CEA-EDF-INRIA Numerical Analysis Summer School (2011): uncertainty quantification for numerical model validation

Inverse problem & calibration of parameters

- Part 2 -

Mathieu Couplet

EDF R&D - Department of Industrial Risk Management

Contents

Contents

Inverse problems as optimization problems

Inverse problem

Let $y^* \in \mathcal{Y}$ and $H : \mathcal{X} \to \mathcal{Y}^\dagger$, find $x^* \in \mathcal{X}$ such that

$$y^* = H(x^*)$$

That is: find $x^* \in H^{-1}(\{y^*\})$, where $H^{-1}(\{y^*\}) = \{x \in \mathcal{X} \,/\, y^* = H(x^*)\}$.

†In practice, H is a numerical code H^{code}.

Optimization problem

Let $\|.\|$ be a norm on $\mathcal{Y}^\dagger$, find

$$x^* = \arg\min_{x \in \mathcal{X}} \|y^* - H(x)\|.$$

This formulation is more general and allows the use of optimization methods (*e.g.* a simulated annealing algorithm, which is a Metropolis-Hastings one).

†especially if $\mathcal{Y}$ is a metric space without norm, another way to generalize the inverse problem is minimizing $\mathrm{d}(y^*, H(x))$ where $\mathrm{d}(.,.)$ is a distance on $\mathcal{Y}$.

Inverse problems as optimization problems

Three possible situations

(S0) $\#H^{-1}(y) = 1$: you are lucky![†]

(S1) $\#H^{-1}(y) > 1$: need to reformulate the problem

(S2) $\forall x \in \mathcal{X} \quad \|y^* - H(x)\| > 0$: no solution?!

[†] In theory... In practice, your problem has to be not ill-conditioned for a "accurate solution" to be computed...

Consequences of (S2)

You have to consider the possibilities that the data y^* could be affected by (measurement) error/noise[†] or that H could be not a perfect model of the reality behaviour. The solution of the optimization problem is then an estimate $\hat{x}$ of x^*.
In fact, even in the situations (S0) or (S1), you often should consider both of these possibilities, which leads to consider probabilistic and statistical tools.
Moreover, $\hat{x}$ depends generally on the chosen norm $\|.\|$: what to choose?

[†] In the following, we will often regard y^* as some data from measurements.

Inverse problems as optimization problems

The euclidean case: GLS (non-linear Generalized Least Squares)

Let $y^* \in \mathbb{R}^P$ and H: $\mathbb{R}^Q \rightarrow \mathbb{R}^P$, considering the inverse problem $y^* = H(x^*)$ and a euclidean-norm $\|.\|_S$ induced by $S \in \mathcal{S}_P^{++}(\mathbb{R})$, that is $\|y\|_S^2 = y^T S\, y$, find

$$\hat{x} = \arg\min_{x \in \mathbb{R}^Q} \|y^* - H(x)\|_S \quad \Leftrightarrow \quad \hat{x} = \arg\min_{x \in \mathbb{R}^q} (y^* - H(x))^T S\, (y^* - H(x))$$

By default, engineers often consider the problem of OLS (non-linear Ordinary Least Squares) and choose the 2-norm, that is $S = I_p$.

Inverse problems as optimization problems

The euclidean case: GLS (non-linear Generalized Least Squares)

Let $y^* \in \mathbb{R}^P$ and H: $\mathbb{R}^Q \to \mathbb{R}^P$, considering the inverse problem $y^* = H(x^*)$ and a euclidean-norm $\|.\|_S$ induced by $S \in \mathcal{S}_P^{++}(\mathbb{R})$, that is $\|y\|_S^2 = y^T S\, y$, find

$$\hat{x} = \arg\min_{x \in \mathbb{R}^Q} \|y^* - H(x)\|_S \quad \Leftrightarrow \quad \hat{x} = \arg\min_{x \in \mathbb{R}^q} (y^* - H(x))^T S\, (y^* - H(x))$$

Note that:

- Let S_1 be in $\mathcal{S}_P^{++}(\mathbb{R})$; the GLS problems with $S = S_1$ and $S = S_\alpha$ (same y and H) are equivalent if $\exists \alpha > 0 \quad S_\alpha = \alpha\, S_1$, which is a necessary[†] and sufficient condition for
$$\exists \phi: \mathbb{R} \to \mathbb{R} \text{ strictly increasing} \quad \forall y \in \mathcal{Y} \quad \|y\|_{S_\alpha} = \phi(\|y\|_{S_1}).$$
- Conversely, the fact that two GLS problems with $S = S_1$ and $S = S_\alpha$ are equivalent does not generally imply this (most strong) condition: *e.g.* $\hat{x}$ does not depend on S in the situation (S0).

[†] Think of the positive homogeneity of norms: $\|\alpha\, y\| = |\alpha|\, \|y\|$.

GLS as MLE (Maximum Likelihood Estimate)

PDF of any P-dimensionnal $Y \sim \mathcal{N}(\mu, \Sigma)$

$$f_{\mathcal{N}}(y; \mu, \Sigma) = (2\pi)^{-P/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2} \underbrace{[y-\mu]^T \Sigma^{-1} [y-\mu]}_{= \|y-\mu\|_{\Sigma^{-1}}^2} \right)$$

Statistical interpretation of the GLS solutions

Therefore, any GLS solution can be interpreted as a MLE:

$$\forall \sigma > 0 \quad \left[\hat{x} = \arg\min_{x \in \mathbb{R}^Q} \|y^* - H(x)\|_S \Leftrightarrow \hat{x} = \arg\max_{x \in \mathbb{R}^Q} \mathcal{L}_{\mathcal{N}}(x, \sigma^2 \Sigma; y^*) \text{ with } \Sigma = S^{-1} \right]$$

where $\mathcal{L}_{\mathcal{N}}(x, \sigma^2 \Sigma; y)$ is the likelihood of parameter (x, σ) associated to the realization y of a random gaussian vector $Y \sim \mathcal{N}(H(x), \sigma^2 \Sigma)$, that is

$$\forall i \in \{1, \cdots, P\} \qquad Y_i = H_i(x) + U_i$$

where the distribution of the "errors" $U = (U_1 \cdots U_n)^T$ is $\mathcal{N}(0, \sigma^2 \Sigma)$.

GLS as MLE (Maximum Likelihood Estimate)

Statistical interpretation of the GLS solutions

In the particular case of OLS ($S = \mathrm{I}_p$):

$$\hat{x} = \arg\max_{x \in \mathbb{R}^Q} \mathcal{L}_{\mathcal{N}}(x,\, \sigma^2\, \mathrm{I}_p;\, y^*)$$

with corresponds to $U \sim \mathcal{N}(0,\, \sigma^2\, \mathrm{I}_p)$, that is the hypothesis of independence and homoscedasticity (same variance) of the "errors".

Except for particular cases (depending on y and H), the use of GLS can be seen as a MLE subject to (implicit) constraints on the covariance matrix of the gaussian vector of the "errors".

The statistical point of view allows to go further, for example:

- assuming or testing non-gaussian errors;
- assuming a gaussian vector U, but estimating all the coefficients of the covariance matrix without any constraint: the appropriate norm to estimate x^* is then not chosen but deduced (many measurements $y^{*,j}$ are needed).

Different practical situations, different statistical models

Now, the door is opened to statistics: what kind of model to chose?
It is useful to distinguish between the different situations:

(1) The data $y^* \in \mathbb{R}^P$ corresponds to P independent 1-dimensional measurements of the response of a physical system caracterized by (or depending on) x^*, possibly in different experimental conditions d_i (then $H_i(x) = h(x, d_i)$ for all $i \in \{1, \cdots, P\}$): it is reasonable to assume i.i.d. "errors" U_i.

(2) The data y^* corresponds to only one p-dimensional measurement ($p = P$), *e.g.* $y_i^* = y^*(t_i)$ where data are acquired at p different times t_i (or at different points in space).

- (2.1) Nevertheless, it is still reasonnable to suppose i.i.d. "errors", *e.g.* the durations between data acquisition (or distances between probes) are "important enough".
- (2.2) The possibility that "errors" do depend from each other cannot be excluded, but it appears reasonnable to assume ergodicity, *e.g.* by using a stationary Gaussian Process indexed by time t.
- (2.3) Otherwise, you should perform other $n-1$ independent measurements $y^{*,j}$, unless you have a very informative prior! Then, you would be in a similar situation than (1) with n i.i.d. p-dimensional "errors" instead of p i.i.d. scalar "errors".

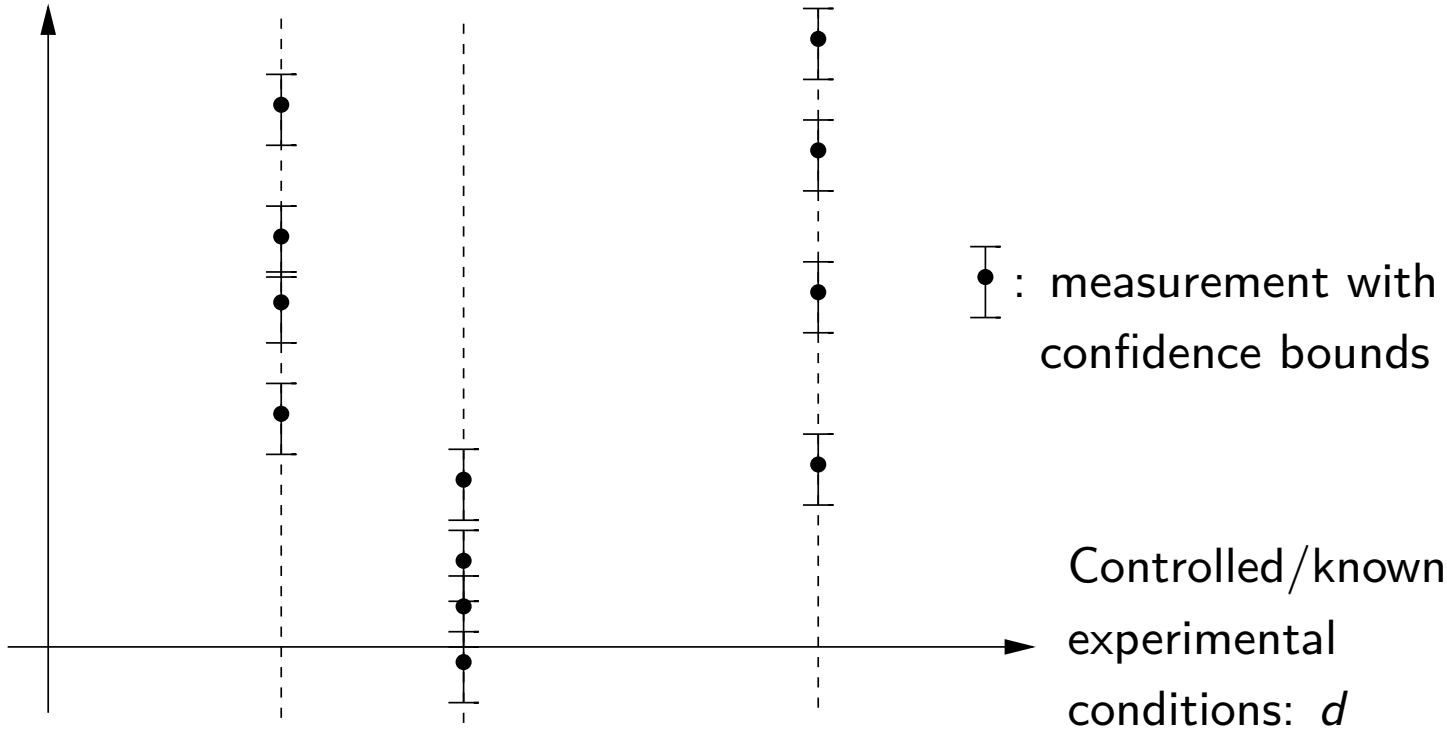
In situation (2), it is always better to perform as many p-dimensional measurements as possible, which enables to compare different statistical models, thus to justify ergodicity or the hypothesis of i.i.d. "errors".

(3) $y^* \in \mathbb{R}^P$ corresponds to n p-dimensional measurements ($P = p \times n$): similar to (2.3).

Statistics as a general framework to deal with inverse problems

Statistics allows to deal with conditions d suffering from measurement errors [Bayarri, Paulo *et al.*, 2007] (R. Paulo, Lecture 4), or other practical situations such as:

Response of the physical system: $y = H^{\text{real}}(x, d)$



For a given d, the variability of the measurements is significant: the uncertain parameter x cannot be reasonably be thought as steady. x has to be consider as an irreductible hazard X: its characterization is naturally a statistical problem and not a "classical" inverse problem.

Statistics as a general framework to deal with inverse problems

Missing data inference problem (in the sense of [Dempster *et al.*, 77])

$$\underbrace{\mathcal{L}(y\,;\theta)}_{\substack{\text{incomplete likelihood}\\ \text{to maximize by tuning } \theta}} = \int_{\{z\in\mathcal{Z}:M(z)=y\}} \underbrace{\mathcal{L}(z;\theta)}_{\text{complete likelihood}} \,\mathrm{d}z$$

where $z \in \mathcal{Z}$ corresponds to the missing data (*e.g.* inputs x or measurements "errors" u). The previous problem can be seen as a such a *missing* (or *incomplete*) *data problem*.

Alternatives to infer X

- Non-stochastic EM (*Expectation Maximization*) algorithms[†] as Circé [de Crécy, 2001], which are convenient and effective if the complete likelihood corresponds to a distribution of the exponential family (*e.g.* linear H, gaussian X and U), but, in general, not applicable if H is non-linear where the density of X is concentrated.
- Stochatic EM[†] [Kuhn, 2003][Celeux, 2007] or MCMC methods from a bayesian approach: in general (non-linear H), need for a metamodel of H, since a lot of evaluations of H are then required.

[†] In the EM approach, the purpose is the computation of a MLE ; see the tutorial of A. Roche, CEA, for an overview.

Contents

GLS as MAP (*Maximum A Posteriori*)

From a (almost-)bayesian[†] point of view, a GLS solution can interpreted as a MAP (like any MLE):

$$\hat{x} = \arg\min_{x \in \mathbb{R}^Q} \|y^* - H(x)\|_S \Leftrightarrow \hat{x} = \arg\underbrace{\max_{\substack{\theta \in \mathbb{R}^{Q+1} \\ \theta = (\theta_x, \theta_\sigma)}} \underbrace{\pi(\theta)}_{\text{prior}} \underbrace{\mathcal{L}_{\mathcal{N}}(\theta_x, {\theta_\sigma}^2 S^{-1}; y^*)}_{\text{likelihood}}}_{\text{MAP}}$$

for example if $\pi(\theta) = \pi(\theta_\sigma)\,\pi(\theta_x)$ (independent marginal priors) and $\pi(\theta_x) \propto 1$ (improper uniform prior for θ_x).

What opportunities come from the bayesian point of view?

In the following, on the basis of linear euclidean examples, we will illustrate the relevance of the bayesian approach and methods to deal with "not well-posed problems".

[†]Viewed from the bayesian decision theory angle, a MAP is only a limit of Bayes estimators.

The linear euclidean case

Assumption: $\exists H \in \mathcal{M}_{P,Q}(\mathbb{R}) \quad H(x) = H\,x$.

Then:

- $\hat{x}$ can be defined as a solution to a linear system:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^Q} \|y^* - H(x)\|_S \quad \Leftrightarrow \quad H^T\,S\,y^* = H^T\,S\,H\,\hat{x}.$$

- $\hat{x}$ is unique whatever y^* is iif the rank of H is maximal, which implies $P \geq Q$ ("as many pieces of data as many unknowns is needed") and $H^T\,S\,H \in \mathcal{S}_P^{++}(\mathbb{R})$.
- If the previous condition holds, the *Gauss-Markov theorem* states that the maximum likelihood estimator $\hat{X}$ corresponding to $\hat{x}$ is a BLUE (Best Linear Unbiased Estimator) of x^* (case of a *linear model* with "error" covariance matrix $\sigma^2\,S^{-1}$ known up to a scale factor σ^2).
- In practice, to compute $\hat{x}$, prefer a QR factorization of $L^T\,H$, where L is a square root or Cholesky factor of S, to a Cholesky factorization of $H^T\,S\,H$ for a better numerical stability.

Well-conditioned problems

Well-posed problem in the sense of Hadamard

1. A solution x^* exists: $H^{-1}(y^*) \neq \emptyset$.
2. The solution is unique: $H^{-1}(y^*) = \{x^*\}$.
3. The solution depends continuously on the data, in some reasonnable topology: H^{-1} is continuous around y^*.

Well-conditioned problem: a practical necessity

In practice, it is necessary to go beyond that latter condition and to ensure that "a small error on y^* may not cause a large error on x^*": the *condition number* of H^{-1} around y^* must be small [†].
If not, you cannot expect to compute an "accurate" solution: the problem need to be *regularized*.

[†] Condition number of f at x: $\lim_{\epsilon \to 0^+} \sup_{\|\delta x\| < \epsilon} \left[\frac{\|f(x+\delta x) - f(x)\|}{\|f(x)\|} \right] / \left[\frac{\|\delta x\|}{\|x\|} \right]$.

Well-conditioned problems

The linear euclidean case

Considering the 2-norm, the condition number $\kappa(A) = |\lambda_{\max}(A)/\lambda_{\min}(A)|$ of the symmetric matrix $A = H^T S H$ must be small. What to do if $\kappa(A)$ is too great?!

Tikhonov regularization

This is the most commonly used method of regularization. The linear euclidean Tikhonov-regularized problem is: find

$$\hat{x}_{x_0,\Lambda} = \arg\min_{x\in\mathbb{R}^Q} \|y^* - Hx\|_S^2 + \|\Lambda(x - x_0)\|_2^2 \Leftrightarrow \underbrace{\left[H^T S H + \Lambda^T \Lambda\right]}_{A_{x_0,\Lambda}} \hat{x}_{x_0,\Lambda} = H^T S y^* + \Lambda^T \Lambda x_0$$

with $\Lambda \in \mathcal{M}_{K,Q}(\mathbb{R})^\dagger$ and $x_0 \in \mathbb{R}^Q$.
For example, choosing by default $\Lambda = \lambda\, \mathrm{I}_Q$, one can try to find the smallest $\lambda \in \mathbb{R}$ such that $\kappa(A_{x_0,\Lambda})$ is "reasonably small".
x_0 appears as some kind of guess of what x^* could be, that is as a prior piece of information! What if one associates a prior distribution $\pi(x)$ on x?

---

†If the rank of Λ is maximal, then $\|\Lambda(x - x_0)\|_2 = \|x - x_0\|_{\Lambda^T\Lambda}$.

Bayesian inference as regularization

We assume the following model:

- Likelihood coming from

$$Y^* = H\,x^* + U \text{ where } U \sim \mathcal{N}(0, {\sigma_U}^2\,S)$$

with $S = \mathrm{I}_P$ (OLS) and known ${\sigma_U}^{\dagger}$:

$$Y^* \sim \mathcal{N}(Hx^*, {\sigma_U}^2\,\mathrm{I}_P);$$

- Prior: $x^* \sim \mathcal{N}(x_0, {\sigma_x}^2\,\mathrm{I}_Q)$.

---

†Variability of the measurement "errors" already estimated.

Bayesian inference as regularization

Model: $Y^* \sim \mathcal{N}(Hx^*, {\sigma_U}^2 \mathrm{I}_P)$ and $x^* \sim \mathcal{N}(x_0, {\sigma_x}^2 \mathrm{I}_Q)$.

Therefore,

$$\begin{aligned}
\pi(x|y^*) &\underset{x}{\propto} \mathcal{L}(x; y^*)\, \pi(x) \\
&\underset{x}{\propto} \exp\left(-\frac{1}{2}\left[\frac{1}{{\sigma_x}^2}(y^* - Hx)^T(y^* - Hx) + \frac{1}{{\sigma_U}^2}(x - x_0)^T(x - x_0)\right]\right) \\
&\underset{x}{\propto} \exp\left(-\frac{1}{2}\left[x^T\Sigma^{-1}x - 2\mu^T\Sigma^1 x\right]\right) \underset{x}{\propto} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)
\end{aligned}$$

with $\Sigma^{-1} = \frac{1}{{\sigma_x}^2}(H^T H + \frac{{\sigma_x}^2}{{\sigma_U}^2}\mathrm{I}_Q)$ and $\mu^T\Sigma^{-1} = \frac{1}{{\sigma_x}^2}{y^*}^T H + \frac{1}{{\sigma_U}^2}{x_0}^T$.

Thus $\mu = \left(H^T H + \frac{{\sigma_x}^2}{{\sigma_U}^2}\mathrm{I}_Q\right)^{-1}\left({y^*}^T H + \frac{{\sigma_x}^2}{{\sigma_U}^2}{x_0}^T\right)$ and

$$x|y^* \sim \mathcal{N}(\hat{x}_{x_0,\Lambda}, \Sigma) \text{ with } \Lambda = \frac{\sigma_x}{\sigma_U}\mathrm{I}_Q.$$

Hence, bayesian inference appears as a way to regularize ill-conditioned problem. Moreover, here, the bayesian approach gives rationale for the choice of Λ and x_0.

Identifiability of statistical models

Statistical model

A statistical model is a triplet $(\mathcal{Y}, \mathcal{T}_{\mathcal{Y}}, \mathcal{P})$ where $\mathcal{T}_{\mathcal{Y}}$ is a σ-algebra on $\mathcal{Y}$ and $\mathcal{P}$ a set of probability distributions on $\mathcal{T}_{\mathcal{Y}}$.

Identifiability of a statistical model

Let $\mathcal{P} = \{P_\theta \ : \ \theta \in \Theta\}$ be a parametrization of $\mathcal{P}^\dagger$, $(\mathcal{Y}, \mathcal{T}_{\mathcal{Y}}, \{P_\theta \ : \ \theta \in \Theta\})$ is said

identifiable iif the mapping $\begin{array}{rccc} \Omega\colon & \Theta & \longrightarrow & \mathcal{P} \\ & \theta & \longmapsto & P_\theta \end{array}$ is injective, that is $\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$.

---

†In practice, defining a statistical model comes to defining a likelihood $\mathcal{L}(\theta; y)$.

Knowing what the "true" distribution $P^* \in \mathcal{P}^\dagger$ is, that is having an infinite amount of data, the identifiability ensures that the inference problem of finding $\theta \in \Theta$ such that $P^* = \Omega(\theta^*)$ satisfies the second condition of Hadamard (uniqueness).

---

†"All models are wrong but some are useful", G.E.P. Box.

Identifiability of statistical models

Consequences of non-identifiability

If the model is not identifiable, it may happen for example that

$$\forall \theta_1 \in \Theta \quad \exists \theta_2 \in \Theta \setminus \{\theta_1\} \quad \forall y \in \mathcal{Y} \quad \mathcal{L}(\theta_1; y) = \mathcal{L}(\theta_2; y)$$

$$\Rightarrow \text{finding } \hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; y^*) \text{ is not a well-posed problem.}$$

Practically, you then cannot decide what is a good estimate of θ^* in general[†] in the non-bayesian (frequentist) framework, where all the decisions relying on θ^* are **only** based on the likelihood.

[†] It is a possibility: you may be able to decide in some cases.

Bayesian inference as regularization against non-identifiability

Bayesian statistical model

Roughly speaking, a *bayesian statistical model* is defined by the "combination" of a parametrized statistical model $(\mathcal{Y}, \mathcal{T}_{\mathcal{Y}}, \{P_\theta \ : \ \theta \in \Theta\})$ with a prior $\pi(\theta)$ which is a probability distribution on $(\Theta, \mathcal{T}_\Theta)$.

Remark

From my point of view,

- $(\mathcal{Y}, \mathcal{T}_{\mathcal{Y}}, \{P_\theta \ : \ \theta \in \Theta\})$ is chosen so as to only model the "true" variability/dispersion/randomness which is postulated by the observer;
- whereas the prior $\pi(\theta)$, chosen in coherence with the available knowledge, represents the epistemic uncertainty (lack of knowledge) on θ^*.

Thus, the use of gaussian processes for deterministic numerical code meta-modelling is intrinsically a bayesian modelling.

Bayesian inference as regularization against non-identifiability

About the well-poseness of the bayesian inference problem

The bayesian inference approach is completely based on the determination of the posterior $\pi(\theta|y^*)$ whose existence and uniqueness are always garanteed (unless the data are incoherent with your model or in some cases of improper prior). Moreover, most of the Bayes rule (minimizing the posterior expected loss) lead to a unique Bayes estimator[†]: the existence and uniqueness of a Bayes estimate $\hat{\theta}$ of θ^* is generally ensured in practice.

[†] Posterior expectation or quantile, typically.

A euclidean linear example, once again

The inference problem (calibration of the simplest code ever)

- Data y^*: P independent 1-dimensional measurements.
- Model: $Y = \tilde{H}\,\tilde{x} + \tilde{U}$ with

$$\tilde{U} \sim \mathcal{N}(0, \sigma^2 \mathsf{I}_P), \quad \tilde{x} = \begin{pmatrix} x \\ \mu_H \\ \mu_U \end{pmatrix} \text{ and } \tilde{H} = \begin{pmatrix} h(d_1) & 1 & 1 \\ \vdots & \vdots & \vdots \\ h(d_P) & 1 & 1 \end{pmatrix}$$

  where
  - x is the scalar parameter of a linear code $H = H^{\text{code}}$ to be calibrated ($H(x,d) = h(d)\,x$);
  - $U_i = \mu_U + \tilde{U}_i \sim \mathcal{N}(\mu_U, \sigma^2 \mathsf{I}_P)$ is the ith measurement error (systematic measurement bias μ_U);
  - μ_H is a systematic[†] bias/difference between the real response $H^{\text{real}}(x,d)$ of the physical system in conditions d and the code prediction $H(x,d)$;
  - d_i is the known/controlled conditions of the observation y_i^*.
- Unknows (3 cases): $\theta = (x,\ \mu_H,\ \mu_U)$, $\theta = (x,\ \mu)$ with $\mu = \mu_H + \mu_U$ or $\theta = x$.

[†] Assumed independent from x and d.

A euclidean linear example, once again

Identifiability: different cases

$$Y_i \sim \mathcal{N}(h(d_i)\,x + \mu_H + \mu_U, \sigma^2 \mathrm{I}_P) \quad \text{(independent, not identically distributed)}$$

(C1) $\theta = (x,\ \mu_H,\ \mu_U)$: non-identifiability (ill-posed OLS problem);
You should entirely calibrate your measurement system before using measurements to calibrate your code.

(C2) $\theta = (x,\ \mu)$ with $\mu = \mu_H + \mu_U$[†]:

(C2.1) $\forall (i,j) \in \{1, \cdots, P\}^2 \quad h(d_i) = h(d_j)$: non-identifiability.
(C2.2) $\exists (i,j) \in \{1, \cdots, P\}^2 \quad h(d_i) \neq h(d_j)$: identifiability.

You should carry out the experiments in different conditions d_i.

(C3) $\theta = x$: identifiability.

In practice, it could be very difficult to carry out experiments in different conditions or to analyse mathematically the identifiability (non-linear code): "potential confonding" (R. Paulo, Lecture 3). Moreover, even in case of non-identifiability, is it impossible to learn from data?

[†] *E.g.* your measurement system was calibrated from a reference ($\mu_U = 0$) or you assume your numerical model H is perfect ($\mu_H = 0$).

A euclidean linear example, once again

Identifiability: a practical (toy) example

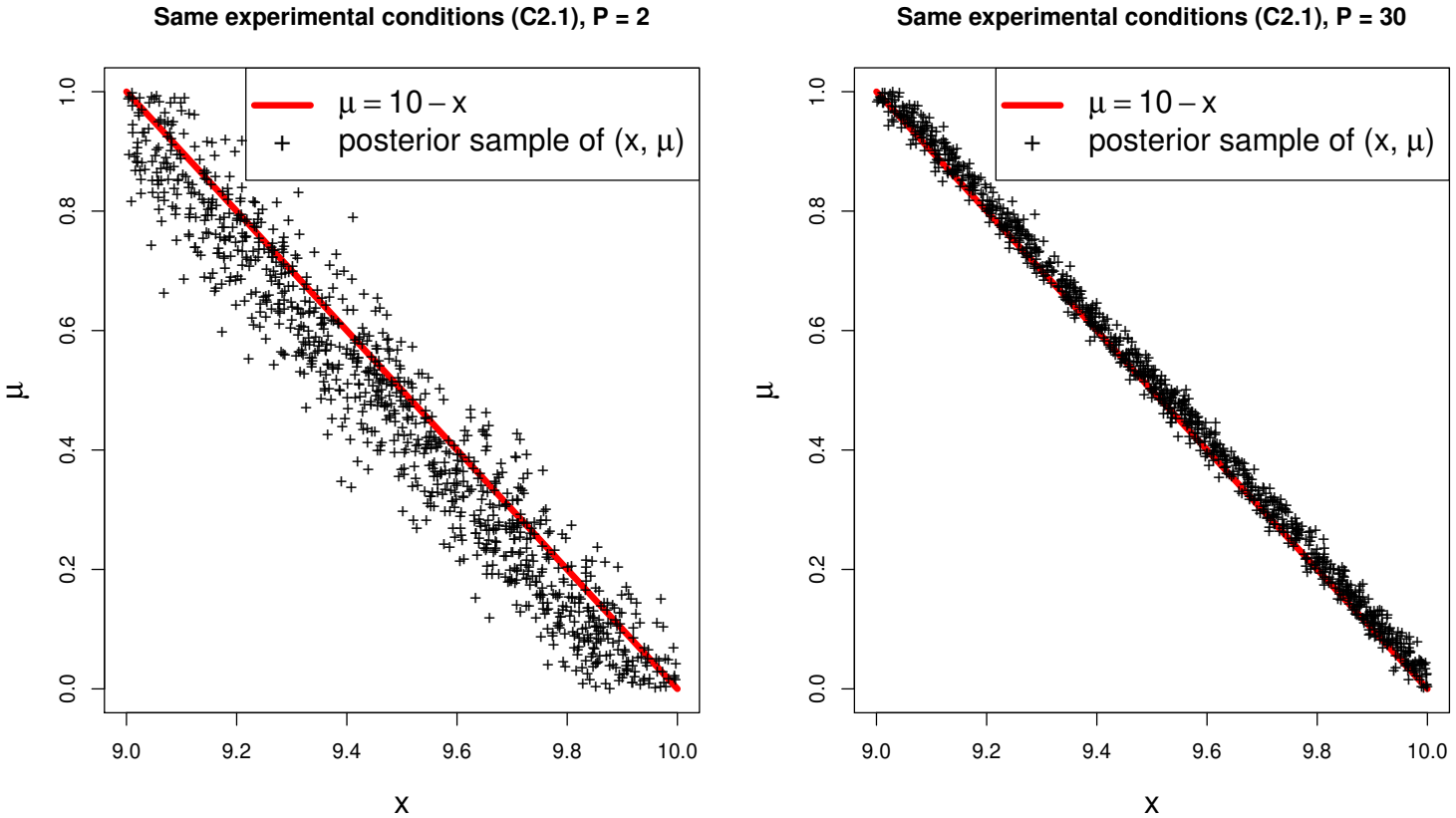
- P generated data according to $Y_i \sim \mathcal{N}(d_i\, x^* + \mu^*, \sigma^2 \mathsf{I}_P)$ $\quad (h(d_i) = d_i)$, with $x^* = 9.5$, $\mu^* = 0.5$, $d_i \in \{1, 2\}$ and $\sigma = 0.1$.
- Uniform proper prior: $x, \mu \sim \mathcal{U}([9, 10] \times [0, 1])$.
- Three data sets[†]:
  - $P = 2$ measurements with $d_1 = d_2 = 1$ (same conditions (C2.1));
  - $P = 30$, $d_1 = \cdots = d_{30} = 1$ (same conditions (C2.1));
  - $P = 30$, $\begin{cases} d_1 = \cdots = d_{15} = 1 \\ d_{16} = \cdots = d_{30} = 2 \end{cases}$ (different conditions (C2.2)).
- Lazy choice: Metropolis-Hastings sampling with isotropic random walk (next slides: $1,000$ draws from $100,000$).

[†] They share as many data as possible (2 or 15).

A euclidean linear example, once again

Identifiability: a practical (toy) example

- $x^* = 9.5$, $\mu^* = 0.5$; uniform prior over $[9, 10] \times [0, 1]$.



Even in case of non-identifiability, you can learn from data with bayesian statistics and carry out computations.

A euclidean linear example, once again

Identifiability: a practical (toy) example

- $x^* = 9.5$, $\mu^* = 0.5$; uniform prior over $[9, 10] \times [0, 1]$.

With same conditions $d_i = 1$, we can learn from data that $x^* + \mu^* = 10$: if we want to use H to predict, the result is interesting if we assume $\mu_U = 0$ (calibrated measurement system), but not if we assume $\mu_H^* = 0$.
In the former case ($x^* + \mu_H^* = 10$), you can predict correctly, but only for $d = 1$.

Generally, you should use a "calibrated" code for prediction **only** in similar conditions than the ones of the data taken into account.

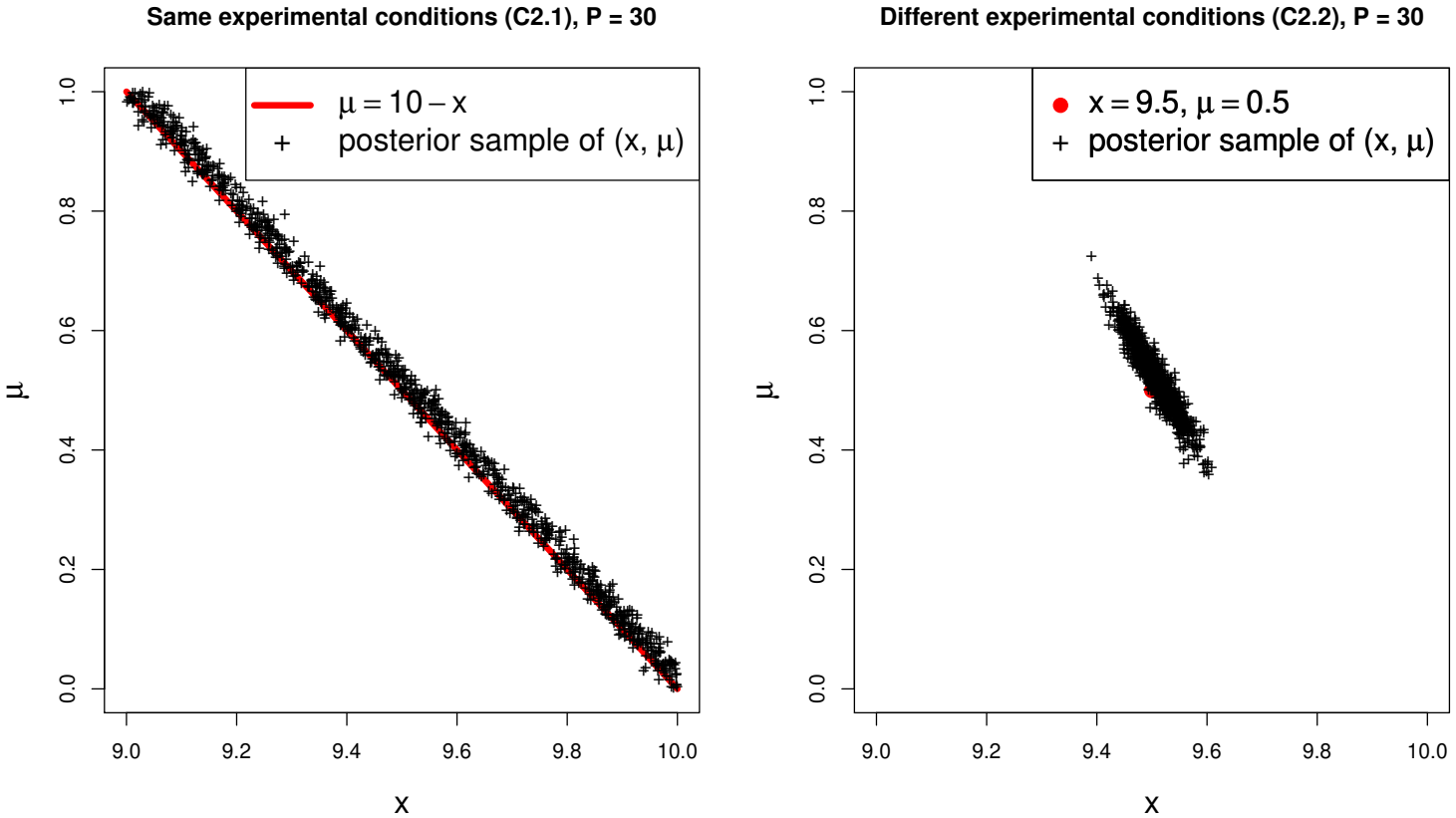
It appears very important to **validate your code after calibration** from experiments in new conditions. **A calibration & validation process should involve three successive steps: calibration (first data set), validation (second data set), then re-calibration (to take benefit of all data sets).**[†]

[†] The situation is similar when meta-modelling; you can also perform k-fold cross validation.

A euclidean linear example, once again

Identifiability: a practical (toy) example

- $x^* = 9.5$, $\mu^* = 0.5$; uniform prior over $[9, 10] \times [0, 1]$.



Although bayesian statistics has a strong power of regularization, you should not neglect the identifiability: it is useful to improve the learning process.

Contents

Conclusions

Except in ideal situations (perfect data, perfect code H, well-posed problem), you should consider statistics to deal with your inverse problems, since it enables:

- to choose other hypothesis for the "errors" (non-gaussian ones) and ways to test and compare them;
- the possibility to estimate confidence intervals or credibility intervals for x^*;
- to deal with various situations of uncertainty (random inputs, not well-known experimental conditions).

In particular, the bayesian statistics methodology is a convenient one:

- to regularize ill-conditioned problem, to deal with non-identifiability;
- to assimilate data sequentially;
- to adjust the decision making (*e.g.* extracting an estimate $\hat{x}$ from the posterior) by the choice of a loss function as relevant as possible;
- to produce interesting results from little data by combining it with expertise.

It is an idealized picture... practice may be something else. In particular, the price to pay may be a great need in computing resources.

Contents

Random inputs: an example (ANR project OPUS)

Description

- Physical system: 50 km of river (Garonne) ;
- Saint-Venant 1D steady eq., composed bed (EDF code Mascaret):

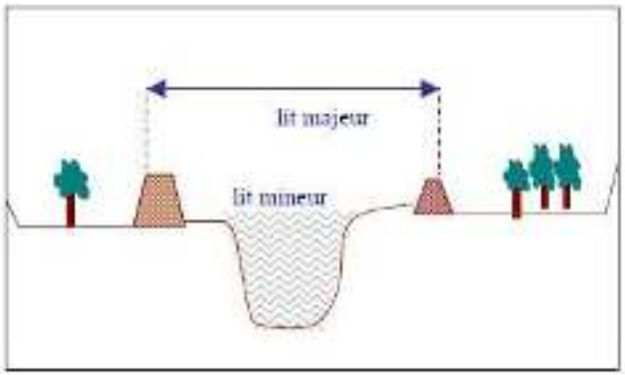
$$z(s) = M(Ks^{\text{minor}}(s), Ks^{\text{major}}(s), Q)$$

où z = height of water, Ks = Strickler coefficient and Q = rate of flow.

- constant Ks all along three river parts, only uncertain along the third:

$$y = \begin{pmatrix} z_{\text{Mas d'Agenais}} \\ z_{\text{Marmande}} \end{pmatrix} = H(Ks, Q) \text{ where } Ks \in \mathbb{R}^2$$

- $n = 100$ measurements



Random inputs: an example (ANR project OPUS)

Find $\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(y\ ; \theta)$ with $\mathcal{L}(y\ ; \theta) = \int_{M^{-1}(z)} \mathcal{L}(z\ ; \theta) \mathrm{d}z$;

E.g. θ corresponds to the unknown parameters of the probability distribution of some random inputs $Z = X$ of a perfect code H and the measurement system was calibrated.

SEM (Stochastic EM) [Celeux & Diebolt, 86]

1. Simulation:
   1. Simulation of $Z \,|\, Y = y$ and $\theta = \theta^k$: $z^k \sim \mathcal{L}(z; \theta^k)$
      $\longrightarrow$ Use of a MCMC method (non-linear H): meta-modelling could be needed
   2. $\theta^{k+1} = \arg\max_{\theta} \mathcal{L}(z^k\ ; \theta^k)$

   $\longrightarrow$ Homogeneous Markov chain which concentrates around the stationnary points of $\mathcal{L}(y\ ; .)$: no ponctual convergence
2. Estimate, generally: $\hat{\theta} = \frac{1}{J-j+1} \sum_{k=j}^{J} \theta^k$

Random inputs: an example (ANR project OPUS)

Application

- $Z = X = Ks \in \mathbb{R}^2$ (Strickler coefficients) with unknown randomness $\mathcal{N}(\mu, \Sigma)$
- We used SAEM-MCMC (Stochastic Approximation EM) [Delyon, Lavielle & Moulines, 99][Kuhn & Lavielle, 03] [Allassonnière & Kuhn, 10] to force the Markov chain to converge ponctually.
- 100 first raws: SEM (chain not forced yet)
- Some results for μ: