X}_n^*, \mathbf{Y}_n^*)$ are available.

- **Calibration of θ by mean-Squares minimization**

$$\widehat{\theta}_{MS} = \underset{\theta \in \Theta}{\text{Argmin}} \frac{1}{n} \sum_{i=1}^{n} (Y_i^* - h(\mathbf{X}_i^*, \theta))^2$$

Pb1: Example of density prediction

Suppose that $(\mathbf{X}_1^*, \mathbf{Y}_1^*), ..., (\mathbf{X}_n^*, \mathbf{Y}_n^*)$ are available.

- **Calibration of θ by mean-Squares minimization**

$$\widehat{\theta}_{MS} = \underset{\theta \in \Theta}{\text{Argmin}} \frac{1}{n} \sum_{i=1}^{n} (Y_i^* - h(\mathbf{X}_i^*, \theta))^2$$

- **Prediction of ρ**
  Compute the probability density of $h(\mathbf{X}, \widehat{\theta}_{MS})$ under $\mathbf{X} \sim \mathbb{P}_{\mathbf{X}}$

$$\boxed{\rightarrow \widehat{f}_{MS}}$$

Pb1: Example of density prediction

Other M-estimators...

- **Kullback-Leibler minimization** $KL(f_1, f_2) = \int_{\mathcal{Y}} \log(\frac{f_1}{f_2})\, f_1$
  - $f_{\mathbf{Y}^*}$ = density of $\mathbf{Y}^*$, f_θ = density of $h(\mathbf{X}, \theta)$
  - **Goal**: Find θ that minimizes $KL(f, f_\theta)$.
- **Two difficulties**
  - f is unknown $\rightarrow$ replaced by $f^n = \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{Y}_i}$
  - f_θ untractable $\rightarrow$ replaced by a **simulation density** (Kernel, projection, etc...) $\left(f_\theta^m = \frac{1}{m}\sum_{j=1}^m K_{b_m}(\cdot - h(\mathbf{X}_j, \theta)), \quad \mathbf{X}_j \underset{i.i.d}{\sim} \mathbb{P}_{\mathbf{X}} \right)$
- **M-estimator**

$$\widehat{\theta}_{KL} = \underset{\theta\in\Theta}{\text{Argmin}}\, KL(f^n, f_\theta^m) = \underset{\theta\in\Theta}{\text{Argmin}} -\frac{1}{n}\sum_{i=1}^n \log(f_\theta^m)(\mathbf{Y}_i^*)$$

- **Prediction**
  Compute the probability density of $h(\mathbf{X}, \widehat{\theta}_{KL})$ under $\mathbf{X} \sim \mathbb{P}_{\mathbf{X}}$

$$\boxed{\rightarrow \widehat{f}_{KL}}$$

EADS

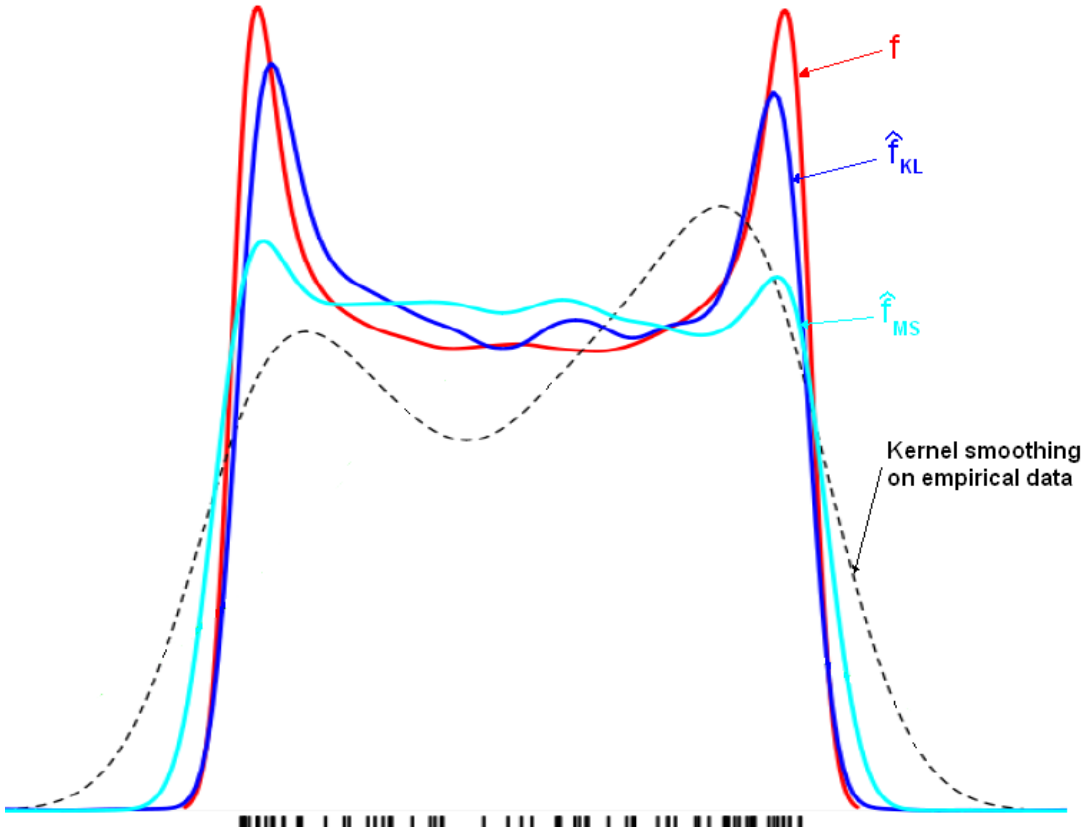
Question ?

What is the "best" estimator of f ,

$\widehat{f}_{MS}$ or $\widehat{f}_{KL}$?

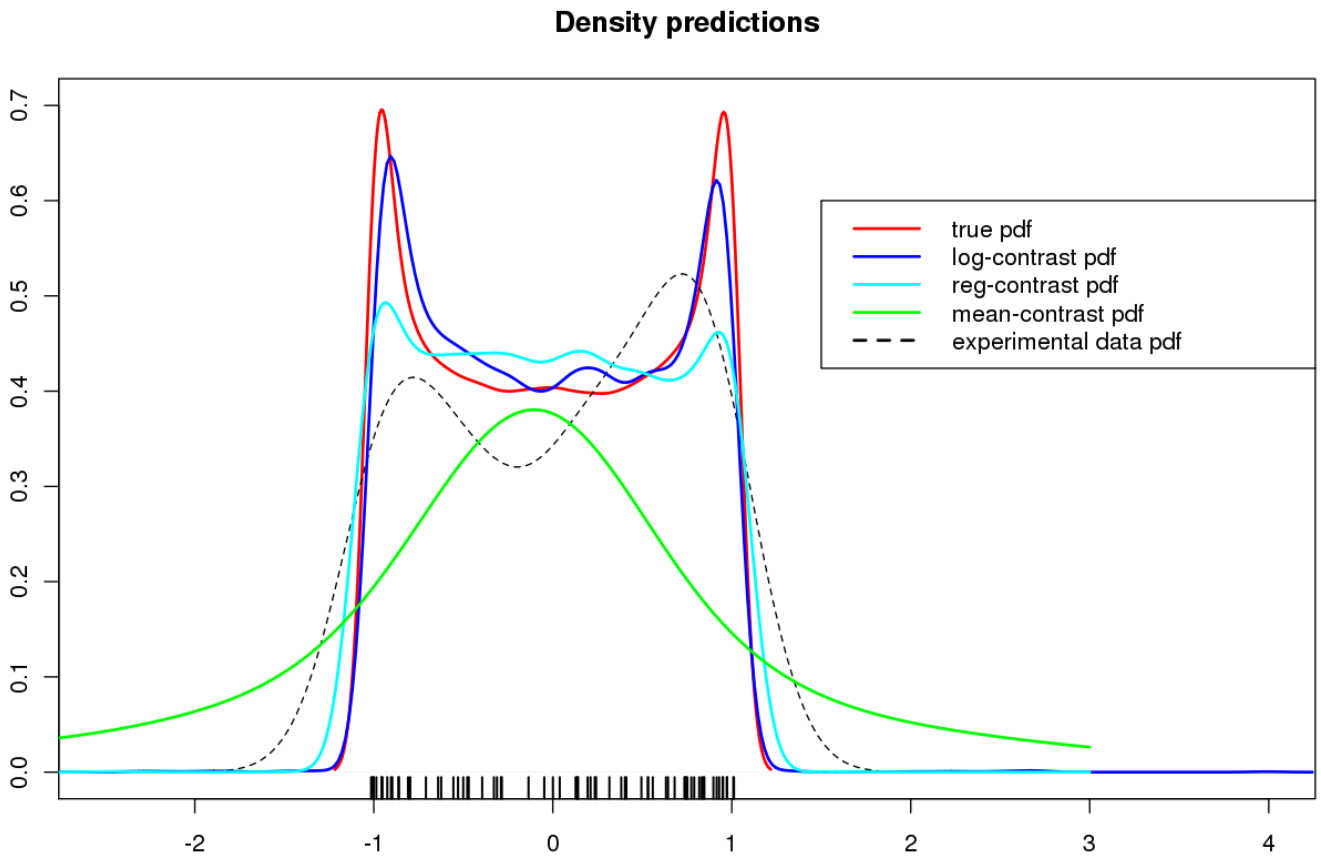
Pb 1: Toy application

- $\mathbf{Y}^* = \sin(\mathbf{X}^*) + 0.01\,\varepsilon, \quad \mathbf{X}^* \perp \varepsilon \sim \mathcal{N}(0,1)$
- $h(\mathbf{X}, \theta) = \theta_1 + \theta_2\, X + \theta_3\, X^3,\ X \sim \mathbb{P}^{\mathbf{x}} = \mathcal{N}(0,1)$
- $n = 50$ and $m = 10^3$



Pb 1: Toy application

- $\mathbf{Y}^* = \sin(\mathbf{X}^*) + 0.01\,\varepsilon, \quad \mathbf{X}^* \perp \varepsilon \sim \mathcal{N}(0,1)$
- $h(\mathbf{X}, \theta) = \theta_1 + \theta_2\, X + \theta_3\, X^3,\ X \sim \mathbb{P}^{\mathbf{x}} = \mathcal{N}(0,1)$
- $n = 50$ and $m = 10^4$



Pb1: Theoretical results from N. Rachdi PhD Thesis [26]

Theorem: Oracle Inequality (Rachdi et al [25])

Under some conditions on the contrast Ψ and under tightness conditions, for all $\varepsilon > 0$, with high probability it holds:

$$0 \leq \mathcal{R}_\Psi(h, \widehat{\theta}) - \inf_{\theta \in \Theta} (\mathcal{R}_\Psi(h, \theta)) \leq \frac{K^\epsilon_{(\widetilde{\rho},\Psi)}}{\sqrt{n}} \left(1 + \sqrt{\frac{n}{m}} \left(K^\epsilon_{(\widetilde{\rho},h)} + B_m\right)\right)$$

where $K^\epsilon_{(\widetilde{\rho},\Psi)}$, $K^\epsilon_{(\widetilde{\rho},h)}$ some *concentration constants* and B_m a bias factor

Pb1: Theoretical results from N. Rachdi PhD Thesis [26]

Theorem: Oracle Inequality (Rachdi et al [25])

Under some conditions on the contrast Ψ and under tightness conditions, for all $\varepsilon > 0$, with high probability it holds:

$$0 \leq \mathcal{R}_\Psi(h, \widehat{\theta}) - \inf_{\theta \in \Theta} \left(\mathcal{R}_\Psi(h, \theta)\right) \leq \frac{K^\epsilon_{(\widetilde{\rho}, \Psi)}}{\sqrt{n}} \left(1 + \sqrt{\frac{n}{m}} \left(K^\epsilon_{(\widetilde{\rho}, h)} + B_m\right)\right)$$

where $K^\epsilon_{(\widetilde{\rho}, \Psi)}$, $K^\epsilon_{(\widetilde{\rho}, h)}$ some *concentration constants* and B_m a bias factor

- Nonasymptotic result, i.e valid for all $n, m \geq 1$
- $\inf_{\theta \in \Theta} \left(\mathcal{R}_\Psi(h, \theta)\right)$ = the minimal *risk* we can achieve for Ψ
  = Modeling error (mesh size ..., model complexity)
- $\frac{K^\epsilon_{(\widetilde{\rho}, \Psi)}}{\sqrt{n}} \left(1 + \sqrt{\frac{n}{m}}(K^\epsilon_{(\widetilde{\rho}, h)} + B_m)\right)$ = Statistical error linked to model complexity and size of the databases

EADS

Pb1: Theoretical results from Rachdi et al [26]

- **Compare** $\mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi^P})$ **and** $\mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi})$

$$\text{study the difference} \quad \mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi^P}) - \mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi})$$

- **By definition of** θ_{Ψ^P}**:** $\mathcal{R}_{\Psi^P}(\theta_{\Psi^P}) - \mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi}) \leq 0$ for all $\widehat{\theta}_{\Psi}$
- **Question :** $\mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi^P}) - \mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi}) \leq$ 0? *a.s? w.h.p?, in* L_1*? $\cdots$ difficult in general*?

Proposition: [Mean squares for mean prediction] (N. Rachdi, JC. Fort 2010)

- **Feature of interest:** $\rho^P = \mathbb{E}(Y)$ --→ Ψ^P : $(\rho, y) \mapsto (\rho - y)^2$
- **Model:** $h(\mathbf{X}, \theta) = \Phi(\mathbf{X}) \cdot \theta, \quad \Phi = (\phi_1, ..., \phi_k)$ orho. w.r.t $P_{\mathbf{X}}$
- **Suppose:** $Y_i = \Phi(\mathbf{X}_i) \cdot \theta^* + \varepsilon_i, \quad \mathbb{E}(\varepsilon_i) = 0$ i.i.d
- **Let 2** Ψ**-estimators:** $\widehat{\theta}_{\Psi^P} = \text{Argmin}_{\theta \in \Theta} \sum_{i=1}^{n} (Y_i - \mathbb{E}\Phi(\mathbf{X}) \cdot \theta)^2$ and $\widehat{\theta}_{\Psi_{reg}} = \text{Argmin}_{\theta \in \Theta} \sum_{i=1}^{n} (Y_i - \Phi(\mathbf{X}_i) \cdot \theta)^2$
- **Result:**

$$\mathbb{E}_{(\mathbf{X}_i, Y_i)_{1..n}} \left(\mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi^P}) - \mathcal{R}_{\Psi^P}(\widehat{\theta}_{\Psi}) \right) \leq 0$$

EADS

Outline

1 Probability and statistics
- Introduction

2 Link with computer experiments
- Uncertainty management

3 Computer experiments
- Introduction
- Our motivation in terms of computer experiment

4 Reduced model techniques
- Introduction
- Polynomial chaos expansion

5 Sensitivity analysis

6 Formalization of the problem in a statistical framework
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 Conclusion and perspectives

8 Bibliography

Some CHALLENGES

CULTURAL challenges

Engineers ARE NOT USED to express the uncertainty in their domain. By the way, only a few of them are trained on the subject !

- Problem to build the probabilistic criteria
- Quantification of the sources of uncertainty

A strong effort is required in basic training and professional training.

TECHNOLOGICAL challenges

The simulation tools are not adapted to evolve towards this revolution !

- Automatization of the computational workflow
- Is the computational budget compatible with the probabilistic criterion? Development of high performance computations capabilities.

CERTIFICATION challenges

The uncertainty management process has to be compatible with certification issues (legal responsability, safety issues, ...)

Advertisement

CHORUS workshops in Paris: Reduction techniques, statistical learning

Journées GDR PMASCOT NUM in March 2014:

Outline

1 Probability and statistics
- Introduction

2 Link with computer experiments
- Uncertainty management

3 Computer experiments
- Introduction
- Our motivation in terms of computer experiment

4 Reduced model techniques
- Introduction
- Polynomial chaos expansion

5 Sensitivity analysis

6 Formalization of the problem in a statistical framework
- Link with statistical learning
- Contrast and risk function
- Mono feature estimation by a single model approach

7 Conclusion and perspectives

8 Bibliography

EADS

- A. Antoniadis, Analyse de sensibilite, Seminaire de Toulouse 2006, www.gdr-mascotnum.fr
- Apostolakis G & Wu JS. (1993), Reliability and decision-making, Chapman & Hall London.
- Aven T. (2003), Foundations of Risk Analysis, John Wiley & Sons, Ltd.1
- Bedford T. & Cooke R. (2001), Probabilistic Risk Analysis - Foundations and Methods, Cambridge University Press.2
- R. Cont, R. Deguest, G. Scandolo (2010), Robustness and sensitivity analysis of risk measurement procedures, Quantitative Finance, Vol 10, No 6 - p593-606.
- Roger G. Ghanem and Pol D. Spanos (1991), Stochastic finite elements: A Spectral Approach, publisher Springer-Verlag - ISBN 0-387-97456-3 or 0-540-97456-3

- De Neufville R. (2003), Architecting/Designing engineering Systems using Real Options, MIT report ESD-WP-2003-01.09 (http://esd.mit.edu).
- De Neufville R. (2004), Uncertainty management for engineering systems planning and design, MIT Engineering
- EsREDA group (2008), Uncertainty in industrial practice: a guide to quantitative uncertainty management, Wiley
- Ghanem R, Spanos PD. (1991), Stochastic Finite Elements: A Spectral Approach. Springer: New York.
- Granger Morgan M. & Henrion M. (1990), Uncertainty - A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press.3
- T. Hastie, R. Tibshirani, J. Friedman (2008) - The elements of statistical learning - Springer Series in Statistics.

- Kleijnen JPC (2009), Design and Analysis of Simulation Experiments
- Knight F.H. (1921), Risk, Uncertainty and Profit, Hart, Schaffner & Marx. Systems Monograph (http://esd.mit.edu/symposium/pdfs/monograph/uncertainty.pdf).
- P. Ladeveze (1999). Non linear Computational Structural Mechanics â New Approaches and Non-Incremental Methods of Calculation, Springer-Verlag, New York.
- Le Maitre O.P. and Knio O.M. (2010), Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics. Scientific Computation.
- Lucor D. (2006), Presentation at CEMRACS'06
- Maday Y., Reduced Basis Method for the Rapid and Reliable Solution of Partial Differential Equations, hal-00112152, version 1.
- GDR CNRS MASCOT-NUM, http://www.gdr-mascotnum.fr

- P. Massart (2007), Concentration inequalities and model selection: Ecole d'Ete de Probabilites de Saint-Flour XXXIII-2003. Springer Verlag.
- H.G. Matthies (2007), Uncertainty Quantification with Stochastic Finite Elements. Encyclopedia of Computational Mechanics.
- Managing Uncertainty in Complex Models (MUCM), http://www.mucm.ac.uk/
- Open Source initiative in uncertainty managempent, http://www.openturns.org
- A. Owen, J.R. Koehler, (1996), Computer Experiments
- N. Rachdi, J-C Fort, T. Klein (2010), Risk bounds for new M-estimation problems, http://hal.archives-ouvertes.fr/hal-00537236.

- N. Rachdi (2011), PhD thesis from University Paul Sabatier of Toulouse.
- J Sacks et al (1989), Design and Analysis of Computer Experiments, Technometrics
- Saltelli A., K. Chan, and E.M. Scott (2000). Sensitivity Analysis. Chichester: Wiley
- Santner, Thomas (2003), The design and analysis of computere experiments -ISBN 0-387-95420-1
- Sobol, I. M. (1993). Sensitivity estimates for non linear mathematical models. Mathematical Modeling and Computational Experiment 1, 407-414
- de Weck O. & Eckert C. (2007), A classification of uncertainty for early product and system design, MIT report ESD-WP-2007-10 (http://esd.mit.edu/WPS).