

Problèmes inverses

G. Perrin

guillaume.perrin@univ-eiffel.fr

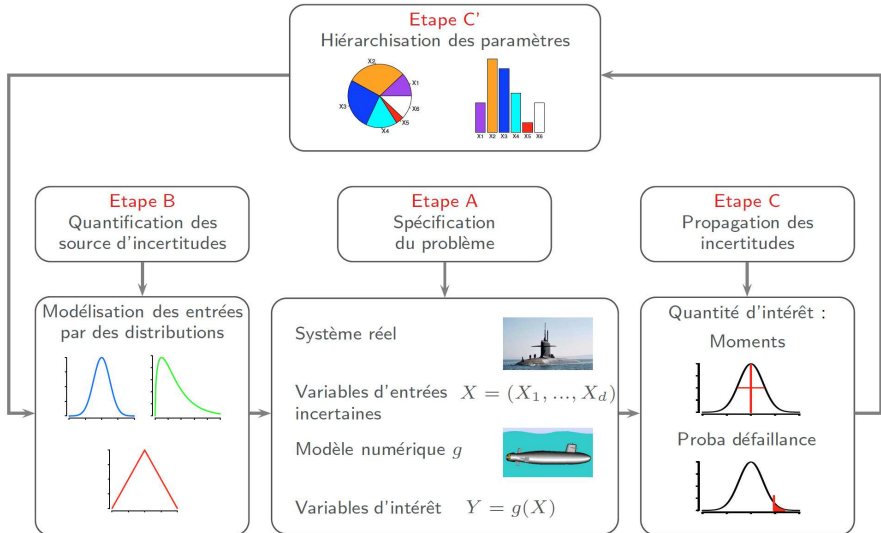
Année scolaire 2022-2023



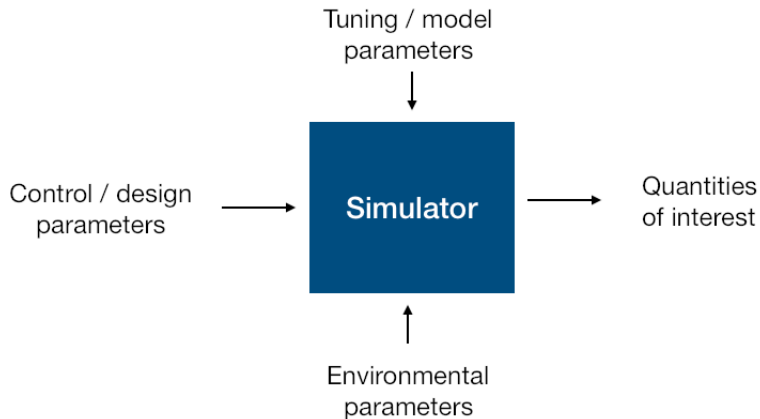
Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire
- 4 Consistance du maximum a posteriori
- 5 Application à la calibration de codes de calcul
- 6 Conclusions et perspectives

Les problèmes inverses au sein de la démarche incertitudes



Les problèmes inverses au sein des problèmes UQ



Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire
- 4 Consistance du maximum a posteriori
- 5 Application à la calibration de codes de calcul
- 6 Conclusions et perspectives

Introduction

Un problème inverse se formule généralement ainsi : on cherche $\mathbf{x} \in \mathbb{X}$, l'entrée d'un modèle, étant donnée une observation (ou une série d'observations) $\mathbf{y} \in \mathbb{Y}$ de la sortie de ce modèle :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}, \epsilon),$$

avec $\mathbf{m} : \mathbb{X} \rightarrow \mathbb{Y}$ une fonction déterministe, \mathbb{X}, \mathbb{Y} deux espaces métriques avec $\|\cdot\|_{\mathbb{X}}$ et $\|\cdot\|_{\mathbb{Y}}$ leurs normes respectives, et ϵ une potentielle erreur.

Exemples

- estimation de paramètres de modèle (**calibration**) à partir de mesures expérimentales bruitées,
- identification de conditions initiales/limites pour des systèmes dynamiques (localisation de sources de polluants, identification de sollicitations,...) à partir de modèles approchés,
- ...

Introduction

Lorsque \mathbf{m} peut s'écrire sous la forme :

$$\mathbf{m}(\mathbf{x}, \boldsymbol{\varepsilon}) = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon},$$

une approche classique de résolution de problème inverse est de le formuler sous la forme d'un problème moindres carrés :

$$\arg \min_{\mathbf{x} \in \mathbb{X}} \|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\mathbb{Y}}^2.$$

- ⇒ De nombreuses méthodes, directes ou itératives, existent pour résoudre ce problème de minimisation.
- ⇒ Cette approche moindres carrés peut néanmoins **échouer** lorsque le problème est **mal posé**, ce qui peut se traduire par le fait que ces méthodes ne convergent pas, semblent converger vers un point hors de \mathbb{X} , ou convergent de manière peu répétable en étant très (trop) sensibles aux valeurs de \mathbf{y} ...

Problème mal posé et régularisation

On dit d'un problème qu'il est **mal posé** (au sens de Hadamard) si :

- il n'admet pas de solution,
- la solution n'est pas unique,
- la solution dépend de manière très sensible à \mathbf{y} .

Pour éviter ces problèmes, il est souvent proposé de **régulariser** le problème des moindres carrés en le remplaçant par (régularisation de Tikhonov) :

$$\arg \min_{\mathbf{x} \in \mathbb{X}} \|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\mathbb{Y}}^2 + \lambda \|\mathbf{x} - \mathbf{x}_0\|_{\mathbb{X}}^2, \lambda > 0.$$

- La pénalisation ajoutée à l'erreur quadratique représente l'a priori que l'utilisateur a sur la structure de la solution.
- Pour λ suffisamment grand, le problème devient fortement convexe, et par conséquent bien posé. Mais si λ est trop grand, on ne résoud plus le problème initial...

Formulation bayésienne

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}.$$

- Une autre façon de résoudre ce problème de problème mal posé est d'adopter un point de vue **bayésien**.
- L'erreur $\boldsymbol{\varepsilon}$ est alors modélisée sous la forme d'un vecteur aléatoire, et l'idée n'est alors plus de chercher une réponse **unique**, mais de chercher une **mesure de probabilité** $\mu(\cdot|\mathbf{y})$ sur \mathbb{X} caractérisant la probabilité relative de l'état \mathbf{x} étant donné \mathbf{y} , appelée mesure **a posteriori** de \mathbf{x} .
- Selon cette approche, le terme de pénalisation s'interprète comme une mesure **a priori**, ce qui permet notamment de donner davantage de structure à la solution.

Formulation bayésienne

Dans la suite, on se limite au cas où

- $\mathbb{X} \subset \mathbb{R}^d$ et $\mathbb{Y} \subset \mathbb{R}^n$,
- la mesure a priori de \mathbf{x} , notée μ^0 , est supposée admettre une densité π_0 sur \mathbb{X} ,
- ε est supposée admettre une densité ρ sur \mathbb{Y} .

Dans ce cas, la vraisemblance (c.a.d. la loi de \mathbf{y} sachant \mathbf{x}) a pour densité $\rho(\mathbf{y} - \mathbf{f}(\mathbf{x}))$. Le théorème de Bayes nous indique alors que la mesure a posteriori $\mu(\cdot|\mathbf{y})$ admet également une densité $\pi(\cdot|\mathbf{y})$ telle que :

$$\pi(\mathbf{x}|\mathbf{y}) = \frac{\rho(\mathbf{y} - \mathbf{f}(\mathbf{x}))\pi_0(\mathbf{x})}{\int_{\mathbb{X}} \rho(\mathbf{y} - \mathbf{f}(\mathbf{x}'))\pi_0(\mathbf{x}')d\mathbf{x}'}, \quad \mathbf{x} \in \mathbb{X}.$$

Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire
- 4 Consistance du maximum a posteriori
- 5 Application à la calibration de codes de calcul
- 6 Conclusions et perspectives

Cas gaussien

Hypothèses : supposons que (résidu) $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ et que (loi a priori de \mathbf{x}) $\mathbf{x} \sim \mathcal{N}(\mathbf{x}_0, \mathbf{\Sigma}_0)$.

On déduit ($\approx \leftrightarrow$ égalité à une constante multiplicative près) :

$$\pi(\mathbf{x}|\mathbf{y}) \approx \exp \left(-\frac{1}{2} \|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\mathbf{\Gamma}}^2 - \frac{1}{2} \|\mathbf{x} - \mathbf{x}_0\|_{\mathbf{\Sigma}_0}^2 \right),$$

avec $\|\mathbf{z}\|_{\mathbf{R}}^2 = \mathbf{z}^T \mathbf{R}^{-1} \mathbf{z} = (\mathbf{R}^{-1/2} \mathbf{z})^T \mathbf{R}^{-1/2} \mathbf{z} = \left\| \mathbf{R}^{-1/2} \mathbf{z} \right\|^2$.

Remarques

- Si l'on se concentre sur le mode de $\pi(\mathbf{x}|\mathbf{y})$, on trouve le Maximum A Posteriori (MAP), équivalent au problème de minimisation moindres carrés régularisé (**le prior jouant le rôle de pénalisation**) :

$$\arg \min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\mathbf{\Gamma}}^2 + \|\mathbf{x} - \mathbf{x}_0\|_{\mathbf{\Sigma}_0}^2.$$

- La loi $\pi(\mathbf{x}|\mathbf{y})$ est a priori non gaussienne (sauf si \mathbf{f} est linéaire), et apporte **beaucoup plus d'information** que le MAP (**confiance sur l'estimation**) !

Problème équilibré

Lorsque $\mathbb{X} = \mathbb{R}^d$ et $\mathbb{Y} = \mathbb{R}^n$ et $d = n$, si :

- f est $\mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^n)$,
- $y = f(x)$ admet une unique solution notée $x^*(y)$,
- il existe $C > 0$ tel que pour $y, \delta \in \mathbb{R}^n$,

$$\|y - f(x^*(y) + \delta)\|^2 \geq C \min(1, \|\delta\|^2),$$

- $\epsilon \sim \mathcal{N}(0, \gamma^2 I_n)$,

alors $\pi(x|y)$ converge faiblement vers $\delta_{x^*(y)}$: l'incertitude **disparaît**, et le prior ne joue **aucun rôle**.

Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire**
- 4 Consistance du maximum a posteriori
- 5 Application à la calibration de codes de calcul
- 6 Conclusions et perspectives

Cas linéaire

Supposons maintenant que $\boldsymbol{\varepsilon} = \mathbf{0}$, $\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Gamma}$ (sans que $\boldsymbol{\varepsilon}$ soit forcément gaussien), et qu'il existe une matrice \mathbf{A} de taille $(n \times p)$ telle que :

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}.$$

Proposition

Si $\mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{A}$ est inversible, alors

$$\hat{\mathbf{x}} := (\mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{y}$$

minimise $\mathbb{E} [\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ parmi tous les estimateurs linéaires non biaisés de \mathbf{x} sachant \mathbf{y} , et on a :

$$\text{Cov}(\hat{\mathbf{x}}) = (\mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{A})^{-1}.$$

$\hat{\mathbf{x}}$ est appelé **BLUE** pour "best linear unbiased estimator".

Cas linéaire gaussien

Si $\boldsymbol{\varepsilon}$ est **gaussien**, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$, on peut également montrer (même si $\mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{A}$ n'est pas inversible ou est mal-conditionnée) que $\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_y, \mathbf{C}_y)$, avec (formule de Woodbury) :

$$\begin{aligned}\boldsymbol{\mu}_y &= (\mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_0^{-1})^{-1} (\mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{y} + \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_0) \\ &= \mathbf{x}_0 + \boldsymbol{\Sigma}_0 \mathbf{A}^T (\boldsymbol{\Gamma} + \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}^T)^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_0),\end{aligned}$$

$$\begin{aligned}\mathbf{C}_y &= (\mathbf{A}^T \boldsymbol{\Gamma}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_0^{-1})^{-1} \\ &= \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0 \mathbf{A}^T (\boldsymbol{\Gamma} + \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}^T)^{-1} \mathbf{A} \boldsymbol{\Sigma}_0.\end{aligned}$$

Exercice : montrer ce résultat.

Influence du bruit

Si $\mathbf{\Gamma} = \gamma^2 \mathbf{\Gamma}_0$, les expressions précédentes se récrivent :

$$\boldsymbol{\mu}_y = \mathbf{x}_0 + \gamma^{-2} \boldsymbol{\Sigma}_0 \mathbf{A}^T (\mathbf{\Gamma}_0 + \gamma^{-2} \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}^T)^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_0),$$

$$\mathbf{C}_y = (\mathbf{A}^T \mathbf{\Gamma}_0^{-1} \mathbf{A} \gamma^{-2} + \boldsymbol{\Sigma}_0^{-1})^{-1}.$$

- L'incertitude a posteriori est plus faible que l'incertitude a priori.
- Si $\gamma \rightarrow +\infty$, $\boldsymbol{\mu}_y$ tend vers \mathbf{x}_0 et \mathbf{C}_y tend vers $\boldsymbol{\Sigma}_0$: on retrouve l'a priori.
- Si $\gamma \rightarrow 0$, on s'attend à ce que le poids du prior diminue vis-à-vis des mesures.

Réduction du bruit : $\Gamma = \gamma^2 \Gamma_0$

- ① Si $n \geq p$ et \mathbf{A} de rang p , alors $\pi(\mathbf{x}|\mathbf{y})$ converge faiblement vers $\delta_{\mathbf{x}^*}(\mathbf{y})$ quand $\gamma \rightarrow 0$, avec :

$$\mathbf{x}^*(\mathbf{y}) := \arg \min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\Gamma_0}^2.$$

\Rightarrow l'incertitude disparaît et le prior ne joue plus.

- ② Si $n < p$, alors \mathbf{A} est une matrice de rang au plus n . On peut alors l'écrire sous la forme $\mathbf{A}_0 \mathbf{Q}_1^T$, avec \mathbf{A}_0 une matrice $(n \times n)$ inversible, et \mathbf{Q}_1 une matrice de taille (p, n) telle que $\mathbf{Q} = [\mathbf{Q}_1 \ \mathbf{Q}_2]$ est une matrice orthogonale. Alors $\pi(\mathbf{x}|\mathbf{y})$ converge faiblement vers $\mathcal{N}(\mathbf{x}_y^+, \mathbf{C}_y^+)$ quand $\gamma \rightarrow 0$, avec :

$$\mathbf{x}_y^+ := \mathbf{Q} \cdot (\mathbf{z}, \mathbf{z}'), \quad \mathbf{C}_y^+ := \mathbf{Q}_2 \mathbf{L}_{22}^{-1} \mathbf{Q}_2^T,$$

$$\mathbf{z} := \mathbf{Q}_1^T \mathbf{A}_0^{-1} \mathbf{y}, \quad \mathbf{z}' := -\mathbf{L}_{22}^{-1} \mathbf{L}_{12}^T \mathbf{z} + \mathbf{L}_{22}^{-1} \mathbf{Q}_2^T \Sigma_0^{-1} \mathbf{x}_0, \quad \mathbf{Q}^T \Sigma_0^{-1} \mathbf{Q} = \begin{pmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{pmatrix}$$

\Rightarrow l'incertitude ne disparaît pas et le prior joue !

Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire
- 4 Consistance du maximum a posteriori**
- 5 Application à la calibration de codes de calcul
- 6 Conclusions et perspectives

Objectif

- L'idée de cette partie est de montrer qu'une approche bayésienne permet de trouver la "bonne" solution notée \mathbf{x}^* **quelque soit le prior choisi** (sous quelques hypothèses techniques) à partir du moment où il y a **beaucoup** de données.
 - On se place dans le cas où :
 - la loi a priori de \mathbf{x} est notée μ^0 (par forcément gaussienne),
 - l'information disponible s'écrit $\mathbf{y} = \mathbf{f}(\mathbf{x}^*) + \boldsymbol{\varepsilon}$, avec $\boldsymbol{\varepsilon}$ de loi connue (par forcément gaussienne) et \mathbf{x}^* inconnu,
 - On observe N observations de \mathbf{y} , notées $\mathbf{y}_1, \dots, \mathbf{y}_N$, associées à N réalisations iid de $\boldsymbol{\varepsilon}$.
- ⇒ on veut alors montrer que la loi de $\mathbf{x} | \mathbf{y}_1, \dots, \mathbf{y}_N$ converge vers $\delta_{\mathbf{x}^*}$ quand $N \rightarrow +\infty$.

Pour simplifier, on suppose que μ^0 et la loi de $\boldsymbol{\varepsilon}$ admettent des densités π_0 et ρ , si bien que :

$$\pi_N(\mathbf{x} | \mathbf{y}) \approx \pi_0(\mathbf{x}) \prod_{i=1}^N \rho(\mathbf{y}_i - \mathbf{f}(\mathbf{x})).$$

Convergence du maximum de vraisemblance

Proposition

Si \mathbb{X} est compact, que \mathbf{f} est continue et injective, et que ρ est continue et strictement positive, alors :

$$\hat{\mathbf{x}}_N \in \arg \max_{\mathbf{x} \in \mathbb{X}} \prod_{i=1}^N \rho(\mathbf{y}_i - \mathbf{f}(\mathbf{x}))$$

converge en probabilité vers \mathbf{x}^* quand N tend vers $+\infty$.

Remarque : l'estimateur du maximum de vraisemblance peut ne pas exister (ou ne pas être unique) pour des valeurs de N trop petites.

Convergence en loi

Proposition

Si de plus, \mathbf{x}^* est dans l'intérieur de \mathbb{X} , alors on peut montrer que sous des conditions de régularité sur \mathbf{f} et ρ (pour que la matrice de Fisher soit bien inversible),

$$\sqrt{N}(\hat{\mathbf{x}}_N - \mathbf{x}^*) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{I}(\mathbf{x}^*)),$$

avec $\mathbf{I}(\mathbf{x}^*)$ la matrice d'information de Fisher de taille $p \times p$.

\Rightarrow **convergence en loi vers une loi gaussienne du maximum de vraisemblance.**

Concentration de la loi asymptotique

Théorème

Sous les conditions précédentes, et si π_0 est continue et strictement positive au voisinage de \mathbf{x}^* , alors pour tout $\varepsilon > 0$:

$$\mathbb{P}(\|\pi_N - \mathcal{N}(\hat{\mathbf{x}}_N, I_N(\mathbf{x}^*)/N)\|_{TV} > \varepsilon) \xrightarrow{N \rightarrow +\infty} 0,$$

où pour toutes mesures μ, ν , la distance en variation totale est donnée par :

$$\|\mu - \nu\|_{TV} = \sum_{B \in \mathcal{B}(\mathbb{R}^p)} |\mu(B) - \nu(B)|.$$

\Rightarrow la loi a posteriori va se concentrer autour de la vraie valeur \mathbf{x}^* .

Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire
- 4 Consistance du maximum a posteriori
- 5 Application à la calibration de codes de calcul**
- 6 Conclusions et perspectives

Calibration de codes de calcul

Notations

- On dispose d'un code numérique paramétré $\mathbf{z} \mapsto f(\mathbf{z}; \mathbf{x})$, où l'on distingue deux types d'entrées :
 - les variables de fonctionnement \mathbf{z} , qui définissent les conditions expérimentales (température, pression, dimensions, CL...).
 - les paramètres de calibration \mathbf{x} , qu'il est nécessaire de définir pour lancer le code (constantes universelles, paramètres numériques, ...).
- On dispose d'un phénomène physique mesurable associé au code, $\mathbf{z} \mapsto y(\mathbf{z})$.
- On note $y^{\text{mes}}(\mathbf{z}^{(n)})$, $1 \leq n \leq N$, N mesures de ce phénomène physique, potentiellement entachées d'une erreur de mesure $\varepsilon^{\text{mes}}(\mathbf{z}^{(n)})$.
- On note $\mathbf{z} \mapsto \varepsilon^{\text{mod}}(\mathbf{z})$ l'erreur de modèle.

Calibration de codes de calcul

On retrouve le formalisme précédent, $\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}$, en introduisant :

$$\mathbf{y} := (y^{\text{mes}}(\mathbf{z}^{(1)}), \dots, y^{\text{mes}}(\mathbf{z}^{(N)})), \quad \mathbf{f}(\mathbf{x}) := (f(\mathbf{z}^{(1)}; \mathbf{x}), \dots, f(\mathbf{z}^{(N)}; \mathbf{x})),$$

$$\boldsymbol{\varepsilon} := (\varepsilon^{\text{mes}}(\mathbf{z}^{(n)}) + \varepsilon^{\text{mod}}(\mathbf{z}^{(1)}), \dots, \varepsilon^{\text{mes}}(\mathbf{z}^{(N)}) + \varepsilon^{\text{mod}}(\mathbf{z}^{(N)})).$$

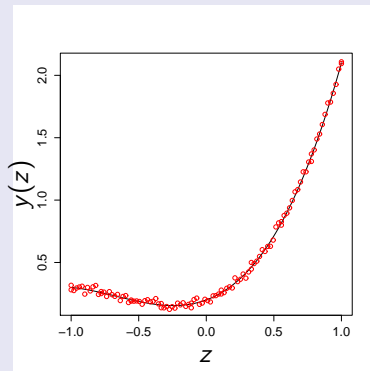
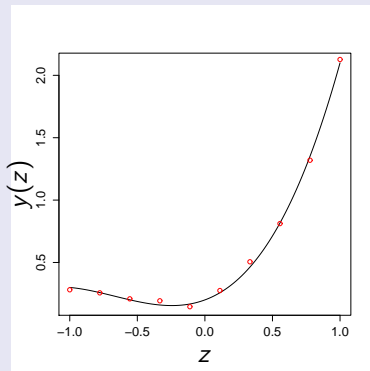
- $\boldsymbol{\varepsilon}$ intègre à la fois l'erreur de mesure et l'erreur de modèle.
- Introduire une erreur de modèle est souvent nécessaire, mais n'est cependant pas sans poser des problèmes d'identifiabilité : plusieurs couples $(\mathbf{x}, \varepsilon^{\text{mod}})$ peuvent conduire à la même somme.
- Le formalisme bayésien introduit à nouveau un cadre formel bien posé pour l'identification de \mathbf{x} , en considérant ε^{mod} comme un processus aléatoire de loi à identifier (souvent choisi gaussien centré, de covariance C).

Importance de l'erreur de modèle

Exemple

$$y(z) = 0.2 + 0.4z + z^2 + 0.5z^3, \quad -1 \leq z \leq 1, \quad \gamma = 0.02.$$

A priori "peu informatif" : $\mathbf{x}_0 = (0, 0, 0, 0)$, $\mathbf{\Sigma}_0 = [I_4]$.



Importance de l'erreur de modèle

Exemple

$$y(z) = 0.2 + 0.4z + z^2 + 0.5z^3, \quad -1 \leq z \leq 1, \quad \gamma = 0.02.$$

	x_1	x_2	x_3	x_4
référence	0.2	0.4	1	0.5
$\mathbf{x} \mathbf{y}, N = 10$	0.19 ± 0.019	0.45 ± 0.050	1.0 ± 0.035	0.45 ± 0.064
$\mathbf{x} \mathbf{y}, N = 100$	0.20 ± 0.0059	0.40 ± 0.017	1.0 ± 0.013	0.51 ± 0.025

TABLE: Sans erreur de modèle : $\mathbf{f} = (1, z, z^2, z^3)$ (IC 95%)

	x_1	x_2	x_3	x_4
référence	0.2	0.4	1	0.5
$\mathbf{x} \mathbf{y}, N = 10$	0.21 ± 0.019	0.77 ± 0.019	0.98 ± 0.035	
$\mathbf{x} \mathbf{y}, N = 100$	0.20 ± 0.0059	0.71 ± 0.0067	1.0 ± 0.013	

TABLE: Avec erreur de modèle : $\mathbf{f} = (1, z, z^2)$ (IC 95%)

Avec erreur de modèle, $\mathbf{x}|\mathbf{y}$ ne converge pas forcément vers β !

Importance de l'erreur de modèle

Exemple

$$y(z) = 0.2 + 0.4z + z^2 + 0.5z^3, \quad -1 \leq z \leq 1, \quad \gamma = 0.02.$$

	x_1	x_2	x_3	x_4
référence	0.2	0.4	1	0.5
$\mathbf{x} \mathbf{y}, N = 10$	0.26 ± 0.52	0.74 ± 0.55	0.81 ± 0.82	
$\mathbf{x} \mathbf{y}, N = 100$	0.25 ± 0.51	0.72 ± 0.52	0.85 ± 0.79	

TABLE: $\mathbf{f} = (1, x, x^2)$, covariance stationnaire (IC 95%)

	x_1	x_2	x_3	x_4
référence	0.2	0.4	1	0.5
$\mathbf{x} \mathbf{y}, N = 10$	0.19 ± 0.027	0.53 ± 0.097	1.1 ± 0.31	
$\mathbf{x} \mathbf{y}, N = 100$	0.20 ± 0.0096	0.40 ± 0.051	0.99 ± 0.23	

TABLE: $\mathbf{f} = (1, x, x^2)$, covariance non stat. en $(x^3)^2$ (IC 95%)

Avec une "mauvaise" covariance, $\mathbf{x}|\mathbf{y}$ ne converge pas forcément vers β !

Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire
- 4 Consistance du maximum a posteriori
- 5 Application à la calibration de codes de calcul
- 6 Conclusions et perspectives**

Conclusions

- Inversion déterministe \neq Inversion bayésienne.
- Bien faire la distinction entre les objectifs de l'estimation (on recherche la vraie valeur \mathbf{x}^* de \mathbf{x}) et de la prédiction (on recherche une valeur de \mathbf{x} permettant la meilleure prédiction de y par $f(\mathbf{z}; \mathbf{x})$ pour un couple (\mathbf{z}, y) non observé).
- Sans erreur de modèle, estimation et prédiction conduisent généralement aux mêmes résultats.
- Avec erreur de modèle, il est primordial (même si c'est difficile) de chercher à proposer une structure pour cette erreur la plus adaptée possible au problème étudié (tout en évitant si possible de mélanger son identification avec celle de $\mathbf{x} \rightarrow$ utilisation de vraisemblance restreinte pour les modèles linéaires).
- Dans une perspective de calibration, la séparation entre erreur de modèle et erreur d'estimation sur \mathbf{x} n'est pas toujours évidente.

Plan de la séance

- 1 Introduction
- 2 Cas gaussien
- 3 Cas linéaire
- 4 Consistance du maximum a posteriori
- 5 Application à la calibration de codes de calcul
- 6 Conclusions et perspectives