Modeling Uncertainties by Simulation

Nabil Rachdi (PhD Student)

nabil.rachdi@eads.net

▶ GdR MASCOT-NUM (17, 18, 19 of March – Avignon) ◀

Advisors:

Jean-Claude Fort (Paris V), Thierry Klein (Toulouse III)
Fabien Mangeant (EADS IW), Régis Lebrun (EADS IW).

Outline

General framework

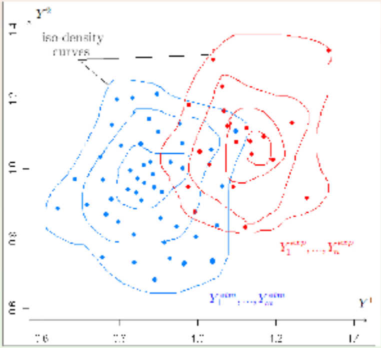
- **Variable of Interest:** Y, density f, probability measure $\mathbb{Q}$ unknown
- **Quantity of Interest (QoI):** Density distribution, mean, threshold probability, quantile etc...
- **Experimental data:** $\mathcal{Y}_n^{exp} = Y_1^{exp}, ..., Y_n^{exp}$ (*a priori* training data) supposed i.i.d from $\mathbb{Q}$
  - Link to history
  - Arise from experiments, complex codes etc...
  - Small number
  - Difficult to obtain
- **Simulated data:** $\mathcal{Y}_m^{sim} = Y_1^{sim}, ..., Y_m^{sim}$ depend on the model h and the parameter $\boldsymbol{\theta}$
  - $Y_i^{sim} = h(X_i, \boldsymbol{\theta}), \quad i = 1, ..., m, \; X_i$ i.i.d random variables (density p_X).
  - $h \in \mathcal{H}$ (set of models), $\boldsymbol{\theta} \in \Theta$ (set of parameters)
- **Goal:** Use Simulated data to improve QoI estimation of Y:

$\Rightarrow$ 1. Calibration procedure: choice of the model h, and parameter $\boldsymbol{\theta}$

$\Rightarrow$ 2. Study of the QoI based on $(Y_1^{exp}, ..., Y_n^{exp}, \widehat{Y}_1^{sim}, ..., \widehat{Y}_m^{sim})$ (*a posteriori* training data) to compare with QoI based on $(Y_1^{exp}, ..., Y_n^{exp})$

Simulated data calibration

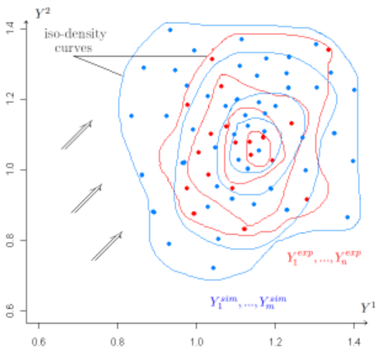
- Illustration of Experimental & Simulated Data: Example of a 2D-Performance $Y = (Y^1, Y^2)$



- Choice of $h \in \mathcal{H}$ and $\theta \in \Theta$? $\Rightarrow$ driven by the QoI

Simulated data calibration

- Illustration of Experimental & Simulated Data: Example of a 2D-Performance $Y = (Y^1, Y^2)$



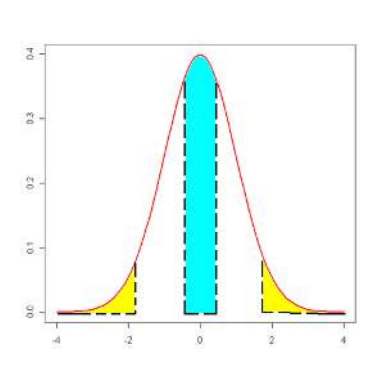
- Choice of $h \in \mathcal{H}$ and $\boldsymbol{\theta} \in \Theta$? $\Rightarrow$ driven by the QoI

- **Quantity of Interest (QoI):**

  Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathbb{D}$ a metric space, and W a random variable defined on $(\Omega, \mathcal{A}, \mathbb{P})$, the QoI of W is defined as the function

$$\begin{array}{rccc} \varphi : & (\Omega, \mathcal{A}, \mathbb{P}) & \longrightarrow & \mathbb{D} \\ & W & \longmapsto & \varphi(W). \end{array}$$

- **Some QoI:**



- $\varphi(W) = \mathbb{E}(W), q_W^\alpha \quad \Rightarrow \mathbb{D} = \mathbb{R}$
- $\varphi(W) = \mathbb{P}(W > s) \quad \Rightarrow \mathbb{D} = [0, 1]$
- $\varphi(W) = f_W \quad \Rightarrow \mathbb{D} = \{\textit{set of distributions}\}$

Choice of $(h, \boldsymbol{\theta})$ for a *QoI* φ

- **Minimization of a criterion :**

$$M(h, \boldsymbol{\theta}) = \mathcal{D}(\varphi_{h,\boldsymbol{\theta}}, \varphi_Y)\,, \qquad \mathcal{D} \text{ : } \textit{distance on } \mathbb{D} \times \mathbb{D}$$

  - $\varphi_{h,\boldsymbol{\theta}}$ and φ_Y are QoI of $h(X, \boldsymbol{\theta})$ and Y (resp.)
  - suppose $(h^*, \boldsymbol{\theta}^*)$ is the unique minimum of M

- **GOAL:**
  - *(Current work)* Minimize $M(h, \boldsymbol{\theta})$ over Θ for fixed $h \in \mathcal{H}$

$$\boldsymbol{\theta}_0(h) = \underset{\boldsymbol{\theta} \in \Theta}{\text{Argmin}}\; M(h, \boldsymbol{\theta})$$

  - *(Later)* Minimize $M(h, \boldsymbol{\theta}_0(h))$ over $\mathcal{H}$

- **We use the form:**

$$M(h,\boldsymbol{\theta}) = \mathcal{D}(\varphi_{h,\boldsymbol{\theta}}, \varphi_Y) \longrightarrow M(h,\boldsymbol{\theta}) = \int_{\mathbb{R}} \gamma_{h,\boldsymbol{\theta}}(y)\, f(y)\, dy$$

  - the function $\gamma_{h,\boldsymbol{\theta}}$ is called contrast of $(h,\boldsymbol{\theta})$:
    $\gamma_{h,\boldsymbol{\theta}} = \Psi\left(\varphi_{h,\boldsymbol{\theta}}\right)$ with Ψ some function
  - recall that f is the density of Y (*unknown*)

- **Example of contrasts:**

  If the *QoI* is the density $\quad \varphi_{h,\boldsymbol{\theta}} = f_{h,\boldsymbol{\theta}}$
  - $y \mapsto \gamma_{h,\boldsymbol{\theta}}(y) = -\ln\left(f_{h,\boldsymbol{\theta}}(y)\right) \quad \Rightarrow \quad M(h,\boldsymbol{\theta}) = K(f, f_{h,\boldsymbol{\theta}})$
  - $y \mapsto \gamma_{h,\boldsymbol{\theta}}(y) = ||f_{h,\boldsymbol{\theta}}||_2^2 - 2f_{h,\boldsymbol{\theta}}(y) \quad \Rightarrow \quad M(h,\boldsymbol{\theta}) = ||f - f_{h,\boldsymbol{\theta}}||_2^2.$
  - etc...

  If the *QoI* is the mean $\quad \varphi_{h,\boldsymbol{\theta}} = \mathbb{E}_X(h(X,\boldsymbol{\theta}))$
  - $y \mapsto \gamma_{h,\boldsymbol{\theta}}(y) = (y - \mathbb{E}_X(h(X,\boldsymbol{\theta})))^2 \quad \Rightarrow \quad M(h,\boldsymbol{\theta}) = \mathbb{E}_Y\left(Y - \mathbb{E}_X(h(X,\boldsymbol{\theta}))\right)^2$

  Etc ...

- **Criterion to minimize:**

$$M(h,\boldsymbol{\theta}) = \int_{\mathbb{R}} \gamma_{h,\boldsymbol{\theta}}(y)\, f(y)\, dy$$

- **Difficulties:**
  - The density function f of Y is unknown
  - For complex models h, the QoI $\varphi_{h,\boldsymbol{\theta}}$ can be unreachable with reasonable CPU time $\rightarrow$ $\gamma_{h,\boldsymbol{\theta}} = \Psi\left(\varphi_{h,\boldsymbol{\theta}}\right)$ unreachable.

- **Alternative: Use of Experimental & Simulated data**
  - Replace f by its *empirical* version $\rightarrow$ $\frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i^{exp}}$ (depends on n-Experimental data $\mathcal{Y}_n^{exp}$)
  - Replace $\gamma_{h,\boldsymbol{\theta}}$ (precisely $\varphi_{h,\boldsymbol{\theta}}$) by its *simulated* version $\rightarrow$ $\gamma_{h,\boldsymbol{\theta}}^m = \Psi\left(\varphi_{h,\boldsymbol{\theta}}^m\right)$ (depends on m-Simulated data $\mathcal{Y}_m^{sim}$).

- **Practical criterion**

$$M(h,\boldsymbol{\theta}) = \int_{\mathbb{R}} \gamma_{h,\boldsymbol{\theta}}(y)\, f(y)\, dy \quad \longleftrightarrow \quad M_{n,m}(h,\boldsymbol{\theta}) := \frac{1}{n}\sum_{i=1}^{n} \gamma_{h,\boldsymbol{\theta}}^m(Y_i^{exp})$$

- **Criterion to minimize in practice:**

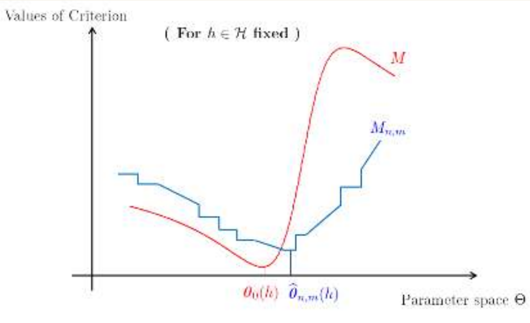
$$M_{n,m}(h,\boldsymbol{\theta}) := \frac{1}{n}\sum_{i=1}^{n} \gamma^{m}_{h,\boldsymbol{\theta}}(Y_i^{exp})$$

- **Estimator of** $\boldsymbol{\theta}_0(h) = \text{Argmin}_{\boldsymbol{\theta}\in\Theta}\ M(h,\boldsymbol{\theta})$

$$\widehat{\boldsymbol{\theta}}_{n,m}(h) = \underset{\boldsymbol{\theta}\in\Theta}{\text{Argmin}}\ M_{n,m}(h,\boldsymbol{\theta})\,.$$

First question: Consistency

$$\widehat{\boldsymbol{\theta}}_{n,m}(h) \underset{\substack{n\to+\infty\\ m\to+\infty}}{\longrightarrow} \boldsymbol{\theta}_0(h)\,?$$



Source of errors

Proposition: Oracle inequality

We prove

$$\underbrace{M(h,\widehat{\boldsymbol{\theta}}_{n,m}(h)) - M(h^*,\boldsymbol{\theta}^*)}_{\textit{risk excess of } (h,\,\widehat{\boldsymbol{\theta}}_{n,m}(h))} \leq 2 \cdot \frac{1}{\sqrt{n}} || \,\mathbb{G}_n \gamma^m_{h,\cdot}\, ||_\Theta + 2 \cdot ||\mathcal{E}^m_h||_\Theta + \Delta_h$$

- **Variance terms :**
  - $\frac{1}{\sqrt{n}} || \,\mathbb{G}_n \gamma^m_{h,\cdot}\, ||_\Theta = \sup_{\boldsymbol{\theta}\in\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \left(\gamma^m_{h,\boldsymbol{\theta}}(Y_i^{exp}) - \mathbb{E}_Y(\gamma^m_{h,\boldsymbol{\theta}}(Y)) \right) \right|$ (deviation)

    $\Rightarrow$ *Estimation Error of Statistical Data* $\rightarrow$ depends on contrast (i.e QoI)
  - $||\mathcal{E}^m_h||_\Theta = \sup_{\boldsymbol{\theta}\in\Theta} || \,\gamma^m_{h,\boldsymbol{\theta}} - \gamma_{h,\boldsymbol{\theta}}\, ||_{1,\,\mathbb{Q}}$, with $||g||_{1,\,\mathbb{Q}} = \int_{\mathbb{R}} |g(y)|\, f(y)\, dy$

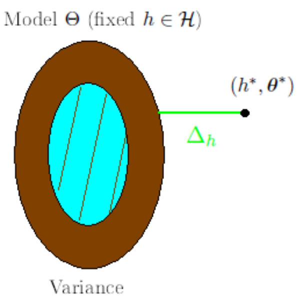
    $\Rightarrow$ *Simulation Error*
- **Bias term :**
  - $\Delta_h = M(h,\boldsymbol{\theta}_0(h)) - M(h^*,\boldsymbol{\theta}^*)$

    $\Rightarrow$ *Approximation Error of the model h*

Representation of the errors

- **Risk minimization :**

$$M(h,\widehat{\theta}_{n,m}) - M(h^*,\theta^*) \le 2 \cdot \frac{1}{\sqrt{n}} ||\mathbb{G}_n \gamma^m_{h,\cdot}||_\Theta + 2 \cdot ||\mathcal{E}^m_h||_\Theta + \Delta_h$$



- **Work:** study the simulation effect on the calibration procedure.

Link with classical methods

Oracle Inequality

Recall $\quad M(h,\boldsymbol{\theta}) = \int_{\mathbb{R}} \gamma_{h,\boldsymbol{\theta}}(y)\, f(y)\, dy$, we have

$$M(h,\widehat{\boldsymbol{\theta}}_n(h)) - M(h^*,\boldsymbol{\theta}^*) \leq \frac{1}{\sqrt{n}} ||\, \mathbb{G}_n \gamma_{h,\cdot}\, ||_{\Theta} + \Delta_h$$

- **For no complex models:** (Linear model etc...)

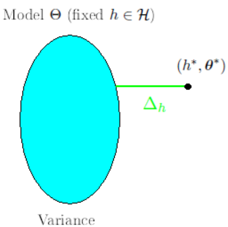
  $\Rightarrow$ the QoI $\varphi_{h,\boldsymbol{\theta}}$ is reachable $\Rightarrow$ Simulation is useless

Advantages

- Well studied
- Maximum Likelihood etc...

Drawback

- Δ_h can be large (due to simplification of h)
  Trade-off Bias-Variance



- **For *Input/Output* data:** $Z_1 = (X_1, Y_1^{exp}), ..., Z_n = (X_n, Y_n^{exp})$

  $z = (x, y) \quad \gamma_{h,\boldsymbol{\theta}}(z) = (y - h(x,\boldsymbol{\theta}))^2 \Rightarrow$ Simulation is useless

Result: Consistency of Calibration

Theorem: Consistency

We prove that $\widehat{\theta}_{n,m}(h) \underset{\substack{n\to+\infty \\ m\to+\infty}}{\longrightarrow} \theta_0(h)$, under some conditions in terms of :

- Model Complexity *(Bracketing Numbers)* $\Rightarrow$ depending on the *quantity of interest*
- Simulation Speed *(Size of Simulated Data set, m)*
- Control of Simulated contrasts *(Modified Lindeberg conditions)*

- **Consequence:** Compute a *QoI* based on $(Y_1^{exp}, \ldots, Y_n^{exp}, \widehat{Y}_1^{sim}, \ldots, \widehat{Y}_m^{sim})$ is better than a *QoI* based on $(Y_1^{exp}, \ldots, Y_n^{exp})$ in some typical cases.
  - Key tool : -*Empirical process theory*- V.d.Vaart (1996,2000), V.d Geer (2000) etc ...
  - We give practical conditions for a wide range of applications.

Theorem

Let $\Gamma_h^m := \{\gamma_{h,\boldsymbol{\theta}}^m,\, \boldsymbol{\theta} \in \Theta\}$ and denote by F_m an envelope function, assume that

- $R_h^m(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}^m\,\gamma_{h,\boldsymbol{\theta}'}^m \;-\; \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}^m\,\mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}'}^m$ converges on $\Theta \times \Theta$
- $\sup_{d(\boldsymbol{\theta},\boldsymbol{\theta}')\leq\delta_m} \mathbb{Q}\left(\gamma_{h,\boldsymbol{\theta}}^m - \gamma_{h,\boldsymbol{\theta}'}^m\right)^2 \underset{m\to+\infty}{\longrightarrow} 0, \quad \forall\,\delta_m \downarrow 0$
- (i) $\quad \mathbb{Q}\,F_m^2 = O(1)$
  (ii) $\quad \mathbb{Q}\,F_m^2\,\mathbb{1}\,\{F_m > \sqrt{n}\,\epsilon\} \underset{n,m\to+\infty}{\longrightarrow} 0 \quad \forall\,\epsilon > 0\,.$
- $$J_{[\,]}\left(\delta_m,\, \Gamma_h^m,\, L_2(\mathbb{Q})\right) \underset{m\to+\infty}{\longrightarrow} 0, \quad \forall\,\delta_m \downarrow 0\,,$$

then $\;||\,\mathbb{G}_n||_{\Gamma_h^m}$ converges $(n, m \to +\infty)$ to the supremum of a centered Gaussian process with covariance function

$$R_h(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}\,\gamma_{h,\boldsymbol{\theta}'} - \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}\,\mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}'}\,.$$

Corollary

If Γ_h^m satisfies to conditions of this Theorem, the calibration procedure is *consistent*, i.e

$$d\left(\widehat{\boldsymbol{\theta}}_{n,m}(h),\, \boldsymbol{\theta}_0(h)\right) \underset{n,m\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0\,.$$

Theorem

Let $\Gamma_h^m := \{\gamma_{h,\boldsymbol{\theta}}^m,\ \boldsymbol{\theta} \in \Theta\}$ and denote by F_m an envelope function, assume that

- $R_h^m(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}^m\,\gamma_{h,\boldsymbol{\theta}'}^m \;-\; \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}^m\,\mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}'}^m$ converges on $\Theta \times \Theta$
- $\sup_{d(\boldsymbol{\theta},\boldsymbol{\theta}')\leq\delta_m} \mathbb{Q}\left(\gamma_{h,\boldsymbol{\theta}}^m - \gamma_{h,\boldsymbol{\theta}'}^m\right)^2 \underset{m\to+\infty}{\longrightarrow} 0, \quad \forall\,\delta_m \downarrow 0$
- (i) $\quad \mathbb{Q}\,F_m^2 = O(1)$
  (ii) $\quad \mathbb{Q}\,F_m^2\,\mathbb{1}\{F_m > \sqrt{n}\,\epsilon\} \underset{n,m\to+\infty}{\longrightarrow} 0 \quad \forall\,\epsilon > 0$ *(Control of Simulated contrasts)*
- $$J_{[\,]}\left(\delta_m,\ \Gamma_h^m,\ L_2(\mathbb{Q})\right) \underset{m\to+\infty}{\longrightarrow} 0, \quad \forall\,\delta_m \downarrow 0 \qquad \textit{(Complexity condition)}$$

then $\;||\,\mathbb{G}_n||_{\Gamma_h^m}$ converges $(n, m \to +\infty)$ to the supremum of a centered Gaussian process with covariance function

$$R_h(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}\,\gamma_{h,\boldsymbol{\theta}'} - \mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}}\,\mathbb{Q}\,\gamma_{h,\boldsymbol{\theta}'}\,.$$

Corollary

If Γ_h^m satisfies to conditions of this Theorem, the calibration procedure is *consistent*, i.e

$$d\left(\widehat{\boldsymbol{\theta}}_{n,m}(h),\ \boldsymbol{\theta}_0(h)\right) \underset{n,m\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0\,.$$

Synthesis

- Risk excess $\leq$ Variance terms + Bias term
- What could happen ?
  - On one hand, a numerician only focuses on minimizing the *bias term* (Δ_h),
  - on the other hand, a statistician can control the *variance term* and ignore the *bias term*.
- We propose a simultaneous approach driven by Simulations

  $\Rightarrow$ Control of variability + Representativity of the model h

- Consequences :
  - The variance $(\Rightarrow \frac{1}{\sqrt{n}}||\mathbb{G}_n\gamma^m_{h,.}||_\Theta + ||\mathcal{E}^m_h||_\Theta)$ depends on

    $\Rightarrow$ the Experimental data
    $\Rightarrow$ the *Quantity of Interest* (contrast)
    $\Rightarrow$ and the Simulated data
  - We expect a better estimation procedure for limited amount of experimental data and complex models.

Example of the *Range* study

Phenomenon : Y = *Range* (distance an aircraft can travel), QoI = density distribution

- ***A priori* training data :** Experimental data, $n = 20$, $\mathcal{Y}_n^{exp} = Y_1^{exp}, \ldots, Y_n^{exp}$

  (obtained from complex model h^* supposed to be the "true")

- **Additional knowledge :** Simulated data, $m = 3000$, $\mathcal{Y}_m^{sim} = Y_1^{sim}, \ldots, Y_m^{sim}$ from

$$h(X, \boldsymbol{\theta}) = \frac{F\,V}{C_s}\,\frac{1}{\theta_1}\log\left(\frac{1}{1-\theta_2}\right)$$

  - Uncertain Inputs $X = (F, V, C_s)^T$
  - Parameters $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta$

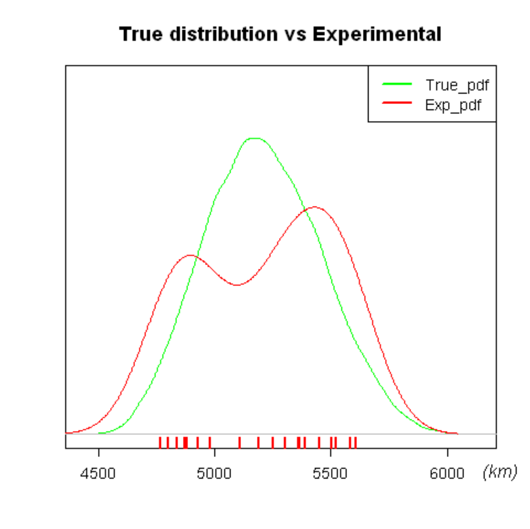
- **Choice of θ ?** $\theta_0(h) = \underset{\boldsymbol{\theta}\in\Theta}{\text{Argmin}} \int_{\mathbb{R}} \gamma_{h,\boldsymbol{\theta}}(y)\, f(y)\, dy$

$$\gamma_{h,\boldsymbol{\theta}} = -\ln(f_{h,\boldsymbol{\theta}}) \quad f_{h,\boldsymbol{\theta}} \leftrightarrow f_{h,\boldsymbol{\theta}}^{m}\ (Kernel) \quad f \leftrightarrow \frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i^{exp}}$$

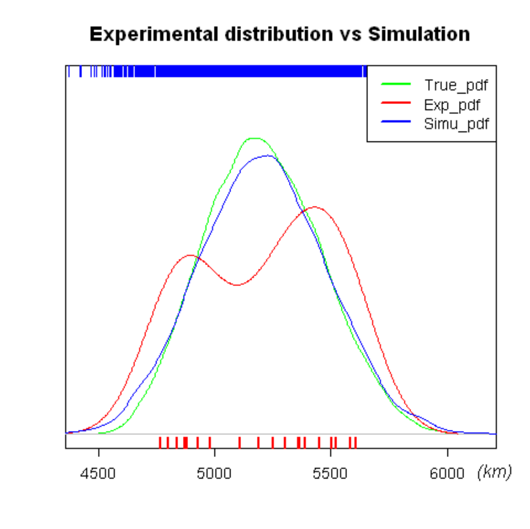
$$\widehat{\boldsymbol{\theta}}_{n,m}(h) = \underset{\boldsymbol{\theta}\in\Theta}{\text{Argmin}} -\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{1}{m}\sum_{j=1}^{m}K_{h_m}(Y_i^{exp} - Y_j^{sim})\right)$$

- ***A posteriori* training data :** $(Y_1^{exp}, \ldots, Y_n^{exp}, \widehat{Y}_1^{sim}, \ldots, \widehat{Y}_m^{sim}) \rightarrow n + m = 3020!$

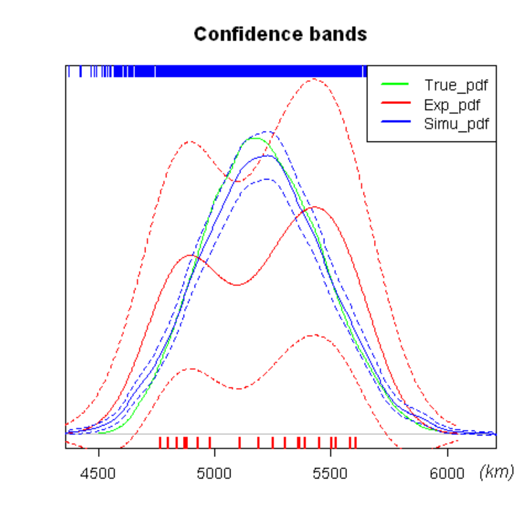
- QoI with *a priori* (Experimental) and *a posteriori* (Experimental + Simulated) training data



- QoI with *a priori* (Experimental) and *a posteriori* (Experimental + Simulated) training data



- QoI with *a priori* (Experimental) and *a posteriori* (Experimental + Simulated) training data



Other example: *Mean* study

Suppose that $Y \sim h(X, \boldsymbol{\theta}_0)$ with $\boldsymbol{\theta}_0 \in \Theta$

Recall that $Y^{sim} = h(X, \boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \Theta$

Let $QoI = \mathbb{E}(Y) = \varphi_Y = \varphi_{h,\boldsymbol{\theta}_0}$

- $\varphi_{h,\boldsymbol{\theta}} = \mathbb{E}_X(h(X, \boldsymbol{\theta}))$, $\quad \mathcal{D}(\varphi_{h,\boldsymbol{\theta}}, \varphi_{h,\boldsymbol{\theta}_0}) = \mathbb{E}(\varphi_{h,\boldsymbol{\theta}} - \varphi_{h,\boldsymbol{\theta}_0})^2$ (quadratic risk)
- let $\varphi(\mathcal{Y}_n^{exp}) = \frac{1}{n} \sum_{i=1}^n Y_i^{exp}$ and

  $$\varphi(\mathcal{Y}_n^{exp}, \mathcal{Y}_m^{sim}) := \frac{1}{n+m}(Y_1^{exp} + \ldots + Y_n^{exp} + \underbrace{Y_1^{sim} + \ldots + Y_m^{sim}}_{\textit{depend on } \boldsymbol{\theta}\ !})$$

- **Question:** Is $\varphi(\mathcal{Y}_n^{exp}, \mathcal{Y}_m^{sim})$ "better" than $\varphi(\mathcal{Y}_n^{exp})$?

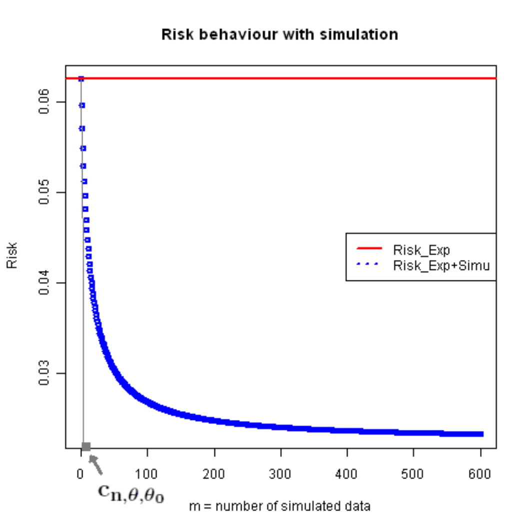
  $\Rightarrow$ turns out to have $\quad \mathbf{R}_{\mathbf{n},\mathbf{m}}(\boldsymbol{\theta}) = \mathbb{E}\left(\varphi(\mathcal{Y}_n^{exp}, \mathcal{Y}_m^{sim}) - \varphi_{h,\boldsymbol{\theta}_0}\right)^2 \leq \mathbf{R}_{\mathbf{n}} = \mathbb{E}\left(\varphi(\mathcal{Y}_n^{exp}) - \varphi_{h,\boldsymbol{\theta}_0}\right)^2$

Lemma

Let $n \in \mathbb{N}$. We show that $\exists\, \Theta_{sim}(n, \boldsymbol{\theta}_0) \subset \Theta$ and $\exists\, c_{n,\boldsymbol{\theta},\boldsymbol{\theta}_0} \in \mathbb{N}$ such that for all $\boldsymbol{\theta} \in \Theta_{sim}(n, \boldsymbol{\theta}_0)$

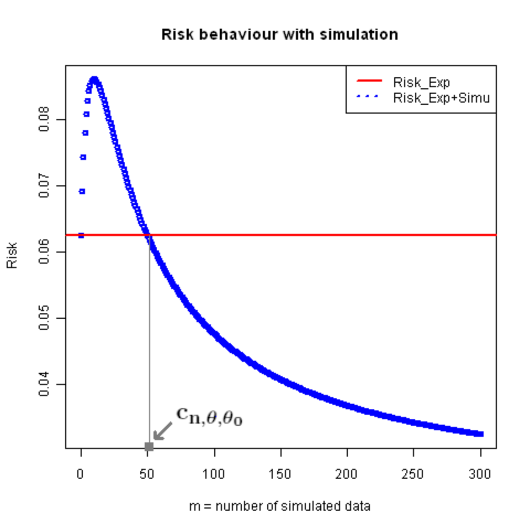
$$\mathbf{R}_{\mathbf{n},\mathbf{m}}(\boldsymbol{\theta}) \leq \mathbf{R}_{\mathbf{n}} \qquad \textit{for all} \quad m \geq c_{n,\boldsymbol{\theta},\boldsymbol{\theta}_0}$$

Mapping of $m \longmapsto \mathbf{R_{n,m}}(\theta)$ and $m \longmapsto \mathbf{R_n}$
For $\Rightarrow$ $\boldsymbol{\theta}$ **in** $\Theta_{sim}(n, \boldsymbol{\theta}_0)$

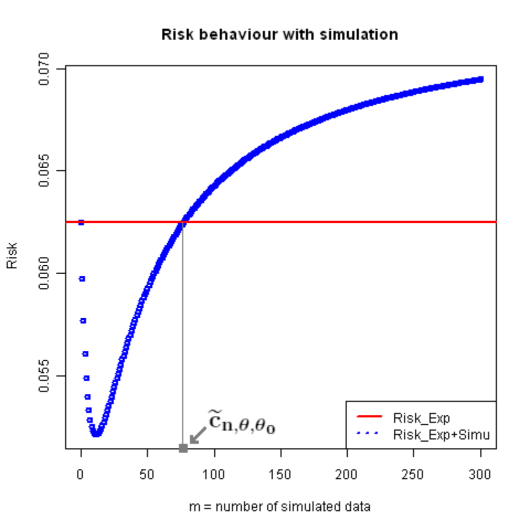


Mapping of $m \longmapsto \mathbf{R_{n,m}}(\theta)$ and $m \longmapsto \mathbf{R_n}$

For $\Rightarrow$ $\boldsymbol{\theta}$ **in** $\Theta_{sim}(n, \boldsymbol{\theta}_0)$

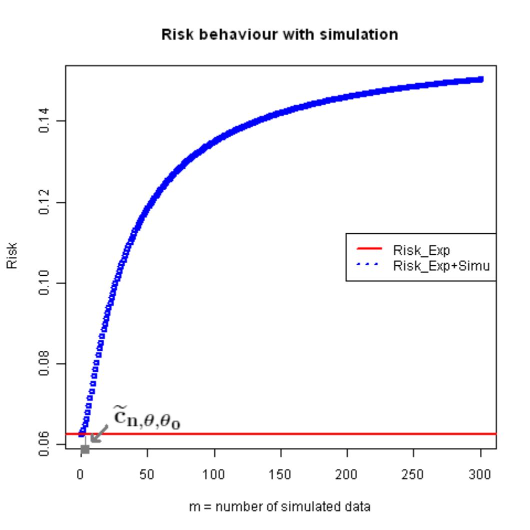


Mapping of $m \longmapsto \mathbf{R_{n,m}}(\theta)$ and $m \longmapsto \mathbf{R_n}$
For $\Rightarrow$ $\boldsymbol{\theta}$ **not in** $\Theta_{sim}(n, \boldsymbol{\theta}_0)$



Mapping of $m \longmapsto \mathbf{R_{n,m}}(\theta)$ and $m \longmapsto \mathbf{R_n}$

For $\Rightarrow$ $\boldsymbol{\theta}$ **not in** $\Theta_{sim}(n, \boldsymbol{\theta}_0)$



Simulate or not to Simulate ?

- The only θ we dispose is $\widehat{\boldsymbol{\theta}}_{n,m}$
- **Question:** Does $\widehat{\boldsymbol{\theta}}_{n,m}$ belong to $\Theta_{sim}(n, \boldsymbol{\theta}_0)$?
- Need of Central Limit Theorem !

Future Work

- Rate of Convergence of the calibration procedure: fonction of n and m ...
  - Impact of Experimental and Simulated data on the estimation
  - For a given *Quantity of Interest* $\Rightarrow$ how many n ? and how many m ?
  - etc ...
- Asymptotic Normality
  - Statistic studies
  - Confidence bands
  - etc...
- Sensitivity of this uncertainty analysis in relation to the *a priori* distribution of X.
- Robustness study: influence of the *QoI* on the *Model Selection*.

Thank you for your attention !