Uncertainty quantification for computer models via Bayesian model averaging

Léo Neufcourt

Michigan State University
Department of Statistics and Probability & FRIB
neufcour@frib.msu.edu

ISNET 5

November 8th 2017

Collaborators : **Vojtech Kejzlar**, *Taps Maiti, Frederi Viens (STT)*
Yannen Jaganathen, Nicolas Michel, Witek Nazarewicz (FRIB)

Introduction

Message : competition between several models induces an additional uncertainty

- Uncertainty decomposition

$$\sigma^2_{prediction} = \sigma^2_{intrinsic} + \sigma^2_{systematic} + \sigma^2_{calibration} + \boldsymbol{\sigma^2_{model}}$$

- UQ for (expensive) computer models *via* Bayesian calibration
- Accounting for uncertainty on the model with Bayesian model averaging : a proof of concept
- Leading example : Gamow shell model

Outline

1. Computer models
2. Bayesian calibration
3. Uncertainty Quantification
4. Bayesian model averaging

Physical system

Physical system (general)

$$y_i = y(x_i) + \epsilon_i$$

- $(x_i, y_i)_{i=1}^{n}$: input and output observations
- $x \mapsto y(x)$: physical process (signal)
- $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon)$: intrinsic randomness + experimental error

Computer model

Computer model : $(x, \theta) \mapsto f(x, \theta)$

$$y(x) = f(x, \theta^*) + \delta(x)$$

- θ^* : *"true"* parameter value, i.e. $f(x, \theta^*)$ is the *true* average of all possible observations
- $\delta(x)$: systematic error

Reference : Kennedy & O'Hagan, Bayesian calibration of computer models (2003)

Notations

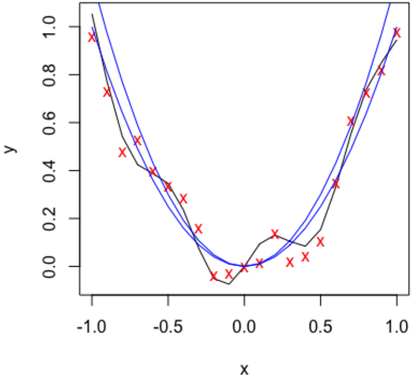


Figure: Observations, true process and computer model

Black box calibration

$$\min_\theta \chi^2(\theta) := \sum_{i=1}^{n} w_i(y_i - f(x_i,\theta))^2$$

- calibrated parameter : θ^0
- optimal value : $f(x,\theta^0)$
- Jacobian : $J^0(x) := \nabla_\theta f(x,\theta_0)$
- Hessian $H^0(x) := H_\theta(f)(x,\theta_0)$

Computational cost : cheap *vs* expensive computer model

$\rightarrow$ computation time of $f(x,\theta)$: if computing $f(x,\theta)$ takes 1 min and θ has dimension d, evaluating of $f(x,\theta)$ on a grid with 100 points per parameter takes $\sim 10^{2d}$ mins

Outline

1 Computer models

2 Bayesian calibration

3 Uncertainty Quantification

4 Bayesian model averaging

Bayesian setup

Inference model : $y_i = f(x,\theta) + \delta(x) + \epsilon_i$

Simplification : $\sigma e_i \leftarrow \epsilon_i + \delta(x_i)$, $e_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ (e.g. $\epsilon_i \ll \delta(x_i)$)

$$y_i := f(x_i,\theta) + \sigma e_i$$

Gaussian process(GP)

- $f(x,\theta) \sim GP(m(x,\theta), V((x,\theta),(x',\theta'))$
- $m(x,\theta) := h(x,\theta)^T\beta$
- $V((x,\theta),(x',\theta')) = \exp\left((x^T,\theta^T)\Omega(x',\theta')^T\right)$
- Hyperparameters : Ω (concentration matrix) and β (regression coefficients)

Bayesian calibration problem : estimate θ and σ

Linearization around the black box optimum :

$f(x,\theta) = f(x,\theta^0) + J^0(x)^T(\theta-\theta^0) + \gamma(x,\theta)$

$m(x,\theta) := f(x,\theta^0) + J^0(x)^T(\theta-\theta^0)$

Bayesian calibration

Bayes formula :

$$p(\theta, \sigma|y) = \frac{p(y|\theta, \sigma)\pi(\theta, \sigma)}{\int p(y|\theta, \sigma)\pi(\theta, \sigma)d\theta d\sigma} \propto p(y|\theta, \sigma)\pi(\theta, \sigma)$$

Prior distributions : (ρ, a^0, b^0, Ω are hyperparameters)

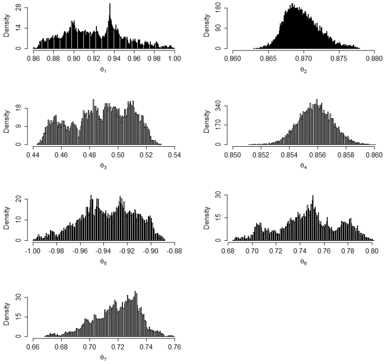
- $\theta \sim \mathcal{N}(\theta^0, \rho \cdot (H^0)^{-1})$
- $\sigma^2 \sim IG(a^0, b^0)$

Posterior distributions :

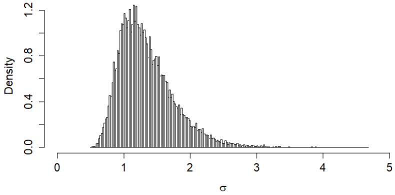
$p(y|\theta, \sigma)$: likelihood (model)

$\pi(\theta, \sigma)$: prior

$$p(\Delta|y) \propto p(y|\Delta)\pi(\Delta) \propto \int p(\Delta|\theta, \sigma)p(y|\theta, \sigma)\pi(\theta, \sigma)d\theta d\sigma$$



(a) θ (calibration parameters)



(b) σ (systematic uncertainty)

Figure: Metropolis samples from the posterior distributions (100,000 Monte Carlo iterations)

Outline

1. Computer models
2. Bayesian calibration
3. Uncertainty Quantification
4. Bayesian model averaging

Decomposition of prediction error

Prediction of the observation y_{new} of a new input x_{new} :

$$
\begin{aligned}
& y_{new} - f(x_{new}, \theta) \\
= \quad & (y_{new} - y(x_{new})) + (y(x_{new}) - f(x_{new}, \theta^*)) + (f(x_{new}, \theta^*) - f(x_{new}, \theta)) \\
= \quad & \epsilon_{new} + \delta(x_{new}) + \xi(x_{new}, \theta)
\end{aligned}
$$

3 sources of error : intrinsic + systematic + calibration

Systematic and calibration errors may not be independent, still as a rule of thumb

$$\sigma^2_{pred} = \sigma^2_{intrinsic} + \sigma^2_{systematic} + \sigma^2_{calibration} = \sigma^2_{\epsilon} + \sigma^2_{\delta} + \sigma^2_{\xi}$$

Bayesian decomposition of uncertainty

Consider the posterior distribution $p(\Delta|y)$ of a general quantity Δ, e.g. $\Delta := y_{new}$

$$\mathbb{V}ar[\Delta|y] = \mathbb{E}[\mathbb{V}ar(\Delta|y,\theta)|y] + \mathbb{V}ar[\mathbb{E}(\Delta|y,\theta)|y]$$

- $\mathbb{V}ar[\mathbb{E}(\Delta|y,\theta)|y] \rightarrow$ parameter uncertainty
- $\mathbb{E}[\mathbb{V}ar(\Delta|y,\theta)|y] \rightarrow$ the rest $=$ intrinsic randomness $+$ experimental uncertainty $+$ systematic uncertainty

Computer model : $y_{new} = f(x_{new},\theta) + \sigma e_{new}$

- calibration uncertainty : $\mathbb{V}ar[f(x_{new},\theta)|y]$
- systematic uncertainty : $\mathbb{E}[\sigma^2|y]$

Accounting for uncertainty on the model

Suppose that p models are available and that the *true* model is given by a random variable M. We have :

$$\mathbb{V}ar[\Delta|y] = \mathbb{E}[\mathbb{V}ar(\Delta|y, M)|y] + \mathbb{V}ar[\mathbb{E}(\Delta|y, M)|y]$$

$$\mathbb{V}ar[\mathbb{E}(\Delta|y, M)|y] \to \text{uncertainty on the model}$$

Accounting for uncertainty on the model can be done via model averaging

Outline

1. Computer models
2. Bayesian calibration
3. Uncertainty Quantification
4. Bayesian model averaging

Setup

Consider the same data $\left(x_i, y_i\right)_{i=1}^{n}$ and p computer models

$$(x, \theta_k) \mapsto f_k(x, \theta_k), k = 1, \dots, p.$$

Simplified Bayesian models:

$$y_i = f_k(x_i, \theta_k) + \sigma_k e_i, i = 1, \dots, n \qquad (\mathcal{M}_k)$$

where

- x_i is a deterministic input
- y_i is the observed output
- $f_k(x_i, \theta_k)$ is the output of the k^{th} computer model
- θ_k and σ_k are parameters
- $e_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$

Bayesian Model Averaging (BMA)

Axiom : when several models are available, an "*honest*" prediction should be given by an average model

$$\hat{y}_a(x_i) := \sum_k 1_{M=k}(f_k(x_i, \theta_k) + \sigma_k e_i)$$

where $p_k := p(M = k)$ represents our subjective belief in model k.

Posteriors distributions of the Bayesian average model :

$$p(\Delta|y) = \sum_k p_k p(\Delta|y, M = k)$$

where $p_k := p(M = k|y) \propto \pi(M = k) \int p(y|\theta, \sigma, M = k)\pi(\theta, \sigma) d\theta d\sigma$

Bayesian uncertainty on the model

BMA prediction : $\mathbb{E}[y_{new}^{BMA}|y] = \sum_k p_k \mathbb{E}[f_k(x_{new}, \theta)|y]$

BMA prediction uncertainty :

$$\mathbb{V}ar[y_{new}|y] = \sum_k p_k \mathbb{V}ar[y_{new}|y, M = k] + \mathbb{V}ar[\mathbb{E}(y_{new}|y, M)|y]$$

Uncertainty on the model :

$$\sigma^2_{model} := (\sigma^{BMA}_{prediction})^2 - \sum_k p_k (\sigma^k_{systematic})^2 - \sum_k p_k (\sigma^k_{calibration})^2$$

Bayesian average of two computer models

→ **Gamow shell model example**

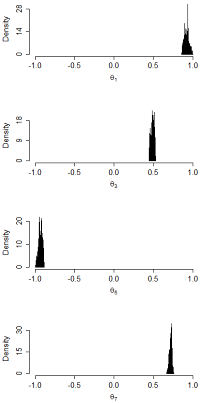
$$f_A(x,\theta) = f(x,\theta^0) + z_A(x)^T(\theta - \theta^0) + \gamma(x,\theta) \text{ (good calibration)}$$
$$f_B(x,\theta) = f(x,\tilde{\theta}^0) + z_B(x)^T(\theta - \tilde{\theta}^0) + \gamma(x,\theta) \text{ (bad calibration)}$$

$\tilde{\theta}^0$: distortion of θ^0, with $(\tilde{\theta}^0 - \theta^0) \sim 1\%$

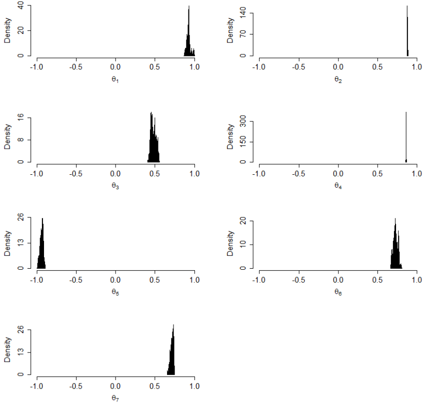
$\gamma(x,\theta)$: GP with mean 0

model	prior	posterior
A	0.5	0.9997
B	0.5	0.0003

Table: Posterior probabilities of models A and B



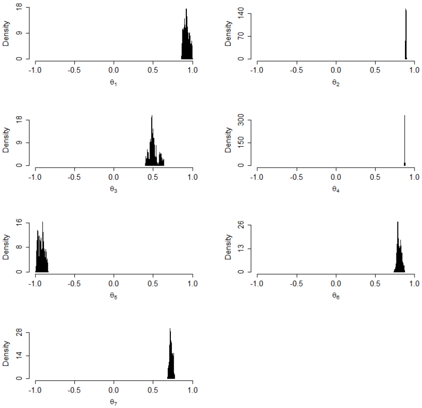
(a) Model A



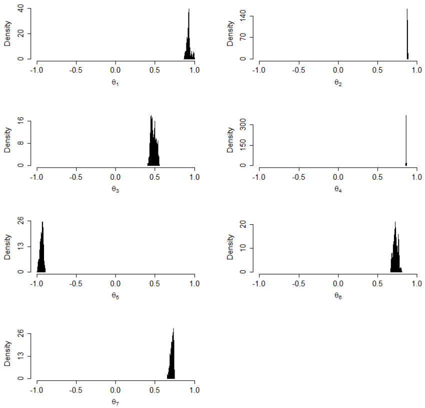
(b) BMA (Bayesian average model)

Figure: Posterior distributions of the calibration parameters

Gamow shell model simulation : Model 2 vs BMA



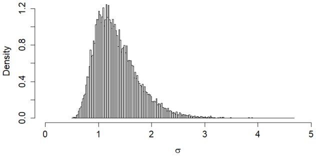
(a) Model B



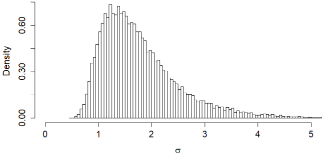
(b) BMA (Bayesian average model)

Figure: Posterior distributions of the calibration parameters

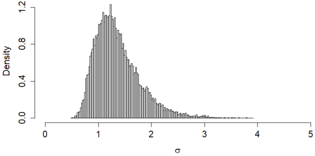
Gamow shell model simulation : systematic uncertainties



(a) Model A



(b) Model B



(c) BMA (Bayesian average model)

Figure: Posterior distributions of the systematic uncertainty

Uncertainty analysis

	error (σ)		
model	prediction	systematic	calibration
A	3.522	1.419	3.224
B	6.489	1.743	6.251
BMA	3.524	**1.420**	3.22

Table: Uncertainty decomposition of models A, B and BMA

$$\boldsymbol{\sigma_{model} = 0.061}$$

Shrinkage of the distortion

- BMA often selects one model
- example with comparable model posteriors : $(\tilde{\theta}^0 - \theta^0) \sim 0.01\%$

model	prior	posterior
A	0.5	0.553
B	0.5	0.447

Table: Posterior probabilities of models A and B

$$\boldsymbol{\sigma_{model} = 0.11}$$

Statistical challenges

Bayesian calibration :

- calibration of Black box emulator for *expensive* computer models
- improve Metropolis sampling
- convergence of Bayesian estimators
- systematic choice of prior

Model averaging : computing the *evidence integral* (Bayes factor)

$$\int_{\theta,\sigma} p(y|\theta,\sigma)\pi(\theta,\sigma)d\theta d\sigma$$

Computing the evidence integral for a (very) good model

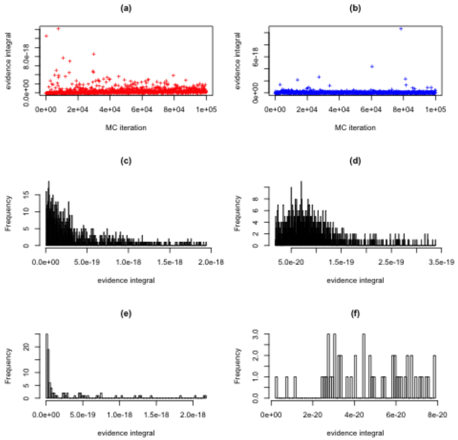


Figure: Monte Carlo estimation of an evidence integral (50 data), with (right) or without (left) importance sampling. (a) and (b) show the estimations for an increasing number of iterations. (b) and (c) show the distributions of the same samples. (e) and (f) give the distribution of 150 Monte Carlo estimates with 10,000 iterations

Comparison of evidence integrals

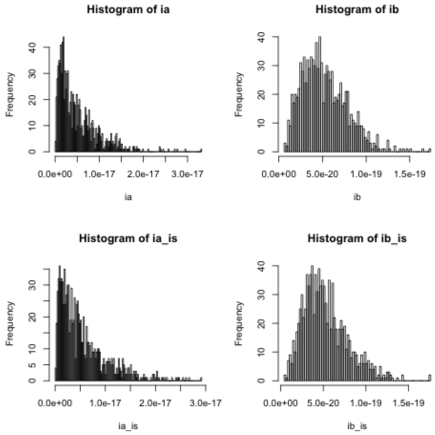


Figure: Distribution of Monte Carlo samples estimating the evidence integrals for A (left) and B (right), with (bottom) or without (top) importance sampling

Thank you!

neufcour@frib.msu.edu