

Stratégie bayésienne pour l'estimation d'un paramètre en thermique

Sujet de travail pratique

Préparation : avant la séance répondez aux questions 1, 5, 6 et 8.

En particulier dans la pratique de l'ingénieur, nombre de phénomènes physiques réels sont modélisés par des équations aux dérivées partielles décrivant l'évolution d'une quantité d'intérêt, avec les variables spatiales et temporelle. C'est le cas par exemple :

- des équations de Maxwell *e.g.*, en électro-magnétisme,
- de la mécanique hamiltonienne,
- des équations de Navier-Stokes pour la mécanique des fluides,
- de l'équation de la chaleur pour l'étude des transferts thermiques.

Pour l'exemple, ce travail concerne les phénomènes thermiques mais sa portée est générique et la démarche s'applique aisément aux divers domaines évoqués précédemment.

1 Un peu de thermique

D'une manière générale, l'étude des phénomènes en question s'appuie sur la température T d'un milieu étudié comme fonction de trois variables d'espace et d'une variable de temps. Pour simplifier les choses on se contente ici d'une seule dimension spatiale notée x : on a donc une fonction d'intérêt $T(x, t)$. Dans ce contexte, l'équation considérée résulte d'un bilan thermique et se met sous la forme :

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} + \beta(T - T_a) = 0. \quad (1)$$

Les deux premiers termes décrivent le phénomène de transport par conduction en reliant évolution temporelle et évolution spatiale. Le troisième terme décrit l'échange par convection avec le milieu ambiant à la température T_a . Cette équation fait intervenir deux coefficients.

1. Le coefficient α est la diffusivité thermique exprimée en $m^2 s^{-1}$. Il se déduit essentiellement de la conductivité thermique et de la capacité thermique. Le premier caractérise l'aptitude à transmettre la chaleur et le second à la stocker.
2. Par ailleurs, β est le coefficient d'échange convectif avec le milieu ambiant. Il est naturellement relié, entre autres choses, aux caractéristiques géométriques (forme, volume, surface, ...) du système. Il s'exprime en s^{-1} .

Ces deux coefficients sont caractéristiques des matériaux considérés. Ce sont les paramètres d'intérêt dans ce travail.

Il existe une énorme quantité de résultats concernant cette famille d'équations. Ils sont en particulier dédiés à l'étude de l'existence et de l'unicité de solutions ainsi qu'aux propriétés de

ces solutions dans divers contextes : géométrie et constitution du système, conditions initiales ou aux bords, ... Il existe également de nombreuses méthodes de résolution que l'on peut ranger dans deux catégories.

1. Des réponses « explicites », comme on peut en obtenir par décomposition en série de Fourier, transformée de Laplace, séparation des variables, ... Elle fournissent une expression de $T(x, t)$ fondée sur des fonctions mathématiques usuelles.
2. Des méthodes « numériques », typiquement par éléments finis ou différences finies, souvent itératives. ... Pour un jeu de localisations spatio-temporelles (x_n, t_n) données, elles produisent directement des valeurs numériques $T(x_n, t_n)$.

Le présent travail ne concernera pas ces questions mais il se focalisera sur la question de l'estimation des paramètres α ou β ou une combinaison des deux, à partir de mesures expérimentales.

La suite des développements pourrait être menée aussi bien dans un cas où on dispose d'une solution « explicite » que dans un cas où on ne dispose que d'une solution « numérique ». Pour la simplicité des développements ultérieurs, on n'aborde que le cas où la solution est « explicite ».

Le problème peut être traité dans sa généralité spatio-temporelle mais pour simplifier un peu plus les développements et se focaliser sur le problème de l'estimation de paramètre lui-même, on considère le cas du régime établi. L'équation devient alors simplement :

$$\frac{\partial^2 T}{\partial x^2} - \omega^2(T - T_a) = 0 \quad \text{avec } \omega^2 = \frac{\beta}{\alpha} \quad (2)$$

qui est une équation différentielle relativement standard. Elle est pilotée par le paramètre ω , exprimé en m^{-1} , et c'est le paramètre d'intérêt que l'on cherche à estimer dans ce travail.

Pour terminer de préciser le contexte, on considère le problème dans un barreau de longueur L donnée, situé entre les positions $x = 0$ et $x = L$. On impose deux conditions aux bords en température : $T(0) = T_0$ et $T(L) = T_L$, les deux températures T_0 et T_L étant données. On montre alors que l'équation se résout en :

$$T(x) = T_a + \frac{(T_L - T_a) \sinh \omega x - (T_0 - T_a) \sinh \omega(x - L)}{\sinh \omega L}. \quad (3)$$

1. Vérifiez que la température $T(x)$ ainsi définie est bien solution de (2) et respecte bien les conditions au bord.
2. Numériquement sous Matlab/Octave, calculez et tracez cette solution en utilisant la routine fournie `CalcTempSol`. On pourra prendre $L = 1.5$, $\omega = 5$, $T_0 = 20$, $T_L = 30$ et $T_a = 10$. Que se passe-t-il si on réduit ou si on augmente ω ? Si on choisit $T_L = T_0$? Et si on choisit $T_L = T_a$? Et $T_0 = T_L = T_a$? Commentez.

2 Mesures, vraisemblance, prior, posterior

2.1 Mesures de températures

Pour estimer le paramètre ω , on réalise N mesures localisées à des abscisses fixées connues x_n , pour $n = 1, \dots, N$. Pour une valeur considérée de ω , on a N températures $T_n = T(x_n)$

données par la solution (3) aux points $x = x_n$. On les notera $T_n = T_n(\omega)$ pour rappeler qu'elles résultent de la solution de l'équation pour la valeur de ω considérée.

Par ailleurs, ces températures sont mesurées mais le système d'acquisition est imparfait et produit des erreurs. On note ces erreurs ε_n pour $n = 1, \dots, N$ et on les fait intervenir sous forme additive. Les mesures μ_n s'écrivent alors comme les températures bruitées

$$\mu_n = T_n(\omega) + \varepsilon_n, \quad \text{pour } n = 1, \dots, N. \quad (4)$$

Pour simplifier les notations, on regroupe les mesures dans le vecteur $\boldsymbol{\mu} = [\mu_1, \dots, \mu_N]$.

Remarque 1 — *Avec certains capteurs ou dans certaines conditions, les températures peuvent être déformées, comme par exemple par une saturation. On aurait alors : $\mu_n = \rho(T_n(\omega)) + \varepsilon_n$, pour $n = 1, \dots, N$. Ce ne sera pas le cas ici et on se basera sur (4).*

3. Récupérez les fichiers fournis `FichierThermique` puis tracez les mesures. Commentez le nombre de mesures, leurs positions ainsi que le niveau de bruit apparent.
4. Comment pourriez vous procéder « à la main » ou bien « par essais et erreurs » pour vous faire une idée de la valeur de ω à partir des mesures et de la forme (3) ?

Dans la suite, on modélise les erreurs de mesure comme des réalisations de variables indépendantes gaussiennes centrées et de variance r_ε . On note $\gamma_\varepsilon = 1/r_\varepsilon$ la précision.

Finalement, on a deux paramètres inconnus : le paramètre lié à la physique du système ω et le paramètre de précision des mesures γ_ε . On les regroupe dans le vecteur $\boldsymbol{\theta} = [\omega, \gamma_\varepsilon]$ et c'est ce vecteur qu'on estime dans la suite.

2.2 Vraisemblance

Cette partie est consacrée à la construction de la vraisemblance de $\boldsymbol{\theta}$ attachée aux mesures $\boldsymbol{\mu}$, c'est-à-dire la densité $f_{M|\Theta}(\boldsymbol{\mu} | \boldsymbol{\theta})$ des mesures $\boldsymbol{\mu}$ pour un paramètre $\boldsymbol{\theta}$ donné.

5a. Expliquez pourquoi la vraisemblance prend la forme factorisée :

$$f_{M|\Theta}(\boldsymbol{\mu} | \boldsymbol{\theta}) = \prod_{n=1}^N \mathcal{N}(\mu_n; T_n(\omega), \gamma_\varepsilon),$$

où $\mathcal{N}(u; m, p)$ est la densité gaussienne de moyenne m et de précision p .

5b. En utilisant son expression :

$$\mathcal{N}(u ; m, p) = (2\pi)^{-1/2} p^{1/2} \exp \left[-p(u - m)^2 / 2 \right],$$

explicitiez cette vraisemblance comme fonction de $\theta = [\omega, \gamma_\varepsilon]$ et des mesures μ . On pourra faire intervenir :

$$\varphi(\omega) = \sum_{n=1}^N [\mu_n - T_n(\omega)]^2, \quad (5)$$

qui représente le carré de la distance entre les données et la sortie modèle pour la valeur du paramètre physique considéré ω . On parle également du carré de la norme de l'erreur de sortie modèle ou erreur de modélisation.

Remarque 2 — Pour les situations où on ne disposerait pas d'une solution explicite (3), il « suffirait » de résoudre (2) de manière numérique pour calculer $T_n(\omega)$, en déduire $\varphi(\omega)$ et donc cette vraisemblance, pour toute valeur de ω considérée.

2.3 Prior

Pour estimer le paramètre θ , on se donne une densité a priori $\pi_\theta(\theta)$. On la choisit de manière à ce qu'elle soit peu informative et qu'elle permette des calculs théoriques et numériques aisés. Pour commencer on fait le choix de la séparabilité :

$$\pi_\theta(\theta) = \pi_\Omega(\omega) \pi_{\Gamma_\varepsilon}(\gamma_\varepsilon),$$

qui, en un sens, permet de ne pas se prononcer a priori sur un lien potentiel entre ω et γ_ε . Par ailleurs, pour chacun des deux paramètres on propose les choix suivants.

— Pour ω on adopte une densité uniforme entre deux valeurs fixées ω_m et ω_M

$$\pi_\Omega(\omega) = \mathcal{U}_{[\omega_m, \omega_M]}(\omega). \quad (6)$$

On suggère de s'inspirer des constatations des questions 2 et 3 pour choisir les bornes. Par exemple, on pourra prendre $\omega_m = 1$ et $\omega_M = 10$.

— Pour γ_ε on propose une densité Gamma de paramètres α_0, β_0 :

$$\pi_\Gamma(\gamma) = \frac{\beta_0^{\alpha_0}}{\Gamma[\alpha_0]} \gamma^{\alpha_0-1} \exp[-\beta_0\gamma] \mathbb{1}_{\mathbb{R}_+}(\gamma). \quad (7)$$

Pour être en adéquation avec l'idée de densité a priori peu informative on se placera dans le cas $\alpha_0 = \beta_0$ et on choisira une valeur faible : par exemple $\alpha_0 = \beta_0 = 10^{-4}$ ce qui correspond à une valeur nominale de 1 et une incertitude relative de 100.

2.4 Posterior

On s'intéresse maintenant à la densité a posteriori $\pi_{\theta|M}(\theta|\mu)$ pour le vecteur des paramètres inconnus $\theta = [\omega, \gamma_\varepsilon]$ étant données les mesures μ .

6. Partant des densités (6) et (7) et de la vraisemblance obtenue question 5, donnez la densité a posteriori $\pi_{\Theta|M}(\boldsymbol{\theta}|\boldsymbol{\mu})$ à un coefficient multiplicatif près. A une constante additive près, donnez également son logarithme, noté $LP(\boldsymbol{\theta})$, pour Log-Posterior.
7. Calculez numériquement et tracez $LP(\boldsymbol{\theta})$ sur une grille de valeurs de $(\omega, \gamma_\varepsilon)$. Sur une autre figure, tracez également la densité a posteriori $\pi_{\Theta|M}(\boldsymbol{\theta}|\boldsymbol{\mu})$ elle-même. On pourra utiliser la fonction `contour` pour un tracé sous forme de lignes de niveau.

3 Posterior et échantillonnage

On se pose alors le problème de la production d'échantillons (de réalisations, de tirages) de $\boldsymbol{\theta}$ sous la densité a posteriori $\pi_{\Theta|M}(\boldsymbol{\theta}|\boldsymbol{\mu})$: il s'agit d'un problème d'échantillonnage stochastique. Compte-tenu de la forme obtenue précédemment (Section 2.4), il est manifeste qu'il ne s'agit pas d'une densité usuelle. Pour l'échantillonner, on va s'appuyer sur une famille particulière d'algorithmes : les algorithmes de *Monte-Carlo par Chaîne de Markov*. Parmi ceux-ci, on va utiliser une structure alternée dite de *Gibbs*. Il s'agit d'une structure générale qui permet de découper le problème de l'échantillonnage d'une densité multidimensionnelle en sous problèmes d'échantillonnage de plus faible dimension (ici scalaire). Plus précisément, on montre qu'on obtient des échantillons de la densité pour $\boldsymbol{\theta} = [\omega, \gamma_\varepsilon]$ en alternant l'échantillonnage de chacune de ses deux composantes ω et γ_ε sous sa densité conditionnelle. Une forme symbolique de l'algorithme est la suivante,

- Initialisation de $\omega^{[0]}$ (un exemple parmi d'autres) :
 - Un échantillon de la loi a priori pour ω .
 - ...
- Boucle pour $k = 1, \dots, K$
 1. Tirer $\gamma_\varepsilon^{[k]}$ sous la densité de $(\gamma_\varepsilon|\boldsymbol{\mu}, \omega^{[k-1]})$
 2. Tirer $\omega^{[k]}$ sous la densité de $(\omega|\boldsymbol{\mu}, \gamma_\varepsilon^{[k]})$

et pour le mettre en œuvre on s'appuie sur les deux densités a posteriori conditionnelles pour $(\omega|\boldsymbol{\mu}, \gamma_\varepsilon)$ et pour $(\gamma_\varepsilon|\boldsymbol{\mu}, \omega)$ que l'on détermine maintenant.

8a. Paramètre du bruit. Déterminez la densité a posteriori conditionnelle pour γ_ε , c'est-à-dire la densité $\pi_{\Gamma_\varepsilon|M,\Omega}(\gamma_\varepsilon|\boldsymbol{\mu}, \omega)$, à un coefficient multiplicatif près. Montrez que c'est une densité gamma et identifiez ses paramètres. Donnez sa moyenne en fonction de ω . A quelle variance $r_\varepsilon = 1/\gamma_\varepsilon$ cela correspond-il ? Commentez.

8b. Paramètre thermique. Déterminez également la densité a posteriori conditionnelle pour ω , c'est-à-dire la densité $\pi_{\Omega|M,\Gamma_\varepsilon}(\omega|\boldsymbol{\mu}, \gamma_\varepsilon)$ en fonction de $\varphi(\omega)$. Cette distribution ne semble pas entrer dans une famille connue (comme la densité Gamma précédente pour γ_ε). On s'appuiera alors sur une étape de *Metropolis-Hastings* comme décrite en Annexe.

On peut maintenant encoder l'algorithme, le faire fonctionner et observer son comportement. L'encodage se réalise à l'aide d'une seule boucle (celle de l'algorithme de Gibbs). On obtient des échantillons de la densité a posteriori.

9. Quelques affichages et quelques réflexions.

9a. Affichez les deux séries de paramètres $\omega^{[k]}$ et $\gamma_\varepsilon^{[k]}$ pour $k = 1, \dots, K$ sous forme de signaux comme fonction de k . Affichez-les également sous forme d'histogrammes. Il s'agit d'histogrammes de quelle densité ?

9b. Représentez également les échantillons en tant que couples $(\omega^{[k]}, \gamma_\varepsilon^{[k]})$ dans le plan comme un nuage de points. On pourra superposer ou comparer ces échantillons à la représentation de la densité a posteriori de la question 7.

On s'intéresse maintenant l'estimateur au sens de la moyenne a posteriori, c'est à dire la valeur $\hat{\theta} = [\hat{\omega}, \hat{\gamma}_\varepsilon]$ du couple $\theta = [\omega, \gamma_\varepsilon]$ définie par

$$\hat{\theta} = E_{\Theta|M} [\Theta|\mu] = \int_{\theta} \theta \pi_{\Theta|M}(\theta|\mu) d\theta$$

ainsi qu'à son calcul numérique.

10. A partir des $(\omega^{[k]}, \gamma_\varepsilon^{[k]})$ calculez une approximation de cet estimateur. Calculez également une approximation des écarts-types a posteriori pour ω et γ_ε . Pour terminer, calculez une approximation de la corrélation a posteriori du couple $(\omega, \gamma_\varepsilon)$.

11. Quelques calculs numériques finaux.

11a. Toujours à partir des $\omega^{[k]}$, calculez une approximation de la probabilité (a posteriori) que ω soit supérieur à sa moyenne. Et que ω soit supérieur à sa moyenne a priori.

11b. Calculez la solution en température donnée par (3) pour la valeur estimée $\omega = \hat{\omega}$ du paramètre physique. Tracez-la sur la figure construite à la question 3 présentant les mesures.

k

4 Posterior et maximisation

On considère maintenant l'estimateur au sens du maximiseur a posteriori, c'est à dire la valeur $\bar{\theta} = [\bar{\omega}, \bar{\gamma}_\varepsilon]$ du couple $\theta = [\omega, \gamma_\varepsilon]$ qui maximise $LP(\theta)$ ou $\pi_{\Theta|M}(\theta|\mu)$:

$$\bar{\theta} = \arg \max_{\theta} \pi_{\Theta|M}(\theta|\mu)$$

et on s'intéresse également à son calcul.

12. A partir des résultats de la partie 2.4 et plus précisément la question 7, déterminez ce maximiseur sur la grille de valeurs. On pourra utiliser la fonction *MaxDeux*. Positionnez-le sur les figures obtenues précédemment.
13. Montrez qu'en fait $\bar{\omega}$ pourrait s'obtenir simplement en minimisant $\varphi(\omega)$ par rapport à ω , séparément de γ_ε . Donnez la conséquence graphique de cela sur les lignes de niveau de la question 7. Proposez un algorithme d'optimisation plus efficace et plus précis que la naïve maximisation sur une grille de valeurs pratiquée précédemment.

5 Un peu plus loin : optimalités et autres incertitudes,...

Les développements précédents ont débouché sur deux estimateurs : la moyenne et le maximiseur a posteriori. Cela amène deux questions.

14a. Selon-vous lequel est le meilleur ? En quel sens, précisément ?

14b. Donnez un autre estimateur possible. Serait-il meilleur que les précédents, en un sens ou en un autre ?

Par ailleurs, dans les développements précédents les positions x_n (où sont réalisées les mesures) sont fixées-données. Naturellement, les « performances » de l'estimation peuvent dépendre de ces positions.

15. Proposez une démarche qui permettrait de choisir ces positions de manière à optimiser ces performances.

Dans les développements précédents, les températures T_0 , T_L et T_a (caractérisant le milieu extérieur) possèdent une valeur nominale connue sans incertitude. On s'intéresse maintenant au cas où ces quantités sont elles-mêmes mal connues : en plus de ces valeurs nominales on dispose d'une incertitude (pour chacune de ces trois températures).

16. Proposez une démarche qui étendrait le travail précédent en permettant d'inclure ces incertitudes, en plus des valeurs nominales.

Annexe : Metropolis-Hastings

Dans la Section 3 à la question 8, l'échantillonnage de γ_ε (question 8a) est réalisé sous une densité qui entre dans une famille clairement reconnue : la famille Gamma, et de plus ses paramètres sont aisément déterminés. Au contraire, la densité pour ω (question 8b) n'entre pas dans une famille identifiée. On s'appuie alors sur une étape dite de Metropolis-Hastings sommairement décrite dans la présente annexe.

Dans la mesure où on ne sait pas tirer directement sous la densité $\pi_{\Omega|M,\Gamma_\varepsilon}(\omega|\boldsymbol{\mu}, \gamma_\varepsilon)$ on réalise un tirage sous une autre densité, dite de proposition et notée p ici. C'est l'étape 1 ci-dessous. L'étape 2 consiste, au hasard, à accepter la proposition ou à dupliquer la valeur courante. Cette

alternative doit être réalisé au hasard avec une probabilité d'acceptation a prescrite et définie par la relation (8). Elle est naturellement fonction d'une part de la valeur proposée et de la valeur courante et d'autre part de la densité de proposition et de la densité visée.

— A l'itération n

1. Proposer une valeur au hasard $\omega^* \sim p(\omega)$
2. Au hasard, faire

$$\begin{cases} \omega^{(n)} = \omega^* & \text{avec probabilité } a \\ \omega^{(n)} = \omega^{(n-1)} & \text{avec probabilité } 1 - a \end{cases}$$

$$\text{avec } a = \min \left(1, \frac{\pi_{\Omega|M, \Gamma_\varepsilon}(\omega^* | \boldsymbol{\mu}, \gamma_\varepsilon)} p(\omega^{(n-1)})}{\pi_{\Omega|M, \Gamma_\varepsilon}(\omega^{(n-1)} | \boldsymbol{\mu}, \gamma_\varepsilon) p(\omega^*)} \right) \quad (8)$$

Dans le cas où on choisit comme densité de proposition la densité a priori, c'est-à-dire $p = \pi_\Omega$ (voir Eq. (6)), vous pourrez voir aisément une simplification importante :

$$\frac{\pi_{\Omega|M, \Gamma_\varepsilon}(\omega^* | \boldsymbol{\mu}, \gamma_\varepsilon)}{\pi_{\Omega|M, \Gamma_\varepsilon}(\omega^{(n-1)} | \boldsymbol{\mu}, \gamma_\varepsilon)} \frac{p(\omega^{(n-1)})}{p(\omega^*)} = \frac{f_{M|\Theta}(\boldsymbol{\mu} | \omega^*, \gamma_\varepsilon)}{f_{M|\Theta}(\boldsymbol{\mu} | \omega^{(n-1)}, \gamma_\varepsilon)}$$

c'est-à-dire que l'on fait apparaître le rapport des vraisemblances pour deux valeurs du paramètre : proposée ω^* et courante $\omega^{(n-1)}$. De plus, compte-tenu de la forme de la vraisemblance (voir la question 5), vous pouvez simplifier à nouveau et obtenir

$$\frac{f_{M|\Theta}(\boldsymbol{\mu} | \omega^*, \gamma_\varepsilon)}{f_{M|\Theta}(\boldsymbol{\mu} | \omega^{(n-1)}, \gamma_\varepsilon)} = \frac{\exp[-\gamma_\varepsilon \varphi(\omega^*) / 2]}{\exp[-\gamma_\varepsilon \varphi(\omega^{(n-1)}) / 2]}$$

qui fait apparaître la norme de l'erreur (5) de la question 5, Section 2.2. Vous pouvez finalement considérer le logarithme népérien \bar{a} de la probabilité a

$$\bar{a} = \log a = \min \left(0, -\frac{1}{2} \gamma_\varepsilon [\varphi(\omega^*) - \varphi(\omega^{(n-1)})] \right)$$

qui se calcule à partir de la différence des erreurs de modélisation pour le paramètre proposé ω^* et le paramètre courant $\omega^{(n-1)}$. Vous constatez que si la valeur proposée produit un modèle en meilleure adéquation aux données que la valeur courante, *i.e.*, $\varphi(\omega^*) < \varphi(\omega^{(n-1)})$ alors $\bar{a} = 0$ et $a = 1$ la valeur proposée est acceptée (avec probabilité 1). Au contraire, lorsque la valeur proposée dégrade l'adéquation par rapport à la valeur courante, elle est quand même susceptible d'être acceptée, avec une probabilité a d'autant plus grande que la dégradation est faible.