

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira de Béjaïa



Faculté des Sciences Exactes

Département De Mathématique

Spécialité Analyse Mathématique

Mémoire de Fin de Cycle Master

Thème :

Approche bayésienne pour la résolution de
problèmes inverses

Réalisé par :

M^r SIGAA SEYF ELISLAM

M^r NOURI IMED

Devant le jury composé de :

Président : *M^r BOUHMILA FATAH*

MCA Univ BEJAIA

Encadreur : *M^{me} BECHIR HALIMA*

MCA Univ BEJAIA

Examinateur : *M^{me} TAS SAADIA*

PROFESSEUR Univ BEJAIA

Promotion : 2019/2020

Remerciements

Remerciements

Nous remercions notre promotrice Mme. BECHIR Halima, pour sa disponibilité et ses précieux conseils et motivations qui nous ont aidé à réaliser ce modeste travail et à faire évoluer de notre mémoire . Nous remercions également l'honorable jury pour avoir consentir à évaluer et à juger notre travail.

On tient a exprimer nos plus sincères remerciements à tous les membres de la faculté des Sciences Exactes en général et aux membres du département de Mathématiques en particulier, ainsi que tous les enseignants pour les peines et les efforts qu'ils se sont donnés durant notre formation.

Nos sincères remerciements vont à tous les étudiants de Master 2 Analyse et Probabilités. Nous tenons également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce modeste travail.

Dédicace

A mes chers parents REBIHA et BOUHAFS, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,

A mes chères sœurs : MARIA, KHOULOUUD et SANAA pour leurs encouragements permanents, et leur soutien moral,

A mon cher frère RADOUNE pour son appui et son encouragement,

A toute ma famille pour leur soutien tout au long de mon parcours universitaire,

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infailible,

Merci d'être toujours là pour moi.

SIGAA SEYF ELISLAM

Dédicace

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,

A ma chères sœurs : YOUSRA pour ses encouragements permanents, et son soutien moral,

A mon cher frère OUSSAMA pour son appui et son encouragement,

A toute ma famille pour leur soutien tout au long de mon parcours universitaire,

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infailible,

Merci d'être toujours là pour moi.

NOURI IMED

Table des matières

Introduction.....	2
Problèmes inverses.....	3
• Rappels d'analyse fonctionnelle.....	4
• Problèmes bien et mal-posés.....	6
• Exemples de problèmes mal-posés.....	7
• Méthode de résolution des problèmes inverses.....	12
• Régularisation par La méthode Tikhonov.....	12
Probabilités Bayésiennes	24
• Rappels.....	24
• Théorème de Bayes.....	25
• Généralisation pour un système complet.....	27
• Exemples.....	28
Application : Test de diagnostic médical.....	30
Conclusion.....	38
Bibliographie.....	39

Introduction

D'après J.B. Keller, deux problèmes sont dits *inverses* l'un de l'autre si la formulation de l'un met l'autre en cause. Cette définition comporte une part d'arbitraire, et fait jouer un rôle symétrique aux deux problèmes considérés. Une définition plus opérationnelle est qu'un problème inverse consiste à déterminer des *causes* connaissant des *effets*. Ainsi, ce problème est l'inverse de celui appelé problème direct, consistant à déduire les effets, les causes étant connues.

La prédiction de l'état futur d'un système physique, connaissant son état actuel, est l'exemple type du problème direct. On peut envisager divers problèmes inverses : par exemple, reconstituer l'état passé du système connaissant son état actuel (si ce système est irréversible), ou la détermination de paramètres du système, connaissant (une partie de) son évolution. Ce dernier problème est celui de *l'identification de paramètres*, qui sera notre principale préoccupation dans la suite.

Une difficulté pratique de l'étude des problèmes inverses est qu'elle demande souvent une bonne connaissance du problème direct, ce qui se traduit par le recours à une grande variété de notions tant physiques que mathématiques. Le succès dans la résolution d'un problème inverse repose en général sur des éléments spécifiques à ce problème.

Notre objectif est de présenter l'approche bayésienne comme une méthode pour traiter les problèmes inverses. Dans le premier chapitre, on présente la notion de problèmes inverses. Le second chapitre est consacré aux probabilités bayésiennes, et on termine avec une application qui offre un modèle montrant l'approche bayésienne comme une méthode pour traiter les problèmes inverses.

Chapitre 01 : Problèmes inverses

Un problème inverse consiste à déterminer des causes d'un phénomène à partir de la connaissance des effets. Ce problème est l'inverse du problème dit direct, consistant à déduire les effets à partir de la connaissance des causes.

Les problèmes inverses peuvent être difficiles à résoudre pour au moins deux raisons différentes : (1) les valeurs différentes des paramètres du modèle peuvent être compatibles avec les données (sachant que la hauteur du mât principal n'est pas suffisante pour calculer l'âge du capitaine), et (2) la découverte des valeurs des paramètres du modèle peut nécessiter l'exploration d'un espace de paramètre énorme (trouver une aiguille dans une botte de foin 100 dimensions est difficile).

Parmi les domaines dans lesquels les problèmes inverses jouent un rôle important, nous pouvons citer: l'imagerie médicale, l'ingénierie pétrolière, l'hydrogéologie, la chimie, le radar et l'acoustique sous-marine, le traitement d'image.

1-Rappel d'analyse fonctionnelle

Définition Soit $A: X \rightarrow Y$ un opérateur (i.e. une application linéaire).

A est dit borné si il existe une constante $C > 0$ telle que $\|Au\| \leq C\|u\| \quad \forall u \in X$

A est dit compact si l'image de la boule unité de X par A est relativement compacte pour la topologie

forte de Y .

Théorème de l'application ouverte Soient X et Y deux espaces de Banach et T un

opérateur borné et surjectif de X sur Y . Alors il existe une constante $c > 0$ telle que

$$T(B_X(0,1)) \supset B_Y(0,c).$$

En particulier, si T est bijectif, alors T^{-1} est continu.

Définition Soit $A \subset X$ un sous-espace fermé d'un espace de Banach X. On dit que B sous-espace de A est un supplémentaire topologique de A si

1. B est fermé
2. $A + B = X$ et $A \cap B = \{0\}$ (que l'on note aussi $A \oplus B = X$).

Dans ce cas, les projecteurs sont linéaires et continus. Tout sous-espace de dimension finie ou de codimension finie admet un supplémentaire topologique (Hahn-Banach).

Définition Soit D un opérateur borné de X sur Y . Un pseudo-inverse de D est un opérateur borné

$$R: Y \rightarrow X \text{ tel que } RDR = R \text{ et } DRD = D.$$

Proposition Un opérateur borné $D: X \rightarrow Y$ admet un pseudo-inverse ssi $\text{im } D$ est fermé, $\text{ker } D$ et $\text{coker } D$ ont des supplémentaires topologiques.

Tout opérateur $D: X \rightarrow Y$ de Fredholm admet un pseudo-inverse. En particulier, en notant T un pseudo-inverse, on a

$$X = \text{ker } D \oplus \text{im } T \text{ et } Y = \text{im } D \oplus \text{ker } T$$

Théorème Soit T un opérateur borné et surjectif de X sur Y .

T admet un inverse à droite $\Leftrightarrow \text{ker } T$ admet un supplémentaire topologique.

Lemme Soit $D: X \rightarrow Y$ un opérateur de Fredholm et $L: Z \rightarrow Y$ un opérateur borné tel que

$D \oplus L: X \oplus Z \rightarrow Y$ est surjectif. Alors $D \oplus L$ a un inverse à droite. De plus la projection

$\Pi: \text{ker}(D \oplus L) \rightarrow Z$ est un opérateur de Fredholm avec $\text{ker } \Pi = \text{ker } D$ et $\text{co ker } \Pi = \text{co ker } D$.

Définition Soit $A: X \rightarrow Y$ un opérateur linéaire borné. On définit $A^*: Y' \rightarrow X'$ par : soit

$v \in Y'$, On définit $g_v: X \rightarrow \mathbb{R}$ par $g_v(u) = \langle v, Au \rangle$ et $|g_v(u)| \leq c \|u\|$.

Par Hahn-Banach, il existe un unique $f: X \rightarrow \mathbb{R}$ ($f \in X'$) tel que $\|f(u)\| \leq c \|u\|$.

On pose alors $A^*v = f$ et A^* vérifie

$$\langle v, Au \rangle_{Y',Y} = \langle A^*v, u \rangle_{X',X} \quad \forall (u, v) \in (X, Y)$$

Soit $D: X \rightarrow Y$ un opérateur borné. Alors

- D^* est borné (et réciproquement)
- $\text{ker } D = (\text{im } D^*)^\perp$
- $\text{im } D$ est fermé $\Leftrightarrow \text{im } D^*$ est fermé

- $\dim \operatorname{coker} D < \infty \Leftrightarrow \dim \ker D^* < \infty$

Définition Un espace de Banach X est dit séparable s'il admet une suite dense.

Proposition Soit X un espace de Banach séparable. Alors tout recouvrement ouvert de X admet un sous-recouvrement dénombrable.

2-Problèmes bien et mal-posés

Définition

Un problème bien-posé au sens de Hadamard a les propriétés suivantes :

1. une solution existe : $\forall b \in F$ il existe une solution $x \in E$ telle que $Ax = b$.
2. une solution unique : $\forall x_1, x_2 \in E$, si $Ax_1 = Ax_2 = b$ alors $x_1 = x_2$.
3. elle dépend de façon continue des données :

$$\forall (x_n) \subset E \text{ tq } b_n = Ax_n \xrightarrow{n \rightarrow \infty} b = Ax \text{ alors } x_n \xrightarrow{n \rightarrow \infty} x.$$

Remarque : On dit que le problème est mal-posé si l'une de ces propriétés n'est pas satisfaite.

Remarque : La non-existence et la non-unicité de la solution d'un problème mal-posé sont sans doute des difficultés sérieuses mais on peut les rétablir. Cependant le manque de continuité est plus problématique, en particulier en vue d'une résolution approchée ou numérique. C'est-à-dire il ne sera pas possible (indépendamment de la méthode numérique) d'approcher d'une manière satisfaisante la solution du problème inverse car les données disponibles seront bruitées donc proches, mais différentes des données réelles.

Définition

Soit E et F deux espaces métriques et $A: E \rightarrow F$:

1. A est injective.
2. A est continue.
3. A^{-1} n'est pas continue.

On dit alors que le problème inverse à savoir : connaissant $g \in F$ trouver $f \in F$ tel que

$f = A^{-1}g$ est mal-posé.

Exemples des problèmes mal-posés

Différentiation et intégration

La différentiation et l'intégration sont deux problèmes inverses l'un de l'autre. Il est plus habituel de penser à la différentiation comme problème direct, et à l'intégration comme problème inverse. En fait l'intégration possède de bonnes propriétés mathématiques qui conduisent à la considérer comme le problème direct. Et la différentiation est le prototype du problème mal posé, comme nous allons le voir.

Exemple 1.1 Considérons l'espace de Hilbert $L^2(\Omega)$, et l'opérateur intégral A définie par

$$Af(x) = \int_0^x f(t) dt$$

Il est facile de voir directement que A est un opérateur linéaire et continue de $L^2(0,1)$ dans lui-même. Cet opérateur est injectif, par contre son image est le sous espace vectoriel

$$\text{Im } A = \{f \in H^1(0,1), f(0) = 0\}$$

Où $H^1(0,1)$ est l'espace de Sobolev. En effet, l'équation

$$Af = g$$

est équivalente à

$$f(x) = g'(x) \text{ et } g(0) = 0$$

L'image de A n'est pas fermée dans $L^2(0,1)$ (bien entendu, elle l'est dans $H^1(0,1)$) En conséquence,

l'inverse de A n'est pas continu sur $L^2(0,1)$, comme le montre l'exemple suivant.

Exemple 1.2 Considérons une fonction $g \in C^1([0,1])$ donnée, et $n \in \mathbb{N}$. Soit

$$g_n(x) = g(x) + \frac{1}{n} \sin(n^2 x)$$

Alors

$$f_n = g'_n$$

$$f_n(x) = g'(x) + n \cos(n^2 x)$$

$$= f + n \cos(n^2 x)$$

nous avons voir que

$$\|g_n - g\| = \sup \left| \frac{1}{n} \sin(n^2 x) \right|$$

$$\lim_{n \rightarrow \infty} \|g_n - g\| = 0$$

Et

$$\|f_n - f\| = \sup |n \cos(n^2 x)|$$

$$\lim_{n \rightarrow \infty} \|f_n - f\| = \infty$$

La différence entre f et f_n peut-être arbitrairement grande, alors même que la différence entre g et g_n est arbitrairement petite. L'opérateur de dérivation (l'inverse de A) n'est donc pas continu, au moins avec ce choix des normes.

Donc la condition de stabilité n'est pas satisfaite alors le problème est mal posé.

Exemple 2

2.1 Soit à résoudre le système:

$$AX = Y$$

$$\text{Où } A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \text{ et } Y = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

On trouve

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

prenons maintenant un second membre Y_* différent de Y ;

$$\text{Soit } Y_* = \begin{pmatrix} 32.1 \\ 22.9 \\ 33.1 \\ 30.9 \end{pmatrix}$$

on vérifie alors que la solution de $AX = Y_*$ est

$$X = \begin{pmatrix} 9.2 \\ -12.6 \\ 4.5 \\ -1.1 \end{pmatrix}$$

On remarque que de très petites perturbations sur Y , ont conduit à de grandes variations.

Dans cet exemple $\text{cond}(A) = 2984.0942$

Ce phénomène de mauvais conditionnement explique en partie la difficulté de prévoir certains phénomènes.

2.2 soit $P: \mathbb{R} \rightarrow \mathbb{R}, x \rightarrow x^2 + 1$

L'équation

$$P(x) = 0$$

n'admet pas de solution dans \mathbb{R} . Donc on a un problème d'existence, c'est un problème mal posé.

En hydrogéologie

Exemple 3 (Transport d'un polluant par un aquifère).

Un milieu poreux est constitué d'une matrice rocheuse, comprenant des pores qui peuvent laisser passer l'eau. Il est essentiellement impossible de décrire l'écoulement d'un fluide dans un tel milieu

hétérogène, dans la mesure où l'on doit prendre en compte des échelles spatiales allant du centimètre (le pore) au kilomètre (le modèle régional), et que la disposition précise des pores n'est de toutes façons pas connue. On utilise alors des modèles physiques simplifiés, le plus courant étant la loi de Darcy, qui relie la hauteur de l'eau dans le milieu, appelée charge piézométrique et notée $h(x, y, z, t)$, à la vitesse de filtration $\bar{q}(x, y, z, t)$. Cette loi exprime que la vitesse est proportionnelle à l'opposé du gradient hydraulique :

$$\bar{q} = -K \text{ grad } h \quad (3.1)$$

où K est le coefficient de conductivité hydraulique. Ce peut en principe être un tenseur, mais nous nous restreindrons au cas où c'est un scalaire. On exprime également la conservation de la masse (on fait l'hypothèse que le milieu est incompressible) :

$$S \frac{\partial h}{\partial t} + \text{div } \bar{q} = f \quad (3.2)$$

où S est le coefficient d'emménagement spécifique, et f est une source (supposée connue).

L'élimination de \bar{q} donne pour h l'équation parabolique :

$$S \frac{\partial h}{\partial t} - \text{div } (K \text{ grad } h) = f \quad (3.3)$$

à laquelle on ajoute des conditions initiales (h donné à $t = 0$) et aux limites (Dirichlet, correspondant à une charge imposée, ou Neumann, correspondant à un flux imposé).

Les problèmes de transport de contaminant font intervenir, en plus de l'écoulement, la façon dont évolue la concentration d'une espèce (composé chimique, hydrocarbure, radionucléide) n portée \vec{z} par l'écoulement. Ce phénomène met en jeu trois mécanismes : la convection (imposée par la vitesse de filtration \bar{q}), la diffusion moléculaire et la dispersion cinématique. Nous ne décrivons pas ces deux derniers mécanismes en détail (pour cela voir les références citées plus haut). La quantité étudiée est la concentration $C(x, y, z, t)$ du polluant, qui obéit à une équation de type convection–diffusion :

$$\frac{\partial C}{\partial t} + \frac{1}{\varepsilon} \operatorname{div}(\bar{q}C) - \operatorname{div}(D \operatorname{grad} C) = f_c \quad (3.4)$$

ou ε est la porosité cinématique (fraction des pores occupés par l'eau en mouvement), D est le tenseur de diffusion (en agrégeant diffusion moléculaire et la dispersion cinématique), et f_c est une éventuelle source de polluant. On ajoute une condition initiale (concentration connue à l'instant initial), et des conditions aux limites.

Le problème direct est constitué par les équations (3.3) et (3.4). Ce problème couplé est théoriquement non-linéaire, à cause du terme $\operatorname{div}(\bar{q}C)$. En pratique, cependant, on peut souvent résoudre d'abord l'équation (3.3), puis (3.4), \bar{q} étant connu.

On mesure, par exemple, la concentration en un certain nombre de points de mesures, et à des instants discrets (il n'est pas réaliste ici de supposer que la mesure est continue en temps). On connaît donc $C(x_o, y_o, z_o, t_o)$; $o = 1, \dots, N_o$. Le problème inverse est alors de chercher la conductivité hydraulique (et dans une moindre mesure les autres paramètres du modèle), connaissant ces mesures. Ce problème est sous-déterminé, car il est rare que l'on ait accès à suffisamment de mesures.

3-Méthode de résolution des problèmes inverses

Régulariser un problème mal-posé, c'est le remplacer par un autre bien posé de sorte que l'erreur commise soit compensée par le gain de stabilité. Cette partie présente une introduction aux méthodes de régularisation, la plus courante étant la méthode de Tikhonov.

La principale difficulté dans l'application d'une méthode de régularisation à un problème particulier est la détermination du paramètre de régularisation lui-même.

Méthode de Tikhonov

La régularisation de Tikhonov est la méthode de régularisation la plus utilisée pour la résolution des problèmes qui ne sont pas bien-posés ainsi que pour les problèmes inverses.

Notion d'opérateur régularisant

On considère l'équation :

$$Ax = b \quad (3.1)$$

où :

A un opérateur : $A: E \rightarrow F$

x paramètre à identifier.

b : Mesures.

Si l'opérateur inverse A^{-1} n'est pas continu et qu'on ne dispose pas de la valeur exacte $b_{\varepsilon x}$ mais de b_δ vérifiant $d(b_{\varepsilon x}, b_\delta) < \delta$ la solution approchée $x_\delta = A^{-1}b_\delta$ ne peut évidemment pas être considérée comme une approximation $x_{\varepsilon x} = A^{-1}b_{\varepsilon x}$.

Définition un opérateur $L(b, \alpha)$ dépendant du paramètre réel positif α est appelé opérateur régularisant pour l'équation $Ax = b$ dans le voisinage $b = b_{\varepsilon x}$ s'il vérifie les propriétés suivantes :

$\exists \delta_1 > 0$ tq $L(b, \alpha)$ soit définit

$$\text{pour } \forall \alpha > 0 \text{ et } \forall b \in F, \|b - b_{\varepsilon x}\| \leq \delta_1$$

$\exists \alpha = \alpha(\delta)$ une fonction tq

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) \leq \delta_1 \text{ tq } \|b_{\varepsilon x} - b_\delta\| < \delta \Rightarrow \|x_{\varepsilon x} - x_\delta\| < \varepsilon \text{ avec } x_\delta = L(b_\delta, \alpha(\delta))$$

Cette définition ne suppose pas que l'opérateur $L(b, \alpha)$ soit univoque et il existe en fait une grande diversité d'opérateurs régularisants. Un tel opérateur L est donc capable de fournir une approximation de x_δ aussi précise que l'on veut, pour peu que l'on dispose d'une précision suffisante sur $b_{\varepsilon x}$.

Le problème de recherche d'une solution approchée de $Ax = b$ stable vis-à-vis de faibles variations du second membre se réduit donc à :

- Chercher un opérateur régularisant.
- Définir le paramètre de régularisation α .

Théorème Soit $A: E \rightarrow F$ un opérateur et soit $L(b, \alpha): F \rightarrow E$ un opérateur défini pour tout $b \in F$ et pour tout $\alpha > 0$ et continu par rapport à b

Si pour tout $x \in E$ on a

$$\lim_{\alpha \rightarrow 0} L(A(x), \alpha) = x$$

Alors $L(b, \alpha)$ est un opérateur régularisant pour $Ax = b$.

Pour l'instant, ces considérations n'ont pas de caractère constructif pour que le concept d'opérateur régularisant soit utile, il faut appliquer des méthodes de construction .

Construction d'opérateurs régularisants

Celle-ci peut découler d'un principe de sélection (Tikhonov & Arsenine). En effet, on suppose que le second membre b en notre possession soit entaché, par rapport au second membre exact $b_{\varepsilon x}$ d'un écart ne dépassant pas une certaine valeurs δ :

$$\|b - b_{\varepsilon x}\|_F \leq \delta$$

On va alors chercher naturellement une solution au problème inverse (3.1) dans le sous ensemble

$E_\delta \subset E$ des éléments x vérifiant :

$$\|Ax - b\|_F \leq \delta \quad (3.2)$$

Fonction stabilisatrice

Mais il ne suffit pas de prendre pour solution de (3.1) un élément quelconque de E_δ une telle solution n'est en général pas continue par rapport à δ . Il faut donc adapter un principe de sélection qui permette pour tout δ suffisamment petit, de choisir un élément x_δ de E_δ de telle sorte que x_δ dépende continûment de δ . On va poser $\Omega(x)$ une fonctionnelle positive continue d'un sous ensemble E_1 de E dense sur F . On dit que Ω est une fonctionnelle stabilisatrice si :

- Ω est définie sur une partie dense E_1 dans E .
- Ω est non-négative continue.
- x_δ appartient au domaine de définition de Ω .
- $\forall d > 0$ l'ensemble des x tels que $\Omega(x) \leq d$ est un compact de E

Existence d'une solution à $\min \|Ax - b\|_2^2 + \alpha\Omega(x)$

Par exemple, on peut prendre :

$$\Omega(x) = \|x - \bar{x}\|_E^2, \bar{x} \in E \text{ fixé, } E_1 = E$$

cela signifie qu'on va chercher, parmi toutes les solutions à δ près de (3.1) celle qui est la plus proche (ou la moins éloignée...) d'une valeur de référence \bar{x} jugée vraisemblable ou représentative à partir de considérations physiques. Il existe une infinité de manières de choisir Ω ; Ce choix traduira de façon mathématique le critère de sélection que l'on souhaite utiliser pour restreindre le champ des solutions approchées de (3.1). Une fois Ω choisie, on prend pour solution régularisée de (3.1) un élément x_δ qui minimise $\Omega(x)$ sur $E_1 \cap E_\delta$ (il est démontré qu'un élément existe). Cet élément x_δ peut être considéré comme le résultat de l'action sur b d'un opérateur $L(b, \delta)$ de δ .

La propriété de toute solution régularisée x_α est donc de minimiser $\Omega(x)$ sous la contrainte (3.2).

Cela revient donc à minimiser le lagrangien

$$\begin{aligned} S_\alpha(x, b) &= \alpha \left[\frac{1}{\alpha} \|Ax - b\|_F^2 + \Omega(x) \right] \\ &= \|Ax - b\|_F^2 + \alpha \Omega(x) \end{aligned}$$

où $\frac{1}{\alpha}$ est le multiplication de lagrange associé à la contrainte (3.2).

Théorème Soit A un opérateur continu de E dans F . Quels que soient $b \in F$ et $\alpha > 0$; il existe un élément $x_\alpha \in E_1$ tel que :

$$\inf_{x \in E_1} S_\alpha(x, b) = S_\alpha(x_\alpha, b)$$

$$(C'est\text{-}\grave{a}\text{-}dire\ S_\alpha(x, b) = \|Ax - b\|_2^2 + \alpha \Omega(x))$$

Donc, quelle que soit la valeur fixée du paramètre α , $S_\alpha(x, b)$ admet un minimum au moins. Pour α fixé, x_α obtenu par un minimisation de S_α ; peut être considéré comme le resultat de l'action d'un opérateur $L_1(b, \alpha)$ sur b .

Stratégie de regularisation

Définition Une stratégie de régularisation est une famille d'opérateurs linéaires bornés

$L_\alpha : F \rightarrow E, \alpha > 0$ tel que

$$\lim_{\alpha \rightarrow 0} \|L_\alpha Ax - x\| = 0, \forall x \in E$$

i.e, l'opérateur $L_\alpha A$ converge simplement vers l'identité I .

Théorème Soit L_α une stratégie de régularisation pour l'opérateur compact $A: E \rightarrow F$, donc

$\dim E = +\infty$. Alors la famille d'opérateurs L_α ne sont pas uniformément bornés : il existe une suite

$\{\alpha_j\}_{j \in \mathbb{N}} \subset \mathbb{R}^+$ telle que

$$\lim_{j \rightarrow \infty} \alpha_j = 0 \text{ et } \lim_{j \rightarrow \infty} \|L_{\alpha_j}\| = \infty$$

i.e il n'y a pas de convergence de $L_\alpha A$ vers l'identité au sens de la norme d'opérateur.

Caractère régularisant de L_1

Théorème Soit x_{ex} solution de $Ax_{ex} = b_{ex}$, alors : $\forall \varepsilon > 0, \forall \beta_1(\delta)$ et $\beta_2(\delta)$ deux fonctions

définies sur $[0, \delta_1]$, non négatives, non décroissantes et continues telles que :

$$\beta_2(0) = 0 \text{ et } \forall \delta \in [0, \delta_1]: \frac{\delta^2}{\beta_1(\delta)} \leq \beta_2(\delta)$$

Il existe $\delta_0 \leq \delta_1$ tq $\forall \delta \leq \delta_0, \forall \alpha \in \left[\frac{\delta^2}{\beta_1(\delta)}, \beta_2(\delta) \right]$; l'inégalité :

$$\|b - b_{\varepsilon x}\|_F \leq \delta$$

entraîne l'inégalité :

$$\|x_\alpha - x_{\varepsilon x}\|_E \leq \varepsilon$$

pour tout les vérifiant δ :

$$\frac{\delta^2}{\beta_1(\delta)} \leq \alpha \leq \beta_2(\delta)$$

Avec $x_\alpha = L_1(b, \alpha)$

L'opérateur L_1 est donc bien régularisant au voisinage de b_{ex} . En fait, L_1 est régularisant pour tout $b \in F$. En rappelant le caractère stabilisateur de $x \rightarrow \|x\|_2^2$, on a bien la validation de la méthode de régularisation de Tikhonov. Cette méthode signifie en clair que :

- L'opérateur $L_1(b, \alpha)$ associé à la minimisation du lagrangien $S_\alpha(x, b)$ est régularisant : une petite perturbation sur le second membre entraîne une petite perturbation sur la solution x_α .
- Et ceci à condition que le paramètre α appartienne à une certaine plage de valeurs.

Remarque Le caractère régularisant de $L_1(b, \alpha)$ est vrai pour toute une gamme de valeurs de α , et non pour une valeur unique.

Une manière de choisir α consiste à imposer la condition (critère de Morosov) :

$$\|Ax - b\|_F = \delta$$

dans laquelle la valeur numérique de δ est fixée empiriquement à partir de l'idée qu'on se fait de la qualité des mesures b et du modèle A . Une autre, dite critère d'Arcangeli, conduit à chercher α solution de :

$$\|Ax - b\|_F^2 = \frac{\delta^2}{\alpha}$$

Quoi qu'il en soit, le problème du meilleur choix de α , qui correspond au compromis optimal entre la satisfaction du modèle ($Ax = b$) et le respect du critère de choix stabilisateur ($\Omega(x)$ pas trop grand), est difficile et semble, d'un point de vue théorique, encore largement ouvert.

Des auteurs (voir Engl and Neubauer ainsi que les références de ces articles) ont étudié la convergence de la solution régularisée x_α vers une quasi-solution en fonction de α et de l'écart δ entre seconde membre approché et exact.

Il y est en particulier signalé que la solution régularisée x_α obtenue par minimisation de $S_\alpha(x, b)$ converge vers la quasi-solution de norme minimale

- Comme $O\left(\delta^{\frac{1}{2}}\right)$ si $A^+b \in \text{Im}(A^T)$
- Comme $O\left(\delta^{\frac{2}{3}}\right)$ si $A^+b \in \text{Im}(A^T A)$

dans le meilleur des cas, c'est à dire avec le meilleur choix possible du paramètre de régularisation α ajusté en fonction δ . Ça indique également que les critères précédents ne sont pas optimaux du point de vue de la vitesse de convergence. Des versions plus élaborées de ces critères, qui reposent sur l'utilisation de la régularisation de Tikhonov itéré, y sont étudiées ; leur convergence est optimale.

En pratique on pourra éventuellement se contenter de choisir de manière intuitive en se rappelant sa signification de paramètre de compromis.

En résumé, la résolution de (3.1) par la méthode de régularisation de Tikhonov utilise les étapes suivantes :

- Choix de la fonctionnelle stabilisatrice Ω et construction de la famille de fonctionnelles $S_\alpha(x, b)$ pour $\alpha > 0$.
- Choix du paramètre de régularisation α , qui doit être strictement positif.
- Recherche par minimisation de $S_\alpha(x, b)$ par rapport à x , de x_α qui sera une solution régularisée pour le problème inverse (3.1).

Le principe de sélection n'est pas la seule façon de régulariser un problème inverse. Une autre voie possible consiste à remplacer le problème direct initial décrit par l'opérateur A , par une famille (indexée par un paramètre $\alpha > 0$) de problèmes directs proches, décrit par des opérateurs A_α proches de A . Les opérateurs A_α doivent avoir un inverse continu L_α régularisant au sens de Tikhonov. Par exemple, le problème inverse linéaire (3.1) peut être remplacé par :

$$A_\alpha x = [A^T A + \alpha I] x = A^T b \quad (3.3)$$

où l'opérateur A_α est par construction inversible. L'opérateur L_α qui est donc donnée par :

$$L_\alpha = [A^T A + \alpha I]^{-1} A^T$$

est régularisant (la solution x_α obtenue par (3.1) tendu vers la quasi-solution de norme minimale quand $\alpha \rightarrow 0$). Il faut d'ailleurs remarquer que (3.3) peut également être obtenue en écrivant que la première variation de la fonctionnelle S_α ci dessous est nulle quand x en réalise le minimum :

$$S_\alpha(x, b) = \|Ax - b\|_F^2 + \alpha \|x\|_F^2$$

Sur cette exemple, les deux approches sont donc équivalentes.

Assurer la fermeture de l'espace de l'image de l'opérateur $A(\text{Im } A)$ pour définir une solution $b \in F$.

Proposer une solution plus robuste (moins sensible au bruit) qu'une inversion généralisée ne répond pas. En dimension finie ou infinie, un régularisateur du problème $Ax = b$ est une famille d'opérateur

$L_\alpha, \alpha \in \mathbb{R}$ tel que

$$\forall \alpha \in \mathbb{R}, L_\alpha \text{ est continu de } F \text{ vers } E$$

$$\forall b \in \text{Im}(A), \lim_{\alpha \rightarrow 0} L_\alpha b = A^T b$$

On choisit L_α pour contraindre l'espace des solutions, et α est le paramètre (coefficient) de régularisation.

Pour des données bruitées

$$b_\varepsilon = (b = Ax) + b_{\text{bruit}}$$

on obtient une solution approchée.

$$x_\varepsilon = L_\alpha b + L_\alpha b_{\text{bruit}}$$

Le choix de α des deux termes antagonistes $L_\alpha b$ et $L_\alpha b_{\text{bruit}}$

Contrôle en dimension

- Minimisation de $\|Ax - b\|_F$ dans un sous-espace.
- Minimisation de $\|Ax - b\|_F$ dans E par une méthode itérative à nombre limité d'itérations.
- Minimisation d'un critère composite on cherche E_α minimisant

$$S_{\alpha(x,b)} = \|Ax - b\|_2^2 + \alpha \Omega(x)$$

On demande à la solution un compromis fidélité aux mesures ($\|Ax - b\|_F$) et fidélité à l'information a priori ($\Omega(x)$)

Présentation de la solution des moindres carrés.

on avoir une solution plus douce, physiquement raisonnable.

$\alpha = 0 \Rightarrow$ moindres carrés.

$\alpha = \infty \Rightarrow$ fidélité parfaite avec l'a priori.

Méthode simple pour la détermination du coefficient de régularisation Une difficulté des méthodes de régularisation réside dans le choix de la valeur du paramètre α . Il existe une grande variété de méthodes permettant d'optimiser ce choix reposant sur une étude attentive du problème posé.

Il existe toutefois une méthode empirique très simple à mettre en œuvre, appelée (en anglais) Generalized Cross-Validation (GCV).

L'idée consiste à mettre de côté l'une des données b du problème et à considérer que la valeur optimale de α conduit à une bonne approximation de b . Le x choisit sera alors celui qui rend optimale cette approximation.

$$Ax = b \quad (\exists x \text{ tq } b \text{ connu})$$

$$A^T Ax = A^T b$$

$$x = (A^T A)^{-1} A^T b$$

On utilise la régularisation de Tikhonov

$$Ax = b \rightarrow \begin{pmatrix} A \\ \alpha I \end{pmatrix} x = \begin{pmatrix} b \\ 0 \end{pmatrix}$$

$$(A^T, \alpha I) \begin{pmatrix} A \\ \alpha I \end{pmatrix} x = (A^T, \alpha I) \begin{pmatrix} b \\ 0 \end{pmatrix}$$

$$\Rightarrow (A^T A + \alpha^2 I) x = A^T b$$

$$A_\alpha x = b_i$$

$$b_i = A^T b$$

$$A_\alpha = A^T A + \alpha^2 I$$

$$\Rightarrow x_\alpha = (A^T A + \alpha^2 I)^{-1} A^T b$$

tq α est le minimum de la fonction

$$GCV(\alpha) = \left[\frac{\|Ax_\alpha - b\|^2}{[\text{tr}(I - AA_\alpha^{-1}A^T)]^2} \right] = \frac{\|(I - AA_\alpha^{-1}A^T)b\|^2}{[\text{tr}(I - AA_\alpha^{-1}A^T)]^2}$$

Chapiter 02 : Probabilités Bayésiennes

Rappel

Définition Probabilité= fonction permettant de « mesurer » la chance de réalisation d'un évènement de $P(\Omega)$ (ou plus généralement d'une tribu A)

Définition Soit (Ω, A) un espace probablisable. Une probabilité sur (Ω, A) est une application satisfaisant les 3 axiomes suivants :

$$0 \leq P(\Omega) \leq 1, A \in A$$

$$P(\Omega) = 1$$

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i), \forall (A_i)_{i \in \mathbb{N}}$$

ensemble dénombrable d'évènements disjoints

Dès lors que P est définie, (Ω, A, P) s'appelle un espace probablisé.

Opérations sur les probabilités

$$P(\emptyset) = 0$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A) \leq P(B) \text{ si } A \leq B$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P\left(\bigcup A_i\right) \leq \sum P(A_i)$$

$$0 \leq P(A_i) \leq 1$$

CP: Si $P(A) = 0$ alors A est presque impossible.

On écrit Si $P(A) = 1$ alors A est presque sûr. On écrit $A = \Omega$ p.s.

Probabilité conditionnelle de A sachant B:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

(probabilité que A se réalise sachant que B se réalise). C'est une probabilité sur B .

Indépendance de deux évènements A et B

$$P(A \cap B) = P(A).P(B)$$

$$P(A/B) = P(A)$$

$$P(B/A) = P(B)$$

Rq : Deux évènements disjoints ne sont pas indépendants.

Indépendance mutuelle d'une séquence d'évènements

$$\forall I \in P(2, \dots, n), P\left(\bigcap_I A_i\right) = \prod_I P(A_i)$$

Théorème de Bayes

pour deux évènements A et B:

$$P(A/B) = \frac{P(B/A).P(A)}{P(B)}$$

Preuve

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \text{ et } P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$\Rightarrow \frac{P(A / B)}{P(B / A)} = \frac{P(A)}{P(B)} \Leftrightarrow \text{formule de Bayes}$$

Le théorème de Bayes vise à calculer les probabilités a posteriori d'un événement en fonction des probabilités a priori de cet événement. A priori et a posteriori s'entendent par rapport à la connaissance d'une information. L'exemple typique est celui du diagnostic : a priori on juge que le patient a une telle probabilité d'avoir la maladie M.

Ce théorème est un outil de modélisation de la démarche diagnostique (en sens large). L'idée de base du théorème de Bayes est de calculer les probabilité a posteriori d'un événement A, sachant qu'un autre événement B s'est produit, en fonction de sa probabilité a priori : on veut donc exprimer $P(A|B)$ en fonction de $P(A)$.

Tout d'abord, si A et B sont indépendants, on vient de voir que $P(A|B) = P(A)$, puisque B n'apporte pas d'information sur A. Que se passe-t-il si B apporte une information ? Dans un tel cas, on a :

$$P(A / B) \neq P(A / \bar{B})$$

La probabilité de A sera différente selon que l'on a l'information que B s'est produit, ou que B ne s'est pas produit.

La clé de démonstration c'est d'écrire que : $B = (B \text{ et } A) \text{ ou } (B \text{ et non-}A)$ Cette égalité rappelle que les événements élémentaires qui définissent B peuvent être classés en deux catégories :

- ceux qui réalisent A (ils forment B et A),
- ceux qui ne le réalisent pas (ils forment B et non-A).

De plus, (B et A) est incompatible avec (B et non-A) parce que A et non-A le sont.

On peut donc écrire :

$$P(A / B) = \frac{P(B / A).P(A)}{P(B \text{ et } A) + P(B \text{ et } \bar{A})} \text{ On a de plus :}$$

$$P(B \text{ et } A) = P(B / A).P(A)$$

$$P(B|\bar{A}) = P(B/\bar{A}) \cdot P(\bar{A})$$

Le théorème de Bayes s'écrit :

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B/A) \cdot P(A) + P(B/\bar{A}) \cdot P(\bar{A})}$$

Généralisation pour un système complet d'évènements A_1, A_2, \dots, A_n :

$$P(B) = \sum_i^n P(B/A_i)P(A_i) \quad P(A_i/B) = \frac{P(B_i/A)P(A_i)}{\sum_i^n P(B/A_i)P(A_i)}$$

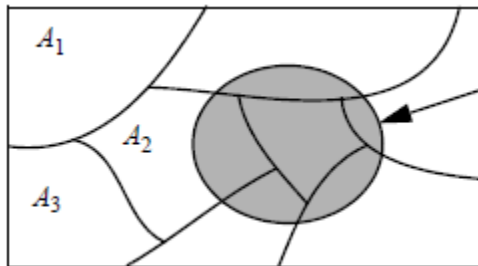
Le théorème est une forme développée de cette formule que nous introduisons maintenant.

Considérons des événements A_1, \dots, A_n tels qu'ils forment une partition de l'ensemble fondamental

E . Par définition, les A_i s'excluent mutuellement et leur union est E :

$$\forall (i \neq j), (A_i \cap A_j = \emptyset), \bigcup_i^n A_i = E$$

Soit B un événement quelconque



De $E = A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n$ et de $B \cap E = B$; on tire $B \cap (A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n)$.

Soit, par distributivité, $B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup \dots \cup (B \cap A_n)$.

En remarquant que les $(B \cap A_i)$ sont exclusifs, puisque les A_i le sont, et en appliquant la 3^{ème} règle

du calcul des probabilités on obtient la formule dite des « probabilités totales » :

$$P(B) = P(B \cap A_1) \cup P(B \cap A_2) \cup P(B \cap A_3) \cup \dots \cup P(B \cap A_n)$$

En appliquant le théorème de la multiplication :

$$P(B) = P(B / A_1).P(A_1) + P(B / A_2).P(A_2) + P(B / A_3).P(A_3) + \dots P(B / A_n).P(A_n)$$

Or, par la forme simple du théorème de Bayes, on a $P(A_i / B) = \frac{P(B / A_i)P(A_i)}{P(B)}$

D'où le théorème de Bayes :

$$P(A_i / B) = \frac{P(B_i / A)P(A_i)}{P(B / A_1).P(A_1) + P(B / A_2).P(A_2) + P(B / A_3).P(A_3) + \dots P(B / A_n).P(A_n)}$$

Exemples

1. On considère 3 cartes à jouer de même forme. Cependant, les deux faces de la première carte ont été colorées en noir, les deux faces de la deuxième carte en rouge tandis que la troisième porte une face noire et l'autre rouge. On mélange les trois cartes au fond d'un chapeau puis une carte tirée au hasard en est extraite et placée au sol. Si la face apparente est rouge, quelle est la probabilité que l'autre soit noire ?

Soient RR, NN et RN respectivement les événements, 'la carte choisie est entièrement rouge', 'entièrement noire' et 'bicolore'. Soit encore R l'événement, 'la face apparente de la carte tirée est rouge'.

On aura :

$$P(RN / R) = \frac{P(RN \cap R)}{P(R)}$$

$$P(RN / R) = \frac{P(R / RN).P(RN)}{P(R / RN).P(RN) + P(R / RR).P(RR) + P(R / NN).P(NN)}$$

$$P(RN / R) = \frac{(1/2).(1/3)}{(1/2).(1/3) + (1).(1/3) + (0).(1/3)} = \frac{1}{3}$$

2. Reprenons l'exemple des résultats au concours des étudiants de Paris VI. Soit R l'événement « un étudiant de Paris VI est reçu ». On a, en notant C_1, C_2, C_3 les 3 anciens CHU Saint Antoine, Pitié et Broussais respectivement :

$$P(R) = P(R / C_1) \cdot P(C_1) + P(R / C_2) \cdot P(C_2) + P(R / C_3) \cdot P(C_3)$$

[noter que c'est la même chose que la somme des probabilités des chemins de l'arbre, qui conduisent à un succès]

Le théorème de Bayes permet de répondre à la question duale. Au lieu de chercher la probabilité d'obtenir un étudiant reçu sachant qu'il venait d'un CHU donné, on cherche la probabilité qu'un étudiant ait été inscrit à un CHU donné sachant qu'il a été reçu (probabilité des causes). Calculons la probabilité qu'un étudiant reçu soit issu du CHU Pitié-Salpêtrière.

$$P(C_2 / R) = \frac{P(R / C_2) \cdot P(C_2)}{P(R / C_1) \cdot P(C_1) + P(R / C_2) \cdot P(C_2) + P(R / C_3) \cdot P(C_3)}$$

$$\text{Avec } P(C_1) = 0,25; P(C_2) = 0,50; P(C_3) = 0,25$$

$$\text{Et } P(R / C_1) = 0,15; P(R / C_2) = 0,20; P(R / C_3) = 0,10$$

$$\text{D'où } P(C_2 / R) = \frac{(0,20) \cdot (0,50)}{(0,15) \cdot (0,25) + (0,20) \cdot (0,50) + (0,10) \cdot (0,50)} = 0,61$$

Ce qui signifie que, dans ce cas, la probabilité qu'un étudiant appartienne à C_2 , s'il est reçu, est plus grande que si l'on ne sait rien (probabilité a priori $P(C_2) = 0,50$). Cette façon de calculer les probabilités des causes connaissant les effets est essentielle en médecine. En effet, le problème du diagnostic peut être posé en ces termes.

Chapiter 03

Application : Test de diagnostic médical

Le test de diagnostic médical représente un problème inverse, qui vise à déterminer les causes de la maladie, les effets étant connus. nous essayons de traiter le problème avec les probabilités Bayésiennes.

Pour illustrer l'importance de ce théorème en termes de modélisation du diagnostic :

- supposons que A soit l'événement *atteint de la maladie M* .
- supposons que B soit l'événement *positif pour le test diagnostic D* .

Le théorème de Bayes permet de calculer la probabilité qu'un patient positif pour le test diagnostic D soit atteint de la M , en fonction de la probabilité de la maladie M dans la population.

$$P(M / D^+) = \frac{P(D^+ / M).P(M)}{P(D^+ / M).P(M) + P(D^+ / \overline{M}).P(\overline{M})}$$

Pour chaque individu examiné, nous nous intéressons aux caractères :

- M = avoir la maladie
- $non-M$ = ne pas avoir la maladie
- T = avoir un résultats positif (+) au test
- $non-T$ = avoir un résultat négatif (-) au test

Dans une phase d'évaluation, le test est appliqué à un groupe M et à un groupe $non-M$. Les groupes sont formés à l'aide d'un test de référence (*gold standard*) dont le résultat est considéré comme sûr.

On détermine les fréquences des quatre résultats possibles indiqués par

$$n_{TM}, n_{T\overline{M}}, n_{\overline{T}M}, n_{\overline{T}\overline{M}}$$

	M	\overline{M}	Total
T	n_{TM}	$n_{T\overline{M}}$	n_T
\overline{T}	$n_{\overline{T}M}$	$n_{\overline{T}\overline{M}}$	$n_{\overline{T}}$
Total	n_M	$n_{\overline{M}}$	n

$Sensibilité = \frac{n_{TM}}{n_M} =$ proportion de '+' parmi les malades.

$Spécificité = \frac{n_{\overline{T}\overline{M}}}{n_{\overline{M}}} =$ proportion de '-' parmi les sains.

Des valeurs élevées de sensibilité et de spécificité indiquent clairement une bonne qualité du test.

	M	\overline{M}	Total
T	950	10	
\overline{T}	50	990	
Total	1000	1000	

$$Sensibilité = \frac{n_{TM}}{n_M} = \frac{950}{1000} = 95\%$$

$$Spécificité = \frac{n_{\overline{T}\overline{M}}}{n_{\overline{M}}} = \frac{990}{1000} = 99\%$$

La sensibilité et la spécificité du test ne sont pas des informations directement utiles à l'individu (le patient) auquel le test est administré. En effet, supposons que la sensibilité et la spécificité d'un certain test soient :

- sensibilité = 95%
- spécificité = 99%

Supposons aussi que le médecin applique ce test à un patient et obtient un résultat positif (+).

Il s'agit de répondre à la question : *Quelle est la probabilité que le patient soit réellement malade ?*

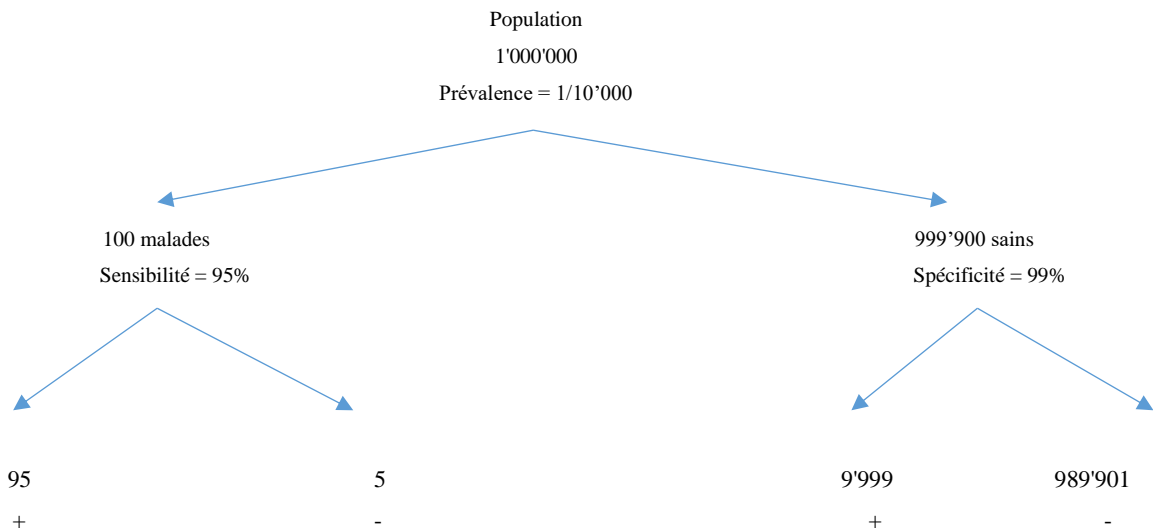
Pour répondre il nous faut une information supplémentaire concernant la fréquence de la maladie M dans la population, c'est-à-dire la *prévalence* de la maladie.

Supposons par exemple, qu'il y ait 1 seul malade par 10'000 habitants (*prévalence de $M = 1/10'000$*).

La prévalence de $1/10'000$ nous permet d'affirmer que dans une population hypothétique de 1'000'000 d'individus, on peut s'attendre à 100 malades et 999'900 sains.

Le test dépiste 95 cas positifs et 5 cas négatifs parmi les malades, car la sensibilité est de 95%.

Le test trouve aussi 9'999 résultats positifs et 989'901 résultats négatifs dans la partie saine de la population (voir schéma suivant).



En conclusion, la proportion de malades parmi les cas positifs est de 95/10'094, ce qui indique que les chances qu'un individu positif au test soit réellement malade sont seulement de 0.0094 (approximativement 1%).

$P(M|T)$ = Probabilité que le patient soit malade donnée le résultat positif du test (sachant que le résultat du test est positif).

$$P(M / T) = \frac{P(T / M) \cdot P(M)}{P(T / M) \cdot P(M) + P(T / \bar{M}) \cdot P(\bar{M})}$$

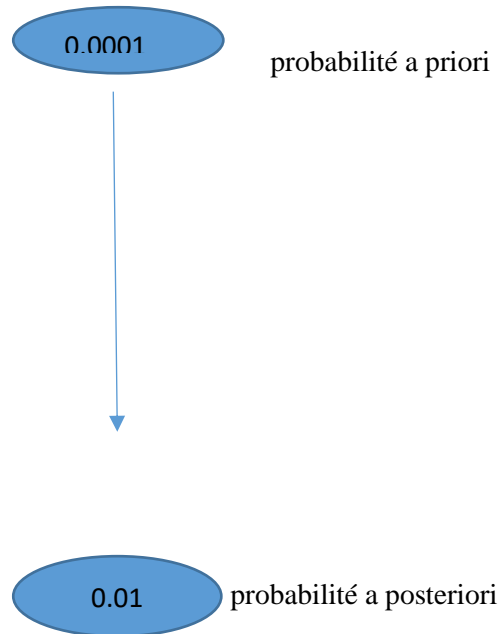
Sensibilité = 0.95 *prévalence* = 0.0001
 1 - *Spécificité* = 0.01 1 - *prévalence* = 0.99999

$P(M|T)$ = Probabilité que le patient soit malade donnée le résultat positif du test (sachant que le résultat du test est positif).

$$P(M / T) = \frac{0,95 \cdot 0,0001}{0,95 \cdot 0,0001 + 0,01 \cdot 0,9999}$$

$$P(M / T) = 0.0094$$

Remarque Les valeurs de $P(T|M)$ et $P(non-T|non-M)$ dans la population sont inconnues. Nous utilisons leurs estimations 95%, respectivement 99%, obtenus dans la phase d'évaluation à l'aide d'un échantillon de 1'000 malades et 1'000 sains.



- $P(T|M)$ = sensibilité du test
- $P(non-T|non-M)$ = spécificité du test
- $P(M|T)$ = valeur prédictive positive du test
- $P(non-M|non-T)$ = valeur prédictive négative du test
- $P(T|non-M)$ = taux de faux positifs (= 1-spécificité)

Attention ! pour certains auteurs : taux de faux positifs = $P(non-M|T)$.

Il est souvent difficile de connaître la prévalence $P(M)$ avec précision. Il convient alors d'examiner le test pour différentes valeurs de $P(M)$. Par exemple, si $P(T|M)=0.95$ et $P(non-T|non-M)=0.99$.

$P(M)$	$P(\bar{M} / T)$	$P(M / \bar{T})$
1/1'000'000	0.9999	0.00000
1/100'000	0.9991	0.00000
1/10'000	0.9906	0.00001

1/1'000	0.9132	0.00005
1/500	0.8401	0.00010
1/200	0.6769	0.00025
1/100	0.5103	0.00051

Le taux de faux négatifs est bon : moins de 5 malades sur 10'000 (~ 0.00051) échappent au test. Par contre, le taux de faux positifs est élevé : sur 100 individus positifs plus de 50 (~ 0.5103) sont sains. La décision de maintenir un tel test dépendra de l'importance de la maladie, des conséquences du test, des coûts des examens complémentaires et de l'éventuel traitement, des chances de succès du traitement, etc. Il est important de savoir qu'on peut réduire les taux d'erreur en combinant (ou en répétant) deux ou plusieurs tests.

Notations :

- M = le patient est malade
- T_1 = le premier test est positif
- T_2 = le deuxième test est positif
- $non-M$ = le patient n'est pas malade
- $non-T_1$ = le premier test est négatif
- $non-T_2$ = le deuxième test est négatif

Hypothèse : Les résultats des deux tests sont indépendants.

Données :

- $P(M)$ = prévalence = 10%
- $P(T_1|M)$ = sensibilité du premier test = 75%
- $P(non-T_1|non-M)$ = spécificité du premier test = 80%
- $P(T_2|M)$ = sensibilité du deuxième test = 90%

$P(non-T_2|non-M)$ = spécificité du deuxième test = 95% Valeur prédictive positive (probabilité a posteriori) du premier test, $P(M|T_1)$?

$$P(M / T_1) = \frac{P(T_1 / M) \cdot P(M)}{P(T_1 / M) \cdot P(M) + P(T_1 / \bar{M}) \cdot P(\bar{M})}$$

$$P(M / T_1) = \frac{0,75 \cdot 0,10}{0,75 \cdot 0,10 + 0,20 \cdot 0,90}$$

$$P(M / T) = 0,294$$

Première approche, grâce à l'indépendance des résultats :

$$P(T_1 \text{ et } T_2 / M) = P(T_1 / M) \cdot P(T_2 / M) = 0,75 \cdot 0,90 = 0,675$$

$$P(T_1 \text{ et } T_2 / \bar{M}) = P(T_1 / \bar{M}) \cdot P(T_2 / \bar{M}) = 0,20 \cdot 0,05 = 0,010$$

et donc :

$$P(M / T_1 \text{ et } T_2) = \frac{P(T_1 \text{ et } T_2 / M) \cdot P(M)}{P(T_1 \text{ et } T_2 / M) \cdot P(M) + P(T_1 \text{ et } T_2 / \bar{M}) \cdot P(\bar{M})}$$

$$P(M / T_1 \text{ et } T_2) = \frac{0,675 \cdot 0,100}{0,675 \cdot 0,100 + 0,010 \cdot 0,900} = 0,882$$

Deuxième approche :

$$P_{T_1}(M / T_2) = \frac{P_{T_1}(T_2 / M) \cdot P_{T_1}(M)}{P_{T_1}(T_2 / M) \cdot P_{T_1}(M) + P_{T_1}(T_2 / \bar{M}) \cdot P_{T_1}(\bar{M})}$$

$$P_{T_1}(M / T_2) = \frac{0,900 \cdot 0,294}{0,900 \cdot 0,294 + 0,050 \cdot 0,706} = 0,882$$

Deuxième approche :

Après avoir obtenu le résultat T_i ,
nous interprétons $P_{T_i}(M) = P(M|T_i)$
comme 'prévalence de M parmi les
résultats T_i '

$$P_{T_1}(M / T_2) = \frac{P_{T_1}(T_2 / M) \cdot P_{T_1}(M)}{P_{T_1}(T_2 / M) \cdot P_{T_1}(M) + P_{T_1}(T_2 / \bar{M}) \cdot P_{T_1}(\bar{M})}$$

$$P_{T_1}(M / T_2) = \frac{0,900 \cdot 0,294}{0,900 \cdot 0,294 + 0,050 \cdot 0,706} = 0,882$$

Le théorème de Bayes est accepté.

Conclusion

Le thème de ce mémoire décrit essentiellement l'utilisation du théorème de Bayes. Nous avons d'abord rappelé quelques outils et résultats d'analyse fonctionnelle puis nous avons donné les principaux aux résultats concernant les problèmes inverses.

Nous avons ensuite présenté le théorème de Bayes comme une méthode permettant le traitement de certains problèmes inverses.

Nous espérons que ce modeste travail, contribuera à éclairer celles et ceux qui pourraient être intéressés par la relation entre les probabilités bayésiennes et les problèmes mal posés.

Bibliographie

- [1] Michel Kern, Problèmes inverses : aspects numériques, Article · September 2002, National Institute for Research in Computer Science and Control.
- [2] F. Carrat, A. Mallet, V. Morice, Université Pierre et Marie Curie, Biostatistique, PACES - UE4 2013 – 2014.
- [3] Lionel Ségui, GdT Ignoutus LAAS-CNRS, Introduction aux problèmes mal posés, 18 janvier 2007.
- [4] Mémoire de master, Résolution Numérique d'un problème inverse. Application à l'imagerie médical, Université de Abdelhamid Ibn Badis de Mostaganem, 06/2017.
- [5] GOUSSEM Rahima, LASMI Razika. Méthode de régularisation de Tikhonov et applications, Mémoire présenté pour l'obtention du diplôme de Master en mathématiques, Promotion : 2015/2016.
- [6] Maria Abdelali. résolution du problème inverse en l'astrographie ultrasonore par une méthode variationnelle. thèse présente en vue de l'obtention du diplôme de philosophiae doctorat. institue de génie biomédical, ecole polytechnique de Montréal.

Abréviations et Notations

Quelques abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

\mathbb{R} : ensemble des nombres réels

A : opérateur linéaire défini sur un espace de Hilbert

A^* : l'opérateur adjoint de l'opérateur A

$\text{Ker}A$: le noyau de l'opérateur A

$\text{Im } A$: l'image de l'opérateur A

H : espace de Hilbert

(\cdot, \cdot) : produit scalaire

$\|\cdot\|$: une norme

