



Université des Sciences et de la Technologie Houari Boumediène
Faculté des Mathématiques
Alger

Mémoire

Présenté pour l'obtention du diplôme de : **Magister**
en : **Mathématiques**
Spécialité : **Recherche Opérationnelle (Mathématiques de gestion)**

Par : ABDELMALEK Feraz

Thème

**LIEN ENTRE LA METHODE D'ANALYSE
HIERARCHIQUE ET LE THEOREME DE BAYES.
APPLICATION AU DIAGNOSTIC MEDICAL**

Soutenu le 14.12.2004, Devant le jury composé de :

M^f AIDER Meziane, Maître de conférences, USTHB

Président

M^f KELLADI Abdelkader, Professeur, USTHB

Directeur de thèse

M^f ABBAS Moncef, Professeur, USTHB

Examineur

M^f MOULAÏ Mustapha, Maître de conférences, USTHB

Examineur

Remerciements

Je tiens à exprimer ici mes vifs remerciements et ma vive reconnaissance à mon promoteur M^r KHELLADI Abdelkader, qui a su diriger ce travail avec beaucoup de patience et beaucoup d'efficacité.

Je tiens à remercier également M^r AIDÈR Meziane qui me fait le plaisir et l'honneur de présider le jury.

C'est aussi un grand plaisir pour moi que M^r ABBAS Moncef et M^r MOULAÏ Mustapha aient bien voulu faire partie du jury et examiner ce travail.

Mes remerciements vont également à M^r SAATY Thomas et M^r VARGAS Louis qui n'ont pas hésité de m'aider et de m'envoyer une documentation qui m'a été d'une grande opportunité.

Table des Matières

Introduction générale	1
Chapitre I : Historiques	
Introduction	4
1. Décisions dans le monde antique.....	4
2. Décisions à partir du 16ème siècle.....	5
3. Probabilités et statistiques.....	6
4. Décisions dans le 20ème siècle	7
5. Subjectivité.....	9
6. Théorie de décision	9
7. Deux écoles statistiques : classique et Bayésienne	10
8. Outils modernes d'aide à la décision	11
9. Applications	11
Chapitre II : Méthodologie de Hiérarchisation en Aide à la Décision	
Introduction	13
1. La méthode d' analyse hiérarchique	14
1.1 Schéma général	14
1.2 Identification et décomposition	14
1.3 Discrimination et comparaisons	16
1.3.1 Les comparaisons par paires	16
1.3.2 L'échelle de comparaison	17
1.3.3 Illustration	19
1.4 Synthèse des priorités	20
1.5 Agrégation des jugements	20
1.5.1 La méthode du vecteur propre	21
A. Description, Théorème de Perron – Frobenius	21
B. Indice et rapport de cohérence	27
C. Illustration	28
D. Remarques	29
E. Importance du choix d'échelle	29
1.5.2 La méthode des moindres carrées	31
1.5.3 La méthode du logarithme des moindres carrées	31
1.5.4 La méthode de Chi Deux	34
1.6 Causes d'incohérence	34
1.7 Transitivité dans les jugements.....	35
2. Analyse de sensibilité	37
2.1 Changement des poids des critères	37
2.2 Changement des poids des alternatives par rapport aux critères	38
3. Le feedback	40
3.1 Evaluation haut – bas	41
3.2 Evaluation bas – haut	41
3.3 L'AHP et le feedback	42
3.4 La supermatrice pour le feedback	43

4. Analytic Network Process ANP (procédure d'analyse par réseau)	43
4.1 Schéma général	43
4.2 Comment calculer une fonction de matrice ? (Formule de Sylvester).....	46
4.3 Etapes de la méthode d'analyse par réseau (ANP)	52
Conclusion	53

Chapitre III : L'approche Bayésienne de l'aide à la décision

Introduction	56
1. Théorème de Bayes	56
2. La vraisemblance.....	58
3. La distribution à posteriori.....	59
4. L'inférence Bayésienne	61
5. La méthode Bayésienne	61
6. La distribution à priori	62
7. Le statut des vraisemblances et la distribution à posteriori	62
8. Qu'est ce qu'une bonne inférence ?	64
9. Implémentation	65
10. Comparaison avec l'inférence classique	66

Chapitre IV : Lien entre AHP et le Théorème de Bayes

1. Introduction	68
2. Origine	68
3. Le théorème de Bayes et le diagnostic médical	69
4. La dépendance entre les symptômes	71
5. Le modèle	72
6. Dépendance entre les symptômes	77
7. Au-delà du théorème de Bayes	79
8. Conclusion	81

Chapitre V : Application

1. Introduction	84
2. Evaluation des priorités des symptômes par rapport aux maladies	85
3. Evaluation des priorités des maladies par rapport aux symptômes	87
4. Evaluation des priorités des symptômes par rapport aux symptômes	89
5. Priorités globales	91
6. Analyse des résultats	92

Conclusion	95
-------------------------	----

Annexe	97
---------------------	----

Bibliographie

Introduction générale

Notre vie est la somme de nos décisions, aussi bien dans nos affaires professionnelles que personnelles. Souvent "Quand est-ce qu'on décide ?" est aussi important que "Qu'est-ce qu'on décide ?". Décider trop rapidement peut être hasardeux, et prendre beaucoup de temps pourra signifier rater des opportunités.

Le décideur, motivé par le besoin de bien contrôler ou de prévoir, souvent se retrouve en face d'un système complexe de composantes inter-reliées, telles que des ressources, des résultats désirés ou des objectifs. Mieux comprendra-t-il la complexité du système, mieux sa décision sera.

Parce que plusieurs théories d'aide à la décision sont en développement et en concurrence, la théorie la plus utile doit être en harmonie avec les besoins et la nature humaine. De plus cette théorie ne devrait pas exiger de longues années d'apprentissage et de formation pour appliquer des techniques que seulement un fanatique pourrait apprécier.

Quand une seule expérience induit une variété de sensations et d'activités différentes alors un certain type d'interprétation sera requis pour combiner ces activités. La manière de les combiner dépend de l'objectif que sont supposés servir, autrement dit, nos objectifs nous dictent où on doit concentrer notre effort, d'où le besoin aux priorités et aux mesures.

Si le décideur dispose de plusieurs alternatives à lesquelles il veut associer des priorités, alors elles doivent être significatives. Par exemple, si ces priorités représentent l'importance relative de valeurs monétaires ou de distances ..etc. alors elles doivent être identiques ou proches de celles qu'un expert en la matière aurait fournies en utilisant ses méthodes de mesures.

Dans le présent travail, on étudie une méthode qui permet d'obtenir des telles priorités.

Dans le premier chapitre, on donne un bref historique d'aide à la décision allant de ses premières significations pour l'être humain du monde antique jusqu'à son état actuel et ses différents domaines d'applications. Au deuxième chapitre, on détaille le fondement théorique

d'une méthode qui s'inscrit parmi les méthodes d'aide à la décision, c'est la méthode d'analyse hiérarchique (Analytic Hierarchy Process AHP) ainsi que ses extensions. L'approche Bayésienne est une autre approche d'aide à la décision largement utilisée dans les situations d'incertitude, ses grandes lignes sont présentées dans le troisième chapitre. Au quatrième chapitre on étudie un modèle basé sur l'AHP avec feedback pour résoudre le problème du diagnostic médical et on mettra en évidence le rôle que joue la méthode Bayésienne dans ce modèle. Une application sur un malade est illustrée dans le cinquième chapitre, suivie d'une conclusion et perspectives.

Chapitre I

Historique

Introduction :

La prise de décision consciente a probablement fait partie des facultés humaines depuis que l'homme a commencé à réfléchir, beaucoup de décisions doivent avoir lieu dans des conditions de manque d'informations et d'incertitude dans les résultats. Tout au début de l'histoire, l'homme a cherché de l'aide pour prendre ses décisions aux forces super-naturelles d'une certaine manière ou autre. Cependant l'amélioration de la compréhension humaine des processus naturels le long du temps a créé les bases d'une approche décisionnelle plus raisonnable.

La foule de découvertes en Europe a établi les fondements des approches mathématiques pour l'aide à la décision et la gestion du risque et d'incertitude, permettant ainsi à l'être humain depuis le dix-septième siècle d'utiliser ou d'abuser dans l'utilisation de ces méthodes pour prendre des décisions de caractère personnelles, politiques ou économiques.

Le développement de la théorie d'utilité, ayant ses racines dans le dix-septième et dix-huitième siècle, a été rétabli (après avoir été délaissée comme étant une science non mesurable au mi-dix-neuvième siècle) avec la publication historique du livre de Von Neuman et Morgenstern sur la théorie des jeux en 1944, le livre a revitalisé la théorie d'utilité et a finalement indiqué le chemin aux méthodes scientifiques pour fournir des mesures pour les nombreux, et souvent, critères conflictuels, dans la prise de décision multi-critères. Depuis cette époque, les méthodes et les théories se sont développées, devenues concurrentes et activement discutées par des milliers de scientifiques et de décideurs dans le monde entier, le nombre de théories a été augmenté par la prolifération de grands organismes, l'augmentation de la communauté académique dans des environnements de plus en plus complexes.

1. Décisions dans le monde antique :

Dans les époques très anciennes, la prise de décision consistait simplement à deviner les conséquences d'un choix imminent, et elle était concentrée sur la prévision des résultats, et pas le procédé de décision ou la technique, en raison d'absence d'information prédictive.

Les décideurs consultaient leurs aînés pour des alternatives et des données expérimentales au sujet de la probabilité de succès pour des choix de décision dans des situations semblables.

Puis à un certain point, cette fonction consultative est devenue la tâche des devins et des astrologues.

Alexandre le grand consultait régulièrement des prêtres la veille des grandes batailles. Toujours en général créateur, Alexandre ne recherchait pas des stratégies innovatrices de batailles chez ses conseillers, ce qu'il a eu besoin était des conseils sur les résultats potentiels de ses stratégies non encore essayées, ces informations pourraient seulement être fournies par ceux qui ont prétendu avoir une "fenêtre" sur le futur.

Les devins fatigués d'être les "caprices" des grands chefs comme Alexandre, se sont retirés dans des temples et dans des ermitages, créant ainsi les premiers lieux de consultation. L'un des plus célèbres était Oracle de Delphi de l'ancienne Grèce, les chercheurs de conseils de cet oracle, voulaient principalement un aperçu dans le futur, ils étaient impatients de savoir les conséquences d'un choix, plutôt que d'analyser des données ou des alternatives, selon les rapports historiques, la plupart des conseils d'oracle étaient suffisamment cryptes et flous pour tenir jusqu'à l'examen. Avec sa grande place dans la société, oracle de Delphi a garanti une position exclusive du marché avec sa clientèle.

Les anciens romains avaient eux aussi leurs oracles, leurs spécialités étaient l'explication des phénomènes naturels tel qu'où et comment la foudre se manifestait, d'autres interprétations de données ont été faites par Haruspicians. Les anciens romains ont ainsi transformé la prévision en affaire de gains rapides.

Le Talmud Babylonien, était un travail bien plus pragmatique, écrit entre 0-500 après JC, servait non seulement de base pour la loi juive, mais traitait des décisions telle que la division de la fortune d'un homme mort entre ses trois épouses.

2. Décisions à partir du 16^{ème} siècle :

Dans la deuxième moitié du seizième siècle, l'Angleterre était la maison de deux contributeurs, les plus brillants, dans l'étude des processus décisionnels, Francis Bacon et William Shakespeare, la contribution de Bacon était dans le développement de méthodes scientifiques, tandis que l'effort de Shakespeare était plus descriptif à travers des tragédies mettant en évidence les conséquences des décisions telle que : Othello, King Lear,

Romeo & Juliet, Hamlet et autres ... Ces approches peuvent être considérées, d'une certaine manière, comme causes à la recherche de théories et stratégies de décision développées dans le vingtième siècle.

Un siècle et demi plus tard, Benjamin Franklin a développé une approche dite de la "feuille de la balance", qui donne une manière simple mais exploitable pour structurer des informations pour les évaluer. Franklin recommande l'établissement d'une liste de deux colonnes correspondantes au pour et au contre de chaque alternative et puis calculer la ligne moyenne des valeurs

Cette technique d'évaluation paraît naïve par rapport aux techniques actuelles, mais elle présente une grande efficacité car elle ne liste pas seulement ce qui est connu à propos d'un choix donné, mais elle détermine aussi les informations manquantes et devant être rajoutées avant de prendre une décision.

3. Probabilités et statistiques :

Le 17^{ème} siècle est un siècle mémorable du point de vue de l'établissement des fondements des statistiques et des probabilités, deux fondements essentiels de la prise de décision moderne. La majorité du développement statistique gravite autour d'objectifs économiques, Dutchman Jan Laet a utilisé les trois catégories suivantes : les ressources, le redressement de la constitution, l'économie et la politique économique pour catégoriser les différents royaumes entre 1624 et 1640, son investigation était limitée à une description des données plutôt qu'une compréhension des causes, il paraît qu'il ne les a pas utilisées pour prendre toute décision utile dans la politique économique, John Graunt en Angleterre a cherché dans les causes et les régularités de l'augmentation de la population en 1662, William Petty publia son livre "les arithmétiques politiques" entre 1681 et 1690 qui traite de l'inférence statistique. Dans le 18^{ème} siècle, des travaux se sont succédés pour analyser et même pour choisir certaines politiques, notamment en ce qui concerne les dettes des gouvernements.

L'autre partie est le calcul des probabilités, le fondement de ses lois revient à un jeune homme italien appelé Gerolamo Cardano au 16^{ème} siècle.

Cardano était plus qu'un mathématicien, il est né en 1501 à Pavia en Italy, il était une personne extraordinairement inventive, son ambition était d'étudier la médecine, mais descendant d'une famille pauvre, il avait à chercher un autre moyen pour vivre et payer les

frais de ses études, sa meilleure chance était les jeux, son père Fazio Cardano était aussi un joueur (ainsi qu'un géomètre connaissant Leonardo da Vinci), mais malheureusement tué dans une dispute durant le jeu à cause de sa triche. Gerolamo n'avait pas l'intention de faire ainsi. En jouant, il a découvert les lois sous-jacentes de la théorie des probabilités d'où la publication, à l'âge de 23 ans, de son livre "Le livre des jeux de chance" *Liber de Ludo Aleae*.

Pour Cardano et ses contemporains, l'idée de calculer des probabilités et des valeurs espérées était centrale, par exemple, un joueur peut parier sur une pièce de monnaie pour qu'elle tombe sur une face, et la connaissance de quelques notions de probabilités pourra augmenter ses chances de gain et lui permettra de gérer son risque.

Daniel Bernoulli en 1738 a proposé l'idée que le décideur peut vouloir maximiser une valeur autre que la richesse. Autour de cette époque, les économistes politiques tels que Bentham ont encouragé le concept d' "utilité" comme étant une mesure de l'excès du plaisir contre le mal, qui devrait être maximisé par les systèmes sociaux et économiques, cependant leur usage de ce concept était plus qualitatif que quantitatif.

Un siècle ou deux plus tard, la théorie de la probabilité est devenue plus développée et appliquée par Bayes, Laplace, Pascal et autres ...

Dans les années 1870, la révolution économique a élevé la théorie d'utilité mathématique à une place centrale dans l'économie, mais elle est encore tombée pour quelques décennies plus tard, en raison de peur qu'elle ne soit pas réellement mesurable.

Tôt dans le XX^{ème} siècle, les méthodes mathématiques axiomatiques ont commencé à pénétrer dans la théorie de décision, dans les travaux de mathématiciens, statisticiens et psychologues tels que Ramsey, de Finetti, Alt et Frisch.

4. Décisions dans le 20^{ème} siècle :

Pour plusieurs, la décennie des années cinquante était l'âge d'or d'aide à la décision, les psychologues sociaux établissaient des données de la ligne de base à propos de la manière avec laquelle les individus ont pris des décisions et ont résolu des problèmes. Les scientifiques se sont mis à étudier comment les équipes de gestion ont travaillé, et ont commencé à construire des théories et des modèles basés sur ces données.

Une méthode appelée "BrainStorming" a été développée par NTL (National Training Laboratories au USA) pour développer les activités d'un groupe afin de stimuler l'interaction et la réflexion sociale. En premier, elle a été créée pour vaincre la répugnance naturelle des gens pour participer dans des groupes ouvertement et honnêtement. BrainStorming est une technique dans laquelle un animateur de groupe demande que les participants offrent des solutions alternatives pour un problème ou une question donnée, la règle du processus est que tous les participants doivent faire une contribution que l'animateur enregistre mot à mot, d'autres participants doivent alors encourager et construire des alternatives sans avoir recours à la négativité. A la fin de la session de BrainStorming, plusieurs données créatives seront enregistrées, aussi bien que les inquiétudes du groupe et quelques-uns des besoins subjectifs. Quand la technique était publiée, plusieurs animateurs se demandaient de ce qu'il faut faire avec les données accumulées, le BrainStorming continue d'être un outil pour générer une grande quantité de données potentiellement utiles, mais il reste à les éditer, les classer et les évaluer dans un sens objectif.

Basé sur ce travail du NTL, Charles Kepner et Benjamin Tregoe ont développé une méthodologie pratique pour la résolution des problèmes d'aide à la décision, Kepner et Tregoe ont conçu un processus pour isoler les sous – problèmes, générer les solutions alternatives et les évaluer pour déterminer la meilleure.

1. Définir le problème.
2. Formuler un objectif global de la décision.
3. Générer les critères.
4. Générer les alternatives.
5. Evaluer chaque alternative par rapport à chaque critère.
6. Comparer les points de chaque alternative et choisir celle ayant le plus grand point.

Cependant, le processus de Kepner-Tregoe n'est jamais devenu la méthodologie universelle d'aide à la décision pour plusieurs raisons, il y avait surtout deux inconvénients : le premier c'est la prise en considération de l'aspect subjectif du problème, la subjectivité joue un rôle actif dans l'établissement des critères pour même la décision la plus mécanique.

Deuxièmement, vu que tous les critères n'ont pas nécessairement la même importance, le processus idéal doit fournir une méthode pour comparer et évaluer les critères d'une certaine manière qui reflète leurs valeurs subjectives et objectives.

5. Subjectivité :

Henri Poincaré affirme en 1906 "...mais de tous ces chemins, qui nous mènera au but le plus rapidement ? qui nous dira lequel choisir ? Nous avons besoin d'une faculté qui nous aidera à percevoir le but de loin, cette faculté est l'intuition. La logique et l'intuition ont une partie nécessaire à jouer, les deux sont indispensables, la logique est l'instrument de la certitude, et l'intuition est l'instrument de l'invention ... "

Carl Jung retrouve la même idée en 1923 quand il écrit dans "Types psychologiques" : " Nous ne devrions pas prétendre de comprendre le monde par l'intelligence seule, nous le craignons de même par nos sentiments. Par conséquent, le meilleur jugement de l'intelligence est seulement la moitié de la vérité ..."

Les conclusions de Henri Poincaré et Carl Jung se sont confirmées dans le développement de la théorie de décision et la théorie des probabilités. La modélisation de notre partie intuitive peut nous aider à accomplir nos décisions qui sont autrement intraitables.

Au début des années cinquantes, un nombre de scientifiques ont essayé de construire un modèle mathématique de la réalité subjective, par exemple, Likert a utilisé une technique d'échelle pour mesurer ce que signifie des mots pour différents gens afin d'évaluer l'opinion publique et mesurer la perception culturelle et les préjugés [ARL].

6. Théorie de décision :

En 1944, Jon von Neuman et Oskar Morgenstern ont publié un livre sur la théorie des jeux, qui l'a incluse dans le domaine d'aide à la décision, la théorie des jeux a permis de réaliser des applications intéressantes en économie, statistiques, psychologie et sciences politiques en théorie d'utilité, elle a mené à la définition des concepts du risque opposé et la recherche des mesures du risque. L'utilité a finalement trouvé une manière pour être mesurée, la raison de son délaissement à la fin du 19^{ème} siècle.

L'université de Princeton et la corporation RAND ont pris la bannière de Von Neuman et Morgenstern, et un travail novateur a été fait permettant le développement des arbres de décisions, les arbres des jeux, et ainsi de suite ...

La théorie des jeux aujourd'hui est utilisée dans l'aide à la décision dans des domaines allant de la stratégie des guerres jusqu'au jeu de cartes et le jeu d'échec.

7. Deux écoles statistiques : classique et Bayésienne :

Depuis que Gauss et Legendre en 1810 ont annoncé la découverte de la distribution normale, les statistiques ont exploité ses pleins avantages et ont harmonisé presque tout avec elle, Jeffreys et d'autres statisticiens tel que De Finetti ont défié l'exploitation de cette loi, et se sont dirigé vers le travail de Reverend Bayes, qui a publié son papier en mi-18^{ème} siècle. Bayes a élaboré une méthode pour combiner les probabilités de deux événements ou plus qui ne sont pas indépendants. Les statistiques classiques ont été largement utilisées durant la seconde moitié du 19^{ème} siècle et le début du 20^{ème} siècle.

D'autres auteurs ont contribué au développement des statistiques Bayésiennes de puis mi-19^{ème} siècle : De Finetti en économie, Zellner en économétrie et Ward Edwards en théorie de décision. Un pas notable dans la méthode Bayésienne en théorie de décision fut réalisé avec la publication en 1961 du livre de Howard Raiffa et Robert Schlaifer "Théorie de décision statistique appliquée" qui défi les règles statistiques classiques d'aide à la décision.

L'université de Chicago était un autre centre de recherche en théorie de décision statistique avec des noms tels que Arnold Zellner, Harry Robert et d'autres ingénieurs comme Ron Howard qui ont commencé à développer des méthodes formelles d'aide à la décision sous incertitude. A la fin des années soixantes et début des années soixantes-dix, l'institut de recherche de Stanford devient un centre principal en recherche de décision analytique avec Jim Matheson, WarnerNorth et Ron Howard, Raiffa et son étudiant Ralph Keeny ont réalisé un travail original en théorie d'utilité multi-attributs, en publiant en 1976 leur livre : "Décisions avec objectifs multiples : préférences et échanges de valeurs", Keeny a continué à développer cette approche et publia son livre "la réflexion concentrée sur les valeurs" en 1992.

8. Outils modernes d'aide à la décision :

L'arrivée des systèmes informatiques modernes et des grandes bases de données à partir de la fin des années 1960 a également servi comme un point de départ pour de nouveaux défis dans l'aide à la décision d'organisation, le besoin d'accès et de traitement de l'information est une partie critique du processus décisionnel, et en dépit de toute l'information, ce n'est que récemment que les outils –aussi tard que le début des années 1990- ont apparu pour faire face aux volumes de données et de l'analyse complexe si nécessaire pour la décision.

Les interfaces graphiques se sont développées le long des dernières années pas seulement parce qu' une interface standard était nécessaire, mais parce que de telles interfaces sont à la fois à usage intuitif et sont très puissantes pour représenter les renseignements. De plus ils fournissent un environnement rationnel plus flexible que tout stylo ou papier, avec une grande vitesse pour assembler les données et une bonne façon pour les visualiser. La combinaison des outils puissants d'analyse et des interfaces graphiques d' utilisateurs devrait mener à des décisions plus compréhensibles et des décideurs mieux informés. On peut citer à titre d'exemple : Expert Choice, Logical Decisions, Criterium Decision Plus, Electre, et Arlington' s early Which & Why products, ainsi que d'autres outils d'aide à la décision.

9. Applications :

En ce qui concerne les applications, La revue de la littérature nous permet de constater que les outils d' aide à la décision ont fait l' objet d' applications diversifiées dans ~~des~~ domaines tels que : l' environnement, l' aménagement du territoire et la gestion de ses ressources naturelles, la planification minière, la gestion énergétique, la gestion des déchets, la localisation, la planification économique, la gestion financière et bancaire, la gestion urbaine et le transport, l' évaluation et la sélection de projets, la gestion de la production et des approvisionnements, la gestion des ressources humaines et matérielles, la gestion des systèmes de défense et la planification militaire, le développement international...

Cette liste n' est certainement pas exhaustive et l' on sait très bien que les applications concrètes d' outils d' aide à la décision ne font pas toujours l' objet d' une publication.

Chapitre II

Méthodologie de Hiérarchisation en Aide à la Décision

Introduction:

Dans ce chapitre on présentera une méthode dont son application réduira l'étude d'un système complexe à une séquence de comparaisons par paires de composantes proprement identifiées.

Cette méthode appelée Méthode d'Analyse Hiérarchique (Analytic Hierarchy Process AHP), élaborée par T. Saaty remonte à 1971, quand il a travaillé sur les problèmes de planification de la contingence pour le Département de Défense au USA, et puis pour rationner la consommation d'électricité pour les industries selon leur contribution à la richesse de la nation.

L'application majeure de cette méthode était l'étude du transport au Soudan en 1973 dirigée par Saaty, son enrichissement théorique ne cessait pas tout le long du chemin et surtout entre 1974 et 1978. Plus tard, les applications sont devenues nombreuses et variées allant de l'analyse pour l'agence de contrôle des armes et le désarmement, et différentes autres études de conflit, jusqu'à l'allocation de ressources selon les priorités pour des intérêts privés, gouvernementaux ou internationaux.

Cette théorie reflète ce qui apparaît comme une méthode innée du fonctionnement de l'esprit humain, disposant d'une multitude d'éléments, constituant une situation complexe, elle les agrège dans des groupes selon leurs propriétés communes, ce modèle permet la répétition de ce processus, donc on considère ces groupes ou leurs propriétés communes comme étant les éléments d'un nouveau niveau du système, ces éléments peuvent être aussi groupés selon un autre ensemble de propriétés qui forment les éléments d'un autre niveau jusqu'à l'arrivée à un seul élément "Top" qui est souvent l'objectif du processus décisionnel.

On vient de décrire ce qu'on appelle une hiérarchie : c'est un système de niveaux stratifiés, chacun consiste en plusieurs éléments ou facteurs. La question centrale est comment les facteurs ou les éléments du niveau le plus bas influent les éléments du niveau le plus haut : l'objectif global ? Du moment que cette influence n'est pas identique par rapport à tous les éléments, on a besoin d'identifier leurs intensités ou priorités.

Les participants dans l'évaluation tendent à trouver que le processus capture leur compréhension intuitive du problème, de plus, les limites psychologiques sont conformes aux conditions de stabilité mathématique des résultats.

Durant le processus décisionnel, il n'y a pas de chose qu'on peut considérer comme étant *La réponse* mais seulement *une réponse*, qui après une amélioration constante devient *La réponse* adressée au décideur.

1. La méthode d' analyse hiérarchique

1.1 Schéma général

La méthode d' analyse hiérarchique (Analytic Hierarchy Process AHP) a été développée par Thomas L.Saaty en 1971, c' est une procédure systématique pour représenter les éléments d' un problème sous forme hiérarchique, elle consiste à diviser le problème en sous problèmes, et puis elle guide le décideur à travers une série de comparaisons par paires pour exprimer l' importance relative d' un élément de cette hiérarchie. Ces jugements seront transformés en nombres et des procédures et principes seront utilisés pour synthétiser les différents jugements afin de dériver des priorités pour les critères et puis pour les solutions alternatives.

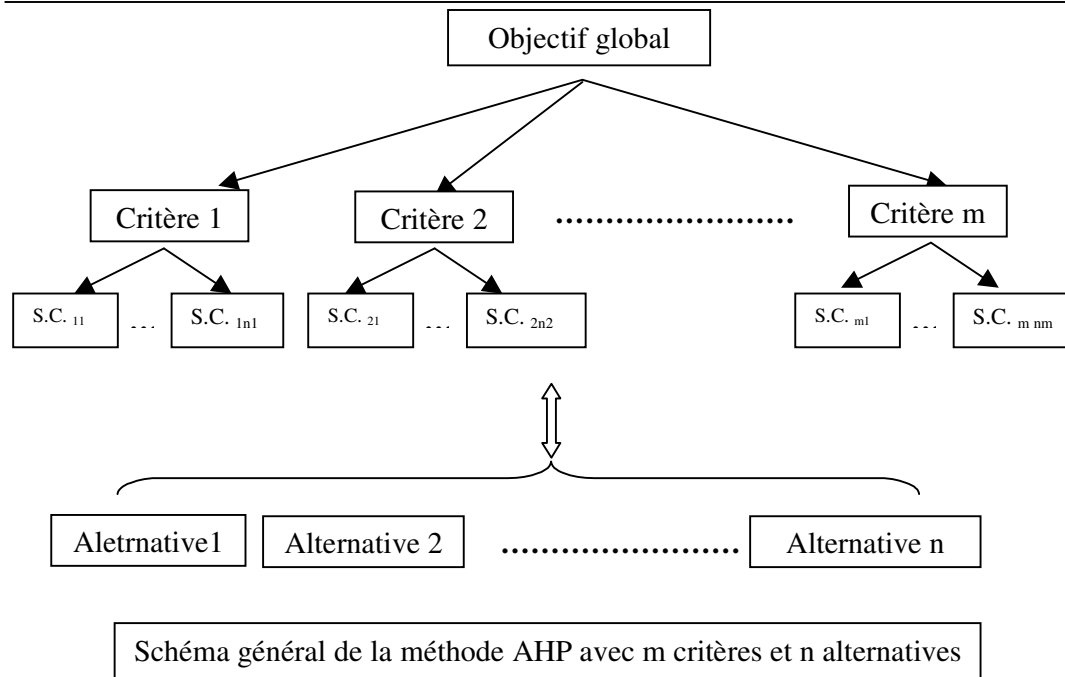
La résolution du problème passe par plusieurs étapes, on commence par l' identification des éléments du problème à prendre en considération, ensuite les comparer et les évaluer pour arriver à prendre une solution et mesurer sa performance. Ce processus sera répété jusqu' à l' obtention d' une solution satisfaisante qui répond à toutes les caractéristiques exigées par le décideur.

Même si le problème comporte des éléments dont la comparaison fait appel aux sentiments et aux choix subjectifs, la technique des comparaisons par paires reste toujours utilisable pour déterminer leurs priorités.

Cette approche nécessite une capacité de réflexion rigoureuse et créative pour identifier les constituants du problème et établir des relations entre eux. Elle est composée de trois étapes, la première est l' identification et la décomposition, la seconde est la discrimination et la comparaison, la dernière est la synthèse des priorités.

1.2 Identification et décomposition

Dans cette étape il s'agit de concevoir une structure hiérarchique du problème. Dans sa forme la plus élémentaire, une hiérarchie est structurée d' un niveau haut correspondant à l'objectif global, et passant par des niveaux intermédiaires correspondants aux critères et éventuellement aux sous critères, jusqu' à l' arrivé au niveau le plus bas correspondant aux alternatives.



S.C. : Sous Critère.

Exemple : Allocation d'énergie

On veut déterminer des priorités de différents grands utilisateurs d'énergie dans la société par rapport à leur contribution dans plusieurs critères.

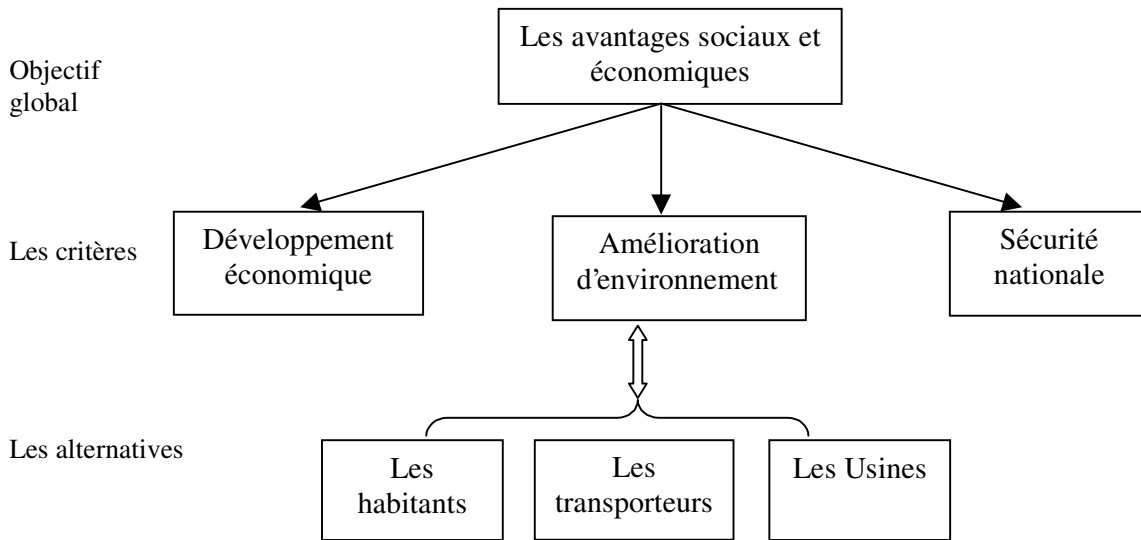
On considère trois utilisateurs d'énergie

- Les habitants
- Les transporteurs
- Les usines

Ces grands utilisateurs vont être évalués selon trois critères :

- La contribution dans le développement économique.
- La contribution dans l'amélioration d'environnement.
- La contribution dans la sécurité nationale.

On construit la hiérarchie suivante :



La loi de la continuité hiérarchique nécessite que les éléments du niveau le plus bas soient comparables par paires par rapport aux éléments du niveau juste le plus haut qui sont eux même comparables par rapport au niveau juste le plus haut, jusqu'à l'arrivée à l'entrée de la hiérarchie. L'objectif est de dériver des priorités aux alternatives qui reflètent le mieux possible leur impact sur l'objectif global.

1.3 Discrimination et comparaisons :

Une fois la représentation hiérarchique du problème est obtenue, on s'intéresse à l'établissement des priorités entre les critères et entre les alternatives par rapport aux critères.

1.3.1 Les comparaisons par paires :

Dans l'AHP, les éléments de chaque niveau sont comparés entre eux par paires par rapport à une propriété commune dans le niveau adjacent supérieur.

Soient A_1, A_2, \dots, A_n : n éléments à comparer entre eux par rapport à une propriété commune.

a_{ij} représente la préférence de l'élément A_i à A_j par rapport à cette propriété commune.

l'ensemble des comparaisons est stocké dans une matrice carrée :

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

avec $a_{ij} = \frac{1}{a_{ji}}$; $i, j = 1, \dots, n$: c'est la propriété de réciprocité car les deux éléments i et j

participent dans ce jugement, Une telle matrice est appelée *matrice réciproque*.

Si deux éléments A_i et A_j ont la même préférence alors $a_{ij} = a_{ji} = 1$, en particulier $a_{ii} = 1$, $i = 1..n$

Tout d'abord les jugements seront exprimés verbalement, en choisissant un jugement parmi : également, légèrement plus, fortement plus, très fortement plus, absolument plus important.

Ces jugements seront transformés en nombres selon une certaine échelle.

1.3.2 L'échelle de comparaison : [SAT77, SAT90]

Souvent, il existe une échelle sous-jacente à un problème donné, et les jugements seront exprimés sous forme de rapport à partir de cette échelle, par exemple, si on compare deux objets A et B de poids respectifs w_A et w_B , donc le rapport w_A / w_B est entré dans la matrice pour l'importance relative de l'objet A par rapport à l'objet B , réciproquement w_B / w_A est entré pour mesurer l'importance relative de l'objet B par rapport à l'objet A .

Soient A_1, A_2, \dots, A_n : n éléments de poids respectifs w_1, w_2, \dots, w_n .

Les comparaisons peuvent être représentées comme suit :

$$\begin{matrix} & A_1 & A_2 & \dots & A_n \\ A_1 & \left(\begin{matrix} 1 & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & 1 & \dots & \frac{w_2}{w_n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & 1 \end{matrix} \right) \\ A_2 & & & & \\ \cdot & & & & \\ \cdot & & & & \\ A_n & & & & \end{matrix}$$

On commence par un élément à gauche de la matrice et on demande au décideur d'exprimer verbalement le degré d'importance de cet élément par rapport à un élément en haut de la matrice.

On répète ce procédé pour toutes les paires.

Généralement dans l'AHP, on utilise l'échelle suivante pour transformer les jugements verbaux en nombres (appelé échelle fondamentale de 1 à 9) [SAT90] :

Valeurs numériques	Définition
1	Egalement important (aucune préférence)
3	Légèrement plus important
5	Fortement plus important
7	Très fortement plus important
9	Absolument plus important (une préférence absolue)
2,4,6,8	Valeurs intermédiaires pour mettre en évidence des compromis
Valeurs inverses	Valeurs Inverses utilisées pour montrer la dominance du second élément par rapport au premier

Pour savoir d'où vient cette échelle, voir l'annexe.

On trouve dans [SAT90] une comparaison de plusieurs échelles, par exemple :

l'échelle : 1, 2, 3

l'échelle : 1, 2, 3, 4, 5

l'échelle : 1..9

l'échelle : 0.1, 0.2, 0.3, ..., 0.7

l'échelle : 1, 2², 3², 4², ..., 9²

l'échelle : 2⁰, 2^{1/2}, 2^{2/2}, 2^{3/2}, ..., 2^{8/2}

..etc.

Sur plusieurs exemples, (dont on connaît les solutions exactes pour quelques-uns) on remarque que l'échelle de 1 à 9 est la meilleure.

1.3.3 Illustration :

Dans la matrice suivante, on compare les critères de l'exemple précédent :

	Develop Econom	Améliorat. D'environn	Sécurité nationale
Developpement Economique	1	5	3
Amélioration D'environnement	1/5	1	1/2
Sécurité Nationale	1/3	2	1

Quand l'économie est comparée avec l'amélioration d'environnement et avec la sécurité nationale selon leur impact socio-politique, l'économie est jugée fortement plus importante dans le premier cas et faiblement plus importante (mais toujours plus importante) dans le second cas, d'où les valeurs 5 et 3 dans la première ligne. La sécurité nationale est également importante et légèrement plus importante que l'amélioration d'environnement.

Les comparaisons par paires des alternatives c-à-d les grands utilisateurs d'énergie par rapport aux critères, donnent les matrices suivantes :

Econom	Habitants	Trans	Usines
Habitants	1	3	5
Transporteurs	1/3	1	2
Usines	1/5	1/2	1

Environnem	Habitants	Trans	Usines
Habitants	1	2	7
Transporteurs	1/2	1	5
Usines	1/7	1/5	1

Sécurité	Habitants	Trans	Usines
Habitants	1	2	3
Transporteurs	1/2	1	2
Usines	1/3	1/2	1

Remarque :

Parfois quand les poids sont des mesures connues telles que des tonnes ou des valeurs monétaires, il est préférable de les utiliser directement que de construire une matrice de jugements. Cependant ce processus peut engendrer des erreurs, particulièrement quand

l'utilité du décideur n'est pas reflétée en terme de rapport, par exemple, pour un homme riche un dollar ou deux peuvent être de la même utilité, malgré que leur rapport montre des utilités différentes.

1.4 Synthèse des priorités :

La dernière étape de l'AHP consiste à établir les priorités globales des alternatives. Supposons qu'on a N alternatives et M critères, Soit α_j la priorité du critère j, $j = 1, \dots, M$ et p_{ij} la priorité de l'alternative i par rapport au critère j, $i = 1, \dots, N ; j = 1, \dots, M$ Alors la priorité globale w_i de l'alternative i par rapport à l'objectif global est donnée par :

$$w_i = \sum_{j=1}^M \alpha_j p_{ij} \quad i = 1, \dots, N$$

1.5 Agrégation des jugements :

La question qui se pose maintenant est : Etant donnée une matrice A des comparaisons par paires, comment les agréger en un seul vecteur de priorités ? Si on disposait des poids de tous les n éléments à comparer w_1, w_2, \dots, w_n

On suppose, sans perte de généralité, que $\sum_{i=1}^n w_i = 1$,

On définit la matrice $W =$
$$\begin{pmatrix} 1 & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & 1 & \dots & \frac{w_2}{w_n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & 1 \end{pmatrix}$$

On remarque que: $W_{ij} = \frac{1}{W_{ji}} \dots \dots \dots (1) \quad \forall i, j = 1, \dots, n$

$W_{ij} = W_{ik} * W_{kj} \dots \dots \dots (2) \quad \forall i, j, k = 1, \dots, n$

Toute matrice remplissant les propriétés (1) et (2) est appelée matrice cohérente. Dans ce qui suit, on verra un aperçu sur différentes méthodes de synthèse.

1.5.1 La méthode du vecteur propre :

A. Description, Théorème de Perron - Frobenius :

C'est la méthode originale démontrée Par Saaty [SAT90].

Si on était dans un cas idéal c-à-d W est une matrice cohérente alors, on remarque que chaque ligne est un multiple de la première ligne, donc la matrice est de rang égal à 1, et elle possède une seule valeur propre non nulle.

De plus $W_{ii} = 1, i = 1, \dots, n$ et la somme de toutes les valeurs propres est égale à la trace de W

($\text{Tr } W = \sum_{i=1}^n W_{ii}$) donc la seule valeur propre non nulle est égale à n .

soit $w = (w_1, w_2, \dots, w_n)^T$, on remarque que $W^*w = n^*w$, donc w est un vecteur propre de W correspondant à la valeur propre maximum n .

La méthode du vecteur propre consiste à estimer W_{ij} par a_{ij} de la matrice des comparaisons par paires.

Comme $W_{ij} > 0$ alors supposons que $a_{ij} > 0$

$$W_{ij} = \frac{1}{W_{ji}} \text{ alors supposons que } a_{ij} = \frac{1}{a_{ji}}$$

Soit A la matrice des comparaisons par paires,

Soit λ_{\max} sa valeur propre maximum et w^* son vecteur propre normalisé correspondant.

Comme A est une approximation de W et sous certaines conditions de cohérence, on suppose que w^* est une approximation de w .

Définition 1 :

Une matrice carrée A d'ordre n est dite réductible si l'ensemble des indices $1, 2, \dots, n$ peut être séparé en deux sous-ensembles complémentaires (sans indices communs) i_1, i_2, \dots, i_μ ; $j_1, j_2, \dots, j_\vartheta$ ($\mu + \vartheta = n$) tels que $a_{i_\alpha j_\beta} = 0$ ($\alpha = 1 \dots \mu$; $\beta = 1 \dots \vartheta$)

Sinon la matrice est dite irréductible.

Par une permutation de la matrice A , nous entendons une permutation des lignes de A combinée avec la même permutation des colonnes.

La définition d'une matrice réductible et d'une matrice irréductible peut être formulée comme suit :

Définition 1' : Une matrice A d'ordre n est dite réductible s'il existe une permutation qui la mette sous la forme $\tilde{A} = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$ où B et D sont des matrices carrées, sinon A est dite irréductible.

Définitions :

Soit A une matrice carrée d'ordre n.

Si on supprime de A, les éléments de la ligne i et de la colonne j, le déterminant de la matrice d'ordre n-1 ainsi obtenu est appelé mineur de a_{ij} et est noté $|M_{ij}|$.

L'expression $(-1)^{i+j} |M_{ij}|$ est appelée cofacteur de a_{ij} et est noté α_{ij} .

La matrice adjointe de la matrice A s'obtient en remplaçant chaque élément de la transposée de A par son cofacteur.

Si tous les éléments de la matrice A sont positifs ou nuls alors on notera : $A \geq 0$.

Lemme : si $A \geq 0$ est une matrice irréductible d'ordre n,

$$\text{alors : } (I + A)^{n-1} > 0.$$

Pour la preuve de ce lemme voir par exemple [GAN66].

En 1907, Perron a découvert une propriété remarquable du spectre (c-à-d des valeurs propres et des vecteurs propres) des matrices positives.[GAN66]

Les démonstrations suivantes sont données pour illustrer la méthode et qui ne se trouve pas aisément dans la littérature.

Théorème 1 (Perron) :

Une matrice positive A d'ordre n, a toujours une valeur propre réelle et positive r qui est une racine simple de l'équation caractéristique et est supérieure aux modules de toutes les autres valeurs propres.

A cette valeur propre "maximale" r, il correspond un vecteur propre $z = (z_1, \dots, z_n)$ de A de coordonnées positives $z_i > 0$, $i = 1, \dots, n$.

Une matrice positive est un cas particulier d' une matrice non négative irréductible, Frobenius a généralisé les théorèmes de Perron en étudiant les propriétés spectrales des matrices non négatives irréductibles.

Théorème 2 (Frobenius) :

Une matrice non négative irréductible A d' ordre n, a toujours une valeur propre réelle positive r qui est une racine simple de l' équation caractéristique. Les modules de toutes les autres valeurs propres ne sont pas supérieurs à r.

A la valeur propre maximale r, il correspond un vecteur propre de coordonnées positives.

de plus, si A possède h valeurs propres $\lambda_0 = r, \lambda_1, \dots, \lambda_{h-1}$ de module r, ces nombres sont tous distincts et sont des racines de l' équation $\lambda^h - r^h = 0$.

Preuve :

Puisque le théorème de Perron est un cas particulier du théorème de Frobenius, nous n' aurons à démontrer que ce dernier.

Soit $x = (x_1, \dots, x_n) \geq 0$ ($x \neq 0$) un vecteur réel fixé,

Posons $r_x = \min_{1 \leq i \leq n} \frac{(Ax)_i}{x_i}$ avec $(Ax)_i = \sum_{j=1}^n a_{ij} x_j$, $i = 1, \dots, n$

Dans la définition du minimum, nous excluons ici les valeurs de i pour lesquelles $x_i = 0$, il est clair que $r_x \geq 0$ et que r_x est le plus grand nombre réel ρ pour lequel $\rho x \leq Ax$.

Nous allons montrer que la fonction r_x prend sa valeur maximale r pour un certain $x \geq 0$:

$$r = r_z = \max_{x \geq 0} r_x = \max_{x \geq 0} \min_{1 \leq i \leq n} \frac{(Ax)_i}{x_i} \dots\dots\dots (1)$$

de la définition de r_x , il suit que lorsqu' un vecteur $x \geq 0$ ($x \neq 0$) est multiplié par un nombre $\lambda > 0$, la valeur de r_x ne change pas, par suite dans la recherche du maximum de r_x nous pouvons nous limiter à l' ensemble fermé M des vecteurs x pour lesquels : $x \geq 0$ et

$$(xx) \stackrel{\text{définition}}{=} \sum_{i=1}^n x_i^2 = 1$$

si la fonction r_x était continue sur M, l' existence d' un maximum serait garantie, cependant bien que continue en tout "point" $x > 0$, r_x peut avoir des discontinuités aux points limites de M où l' une des coordonnées s' annule.

nous introduirons donc au lieu de M, l' ensemble N de tous les vecteurs y de la forme

$$y = (I + A)^{n-1} x \quad (x \in M).$$

l' ensemble N comme M, est borné et fermé et n' est constitué, en vertu du lemme précédent, que par des vecteurs positifs.

De plus, en multipliant les deux membres de l' inégalité $r_x x \leq A.x$ par $(I+A)^{n-1} > 0$, nous obtenons :

$$r_x y \leq Ay$$

(avec $y = (I + A)^{n-1}.x$), d' où d' après la définition de r : $r_x \leq r_y$

par conséquent, dans le calcul du maximum de r_x , nous pouvons remplacer M par l' ensemble N constitué uniquement de vecteurs positifs.

Sur l' ensemble borné et fermé N , la fonction r_x est continue et prend sa plus grande valeur pour un certain vecteur $z > 0$.

Tout vecteur $z \geq 0$ pour lequel $r_z = r$ (2) Sera appelé vecteur extrémal.

Nous allons maintenant montrer que :

1°) le nombre r défini par (1) est positif et est une valeur propre de A .

2°) Tout vecteur extrémal z est positif et est un vecteur propre de A pour la valeur propre r

c-à-d : $r > 0$; $z > 0$; $A.z = r.z$ (3)

en effet : si $u = \underbrace{(1,1,\dots,1)}_{u \text{ fois}}$, $r_u = \min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}$ mais alors $r_u > 0$ car aucune ligne d' une matrice

irréductible $n \times n$ est constitué entièrement par des zéros, par suite $r > 0$ puisque $r \geq r_u$.

posons maintenant $x = (I + A)^{n-1} z$ (4)

alors en vertu du lemme précédent : $x > 0$. Supposons que $A.z - r.z \neq 0$, d' après de (2) et (4) et

$(I + A)^{n-1} > 0$ nous obtenons alors successivement :

$$A.z - r.z \geq 0, (I + A)^{n-1}(A.z - r.z) > 0, A.x - r.x > 0.$$

la dernière inégalité contredit la définition de r car elle indiquerait que $A.x - (r + \epsilon) x > 0$

pour un $\epsilon > 0$ suffisamment petit, c-à-d : $r_x \geq r + \epsilon > r$.

on a donc $A.z = r.z$,

mais alors : $0 < x = (I + A)^{n-1}.z = (1 + r)^{n-1}.z$ de sorte que $z > 0$.

Nous allons maintenant montrer que les modules de toutes les valeurs propres ne sont pas supérieurs à r

Dans ce qui suit C^+ désigne la matrice mod C obtenue à partir de C lorsque tous les éléments sont remplacés par leurs modules.

Posons $A.y = \alpha.y$ ($y \neq 0$) (5)

En considérant les modules des deux membres de cette équation, nous obtenons :

$$|\alpha|y^+ \leq Ay^+ \dots\dots(6)$$

d' où $|\alpha| \leq r_{y^+} \leq r$

soit y un vecteur propre correspondant à $r : A.y = r.y (y \neq 0)$

alors en faisant $\alpha = r$ dans (5) et (6), nous parvenant à la conclusion que y^+ est un vecteur extrémal de sorte que : $y^+ > 0$ c-à-d $y = (y_1, \dots, y_n)$ où $y_i \neq 0 \ i = 1, \dots, n$

d' où, il suit qu' à la valeur propre r , il correspond une seule direction caractéristique, car s' il y avait deux vecteurs propres linéairement indépendant z et z_1 , nous pourrions choisir deux nombres c et d tels que le vecteur propre $y = c.z + d.z_1$ ait une coordonnée nulle, ce qui est impossible, comme nous l' avons montré.

Considérons maintenant la matrice adjointe de la matrice caractéristique $\lambda.I - A$

$B(\lambda) = \Delta(\lambda)(\lambda I - A)^{-1}$ où $\Delta(\lambda)$ est le polynôme caractéristique de A et $B_{ij}(\lambda)$ le complément algébrique de l' élément $\lambda \delta_{ji} - a_{ji}$ dans le déterminant $\Delta(\lambda)$, du fait qu' un seul vecteur propre $z = (z_1, \dots, z_n)$ avec $z_i > 0 \ i = 1, \dots, n$ correspond à la valeur propre r (à un facteur près) il suit que $B(r) \neq 0$ et que dans toute colonne non nulle de $B(r)$, tous les éléments sont différents de zéro et sont du même signe, la même chose est vraie pour les lignes de $B(r)$, puisque dans le raisonnement précédant, A peut être remplacé par A^T ; de ces propriétés des lignes et des colonnes de A , il suit que tous les $B_{ij}(r)$ ($i, j = 1, \dots, n$) sont différents de zéro et sont du même

signe, par suite : $\sigma \Delta'(r) = \sigma \sum_{i=1}^n B_{ii}(r) > 0$

ce qui signifie que $\Delta'(r) \neq 0$ et que r est une racine simple de l' équation caractéristique $\Delta(\lambda) = 0$.

Puisque r est la racine maximale de $\Delta(\lambda) = \lambda^n + \dots$; $\Delta(\lambda)$ croit pour $\lambda \geq r$.

Par suite $\Delta'(\lambda) > 0$ et $\sigma = 1$ et $B_{ij}(r) > 0 \ i, j = 1, \dots, n$

Théorème 3 :

La valeur propre maximum de A (la matrice des comparaisons par paires) est réelle et positive, $w_i^* > 0, i = 1, \dots, n$

Preuve :

Découle du théorème de Perron

Dans ce qui suit, la valeur propre maximale d'une matrice sera notée λ_{\max} .

Théorème 4 :

$$\lambda_{\max} \geq n.$$

Preuve :

Comme w est un vecteur propre correspondant à λ_{\max} alors :

$$(A - \lambda_{\max} I) w = 0.$$

Pour une ligne i donnée ($i = 1, \dots, n$)

$$\sum_{j=1}^n a_{ij} w_j - \lambda_{\max} w_i = 0 \Leftrightarrow \lambda_{\max} = \sum_{j=1}^n a_{ij} \frac{w_j}{w_i}$$

faisant la somme sur tous les i ($i = 1, \dots, n$), on aura :

$$n\lambda_{\max} = n + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \frac{w_j}{w_i}$$

comme $a_{ij} = \frac{1}{a_{ji}}$ on obtient :

$$\lambda_{\max} = 1 + \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n \left(a_{ij} \frac{w_j}{w_i} + \frac{1}{a_{ij} \frac{w_j}{w_i}} \right)$$

$$\lambda_{\max} = n + \sum_{i=1}^n \sum_{j=i+1}^n \frac{(a_{ij} w_j - w_i)^2}{n a_{ij} w_i w_j} \quad \text{car } x + \frac{1}{x} = 2 + \frac{(x-1)^2}{x}, x \neq 0$$

on remarque bien que $\lambda_{\max} \geq n$ et on a l'égalité quand $a_{ij} = \frac{w_i}{w_j}$ (le cas d'une cohérence parfaite).

B. Indice et rapport de cohérence :

Il est important et utile de savoir à qu'elle degré de cohérence sont nos jugements dans la matrice des comparaisons par paires.

Pour cela, on définit l'indice de cohérence (Consistency Index CI) pour une matrice positive

et réciproque de taille $n * n$: $CI = \frac{\lambda_{\max} - n}{n - 1}$,

Comme $\lambda_{\max} \geq n$ alors $CI \geq 0$.

Dans le cas idéal λ_{\max} sera égal à n , donc il s'agit de mesurer l'éloignement par rapport à n sur les $n-1$ valeurs restantes.

On définit l'Indice de Cohérence Aléatoire Moyen ICAM (Mean Random Consistency Index

MRCI) comme suit : $MRCI = \frac{\overline{\lambda_{\max}} - n}{n - 1}$ avec $\overline{\lambda_{\max}}$ est la moyenne des valeurs propres des matrices de taille $n * n$ positives, réciproques, de valeurs dans l'ensemble $\{1/9, 1/8, 1/7, \dots, 1, 2, \dots, 9\}$.

Le rapport de cohérence (Consistency Ratio CR) est alors définit comme suit : $CR = \frac{CI}{MRCI}$.

Si la matrice est cohérente alors $\lambda_{\max} = n$ donc $CI = 0$ et $CR = 0$.

Dans le cas où les jugements on été faits d'une manière aléatoire, CR s'approcherait de 1, alors CR mesure l'éloignement de nos jugements par rapport aux jugements aléatoire.

Saaty suggère qu'un CR inférieur ou égal à 10 % est acceptable.

Des jugements parfaitement cohérents produisent un rapport de cohérence nul tandis que des jugements aléatoires donnent, en moyenne, un rapport de cohérence égal à 1.

Voici quelques valeurs de MRCI pour différentes valeurs de n :

n	3	4	5	6	7	8	9	10
MRCI	0.5245	0.8830	1.1085	1.2493	1.3405	1.4042	1.4511	1.4857

C. Illustration :

Appliquons la méthode du vecteur propre sur l'exemple d'allocation d'énergie précédent :

Les priorités des alternatives sont :

Développement économique : 0.6483

Amélioration d'environnement : 0.122

Sécurité Nationale : 0.2296

CI = 0.0018 ; CR = 0.0035

Les priorités des grands utilisateurs d'énergie sont :

Par rapport au **développement économique** :

Les habitants : 0.6483

Les transporteurs : 0.2296

Les usines : 0.122

CI = 0.0018 ; CR = 0.0035

Par rapport à l'**amélioration d'environnement** :

Les habitants : 0.5917

Les transporteurs : 0.3332

Les usines : 0.0751

CI = 0.0071 ; CR = 0.0136

Par rapport à la **sécurité nationale** :

Les habitants : 0.5396

Les transporteurs : 0.297

Les usines : 0.1634

CI = 0.0046 ; CR = 0.0087

	Develop Econom (0.6483)	Améliorat. D'environn (0.122)	Sécurité Nationale (0.2296)
Les habitants	0.6483	0.5917	0.5396
Les transporteurs	0.2296	0.3332	0.297
Les usines	0.122	0.0751	0.1634

Tableau récapitulatif

Les priorités globales sont :

Les habitants : 0.6165

Les transporteurs : 0.2577

Les usines : 0.1258

D. Remarques :

1. Selon Saaty, un rapport de cohérence plus grand que 10% nous indique qu'il faut :
 - a. Chercher plus d'informations.
 - b. Chercher des jugements erronés résultant d'un manque de concentration, et qui ont engendré cette grande incohérence.
 - c. Sinon, conclure que cet aspect particulier du problème contient plus que la quantité moyenne d'incohérence.
2. Il est important de noter qu'être cohérent n'est pas l'objectif du processus décisionnel, une faible incohérence est nécessaire mais n'est pas suffisante pour une bonne décision. Il est possible d'être parfaitement cohérent mais avec une décision incorrecte. Le plus important est d'être exacte et juste que d'être cohérent.
3. La valeur de 10 % de CR n'est pas une règle absolue, mais plutôt une indication.
4. Plus les éléments à comparer sont non homogènes plus le CR devient élevé.

E. Importance du choix d'échelle :

Considérons l'exemple suivant [SAT90] :

Soient les six villes suivantes : Montreal, Chicago, San Francisco, London, Caire, Tokyo.

Leurs distances par rapport à Philadelphie, ont été comparées par paires par un voyageur expérimenté (un pilote par exemple).

Les jugements verbaux sont donnés dans la matrice suivante :

Philadelphie	Caire	Tokyo	Chicago	San Francisco	London	Montreal
Caire	E	-	E(TFr,A)	Fb	Fb	TFr
Tokyo	Fb	E	A	Fb	Fb	A
Chicago	-	-	E	-	-	E(E,Fb)
San Francisco	-	-	E(Fr,TFr)	E	-	E(Fr,TFr)
London	-	-	Fr	Fb	E	E(Fr,TFr)
Montreal	-	-	-	-	-	E

E : égal - Fb : faiblement - Fr : Fortement - TFr : Très fortement - A : Absolument

E(_ , _) : entre les deux jugements entre parenthèse.

Les distances réelles par rapport à Philadelphie sont :

Caire	Tokyo	Chicago	San Francisco	London	Montreal
5729	7449	660	2732	3658	400
0.278	0.361	0.032	0.132	0.177	0.019

La dernière ligne correspond au rapport de chaque distance sur la somme de toutes les distances.

En utilisant l'échelle fondamentale de 1 à 9 on trouve les priorités suivantes :

Caire	Tokyo	Chicago	San Francisco	London	Montreal
0.263	0.397	0.033	0.116	0.164	0.027

$$\lambda_{\max} = 6.45, CI = 0.09, CR = 0.07.$$

Si on utilise une autre échelle, par exemple l'échelle 10, 20, 30, 40, 50, 60, 70, 80 et 90 alors les priorités deviennent :

Caire	Tokyo	Chicago	San Francisco	London	Montreal
0.21	0.707	0.002	0.019	0.061	0.001

$$\lambda_{\max} = 13,72, CI = 1.544, CR = 1.23 ? !$$

d'où l'importance de l'échelle.

si on utilise l'échelle suivant :

$$1, 9^{1/8}, 9^{2/8}, 9^{3/8}, 9^{4/8}, 9^{5/8}, 9^{6/8}, 9^{7/8}, 9$$

on aura :

Caire	Tokyo	Chicago	San Francisco	London	Montreal
0.248	0.342	0.044	0.151	0.175	0.039

$$\lambda_{\max} = 6.097, \text{ CI} = 0.0194, \text{ CR} = 0.015.$$

Alors CI est meilleur que celui de l'échelle fondamentale de 1 à 9 mais les priorités sont plus mauvaises que celles de ce dernier, ceci montre bien que l'objectif n'est pas d'avoir un CI minimum mais plutôt des priorités qui reflètent le maximum possible les choix du décideur.

1.5.2 La méthode des moindres carrées

Il s'agit de résoudre le programme mathématique suivant :

$$\left\{ \begin{array}{l} \text{Min} \sum_{i=1}^n \sum_{j=1}^n \left(a_{ij} - \frac{w_i}{w_j} \right)^2 \\ \sum_{i=1}^n w_i = 1 \\ w_i > 0 \quad i = 1, \dots, n \end{array} \right.$$

1.5.3 La méthode du logarithme des moindres carrées :

$$\left\{ \begin{array}{l} \text{Min} \sum_{i=1}^n \sum_{j=i+1}^n \left[\log a_{ij} - \left(\log \frac{w_i}{w_j} \right) \right]^2 \\ \prod_{i=1}^n w_i = 1 \\ w_i > 0 \quad i = 1, \dots, n \end{array} \right. \quad (1)$$

Théorème :

La solution optimale de ce programme mathématique est $w_i = \sqrt[n]{\prod_{j=1}^n a_{ij}} \quad i = 1, \dots, n$

Preuve :

Soit $r_{ij} = \log a_{ij}$, $i, j = 1, 2, \dots, n$ et $x_i = \log w_i$ $i = 1, 2, \dots, n$.

Comme $w_i = e^{x_i}$, la condition $\prod_{i=1}^n w_i = 1$ devient $\sum_{i=1}^n x_i = 0 \dots (*)$

Le programme précédent devient :

$$\begin{cases} \text{Min} \sum_{1 \leq i < j \leq n} (r_{ij} - x_i + x_j)^2 \\ \sum_{i=1}^n x_i = 0 \end{cases} \quad (2)$$

d'après (*) : $x_n = -\sum_{i=1}^{n-1} x_i$, on remplaçant x_n dans ce programme on obtient un programme de

minimisation avec n-1 variables et sans contrainte :

$$\text{Min} f(x_1, x_2, \dots, x_{n-1}) = \sum_{1 \leq i < j \leq n-1} (r_{ij} - x_i + x_j)^2 + \sum_{i=1}^{n-1} \left(r_{in} - x_i - \sum_{l=1}^{n-1} x_l \right)^2$$

dérivons f par rapport à x_k :

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{k-1} 2(r_{ik} - x_i + x_k) + \sum_{i=k+1}^{n-1} 2(-r_{ki} - x_i + x_k) + \sum_{\substack{i=1 \\ i \neq k}}^{n-1} 2 \left(-r_{in} + x_i + \sum_{l=1}^{n-1} x_l \right) - 4 \left(r_{kn} - x_k - \sum_{l=1}^{n-1} x_l \right)$$

la matrice originale A est réciproque $\left(a_{ij} = \frac{1}{a_{ji}} \right) \forall i, j = 1, \dots, n$ ce qui implique que $r_{ij} = -r_{ji}$,

et $r_{ii} = 0$, d' où $\frac{\partial f}{\partial x_k} = 2 \left[n \left(x_k + \sum_{i=1}^{n-1} x_i \right) + \sum_{i=1}^n r_{ik} - \sum_{i=1}^n r_{in} \right]$.

On observe que si on prend $x_i = \frac{\sum_{j=1}^n r_{ij}}{n}$ $i = 1, 2, \dots, n$ alors toutes les dérivés partielles

s' annulent, ce qui est dû au fait que $\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} r_{ij} = 0$,

Considérons la dérivé seconde de f pour s' assurer qu' il s' agit d' un minimum.

$$\begin{aligned} \frac{\partial^2 f}{\partial x_k^2} &= 4n \quad \text{si } k = 1, 2, \dots, n-1 \\ \frac{\partial^2 f}{\partial x_k \partial x_l} &= 2n \quad \text{si } k \neq l, k, l = 1, 2, \dots, n-1 \end{aligned}$$

soit D_N le déterminant de la matrice suivante :

$$\begin{pmatrix} 4n & 2n & \dots & 2n \\ 2n & 4n & \dots & 2n \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 2n & 2n & \dots & 4n \end{pmatrix}_{N \times N}$$

soustrayons la première ligne des autres lignes, on obtient :

$$\begin{pmatrix} 4n & 2n & 2n & \dots & 2n \\ -2n & 2n & 0 & \dots & 0 \\ -2n & 0 & 2n & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ -2n & 0 & 0 & \dots & 2n \end{pmatrix}_{N \times N}$$

on peut trouver facilement que $D_2 = 12.n^2$ et par récurrence on montre que $D_N = 2n D_{N-1} + (2.n)^N$.

D_N est positive pour tout $N \geq 2$, ce qui signifie que la matrice des secondes dérivées est

définie positives, donc $x_i = \frac{\sum_{j=1}^n r_{ij}}{n}$ $i = 1, 2, \dots, n$ est une solution optimale pour (2)

d' où $\log w_i = \frac{\sum_{j=1}^n r_{ij}}{n}$ $i = 1, 2, \dots, n$ ou encore $w_i = \sqrt[n]{\prod_{j=1}^n a_{ij}}$ $i = 1, 2, \dots, n$

Cette méthode présente l' avantage de pouvoir obtenir les priorités par une méthode très simple.

1.5.4 La méthode de Chi Deux

$$\left\{ \begin{array}{l} \text{Min} \sum_{i=1}^n \sum_{j=1}^n \frac{\left(a_{ij} - \frac{w_i}{w_j} \right)^2}{\frac{w_i}{w_j}} \\ \sum_{i=1}^n w_i = 1 \\ w_i > 0 \quad i = 1, \dots, n \end{array} \right.$$

On remarque bien qu'il y a différentes méthodes d'agrégation des jugements et qui donnent des résultats différents pour une même matrice de comparaisons par paires (la méthode des moindres carrées donnée dans 1.5.2 peut donner pour la même matrice des priorités différentes et un ordre différent [SAT01]). On trouve dans [SAT90], [SAT01] des comparaisons de ces méthodes, ainsi que des arguments en faveur de la méthode du vecteur propre.

1.6 Causes d'incohérence :

- **Manque d'informations :**

Si on dispose de peu d'informations ou on ne dispose plus d'informations sur les facteurs à comparer alors nos jugements auront une tendance d'être aléatoires et une grande incohérence en résulte. Parfois on pense qu'on a plus d'informations que ce qu'on a réellement, et par suite il est utile de détecter si un manque d'informations existe.

Dans certaines situations on procède directement dans le processus de jugements sans dépenser de l'argent et du temps pour collecter des informations additionnelles dans le but de savoir si réellement cette information a un impact sur la décision.

- **Manque de concentration :**

Le manque de concentration peut être dû à la fatigue durant le processus de jugements ou le décideur n'est pas réellement intéressé par la décision.

- **Structure inadéquate du modèle :**

Par exemple si les éléments ne sont pas suffisamment homogènes pour être dans le même niveau de la hiérarchie ou la structure hiérarchique elle-même ne convient pas au problème.

- **Le monde réel n'est pas toujours cohérent.**

1.7 Transitivité dans les jugements :

Il est utile de noter que les jugements peuvent violer la transitivité c-à-d si l'importance relative de C_1 est plus grande que celle de C_2 qui est plus grande que celle de C_3 alors l'importance relative de C_3 peut être plus grande que celle de C_1 .

Une illustration simple et intéressante est offerte par les tournois car elle montre bien l'incohérence et le manque de transitivité dans les jugements : une équipe sportive C_1 gagne une autre équipe C_2 qui a gagné une troisième équipe C_3 , mais on peut très bien avoir C_3 gagne C_1 ? d'où l'absence de transitivité, c'est un phénomène qu'on accepte dans les jugements et rien ne peut être fait.

Dans une autre expérience, 62 étudiants ont été demandé de choisir, pour un mariage, entre trois jeunes filles x , y et z

Par rapport à l'intelligence : leur ordre est $x - y - z$

Par rapport à l'apparence : leur ordre est $y - z - x$

Par rapport à la richesse : leur ordre est $z - x - y$

x est décrite comme très intelligente, d'un regard acceptable et de bon revenu.

y est intelligente, très bon regard mais pauvre

z est assez intelligente, bon regard et riche.

Toutes les jeunes filles sont acceptables et aucune n'est vraiment à la fois pauvre, moche ou stupide pour être automatiquement éliminées.

Après avoir effectué les préférences, on remarque bien qu'elles sont intransitives car :

x est préférable à y avec 39 contre 23

y est préférable à z avec 57 contre 5

z est préférable à x avec 33 contre 29

ces expériences montrent bien qu'il existe parfois une circulation dans nos préférences.

Voici quelques critiques de l'intransitivité par certains chercheurs :

[**May 1954**] a étudié l'idée que l'intransitivité entre les préférences peut être un phénomène naturel et pas une conséquence d'erreur dans les jugements, il conclut qu'il n'y a pas d'issue pour éviter de considérer que l'intransitivité est un phénomène naturel.

[**Fishburn 1991**] conclut une discussion sur l'intransitivité des préférences comme suit : "La transitivité était la pierre angulaire des notions traditionnelles d'ordre et de rationalité dans la théorie de décision.

Trois axes de recherche durant les quelques décennies passées ont eu une tendance à défier son statut. Premièrement, une variété d'expériences et d'exemples qui sont souvent basés sur des comparaisons binaires entre les alternatives suggèrent que parfois les gens violent la transitivité et ils peuvent avoir de bonnes raisons pour le faire. Deuxièmement, les résultats théoriques montrent que, dans plusieurs situations, la transitivité n'est pas essentielle pour l'existence d'une alternative préférée le maximum. Troisièmement, de nouveaux modèles qui ne nécessitent pas la transitivité ont été développés et parfois axiomatisés, comme alternatives aux méthodes traditionnelles moins flexibles."

[**Saaty**] :

"L'esprit conscient absorbe les nouvelles idées par une analyse pour comprendre comment sont similaires à des idées familières, les idées sont aussi reliées à des activités courantes ou futures, et appliquées à des situations concrètes pour tester leur compatibilité avec ce qui est déjà connu comme étant praticables, les idées peuvent être acceptées comme une partie cohérente de la compréhension existante ou non cohérente avec ce qui est déjà connu ou accepté, dans ce cas, le système de compréhension et la pratique sont étendus et ajustés pour inclure les nouvelles idées, si l'ajustement des anciennes idées pour harmoniser les nouvelles est brutal alors l'incohérence causée par ces nouvelles idées est importante, ce qui nécessite un ajustement considérable dans les anciennes et par suite les relations entre ces dernières ne restent plus intuitivement reconnaissables.

Ces changements majeurs ne peuvent pas être faits chaque heure, jour ou semaine parce qu'ils prennent beaucoup de temps pour interpréter et assimiler les relations, d'où l'incohérence venant d'une exposition à des nouvelles idées peut être dérangeante et pénible.

Notre biologie a reconnu ceci et a développé des solutions pour filtrer l'information d'une façon à faire peu d'ajustement dans ce qu'on connaît déjà quand on a une nouvelle ou une meilleure idée, absorbant ces nouvelles idées en les interprétant du point de vue avantageux des relations déjà établies".

2. Analyse de sensibilité

Soient A_1, \dots, A_M : M Alternatives et C_1, \dots, C_N : N critères.

Soit w_j le poids du critère C_j , $j = 1, \dots, N$.

et p_{ij} le poids de l'alternative i par rapport au critère j , $i = 1, \dots, M$; $j = 1, \dots, N$

on suppose que $\sum_{i=1}^M p_{ij} = 1$; $j = 1, \dots, N$ et $\sum_{j=1}^N w_j = 1$.

On a donc le poids total de l'alternative i : $P_i = \sum_{j=1}^N p_{ij} w_j$ $i = 1, \dots, M$ (1).

On suppose, sans perte de généralité, que les alternatives sont arrangées dans l'ordre suivant :

$$P_1 \geq P_2 \geq \dots \geq P_M$$

2.1 Changement des poids des critères :

Soit δ_{st}^k ; $1 \leq s < t \leq M$; $1 \leq k \leq N$ le changement minimum dans le poids w_k du critère C_k pour lequel l'ordre des alternatives A_s et A_t sera renversé.

Soit $\delta_{st}^k = \frac{\delta_{st}^k}{w_k} * 100$; $1 \leq s < t \leq M$; $1 \leq k \leq N$

Théorème :

La quantité δ_{st}^k , avec laquelle le poids w_k du critère C_k doit être modifié (après normalisation) pour que l'ordre des alternatives A_s et A_t sera renversé est donné par :

$$\begin{cases} \delta_{st}^k < \frac{P_t - P_s}{p_{tk} - p_{sk}} * \frac{100}{w_k} & \text{si } p_{tk} > p_{sk} \\ \delta_{st}^k > \frac{P_t - P_s}{p_{tk} - p_{sk}} * \frac{100}{w_k} & \text{si } p_{tk} < p_{sk} \end{cases}$$

Preuve :

Posons w_k^* le poids modifié du critère C_k : $w_k^* = w_k - \delta_{st}^k$

pour garder la somme des priorités des critères égale à 1, il est nécessaire de normaliser toutes les priorités :

$$w_i' = \frac{w_i}{w_k^* + \sum_{j \neq k} w_j} \quad i = 1, 2, \dots, N ; i \neq k \quad (2)$$

$$w_k' = \frac{w_k^*}{w_k^* + \sum_{j \neq k} w_j} \quad (3)$$

soient P_s' et P_t' les nouvelles priorités globales des alternatives A_s et A_t respectivement.

Comme on veut que l'ordre de ces deux alternatives soit renversé alors :

$$P_s' < P_t'$$

D'après (1) :

$$\sum_{j=1}^N w_j' p_{sj} = P_s' < P_t' = \sum_{j=1}^N w_j' p_{tj}$$

En utilisant les formules (2) et (3) on a :

$$\frac{w_k^* p_{sk}}{w_k^* + \sum_{j \neq k} w_j} + \frac{\sum_{j \neq k} w_j p_{sj}}{w_k^* + \sum_{j \neq k} w_j} < \frac{w_k^* p_{tk}}{w_k^* + \sum_{j \neq k} w_j} + \frac{\sum_{j \neq k} w_j p_{tj}}{w_k^* + \sum_{j \neq k} w_j}$$

$$w_k^* p_{sk} + \sum_{j \neq k} w_j p_{sj} < w_k^* p_{tk} + \sum_{j \neq k} w_j p_{tj} \quad (4)$$

d'après (1) et (4) :

$$-\delta_{st}^k p_{sk} + \underbrace{\sum_{j=1}^N w_j p_{sj}}_{P_s} < -\delta_{st}^k p_{tk} + \underbrace{\sum_{j=1}^N w_j p_{tj}}_{P_t}$$

$$\delta_{st}^k (p_{tk} - p_{sk}) < P_t - P_s$$

divisant par $(p_{tk} - p_{sk})$:

$$\delta_{st}^k < \frac{P_t - P_s}{p_{tk} - p_{sk}} \quad \text{si } p_{tk} > p_{sk}$$

$$\delta_{st}^k > \frac{P_t - P_s}{p_{tk} - p_{sk}} \quad \text{si } p_{tk} < p_{sk}$$

multipliant par $100 / w_k$, on obtient le résultat.

2.2 Changement des poids des alternatives par rapport aux critères :

Soit Δ_{st}^k ; $1 \leq t < s \leq M$; $1 \leq k \leq N$ la valeur minimum avec laquelle on modifie p_{sk} pour

renverser l'ordre des alternatives A_s et A_t . Soit encore $\Delta_{st}^k' = \frac{\Delta_{st}^k}{p_{sk}} * 100$

Théorème :

La quantité Δ_{st}^k avec laquelle p_{sk} doit être modifié pour renverser l'ordre des alternatives A_s et A_t est donnée par :

$$\Delta_{st}^k = \frac{P_s - P_t}{P_s - P_t + w_k (p_{tk} - p_{sk} + 1)} * \frac{100}{p_{sk}}$$

de plus Δ_{st}^k doit être inférieure à 100.

Preuve :

Soit p_{sk}^* la nouvelle valeur de p_{sk} : $p_{sk}^* = p_{sk} - \Delta_{st}^k$ (5)

On a, par hypothèse, avant modification : $P_t \geq P_s$.

Après avoir changé p_{sk} à p_{sk}^* , on aura : $P_t' < P_s'$.

$$p_{sk}' = \frac{p_{sk}^*}{p_{sk}^* + \sum_{j \neq s} p_{jk}} \quad (6)$$

Après normalisation, on obtient :

$$p_{ik}' = \frac{p_{ik}}{p_{sk}^* + \sum_{j \neq s} p_{jk}} \quad i = 1, 2, \dots, M \quad ; \quad i \neq s \quad (7)$$

Comme $\sum_{i=1}^M p_{ik} = 1$, et d'après (5), on peut écrire :

$$p_{sk}^* + \sum_{j \neq s} p_{jk} = \sum_{j=1}^M p_{jk} - \Delta_{st}^k = 1 - \Delta_{st}^k$$

les équations (6) et (7) deviennent :

$$p_{sk}' = \frac{p_{sk}^*}{1 - \Delta_{st}^k} = \frac{p_{sk} - \Delta_{st}^k}{1 - \Delta_{st}^k} \quad (8)$$

$$p_{ik}' = \frac{p_{ik}}{1 - \Delta_{st}^k} \quad i = 1, 2, \dots, M \quad ; \quad i \neq s \quad (9)$$

comme $P_t' < P_s'$ alors :

$$p_{tk}' w_k + \sum_{j \neq k} p_{tj} w_j = P_t' < P_s' = p_{sk}' w_k + \sum_{j \neq k} p_{sj} w_j$$

$$p_{tk}' w_k + (p_{tk} - p_{tk}) w_k + \sum_{j \neq k} p_{tj} w_j < p_{sk}' w_k + (p_{sk} - p_{sk}) w_k + \sum_{j \neq k} p_{sj} w_j$$

$$p_{tk}'w_k - p_{tk}w_k + \underbrace{\sum_{j=1}^N p_{ij}w_j}_{P_t} < p_{sk}'w_k - p_{sk}w_k + \underbrace{\sum_{j=1}^N p_{sj}w_j}_{P_s}$$

$$p_{tk}'w_k - p_{tk}w_k + P_t < p_{sk}'w_k - p_{sk}w_k + P_s \quad (10)$$

en remplaçant (8) et (9) dans (10) on trouve :

$$\frac{p_{tk}w_k}{1 - \Delta_{st}^k} - p_{tk}w_k + P_t < \frac{(p_{sk} - \Delta_{st}^k)w_k}{1 - \Delta_{st}^k} - p_{sk}w_k + P_s$$

ou encore :

$$\Delta_{st}^k < \frac{P_s - P_t}{P_s - P_t + w_k(p_{tk} - p_{sk} + 1)} \quad (11)$$

Notons que le dénominateur de la partie droite est toujours positif à cause de (10).

Il est clair que : $0 \leq p_{sk}' \leq 1$ ou $\Delta_{st}^k \leq p_{sk}$ (12)

Multipliant (11) et (12) par $100 / p_{sk}$, le résultat en découle.

3. Le feedback :

Dans quelques problèmes de décisions, l'importance des critères peut dépendre des alternatives à comparer, cette dépendance peut être traitée soit par un calcul formel du feedback soit intuitivement par le décideur. Prenons un exemple, un gestionnaire veut construire un pont, il décide de choisir parmi les propositions selon deux critères, la sécurité et l'esthétique. Il est clair que la sécurité est beaucoup plus importante que l'esthétique.

On dispose de deux alternatives pour ce nouveau pont :

Un pont A : très sécurisé et très beau.

Un pont B : doublement sécurisé que A mais moche

Selon les préférences des critères, le pont B qui est moche sera directement choisi, et dans ce cas comment les citoyens qui utiliseront ce pont au moins deux fois par jour vont juger la décision de ce gestionnaire ?

On verra dans ce qui suit différentes approches de traitement de cet exemple.

3.1 Evaluation haut - bas :

C'est l'approche classique de l'AHP, on évalue les éléments de la hiérarchie du haut en bas, tout d'abord on évalue les critères : supposons que la sécurité est jugée absolument plus importante que l'esthétique, d'où les priorités 0.9 et 0.1 pour les critères : sécurité et esthétique respectivement.

Les comparaisons par paires des alternatives par rapport aux critères sont comme suit :

Par rapport à la sécurité : le pont B est 2 fois meilleur que le pont A.

Par rapport à l'esthétique : le pont A est 6 fois meilleur que B.

	Sécurité (0.9)	Esthétique (0.1)
Pont A	0.333	0.857
Pont B	0.667	0.143

La composition de ces priorités avec les priorités des critères donne :

Le pont A : 0.386

Le pont B : 0.614

D'où le choix du pont B !

3.2 Evaluation bas - haut :

Dans ce cas, on commence par l'évaluation des alternatives par rapport aux critères avant d'évaluer l'importance des critères eux même. Donc le décideur remarque que le pont B est plus fort que A mais les deux excèdent les normes standards de sécurité, de plus A est beau tandis que B ne l'est pas. Par conséquent lors de l'évaluation de l'importance relative des critères, le décideur juge raisonnablement que l'esthétique dans ce cas est plus importante que la sécurité.

Alors jugeons l'esthétique doublement préférable que la sécurité, d'où les priorités 0.333 et 0.667 pour la sécurité et l'esthétique respectivement.

Les priorités globales deviennent :

Le pont A : 0.683

Le pont B : 0.317

Connaissant que les deux ponts sont plus que sécurisés, le décideur doit réévaluer ces jugements, et ainsi on aboutira à un choix correct : le pont A

3.3 L’AHP et le feedback :

On décrira maintenant une approche plus formelle, au lieu de demander au décideur l’importance des critères sécurité et esthétique par rapport à l’objectif global, on lui demandera de les évaluer par rapport aux alternatives (pont A et pont B)

Par rapport au pont A : la sécurité est 2 fois plus importante que l’esthétique.

Par rapport au pont B : l’esthétique devient absolument plus important que la sécurité.

	Pont A	Pont B
Sécurité	0.667	0.1
Esthétique	0.333	0.9

processus itératifs :

Supposons que tout d’abord, il n’y a pas de préférences entre les alternatives (ou ils ont la même préférence) donc les priorités des critères seront 0.383 et 0.617

maintenant, on détermine les priorités des alternatives par rapport aux critères, on aura 0.657 et 0.343

Alternatives Ou critères	Pont A ou Sécurité	Pont B ou Esthétique
2 ^{ème} Alt	0.610	0.390
3 ^{ème} Obj	0.446	0.554
3 ^{ème} Alt	0.624	0.376
4 ^{ème} Obj	0.454	0.546
4 ^{ème} Alt	0.619	0.381
5 ^{ème} Obj	0.451	0.549
5 ^{ème} Alt	0.621	0.379
6 ^{ème} Obj	0.452	0.548
6 ^{ème} Alt	0.620	0.380
7 ^{ème} Obj	0.451	0.549
7 ^{ème} Alt	0.621	0.379

Ces résultats obtenus par cette approche de feedback sont similaires à ceux de l'approche intuitive dans laquelle les préférences des critères ont été faites par rapport à l'objectif global après avoir évalué les alternatives.

3.4 La super matrice pour le feedback :

Une super matrice est une matrice dont ses éléments sont eux même des matrices de priorités en colonnes, donc au lieu de procéder itérativement, on utilise la super matrice construite comme suit :

$$\begin{array}{r}
 \begin{array}{l}
 \text{critères} \\
 \text{alternatives}
 \end{array}
 \left\{ \begin{array}{l}
 \text{sécurité} \\
 \text{esthétique} \\
 \text{pont A} \\
 \text{pont B}
 \end{array} \right.
 \begin{pmatrix}
 & \begin{array}{cc} \text{critères} & \\ \text{sécurité} & \text{esthétique} \end{array} & \begin{array}{cc} \text{alternatives} & \\ \text{pont A} & \text{pont B} \end{array} \\
 0 & 0 & 0.667 & 0.1 \\
 0 & 0 & 0.333 & 0.9 \\
 0.333 & 0.857 & 0 & 0 \\
 0.667 & 0.143 & 0 & 0
 \end{pmatrix}
 \end{array}$$

On placera les vecteurs des priorités des alternatives par rapport aux critères ainsi que les priorités des critères par rapport aux alternatives.

Les priorités limites sont obtenues en élevant la super matrice à des puissances infinies, ce qui

donne :

$$\begin{pmatrix}
 0.452 & 0.452 & 0 & 0 \\
 0.548 & 0.548 & 0 & 0 \\
 0 & 0 & 0.621 & 0.621 \\
 0 & 0 & 0.379 & 0.379
 \end{pmatrix}$$

Qui sont similaires aux résultats de l'approche précédente.

4. Analytic Network Process ANP (procédure d'analyse par réseau)

4.1 Schéma général :

La procédure d'analyse par réseau ANP est une généralisation de l'AHP, elle traite les problèmes de décisions sans faire des hypothèses d'indépendance entre les éléments de deux

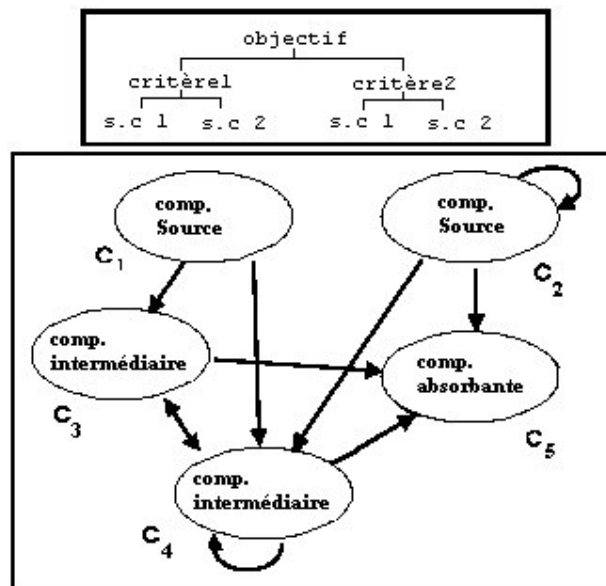
niveaux différents, ni sur les éléments d'un même niveau, en effet l'ANP utilise un réseau sans avoir besoin à spécifier des niveaux.

l'ANP construit deux parties, la première est appelée une hiérarchie de contrôle : c'est une hiérarchie de critères et de sous critères gouvernant les interactions du système.

La seconde partie est un réseau contenant des groupes d'éléments avec leurs influences, ce réseau peut varier d'un critère à un autre.

Les critères de la hiérarchie de contrôle servent à comparer les groupes ou les composantes du système tandis que les sous critères seront utilisés pour comparer les éléments des composantes.

La question générique est : Etant donné un élément ou une composante du réseau, de combien un élément d'une paire (une composante) influe l'autre élément (la composante) selon un sous critère (un critère) de la hiérarchie de contrôle ?



Exemple d'un réseau à hiérarchie de contrôle.

Supposons qu'on a un système de N groupes ou composantes où les éléments dans chaque composante ont une influence ou sont influencés par d'autres éléments d'une autre composante selon un certain critère de la hiérarchie de contrôle.

Supposons que chaque composante C_h , $h = 1, \dots, N$ contient n_h éléments notés $e_{h_1}, e_{h_2}, \dots, e_{h_{n_h}}$.

L'impact d'un ensemble d'éléments d'une composante, par rapport à un autre élément est représenté par un vecteur de priorités obtenu par des comparaisons par paires usuelles

La super matrice est de la forme suivante :

$$W = \begin{matrix} & & C_1 & & C_2 & & \dots & & C_N \\ & & e_{11} & \dots & e_{1n_1} & e_{21} & \dots & e_{2n_2} & \dots & e_{N1} & \dots & e_{Nn_N} \\ C_1 & e_{11} & \dots & W_{11} & & W_{12} & & \dots & & W_{1N} \\ & e_{1n_1} & & & & & & & & & & \\ C_2 & e_{21} & \dots & W_{21} & & W_{22} & & \dots & & W_{2N} \\ & e_{2n_2} & & & & & & & & & & \\ \dots & \dots & & \dots & & \dots & & \dots & & \dots \\ C_N & e_{N1} & \dots & W_{N1} & & W_{N2} & & \dots & & W_{NN} \\ & e_{Nn_N} & & & & & & & & & & \end{matrix}$$

avec la matrice W_{ij} est de la forme :

$$W_{ij} = \begin{pmatrix} w_{i1}^{j1} & \dots & w_{i1}^{jn_j} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ w_{i_{n_i}}^{j1} & \dots & w_{i_{n_i}}^{jn_j} \end{pmatrix}$$

avec chaque colonne est un vecteur normalisé de priorités des éléments en lignes par rapport à l'élément en colonne correspondante.

Cette matrice W n'est pas forcément stochastique (c-à-d la somme des éléments de chaque colonne n' est pas forcément égale à 1). Pour la rendre stochastique, il faut comparer les composantes eux même (et pas leurs éléments) qui sont à gauche de la super matrice W par rapport à chaque composante en haut de W selon un certain critère de la hiérarchie de contrôle. Les blocks de chaque colonne de W seront multipliés par ces dernières priorités, permettant ainsi d'avoir la somme des éléments de cette colonne de W égal à 1.

On répétant ce procédé pour toutes les colonnes, on obtiendra une super matrice stochastique. Les priorités des composantes servent à pondérer les blocks de la super matrice pour assurer qu' elle soit stochastique.

L' objectif est de savoir le comportement de W à long terme. Ceci peut être obtenu par le calcul de W^∞ (les priorités limites)

mais est-ce-que W^∞ converge ?

le comportement de W^∞ dépend des notions d'**Irréductibilité** et de **Primitivité**.

Définition : Une matrice est dite primitive si et seulement si il existe un entier $m > 0$ tel que $A^m > 0$. Sinon elle est dite imprimitive.

4.2 Comment calculer une fonction de matrice ? (Formule de Sylvester)

Il arrive parfois d'avoir besoin à calculer une certaine fonction d'une matrice A (par exemple $f(A) = A^{-1}$), dans notre cas on veut calculer $f(A) = A^k$.

Il est utile d'avoir une méthode générale pour la détermination de $f(A)$.

Considérons le polynôme : $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$.

Cette fonction peut être déterminée par la connaissance de ces valeurs en $(n + 1)$ points distincts.

La formule suivante (due à Lagrange) nous permet de calculer la valeur du polynôme en tout point, par la connaissance de ses valeurs en $n + 1$ points x_1, \dots, x_{n+1} .

$$f(x) = \sum_{i=1}^{n+1} f(x_i) \frac{(x-x_1)(x-x_2)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_{n+1})}{(x_i-x_1)(x_i-x_2)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_{n+1})}$$

$$= \sum_{i=1}^{n+1} f(x_i) \frac{\prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x-x_j)}{\prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x_i-x_j)}$$

On remarque bien que, si on remplace x par x_i dans les deux membres de cette équation, on trouvera : $f(x_i) = f(x_i)$.

Si au lieu de considérer une variable x , on considère une matrice A alors on écrira :

$$f(A) = \sum_{i=1}^{n+1} f(x_i) \frac{\prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x_j I - A)}{\prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x_j - x_i)}$$

pour la matrice, les points correspondent aux valeurs propres de A.

donc pour les valeurs propres distinctes de A : $\lambda_1, \lambda_2, \dots, \lambda_n$, on a :

$$f(A) = \sum_{i=1}^n f(\lambda_i) Z(\lambda_i) \text{ avec } Z(\lambda_i) = \frac{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j I - A)}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)}$$

si en particulier $f(A) = A^k$, et toutes les valeurs propres de A sont distinctes alors

$$A^k = \sum_{i=1}^n \lambda_i^k Z(\lambda_i) = \sum_{i=1}^n \lambda_i^k \frac{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j I - A)}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)}$$

Dans cette formule, chaque terme consiste en un produit d'une fonction d'une valeur propre de A et d'un polynôme en A de degré n-1.

Si on veut calculer, par exemple, W^k quand $k \rightarrow \infty$, il est suffisant d'avoir la limite de λ_i^k .

$$k \rightarrow \infty ; i=1, \dots, n$$

On peut remplacer λ_i par sa représentation $r \cdot e^{i\theta}$. et donc la limite de λ_i^k dépend seulement de r^k car $e^{i\theta \cdot k} = \cos \theta \cdot k + i \sin \theta \cdot k$. est borné. Et on a :

$$\lim_{k \rightarrow \infty} r^k = \begin{cases} 0 & \text{si } r < 1 \\ 1 & \text{si } r = 1 \\ \infty & \text{si } r > 1 \end{cases}$$

Quand une matrice non négative est stochastique alors sa plus grande valeur propre est réelle et égale à 1 et toutes les autres valeurs propres ont des modules qui n'excèdent pas 1 (il peut y avoir certains modules égaux à 1), donc à l'infini les seuls termes qui restent sont ceux qui correspondent à $\lambda_{\max} = 1$ ou à des valeurs complexes de modules égaux à 1.

Si la matrice W est non négative et primitive alors elle aura le même comportement.

Remarquons que dans la formule de Lagrange, les termes $(x - x_i)$ et $(x_i - x_i)$ ont été exclus une seule fois car tous les points x_i étaient distinctes.

Tandis que pour une matrice, il se peut qu'une valeur propre soit multiple (de multiplicité k par exemple) et dans ce cas il faut la supprimer k - 1 fois supplémentaires, c-à-d dériver k - 1 fois et diviser par (k - 1) ! pour chaque valeur propre multiple.

D'où la formule confluente du théorème de Sylvester :

$$f(A) = \sum_{i=1}^k T(\lambda_i) = \sum_{i=1}^n \frac{1}{(m_i - 1)!} \frac{d^{m_i-1}}{d\lambda^{m_i-1}} f(\lambda)(\lambda I - A)^{-1} \left. \frac{\prod_{i=1}^n (\lambda - \lambda_i)}{\prod_{i=m_{i+1}}^n (\lambda - \lambda_i)} \right|_{\lambda=\lambda_i}$$

avec k est le nombre de valeurs propres distinctes et m_i est la multiplicité de λ_i .

mais il faut détailler $(\lambda.I - A)^{-1}$.

On sait que $A.Adj(A) = (\det A)I$ avec $Adj(A)$ dénote la matrice adjointe de A et $\det(A)$ dénote le déterminant de A .

Posant $F(\lambda) = (\lambda I - A)^{-1}$ on obtient $(\lambda I - A).F(\lambda) = \Delta(\lambda)I$ avec $\Delta(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)$

le polynôme caractéristique de A . ou encore : $(\lambda I - A)^{-1} = \frac{F(\lambda)}{\Delta(\lambda)}$.

Donc $T(\lambda_i) = f(\lambda_i)Z_{m_i-1}(\lambda_i) + f'(\lambda_i)Z_{m_i-2}(\lambda_i) + \frac{f''(\lambda_i)}{2!}Z_{m_i-3}(\lambda_i) + \dots + \frac{f^{m_i-1}(\lambda_i)}{(m_i-1)!}Z_0(\lambda_i)$

$$Z_{m_i}(\lambda_i) = \frac{1}{m_i!} \left. \frac{d^{m_i}}{d\lambda^{m_i}} F(\lambda) \right|_{\lambda=\lambda_i}$$

avec $F^m(\lambda_i) = m!(-1)^{n-m-1}(\lambda_i I - A)^{m-1} \prod_{j \neq i} (\lambda_j I - A)$ est la dérivé de $F(\lambda)$ de l'ordre m et

$$\Delta_{m_i}(\lambda) = \prod_{j \neq i} (\lambda - \lambda_j)$$

Revenons maintenant à $\lim_{k \rightarrow \infty} W^k$;

On distingue 6 cas :

Si W est irréductible stochastique (par conséquent $\lambda_{\max} = 1$ est une valeur propre simple de W) alors :

	Pas d'autres valeurs propres complexes de module égale à 1 (primitive)	Il existe d'autres valeurs propres complexes de module égale à 1 (imprimitive de périodicité C)
	Cas A A l'infini toutes les colonnes de W sont identiques et correspondent au vecteur propre principale correspondant à λ_{\max}	Cas A' $W^\infty = \frac{1}{c} (I - W^c)(I - W)^{-1} (W^c)^\infty$

Si la matrice est réductible stochastique, alors :

$$W^\infty =$$

	Pas d'autres valeurs propres complexes de modules égales à 1	Il existe d'autres valeurs propres complexes de modules égales à 1
$\lambda = 1$ v.p. simple	Cas B $\frac{(I-W)^{-1}\Delta(1)}{\Delta'(1)} = \frac{adj(I-W)}{\Delta'(1)} \text{ normalisé}$	Cas B' $\frac{1}{c}(I-W^c)(I-W)^{-1}(W^c)^\infty$
$\lambda = 1$ v.p. multiple	Cas C $n_1 \sum_{k=0}^{n_1} (-1)^k \frac{n_1!}{(n_1-k)!} \frac{\Delta^{(n_1-k)}(\lambda)}{\Delta^{(n_1)}(\lambda)} (\lambda I - W)^{-k-1} \Big _{\lambda=1}$	Cas C' $\frac{1}{c}(I-W^c)(I-W)^{-1}(W^c)^\infty$

Les six cas précédents peuvent être groupés en 4 cas :

Posons $\lambda_{\max} = \lambda_1$.

	Matrice propre $ \lambda_i < 1 ; \forall i > 1$ une matrice primitive stochastique est dite propre	Matrice Impropre La matrice possède des valeurs propres complexes de modules égales à 1 qui sont des racines de 1
$\lambda = 1$ v.p. simple	Cas 1 $\frac{(I-W)^{-1}\Delta(1)}{\Delta'(1)} = \frac{adj(I-W)}{\Delta'(1)} \text{ normalisé}$	Cas 2 $\frac{1}{c}(I-W^c)(I-W)^{-1}(W^c)^\infty$
$\lambda = 1$ v.p. multiple	Cas 3 $n_1 \sum_{k=0}^{n_1} (-1)^k \frac{n_1!}{(n_1-k)!} \frac{\Delta^{(n_1-k)}(\lambda)}{\Delta^{(n_1)}(\lambda)} (\lambda I - W)^{-k-1} \Big _{\lambda=1}$	Cas 4 $\frac{1}{c}(I-W^c)(I-W)^{-1}(W^c)^\infty$

Dans le cas (1), appliquons la formule de Sylvester pour W^k et $\lambda_{\max} = \alpha$ une valeur propre simple et W primitive, donc $|\lambda_i| < \alpha$ pour les autres valeurs propres.

D'où $W^k = \frac{(I\alpha - W)^{-1}}{\Delta'(\alpha)} \alpha^k$ plus d'autres termes qui dépendent de λ_i ($|\lambda_i| < \alpha$).

Comme W est stochastique on a $\alpha = 1$ et donc les autres termes s'annulent à l'infini et il ne

reste que : $W^\infty = \frac{(I-W)^{-1}\Delta(1)}{\Delta'(1)} = \frac{adj(I-W)}{\Delta'(1)}$.

Dans (2) et (4) :

La matrice est imprimitive donc cyclique de périodicité C.

Prenons l'exemple suivant :

$$\begin{aligned}
 W &= \begin{pmatrix} 0 & W_{12} & 0 \\ 0 & 0 & W_{23} \\ W_{31} & 0 & 0 \end{pmatrix}; W^2 = \begin{pmatrix} 0 & 0 & W_{12}W_{32} \\ W_{23}W_{31} & 0 & 0 \\ 0 & W_{31}W_{12} & 0 \end{pmatrix} \\
 W^3 &= \begin{pmatrix} W_{12}W_{23}W_{31} & 0 & 0 \\ 0 & W_{23}W_{31}W_{12} & 0 \\ 0 & 0 & W_{31}W_{12}W_{23} \end{pmatrix} \\
 W^{3k} &= \begin{pmatrix} (W_{12}W_{23}W_{31})^k & 0 & 0 \\ 0 & (W_{23}W_{31}W_{12})^k & 0 \\ 0 & 0 & (W_{31}W_{12}W_{23})^k \end{pmatrix} \\
 W^{3k+1} &= \begin{pmatrix} 0 & (W_{12}W_{23}W_{12})^k W_{12} & 0 \\ 0 & 0 & (W_{23}W_{31}W_{12})^k W_{23} \\ (W_{31}W_{12}W_{23})^k W_{31} & 0 & 0 \end{pmatrix} \\
 W^{3k+2} &= \begin{pmatrix} 0 & 0 & (W_{12}W_{23}W_{31})^k W_{12}W_{23} \\ (W_{23}W_{31}W_{12})^k W_{23}W_{31} & 0 & 0 \\ 0 & (W_{31}W_{12}W_{23})^k W_{31}W_{12} & 0 \end{pmatrix}
 \end{aligned}$$

on remarque que la matrice n'a pas de limite unique mais peut avoir trois formes différentes à l'infini, donc on prend la moyenne de toutes les formes possibles.

En effet, supposons que la matrice W est de périodicité C ($C \geq 2$), alors :

$$\begin{aligned}
 W^\infty &= \frac{1}{C} \left[(W^C)^\infty + (W^{C+1})^\infty + \dots + (W^{2C-1})^\infty \right] \\
 &= \frac{1}{C} [I + W + \dots + W^{C-1}] (W^C)^\infty \\
 &= \frac{1}{C} (I - W^C)(I - W)^{-1} (W^C)^\infty.
 \end{aligned}$$

Dans le cas (3) : si la matrice est réductible alors il faut chercher si $\lambda_{\max} = 1$ est une racine simple de l'équation caractéristique ou pas.

Si 1 est une racine simple alors, on applique directement le cas 1 sinon , il faut appliquer la formule de Sylvester en considérons $\lambda_{\max} = 1$ comme étant une valeur propre multiple de multiplicité n_1 .

$$W^\infty = n_1 \left. \frac{d^{(n_1-1)}}{d\lambda^{(n_1-1)}} (\lambda I - W)^{-1} \Delta(\lambda) / \frac{d^{n_1}}{d\lambda^{n_1}} \Delta(\lambda) \right|_{\lambda=1}$$

$$= n_1 \sum_{k=0}^{n_1} (-1)^k \frac{n_1!}{(n_1 - k)!} \frac{\Delta^{(n_1-k)}(\lambda)}{\Delta^{(n_1)}(\lambda)} (\lambda I - W)^{-k-1} \Big|_{\lambda=1}$$

Une règle simple pour chercher la matrice limite pour une matrice carrée W d'ordre n est de calculer $(I + W)^{n-1}$.

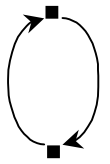
Si $(I + W)^{n-1} > 0$ alors W est irréductible et par suite $\lambda_{\max} = 1$ est une valeur propre simple donc appliquer l'un des cas (1) ou (2) selon que la matrice est primitive ou pas.

Si le groupe des alternatives n'a pas d'impact sur les critères (pas de feedback des alternatives sur les critères) alors il est préférable de ne pas l'ajouter dans la super matrice, car si W est imprimitive, le calcul de la moyenne ne fait pas intervenir le groupe des alternatives.

Les priorités des critères et éventuellement les sous-critères sont ensuite utilisées pour déterminer les priorités des alternatives.

Dans le cas où $\lambda_{\max} = 1$ est une valeur propre multiple, on peut utiliser le fait que l'augmentation des connections entre les groupes et la création des boucles pour certains groupes (avec une matrice identité) peuvent rendre la matrice primitive et donc simplifie les calculs.

Exemple :

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$


$$A^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, A^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

ou encore $A^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $A^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

on remarque bien que c'est une matrice imprimitive de périodicité 2.

mais si on ajoute une boucle :

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$


$$A^2 = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}, A^3 = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}, \dots, A^k = \begin{pmatrix} k & 1 \\ 1 & 0 \end{pmatrix}$$

d'où la matrice est primitive et A^k tend à une limite unique.

Saaty propose, pour simplifier les calculs de créer des connections artificielles (entre des groupes différents) ou un feedback dans un groupe donné mais à condition de donner des priorités faibles par rapport aux connections réelles du réseau lors de l'établissement des comparaisons par paires.

4.3 Etapes de la méthode d'analyse par réseau (ANP) :

L'ANP construit deux parties, la hiérarchie de contrôle qui contient les critères de contrôle et les sous critères qui gouvernent les interactions du système, et le réseau qui contient les groupes ou composantes et leurs éléments. Ces grandes étapes sont :

1. Déterminer la hiérarchie de contrôle, ses critères et sous critères pour comparer les éléments et les groupes du réseau.
2. Comme le réseau peut varier d'un critère à un autre, alors pour chaque critère ou sous critère, déterminer les groupes du réseau et leurs éléments.
3. Pour bien organiser le développement du réseau, dénombrer et arranger les groupes et leurs éléments d'une manière convenable, les éléments de chaque groupe doivent être homogènes. Il est conseillé d'utiliser les mêmes noms pour les mêmes objets par rapport à tous les critères.
4. Choisir l'approche à utiliser dans l'analyse de chaque groupe ou élément : Influencé par un autre groupe ou élément ou influant un autre groupe ou élément par rapport à un critère de la hiérarchie de contrôle. Ce choix doit rester le même pour tous les critères et les sous critères.
5. Pour chaque critère, réaliser les comparaisons par paires des groupes du réseau, les priorités obtenues seront utilisées pour pondérer les blocks de colonnes de la super matrice correspondant à ces groupes pour la rendre stochastique. Mettre zéro s'il n'y a pas d'influence.
6. Comparer par paires les éléments dans chaque groupe, Si la structure est une hiérarchie alors l'objectif global est le critère de la hiérarchie de contrôle, dans ce cas les comparaisons par paires seront réalisées entre les éléments du même niveau par rapport à

- un élément du niveau adjacent supérieur. Si la structure est un réseau, alors comparer les éléments d'un groupe par rapport à chaque élément d'un groupe relié à lui.
7. Construire la super matrice en plaçant les groupes en ordre et les éléments de chaque groupe verticalement à gauche et horizontalement en haut. Mettez dans la bonne position les priorités locales (comme étant une partie d'une colonne de la super matrice) obtenues à partir des comparaisons par paires.
 8. Rendre la super matrice stochastique en utilisant les priorités obtenues dans 5.
 9. Calculer les priorités limites selon que la matrice est irréductible ou réductible, primitive ou imprimitive.
 10. Multiplier les priorités des alternatives par les priorités du critère de contrôle.
 11. Synthétiser les priorités globales des alternatives par rapport à tous les critères de contrôle.

Il existe des problèmes où les critères de contrôle sont : les bénéfiques, les opportunités, les coûts et les risques. Dans ce cas, les priorités globales d'une alternative est le produit des priorités locales par rapport aux bénéfiques et les opportunités divisé par le produit des priorités locales par rapport aux coûts et aux risques.

Conclusion :

L'AHP est utile pour l'analyse et l'aide à la décision dans des problèmes divers de caractères sociaux, politiques ou économiques ..etc., elle traite un problème d'une manière simple car elle nécessite que les éléments de chaque niveau soient homogènes, de taille de plus en plus réduite en descendant du niveau le plus haut vers le niveau le plus bas de la hiérarchie.

L'AHP peut prendre en considération le feedback des éléments des niveaux les plus bas vers les niveaux les plus hauts, permettant ainsi d'élargir le champ d'application de la méthode et un traitement plus élaboré des problèmes déjà traités.

Les plus grandes difficultés de l'AHP résident dans l'établissement des jugements. Si le problème à traiter est complexe et nécessite une analyse attentive, alors l'établissement des jugements va demander beaucoup de temps et d'effort. Si le réalisateur des jugements sent fatigué, il vaut mieux se reposer avant de continuer l'évaluation.

De plus, il est utile de refaire certains jugements (après un certain temps) pour s'assurer que la personne ayant établi les jugements n'a pas considérablement changé ses jugements.

Il est clair que l'utilisation de l'AHP comme toute autre méthode nécessite une connaissance, au moins partielle, du problème considéré car cette connaissance joue un rôle très déterminant dans la construction de la hiérarchie pour l'AHP et la construction du réseau pour l'ANP, ainsi que dans la réalisation des comparaisons par paires. De plus chaque élément d'un certain niveau doit être comparable par rapport à certains éléments du niveau Adjacent supérieur, ceci dit, il n'est pas nécessaire que la hiérarchie soit complète c-à-d chaque élément d'un niveau joue le rôle d'un critère pour tous les éléments du niveau adjacent inférieur, donc la hiérarchie peut être décomposée en sous hiérarchies disjointes, ne partageant en commun qu'un seul élément d'un niveau haut, par exemple, les activités de deux divisions séparées d'une organisation donnée peuvent être structurées séparément.

L'AHP peut être utilisée dans le cas où plusieurs personnes devraient participer dans les jugements, et dans ce cas les jugements des différents juges doivent être agrégés selon des priorités des juges eux même, ces priorités peuvent être dérivées par des comparaisons par paires des juges selon certains critères qu'on définit.

Chapitre III

L'Approche Bayésienne de l'Aide à la Décision

Introduction

Un des problèmes fondamentaux adressé dans l'étude des statistiques est l'inférence. Un ensemble de données est observé et l'on souhaite étudier une ou plusieurs caractéristiques du système donnant lieu à ces observations.

Cette dernière expression a un sens large parce que les circonstances dans lesquelles l'inférence est faite sont très variées.

Par exemple un sondage d'opinion produit des données relatives à certaines personnes d'un échantillon qui répondent "oui" ou "non" sur une question donnée. Le système donnant ces observations est constitué d'une population d'individus, d'un mécanisme pour la sélection d'un échantillon et un mécanisme pour donner la réponse "oui" ou "non" pour chaque individu de cet échantillon. Typiquement, ce dernier consiste en une question directe à laquelle l'individu répondra "oui" s'il possède une certaine caractéristique.

La population est un ensemble d'individus bien définis, mais l'élément inconnu du système est la proportion θ , des individus dans la population qui possèdent cette caractéristique.

L'objectif de l'inférence est alors d'utiliser les observations de l'échantillon pour étudier cette inconnue θ .

Le problème d'inférence est un sujet qui a eu une importance considérable depuis le XVIII^{ème} siècle, début de la théorie des probabilités. Dès lors, plusieurs travaux ont été engagés donnant lieu à plusieurs noms. Cette théorie est appelée l'inférence fréquentiste ou classique.

Depuis environ 1960, on s'intéresse à l'inférence Bayésienne développée par la suite et devenant une approche concurrente à l'approche fréquentiste.

1. Théorème de Bayes :

Considérons deux événements A et B

D'après l'égalité :

$$P(A, B) = P(A) \cdot P(B / A) = P(B) \cdot P(A / B)$$

On aura :

$$P(B / A) = \frac{P(B) \cdot P(A / B)}{P(A)} \dots\dots\dots(1) \text{ Théorème de Bayes}$$

On peut interpréter le théorème de Bayes comme suit :

On s'intéresse à un événement B, et on commence avec une probabilité d'occurrence à priori P(B), puis on observe l'occurrence d'un autre événement A.

La probabilité d'apparition de B quand A est observé est appelée la probabilité à posteriori de B, autrement dit le théorème de Bayes peut être vu comme étant une mise à jour de la probabilité à priori de B, la mise à jour consiste en la multiplication par le rapport $P(A / B) / P(A)$, il montre aussi comment la probabilité varie en fonction de nouvelles informations.

Remarque :

L'occurrence de A va augmenter la probabilité de B si $P(A / B) > P(A)$

On sait que :

$$P(A) = P(A / B).P(B) + P(A / B^c).P(B^c)$$

avec B^c dénote l' événement complémentaire de B.

Ou encore

$$P(A / B) - P(A) = \{P(A / B) - P(A / B^c)\}.P(B^c)$$

donc $P(A / B) > P(A)$ si et seulement si $P(A / B) > P(A / B^c)$

Dans le cas général, soient B_1, B_2, \dots, B_n un ensemble d'événements exclusifs et exhaustifs [ANT94] alors :

$$P(B_r / A) = \frac{P(B_r)P(A / B_r)}{P(A)} = \frac{P(B_r)P(A / B_r)}{\sum_{i=1}^n P(A / B_i).P(B_i)} \dots\dots\dots(2)$$

l'expression (2) est une généralisation de (1).

On peut voir les $B_i, i = 1, \dots, n$ comme un ensemble d'hypothèses, une et une seule est vraie. L'observation de A changera la probabilité à priori de B_r en sa probabilité à posteriori $P(B_r / A)$ Le dénominateur de (2) est une pondération des probabilités $P(A / B_i)$ par les poids $P(B_i)$.

$$\left(\sum_{i=1}^n P(B_i) = 1 \right).$$

L'hypothèse pour laquelle la probabilité est augmentée le maximum par A est celle qui a $P(A / B_i)$ le maximum.

2. La vraisemblance :

Les probabilités $P(A / B_i)$ sont appelées vraisemblances, $P(A / B_i)$ est la vraisemblance donnée à B_i par A .

Pour mieux comprendre cette notion de vraisemblance, prenons un exemple :

Je vois à travers ma fenêtre un objet long avec des branches couvertes par des feuilles vertes, pourquoi je pense que c'est un arbre ? parce que généralement les arbres ressemblent à ceci et je ne pense pas que c'est un poteau ou autre chose.

Formellement, soit A l'événement : "je vois un objet long avec des branches couvertes par des feuilles vertes "

B_1 : "c'est un arbre "

B_2 : "c'est un poteau "

B_3 : "c'est une autre chose "

L'expression "les arbres généralement ressemblent à ceci" implique que $P(A / B_1)$ est très proche de 1.

Tandis que " les poteaux rarement ressemblent à ceci" implique que $P(A / B_2)$ est très proche de 0

Donc je crois que cet objet est un arbre ou il a une très grande probabilité à posteriori d'être un arbre.

Supposons maintenant que cet arbre est suffisamment loin pour le confondre avec un autre objet, une armoire par exemple. Dans ce cas cette hypothèse aura la même vraisemblance que celle d'un arbre, mais je ne vais pas croire que c'est une armoire car sa probabilité à priori est très faible.

On remarque bien que le théorème de Bayes est dans un accord complet avec le raisonnement naturel.

La probabilité à posteriori est le produit de la vraisemblance et de la probabilité à priori, autrement dit le théorème de Bayes combine deux sources d'informations : l'information à priori représentée par la probabilité à priori et la nouvelle information exprimée par la vraisemblance pour obtenir une information totale.

3. La distribution à posteriori :

Tandis qu'il y a des situations pour lesquelles les caractéristiques inconnues sont exprimées en termes d'ensemble d'hypothèses discrètes, les situations les plus fréquentes sont caractérisées par des inconnues continues appelées paramètres.

Dans l'exemple précédant, le paramètre θ est compris entre 0 et 1, et on peut utiliser plusieurs types d'inférences, par exemple si on veut donner à θ une certaine valeur alors on peut faire une estimation ponctuelle, ou encore déterminer un intervalle contenant cette inconnue à l'aide d'une estimation par intervalle. L'autre type est le test d'hypothèses permettant d'accepter ou de rejeter une certaine hypothèse à propos de θ , par exemple $\theta > 0.5$ est vraie

Le théorème de Bayes se généralise très simplement au cas continu comme suit :

$$f(\theta / x) = \frac{f(x / \theta) f(\theta)}{\int_{\Theta} f(x / \theta) f(\theta) d\theta} \dots\dots\dots(3)$$

si les données sont discrètes et l'inconnue est continue alors $f(x)$ et $f(x / \theta)$ sont des fonctions de densités discrètes, l'intégrale dans (3) sera remplacée par une somme.

La relation (3) s'interprète comme précédemment, c'est la combinaison de deux sources d'informations : l'information à priori représentée par la probabilité à priori, et la nouvelle information exprimée par la vraisemblance.

La densité à posteriori pour certaines valeurs va être faible si leur probabilité à priori ou leur vraisemblance est faible.

Par contre, les valeurs appréciables de la densité à posteriori seront celles pour lesquelles ni l'à priori ni la vraisemblance sont faibles, surtout s'ils sont bien supportés par ces deux sources d'informations.

Exemple :

Soit X la variable aléatoire représentant le nombre de succès dans une série de n jets avec la probabilité θ de succès dans chacun.

X suit la loi binomiale (discrète)

$$f(x / \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} ; x = 0, 1, \dots, n$$

c'est la vraisemblance de θ

supposons que θ possède une loi Beta comme loi à priori

$$f(\theta) = \frac{1}{B(p, q)} \theta^{p-1} (1-\theta)^{q-1}$$

avec $0 \leq \theta \leq 1$; $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ et $\Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx \quad a > 0$ (Fonction Γ)

Appliquons le théorème de Bayes :

$$\begin{aligned} f(x) &= \int_{\theta} f(\theta) f(x/\theta) d\theta \\ &= \binom{n}{x} \frac{1}{B(p, q)} \int_0^1 \theta^{p+x-1} (1-\theta)^{q+n-x-1} d\theta \\ &= \binom{n}{x} \frac{B(p+x, q+n-x)}{B(p, q)} \end{aligned}$$

d'après la formule (3) on a :

$$f(\theta/x) = \frac{1}{B(p+x, q+n-x)} \theta^{p+x-1} (1-\theta)^{q+n-x-1}$$

on remarque bien que la distribution à posteriori est encore une distribution Beta.

La vraisemblance atteint son maximum à :

$$\frac{\partial \log f(x/\theta)}{\partial \theta} = 0 \Leftrightarrow x - n\theta = 0 \Leftrightarrow \theta = \frac{x}{n}$$

La distribution à priori atteint son maximum à :

$$\frac{\partial \log f(\theta)}{\partial \theta} = 0 \Leftrightarrow \theta = \frac{p-1}{p+q-2} \quad (p, q > 1)$$

La distribution à posteriori atteint son maximum à :

$$\frac{\partial \log f(\theta/x)}{\partial \theta} = 0 \Leftrightarrow \theta = \frac{x+p-1}{n+p+q-2} = a \left(\frac{x}{n} \right) + (1-a) \left(\frac{p-1}{p+q-2} \right) \dots\dots\dots(4)$$

avec $a = \frac{n}{n+p+q-2}$.

L'expression (4) réalise un compromis entre les maxima des deux sources d'informations.

4. L'inférence Bayésienne :

Après avoir obtenu la densité à posteriori $f(\theta / x)$, il s'agit maintenant de tirer des inférences sur le paramètre inconnu.

La question qui se pose est : après l'observation de x , quelle nouvelle information possédent-nous à propos de θ ?

La réponse la plus simple est de proposer la distribution à posteriori tout entière, car elle contient toute l'information dont on dispose sur θ , mais malheureusement, dans une forme très compacte et difficile à comprendre, Il est utile de proposer encore son graphe et/ou d'autres indicateurs tels que la moyenne, la variance, le mode, la médiane ..etc.

Par exemple, si la distribution à posteriori est une loi normale $N(m, v)$ alors son graphe montre une courbe unimodale, symétrique, centrée en m , donc la valeur la plus probable pour θ est m , car la densité est maximum en m , de plus θ a la même probabilité d'être plus grande ou plus petite que m , donc m est le mode, la médiane et la moyenne à posteriori.

La mesure de dispersion est un indicateur de force et de certitude de notre information, si la dispersion est grande, alors les autres valeurs proches de θ ont une grande probabilité et donc notre connaissance est faible, et inversement.

5. La méthode Bayésienne :

La méthode Bayésienne se résume aux étapes principales suivantes :

1. **Vraisemblance** : établir la vraisemblance $f(x / \theta)$ décrivant le processus générant l'observation x par rapport au paramètre inconnu θ fixé.
2. **La loi à priori** : établir une loi de probabilité de θ exprimant notre information à priori.
3. **La loi à posteriori** : appliquer le théorème de Bayes pour dériver la densité à posteriori $f(\theta / x)$ qui exprimera notre connaissance à propos du paramètre, après l'observation de x .
4. **L'inférence** : dériver des inférences sur θ à partir de la densité à posteriori, cela consiste généralement à tirer des informations à partir de cette densité comme la moyenne, le mode, la médiane ..etc. tracer la courbe de $f(\theta / x)$ ou encore faire des estimations ponctuelles, estimations par intervalles, tests d'hypothèses ..etc.

6. La distribution à priori :

Il se peut que la question la plus importante concerne la distribution à priori, et comment peut elle être obtenue ?

Pour répondre à cette question revenant à la notion fondamentale d'une probabilité.

La théorie d'inférence classique est basée sur l'interprétation fréquentiste de la probabilité qui définit une probabilité d'un événement comme étant la fréquence limite d'occurrence de cet événement dans une séquence infinie d'essais.

Les règles classiques d'inférence sont jugées à travers leur comportement à long terme. Plusieurs défenseurs de l'inférence classique sont peu disposés à n'accepter aucune notion de probabilité autre que la probabilité fréquentiste, à moins que la probabilité à priori ne puisse être formulée comme une probabilité fréquentiste, une autre notion ne peut exister. De plus ils rejettent toute praticabilité de l'approche Bayésienne dans l'inférence. L'inférence classique est souvent appelée inférence fréquentiste à cause de sa forte association à la probabilité fréquentiste.

La probabilité à priori représente la connaissance à priori du décideur à propos de θ avant l'observation des données.

Etant donné un phénomène aléatoire suivant une loi binomiale de paramètre inconnu θ . Dans les statistiques classiques x / n est un estimateur de θ , applicable dans tous les problèmes dans lesquels X possède une distribution binomiale. Tandis que dans une approche Bayésienne, chaque problème est unique puisque dans chacun le décideur possède une connaissance particulière en ce paramètre. et donc même si la vraisemblance est la même dans tous ces problèmes, les lois à priori se diffèrent, ce qui conduit à une analyse Bayésienne différente dans chaque cas.

7. Le statut des vraisemblances et la distribution à posteriori :

En inférence classique $f(x / \theta)$ représente le modèle statistique du processus générant les données. Un statisticien classique regarde souvent le processus comme étant un échantillonnage aléatoire à partir d'une population de données et donc il interprète $f(x/\theta)$ comme une probabilité fréquentiste.

[Good 1959] indique que dans la plupart des cas de telles probabilités auraient été mieux appelées tautologie, le modèle suppose que les données ont cette distribution sachant le paramètre inconnu. Dans la plupart des cas, il n'est pas possible de vérifier que c'est la vraie distribution de probabilité fréquentiste sans faire un nombre infini d'observations, donc son statut réel est comme étant une conséquence tautologique du modèle et toutes les inférences sont conditionnées par la justesse de ce modèle.[ANT94]

D'un point de vue subjectif $f(\theta)$, $f(x / \theta)$ et $f(\theta / x)$ sont toutes des distributions de probabilités subjectives. Toute distribution de probabilité subjective peut être obtenue par l'évaluation du degré de croyance en la variable aléatoire en question pour prendre certaines valeurs. Ce processus est utilisé pour déterminer la distribution à priori, mais il peut être aussi utilisé pour déterminer la distribution à posteriori, c-à-d $f(\theta / x)$ peut être directement évaluée en terme du degré de croyance en θ pour prendre certaines valeurs basées sur toute l'information disponible, qui contient les valeurs observées x des données et l'information à priori.

Ceci paraît contraire à la méthode Bayésienne qui dérive la distribution à posteriori à partir du théorème de Bayes. Comment les défenseurs de la probabilité subjective concilient - ils ce conflit ?

Une probabilité subjective doit être le résultat d'une évaluation très soignée de toute l'information, pour cela, l'individu doit tenir compte de toutes les implications logiques de la théorie des probabilités; par exemple : si une probabilité subjective $P(A) = p$ d'un événement A est établie, alors la probabilité de son complémentaire est impliquée et aurait pu être différente d'une évaluation soignée de l'événement complémentaire.

Supposons maintenant qu'on a deux événements exclusifs A et B , donc $P(A \text{ ou } B) = P(A) + P(B)$, on attribuant des probabilités, il est souvent plus difficile d'établir $P(A \text{ ou } B)$ que d'établir $P(A)$ et $P(B)$, alors il est naturel d'utiliser la formule pour déterminer $P(A \text{ ou } B)$ autrement dit : trouver subjectivement $P(A)$ et $P(B)$ et puis $P(A \text{ ou } B)$ sera leur somme.

Des méthodes similaires à celle ci se trouvent dans la théorie de la probabilité subjective, et la méthode Bayésienne a pour les subjectivistes exactement ce statut.

Généralement, il est difficile de mesurer exactement la probabilité à posteriori par l'évaluation à la fois des informations à priori et des données. Le théorème de Bayes permet de décomposer cette évaluation en tâches séparées et plus simples par l'évaluation des probabilités à priori et des vraisemblances séparément, puis les combiner par le théorème de

Bayes et c'est dans ce contexte que la méthode Bayésienne est acceptée pour assigner des probabilités à posteriori subjectives.

8. Qu'est ce qu'une bonne inférence ?

La théorie d'inférence classique est très concernée par la construction de bonnes règles d'inférence. par exemple, comme toute fonction $T(x)$ des données peut servir comme estimateur classique du paramètre θ , il est nécessaire d'identifier des critères pour comparer ces estimateurs et dire qu'un estimateur est meilleur qu'un autre. L'objectif dans l'approche Bayésienne est différent car on cherche à extraire des informations concernant θ à partir de sa distribution à posteriori, et de la présenter sous forme utile via l'utilisation effective de ses caractéristiques.

Une bonne inférence contribue effectivement à apprécier des informations sur θ qui sont favorisées par la distribution à posteriori.

La théorie classique d'estimation considère un estimateur comme bon s'il est sans biais et possède une faible variance ou généralement la moyenne du carré d'erreur est faible où l'erreur est la différence entre θ et son estimation. Cette moyenne est calculée sur la distribution de l'estimateur.

Dans l'inférence Bayésienne et comme θ est une variable aléatoire alors il est plus approprié de calculer la moyenne du carré d'erreur par rapport à la distribution à posteriori du paramètre θ .

$$\begin{aligned} E((t-\theta)^2 / x) &= E(t^2/x) - E(2t\theta / x) + E(\theta^2 / x) \\ &= t^2 - 2.t.E(\theta / x) + E(\theta^2 / x) \\ &= \{t - E(\theta / x)\}^2 + \text{var}(\theta / x) \end{aligned}$$

donc $t = E(\theta / x)$ la moyenne à posteriori est l'estimateur qui minimise la moyenne du carré d'erreur, d'où la moyenne à posteriori peut être vue comme un estimateur de θ qui minimise la moyenne du carré d'erreur, ceci est différent de son rôle plus naturel comme une caractéristique de la distribution à posteriori.

9. Implémentation :

Il y a différentes difficultés dans l'implémentation pratique de la méthode Bayésienne. Nous avons discuté l'établissement de la distribution à priori, d'autres difficultés surviennent lors du calcul des différentes quantités requises.

Premièrement, pour appliquer le théorème de Bayes, il faut calculer l'intégrale du dénominateur.

Deuxièmement, le processus d'inférence peut nécessiter le calcul d'autres intégrales ou autres opérations dans la distribution à posteriori.

Ces calculs peuvent être difficiles à effectuer dans la pratique.

Exemple :

Soit $X \sim N(\theta, 1)$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$$

supposons que θ suit une loi de Student à d degré de liberté.

$$f(\theta) = \frac{1}{\sqrt{d} B\left(\frac{d}{2}, \frac{1}{2}\right)} \left(1 + \frac{\theta^2}{d}\right)^{-\frac{d+1}{2}}.$$

Si nous essayons d'appliquer le théorème de Bayes, le dénominateur est l'intégrale du produit de ces deux expressions.

Cette intégrale ne peut être calculée directement, donc on ne peut pas déterminer explicitement la distribution à posteriori ainsi que d'autres caractéristiques comme la moyenne à posteriori ..etc.

En cas général, si une vraisemblance arbitraire est combinée avec une loi à priori arbitraire alors le produit peut facilement être mathématiquement intraitable, il ne sera pas possible d'intégrer par rapport à θ pour établir la distribution à posteriori. Une solution à ce problème est de faire appel à une intégration numérique pour obtenir une approximation de la valeur de cette intégrale. Et avec des calculateurs suffisamment puissants, on peut obtenir de bonnes approximations, mais cette approche a ses limites.

10. Comparaison avec l'inférence classique :

Une question naturelle qui se pose est : "Pourquoi doit on utiliser l'inférence Bayésienne comme opposante à l'inférence classique ?"

Il y a plusieurs réponses, quelques arguments en faveur de l'approche Bayésienne sont : l'approche Bayésienne est fondamentalement solide est fondée, très flexible, elle produit des inférences claires et directes et elle utilise toutes les informations disponibles, contrairement à l'approche classique qu'ignore l'information à priori.

Mais les arguments ne sont pas d'un seul côté, car l'approche Bayésienne présente plus de difficultés dans son implémentation que l'approche classique.

Une critique importante concerne le sens d'inférence :

La réponse Bayésienne à la question : " Est-ce-que θ est inférieur ou égal à 0.5 ?" (par exemple) est de calculer $P(\theta < 0.5 / x)$ qui donnera la probabilité que ceci est vraie.

Contrairement, dans un test d'hypothèse classique, pour rejeter cette hypothèse ($\theta < 0.5$) à un niveau de confiance de 95 % (par exemple) il faut que : si $\theta < 0.5$ alors {la probabilité que x est dans un ensemble critique C } est inférieur ou égale à 0.5.

C'est une probabilité de x et non plus de θ .

Toutes les inférences classiques sont de ce type, c-à-d des probabilités sur x sachant θ , mais interprétées comme des probabilités de θ .

La même remarque s'applique sur les intervalles de confiance, avoir un niveau de confiance de 90 % (par exemple) que θ est dans un intervalle donné est une probabilité sur x , contrairement à l'approche Bayésienne qui définit la probabilité sur le paramètre inconnu θ .

C'est ici où réside la différence centrale : l'approche Bayésienne considère l'inconnu comme étant une variable aléatoire possédant une loi de probabilité, tandis que l'inférence classique le traite comme un paramètre fixé même s'il est inconnu.

Malgré l'existence d'un parallélisme fort entre les deux théories (estimation, test, ..etc.) il s'agit d'une différence profonde dans l'interprétation des inférences fournées par ces deux démarches.

Chapitre IV

Lien entre l'AHP et le Théorème de Bayes

1. Introduction :

Grâce à la méthode d'analyse hiérarchique (AHP) et sa généralisation en la méthode d'analyse par réseau (ANP), plusieurs problèmes multicritères à la décision ont pu être résolus ou ont eu une résolution plus élaborée. Cependant, on peut se retrouver en face d'un problème d'aide à la décision multicritères sous incertitude, et de ce fait l'outil statistique et les méthodes probabilistes doivent être utilisées en parallèle avec les méthodes multicritères sous certitude pour prendre en considération cet aspect particulier du problème. Dans ce chapitre, on présentera un modèle qui permettra d'utiliser à la fois les deux méthodologies déjà présentées (la méthodologie de hiérarchisation et la méthode bayésienne d'aide à la décision) et nous permettra ainsi de combiner les avantages des deux approches.

2. Origine :

En médecine, les observations sur lesquelles se basent les décisions médicales sont très imparfaites, elles peuvent être floues, ambiguës, incomplètes, incertaines ...etc. Le processus décisionnel est donc sous incertitude, plusieurs éléments participent à cette situation :

- Les observations sont ambiguës car le malade peut exprimer une plainte et le médecin en entend une autre, de plus des observateurs différents ont des seuils de perception différents dans l'appréciation de la présence ou l'absence d'un signe.
- Elles sont incertaines car la description de l'état présent n'est jamais parfaite, soit par manque de moyen ou de temps (urgence), soit par défaut de mesure ou mauvaise interprétation d'un symptôme, d'un signe ou d'un résultat, de plus les connaissances cliniques sont l'expression d'observations statistiques sur des échantillons de patients présentant des maladies plus ou moins fréquentes, ayant des formes cliniques différentes et ne s'exprimant pas toujours de la même symptomatologie, partageant certains signes avec d'autres maladies ou présentant des réponses variables à un traitement donné.
- La reproductibilité des observations médicales est fonction des méthodes et mesures des observations et le sujet observé.
- Les observations sont floues parce qu'il n'y a pas de vocabulaire standardisé communément utilisé par la communauté médicale et répondant à des définitions clairement exprimées. L'utilisation pour un même concept de termes proches mais pas

synonymes ou pour des concepts voisins du même terme est source d'ambiguïté et d'imprécision sémantique.

- Les observations peuvent dépendre d'eux même rendant difficile la détermination de la maladie qui les engendrent.

Plusieurs méthodes ont été élaborées pour participer dans les diagnostics médicaux et faire face aux difficultés ci-dessus, on particulier : **[Net]**

- Les méthodes mathématiques : elles concernent généralement la simulation du comportement des populations, elles peuvent être utilisées pour décrire le système réel mais surtout pour prédire son évolution future ou son comportement face à des modifications des paramètres d'entrés.
- Les méthodes neuromémitiques : qui sont basées sur les réseaux de neurones, elles sont adaptées aux problèmes de classifications diagnostiques à condition que l'on dispose d'une base de cas suffisante en nombre et en variété pour le processus d'apprentissage. dans cette situation, la couche d'entrée correspond aux signes et la couche de sortie aux diagnostics.
- Les méthodes algébriques qui font appel à l'algèbre de boole qui concerne les variables ne pouvant prendre que deux états : Vrai / Faux ou Présent / Absent. L'algèbre de boole vise à reproduire le raisonnement médical et à formaliser les connaissances aux moyens d'arguments binaires en Oui ou Non, et d'opérateurs logiques Non, Et, Ou, Implique. Cependant, l'exploitation de cette méthode exclut l'aspect d'incertitude du problème qui est fréquent en médecine.
- Les méthodes symboliques de l'intelligence artificielle dont l'objectif est de faire traiter par l'ordinateur des problèmes usuellement résolus par l'homme et dont la solution exige des connaissances de la perception, du raisonnement et d'apprentissage.
- Les méthodes décisionnelles utilisant la théorie de décision et la théorie d'utilité.

3. Le théorème de Bayes et le diagnostic médical :

Le théorème de Bayes produit un paradigme pour la mise à jour de l'information sur le diagnostic médical, codée sous forme d'une probabilité, elle suppose que les décisions avec incertitude ne peuvent être faites qu'avec l'aide d'informations sur l'environnement dans

lequel le diagnostic est fait, autrement dit, avec la connaissance du médecin, la communauté médicale et les travaux du passé qui ont permis de réaliser avec succès des diagnostics et des traitements des patients en se basant sur des symptômes observés de maladies connues.

Le théorème de Bayes met à jour cette information en reliant, sous forme de probabilité conditionnelle, les causes (inconnues) et les résultats qui sont les observations ou les symptômes utilisés pour déterminer les maladies.

En effet, le théorème de Bayes établit que la probabilité à posteriori d'une maladie D_i , lorsque le signe S est présent, est fonction de la probabilité à priori de la maladie, et de la probabilité conditionnelle d'observer le signe lorsque la maladie est présente :

$$P(D_i / S) = \frac{P(D_i)P(S / D_i)}{\sum_{j=1}^N P(D_j)P(S / D_j)}$$

la méthode Bayésienne tient compte des signes positifs et négatifs pour déterminer la maladie la plus probable.

La détermination des probabilités a priori (incidence de la maladie) et des vraisemblances (fréquence d'un signe dans une maladie) suppose l'exploitation d'une base de données médicales correctement constituée, enregistrant de façon non biaisée les signes et le diagnostic de chaque cas. Les performances sont généralement acceptables, avec une concordance observée avec les experts de plus de 70 %.

Le système Bayésien le plus connu est celui développé par l'équipe de DeDombal à Leeds (Angleterre), qui a fait l'objet d'évaluations dans différentes situations de soins. Il s'agit d'un système d'aide au diagnostic des douleurs abdominales aiguës pouvant nécessiter un geste chirurgical. Son utilisation permet au médecin de poser plus souvent le bon diagnostic que sans son assistance. D'autre part, les faux négatifs (diagnostic médical conduisant à une abstention chirurgicale entraînant une aggravation) et les faux positifs (intervention à tort) sont moins fréquents.

[Pew01] : "La connaissance de la prévalence des maladies ainsi que la performance de l'anamnèse des signes cliniques et des tests en matière de diagnostic constituent des outils précieux pour la pose d'un bon diagnostic. Les grands secrets de l'art médical ne nous sont bien entendu pas révélés pour autant, le théorème de Bayes nous permet de franchir un pas de plus sur la voie de l'intuition, par définition difficilement saisissable et sujette à erreur, en donnant à celle-ci un fondement plus rationnel, mais il ne doit ni ne peut remplacer le fruit d'une expérience de longues années ..."

4. La dépendance entre les symptômes :

Une supposition simplifiante et satisfaite dans la vie pratique est que les symptômes sont indépendants entre eux, certains chercheurs tels que Dedambal [DED72], Adams [ADM86] et Spiegelhalter [SPI84] croient que cette supposition ne décroît pas la qualité du diagnostic, d'autres chercheurs remettent en cause l'utilisation du théorème de Bayes, Jonson [JON91] écrit : "Est-ce-que les conditions fondamentales du calcul Bayésien réellement existe dans le diagnostic de chaque jour ? Peut une personne recommander l'utilisation de l'utilité de l'analyse de décision pour sélectionner des tests durant la gestion du diagnostic d'un patient particulier ? ...La distance entre le modèle et la réalité, causé par le hasard des paramètres cliniques motive la question : est-ce-que les hypothèses du diagnostic réellement possèdent une crédibilité comparable à la probabilité dans le sens mathématique ? Il y a une circonstance additionnelle qui augmente le doute à propos de ceci...selon l'interprétation de probabilité par la fréquence de prévalence, la probabilité qu'un patient avec, un certain résultat de test a une certaine maladie, dépend de la fréquence relative des patients avec cette collection de résultats de test et la distribution des maladies entre eux, les combinaisons des résultats des tests seuls ne nous permettra pas l'estimation de la probabilité des hypothèses du diagnostic."

L'indépendance est supposée pour rendre possible l'estimation des probabilités qu'un patient a une ou plusieurs maladies sachant différents symptômes donnés, parce que l'information sur les symptômes et leurs combinaisons n'est pas bien connue et facilement reconnu pour plusieurs maladies.

La version simplifiée du théorème de Bayes consiste à estimer la probabilité à posteriori sous l'hypothèse que les symptômes sont statistiquement indépendants, parce que les données peuvent être disponibles pour une combinaison de symptômes.

Notons que la supposition d'indépendance ne rend pas nécessairement le diagnostic avec le théorème de Bayes moins exacte. Gammerman et Thatcher [GAM91] comparent l'exactitude du diagnostic en utilisant le théorème de Bayes avec et sans l'hypothèse d'indépendance dans un échantillon de 6000 patients souffrant de douleurs abdominales aiguës, ils estiment d'après des données passées la probabilité que ces patients ont une certaine maladie – sachant leurs symptômes – avec les exactitudes suivantes.

Diagnostic par	Exactitude
Médecin	76 %
Bayes simple (avec l'hypothèse d'indépendance)	74 %
Bayes propre (sans l'hypothèse d'indépendance)	65 %

Ce tableau indique qu'une modélisation mathématique plus élaborée n'augmente pas forcément la qualité des résultats.

5. Le modèle :

Il y a un besoin dans le diagnostic médical d' un modèle qui prend en considération à la fois les données statistiques et les jugements d'experts (dans ce cas les médecins) à propos du patient.

Quand les données statistiques sont disponibles mais pas les jugements d'experts alors ce modèle doit reproduire les résultats obtenus par l'approche Bayésienne, quand les jugements d'experts sont aussi présents, alors le modèle doit combiner les deux informations pour mieux déterminer la maladie qui décrit bien les symptômes observés.

Ce modèle doit être une généralisation du modèle basé sur les données statistiques et le modèle basé sur les jugements, donc il doit utiliser une mesure qui traite et combine les jugements et les données statistiques.

L'AHP peut être utilisée dans le contexte du théorème de Bayes pour relier les probabilités à priori et les probabilités des symptômes.

Utiliser la super matrice :

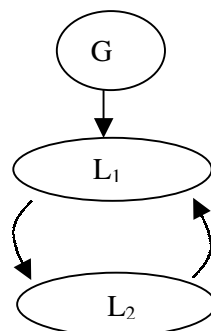


Figure 1

Considérons le réseau de la figure 1 qui contient trois nœuds.

Soit $P_1 = P(L_1)$ le vecteur colonne de l'impact de G sur le nœud L_1

Soit $P_{21} = P(L_2 / L_1)$ la matrice stochastique d'impact du nœud L_1 sur le nœud L_2

Soit $P_{12} = P(L_1 / L_2)$ la matrice stochastique d'impact du nœud L_2 sur le nœud L_1

Dans le contexte du théorème de Bayes, P_{21} correspond aux vraisemblances et P_1 aux probabilités à priori.

Le théorème de Bayes détermine P_{12} en fonction de P_1 et P_{21}

On utilisera P_1 , P_{21} et P_{12} pour obtenir le vecteur des priorités limites du réseau de la figure 1

Pour ceci, soit la super matrice correspondante

$$W = \begin{matrix} & \begin{matrix} G & L_1 & L_2 \end{matrix} \\ \begin{matrix} G \\ L_1 \\ L_2 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 \\ P_1 & 0 & P_{12} \\ 0 & P_{21} & 0 \end{pmatrix} \end{matrix}$$

Combinaison entre l' AHP et le théorème de Bayes

Théorème 1 :

Soit W la super matrice stochastique ci-dessus, il existe deux vecteurs \bar{a} et \bar{b} uniques qui satisfont :

$$P_{12} \cdot \bar{b} = \bar{a} \quad \text{et} \quad P_{21} \cdot \bar{a} = \bar{b}$$

et ce sont les vecteurs propres principaux des matrices $(P_{12} \cdot P_{21})$ et $(P_{21} \cdot P_{12})$.

Preuve :

W est une matrice stochastique périodique de périodicité 2, donc :

$$W^\infty = \frac{1}{2}(I + W)(W^2)^\infty$$

$$(W^2)^k = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (P_{12}P_{21})^k & 0 \\ (P_{21}P_{12})^{k-1}P_{21}P_1 & 0 & (P_{21}P_{12})^k \end{pmatrix}$$

et donc :

$$W^\infty = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ P_{12}BP_{21}P_1 & A & P_{12}B \\ BP_{21}P_1 & P_{21}A & B \end{pmatrix}$$

avec $A = \lim_{k \rightarrow \infty} (P_{12}P_{21})^k$ et $B = \lim_{k \rightarrow \infty} (P_{21}P_{12})^k$

$(P_{12}P_{21})$ et $(P_{21}P_{12})$ sont irréductible et primitives par construction, il s'ensuit que les colonnes de A et de B sont identiques, autrement dit :

soient $e_n^T = (1,1,\dots,1)$ et $e_m^T = (1,1,\dots,1)$ deux vecteurs de dimension n et m respectivement dont leurs composantes valent toutes 1.

$A = \bar{a} e_n^T$ et $B = \bar{b} e_m^T$ avec $\bar{a} = (a_1, \dots, a_n)$ et $\bar{b} = (b_1, \dots, b_m)$

et de plus \bar{a} et \bar{b} sont les vecteurs propres principaux de $(P_{12}P_{21})$ et $(P_{21}P_{12})$ respectivement.

ou encore : $(P_{12}P_{21})\bar{a} = \bar{a} \Leftrightarrow (P_{21}P_{12})P_{21}\bar{a} = P_{21}\bar{a}$

(on rappelle que la valeur propre principale d'une matrice stochastique est égale à 1)

comme le vecteur propre principal de $(P_{21}P_{12})$ est unique, il s'ensuit que $P_{21}\bar{a} = \bar{b}$.

De manière analogue, on trouvera : $P_{12}\bar{b} = \bar{a}$

Sous une forme matricielle, on a : $\begin{pmatrix} 0 & P_{12} \\ P_{21} & 0 \end{pmatrix} \cdot \begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix} = \begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix}$

Ce qui signifie que $\begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix}$ est le vecteur propre principal de $\begin{pmatrix} 0 & P_{12} \\ P_{21} & 0 \end{pmatrix}$

Le résultat suivant exprime la relation entre l' AHP et le théorème de Bayes.

Théorème 2 :

Soit W la super matrice de la figure 1,

Il existe une matrice P_{12} qui vérifie :

$$(P_{12}P_{21})P_1 = P_1 \text{ et } (P_{21}P_{12})P_2 = P_2 \text{ avec } P_2 = P_{21}P_1.$$

Preuve :

Si $P_{12} = (\Delta P_1) \cdot P_{21}^T \cdot (\Delta P_2)^{-1} \dots \dots (1)$

avec

$$\Delta P_1 = \begin{pmatrix} P_{1_1} & \dots & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & \dots & P_{1_n} \end{pmatrix} \text{ et } \Delta P_2 = \begin{pmatrix} P_{2_1} & \dots & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & \dots & P_{2_m} \end{pmatrix}$$

alors multipliant (1) à droite par $(\Delta P_2)e_m$ on obtient :

$$P_{12}P_2 = (\Delta P_1)P_{21}^T e_m \Leftrightarrow P_{12}P_2 = (\Delta P_1)e_n \Leftrightarrow P_{12}P_2 = P_1$$

De même, faisant la transposition dans (1) et multipliant à droite par (ΔP_2) et à gauche par e_n

on obtient : $(\Delta P_2)P_{12}^T e_n = P_{21}P_1 \Leftrightarrow P_{21}P_1 = P_2$

D'où $(P_{12}P_{21})P_1 = P_1$ et $(P_{21}P_{12})P_2 = P_2$

Sous forme matricielle, on a : $\begin{pmatrix} 0 & P_{12} \\ P_{21} & 0 \end{pmatrix} \cdot \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}$

D'après les théorèmes 1 et 2 :

Dans W^∞ , A et B sont des matrices de colonnes toutes identiques et égales à P_1 et P_2 respectivement. et donc on vérifie facilement que W^∞ est de colonnes toutes identiques et égales à $(0, P_1, P_2)^T$.

Il s'ensuit que si P_{12} est la matrice des probabilités à posteriori alors P_1 et P_2 sont les vecteurs propres principaux de $(P_{12}P_{21})$ et $(P_{21}P_{12})$ et donc $(0, P_1, P_2)^T$ est le vecteur de priorités limites correspondant au réseau de la figure 1.

De plus

La matrice P_{12} n'est pas la seule qui permet d'obtenir ce vecteur comme vecteur de priorité limite.

En effet :

Soit le système suivant :

$$\begin{pmatrix} 0 & Q \\ P_{21} & 0 \end{pmatrix} \cdot \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}$$

avec

$Q \neq (\Delta P_1) P_{21}^T (\Delta P_2)^{-1}$ la matrice des probabilités à posteriori.

Alors :

$$Q \cdot P_2 = P_1$$

$$P_{21} \cdot P_1 = P_2$$

$$\text{et } P_{12} \cdot P_2 = P_1$$

$$P_{21} \cdot Q \cdot P_2 = P_{21} \cdot P_{12} \cdot P_2 \Leftrightarrow P_{21} (Q - P_{12}) P_2 = 0 \dots \dots \dots (*)$$

Toute matrice Q vérifiant (*) permet d'obtenir le même résultat que P_{12} , c-à-d que $(0, P_1, P_2)^T$ est le vecteur de priorité limite du réseau de la figure 1.

Exemple :

$$\text{Prenons } P_1 = \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} \text{ et } P_{21} = \begin{pmatrix} 0.2 & 0.6 \\ 0.8 & 0.4 \end{pmatrix}$$

$$\text{Alors } P_2 = \begin{pmatrix} 0.48 \\ 0.52 \end{pmatrix} \text{ et } P_{12} = \begin{pmatrix} 0.125 & 0.462 \\ 0.875 & 0.538 \end{pmatrix}$$

Si on prend $Q = \begin{pmatrix} 0.6 & 0.5538 \\ 0.4 & 0.4462 \end{pmatrix}$, Q satisfait (*) et permet de donner les mêmes priorités

limites que P_{12} , les vecteurs propres de (QP_{21}) et $(P_{21}Q)$ sont P_1 et P_2 respectivement.

Soit $L_1 = \Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ l'espace des états de la nature

et $L_2 = X = \{x_1, x_2, \dots, x_m\}$ l'espace des observations.

Soit $P_{12} = P(\Theta / X) = P(\theta = \theta_i / x = x_j)$ $i = 1, \dots, n$; $j = 1, \dots, m$ une $n \times m$ matrice stochastique des probabilités à posteriori.

Soit $P_{21} = P(X / \Theta) = P(x = x_j / \theta = \theta_i)$ $i = 1, \dots, n$; $j = 1, \dots, m$ une $m \times n$ matrice stochastique des vraisemblances.

Soit $P_1 = P(\Theta)$ un vecteur colonne des probabilités à priori.

L_1 correspond à l'ensemble des maladies et L_2 à l'ensemble des symptômes.

Soit $P_2 = P_{21} * P_1$.

Si P_{12} correspond à la matrice Bayésienne c-à-d $P_{12} = \Delta P(\theta) \cdot P_{21}^T \cdot \Delta P(X)^{-1}$ avec

$$\Delta P(\theta) = \begin{pmatrix} P(\theta_1) & 0 & \dots & 0 \\ 0 & P(\theta_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & P(\theta_n) \end{pmatrix} \quad \text{et} \quad \Delta P(X) = \begin{pmatrix} P(x_1) & 0 & \dots & 0 \\ 0 & P(x_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & P(x_m) \end{pmatrix}$$

Alors d'après les théorèmes 1 et 2 : le vecteur $(0, P(\Theta), P(X))^T$ est le vecteur des priorités limites du réseau de la figure 1.

Cette condition est suffisante pour avoir $(0, P(\Theta), P(X))^T$ comme un vecteur de priorités limites, mais pas nécessaire, car il peut y exister d'autres matrices différentes de celle de la formule de Bayes, et qui permet d'aboutir aux mêmes priorités limites.

Cela signifie que P_{12} n'est pas seulement la matrice des probabilités à posteriori, mais elle peut avoir un autre sens que celui d'une probabilité, par exemple une importance, dominance, préférence ..e tc.

Remarque :

Il est utile de noter que $P(X) = P(X / \Theta).P(\Theta)$ s'obtient de deux manières différentes et indépendantes :

1. D'après la définition des probabilités conditionnelles.
2. D'après les théorèmes 1 et 2 qui ne sont pas basés sur des probabilités conditionnelles, mais plutôt sur le modèle de la figure 1 représentant des influences entre les sommets du réseau.

De plus, l'avantage apporté par ce modèle est de retrouver un résultat déjà connu, mais à partir d'une approche plus générale (AHP avec feedback) basée sur des rapports d'échelles utilisés dans l'aide à la décision, car les rapports d'échelles et les probabilités peuvent, les deux, être dérivés sous forme de priorités.

6. Dépendance entre les symptômes :

Supposons qu'on a un patient qui présente certains symptômes et signes cliniques, le médecin consultant, à base de ces symptômes et d'autres tests, envisage un ensemble de maladies suspectées.

Le modèle devient le suivant :

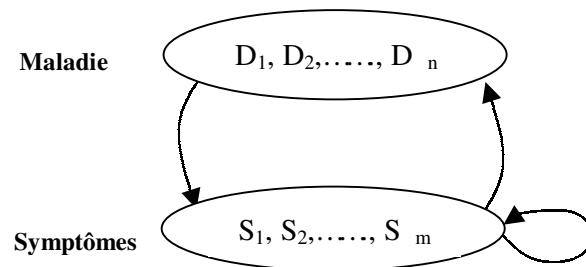


Figure 2

La super matrice associé est : $W = \begin{pmatrix} 0 & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$

On note ici la présence de la dépendance entre les symptômes (cette dépendance n'était pas prise en considération dans le modèle de la figure 1).

Un modèle basé sur les jugements d'experts et les données statistiques et prenant en considération la dépendance entre les symptômes pourra donner des résultats meilleurs et plus fondés qu'un modèle basé sur les données statistiques seulement.

Les priorités dans W_{21} indiquent qu'elle maladie est plus probable pour causer les symptômes observés.

Les priorités dans W_{12} indiquent quel symptôme caractérise le plus une maladie donnée.

W_{12} et W_{21} peuvent être obtenues à travers des jugements d'experts (ici les médecins) ou bien en utilisant des données statistiques dans quel cas la matrice W_{12} correspondra aux probabilités à posteriori par rapport à W_{21} .

W_{22} représente l'influence des symptômes entre eux établi à travers des jugements, les priorités de chaque colonne de W_{22} représentent les probabilités que les symptômes soient présents avec le symptôme correspondant à cette colonne.

Les symptômes considérés dans le modèle ne sont pas tous les symptômes possibles pour une maladie donnée, mais plutôt ceux observés chez le patient.

Les matrices W_{12} et W_{22} sont stochastiques, ce qui implique que la super matrice ne l'est pas. Dans le cas général (le modèle de l'ANP) il faut attribuer des poids à tous les groupes du réseau à l'aide des comparaisons par paires, on posant la question générique suivante : Etant donné un groupe A, quel est le groupe le plus important parmi deux groupes B et C par rapport à A, et de combien ?

Dans notre modèle, et pour rendre W stochastique, il faut pondérer les deux groupes des maladies et des symptômes par les poids α_1 et α_2 respectivement avec $\alpha_1 + \alpha_2 = 1$.

Pour ceci, on pose la question suivante : "Est-ce-que la connaissance d'un symptôme et son intensité pour identifier une maladie est une conséquence de la connaissance de la maladie ou de la connaissance des autres symptômes".

Si les deux contribuent de la même intensité, alors $\alpha_1 = \alpha_2 = 0.5$

On note ici que la question est posée par rapport au groupe des symptômes seulement, car seuls les blocks correspondants aux symptômes dans W nécessitent une pondération.

La matrice W est irréductible alors W après pondération reste irréductible.

L'objectif est de trouver les priorités limites des maladies, ceci s'obtient par l'obtention du vecteur propre de W , autrement dit :

Résoudre :

$$\begin{pmatrix} 0 & \alpha_1 W_{12} \\ W_{21} & \alpha_2 W_{22} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

Ce vecteur s'obtient en élevant la matrice W à des puissances infinies.

Donc :

$$\begin{cases} (\alpha_1 W_{21} W_{12} + \alpha_2 W_{22}) w_2 = w_2 \\ w_1 = \alpha_1 W_{12} w_2 \end{cases}$$

$$\begin{cases} w_2 = \left[\lim_{k \rightarrow \infty} (\alpha_1 W_{21} W_{12} + \alpha_2 W_{22})^k \right] \cdot e^T \dots\dots\dots (*) \\ w_1 = \alpha_1 W_{12} w_2 \end{cases}$$

avec $e = (1, 1, \dots, 1)$.

Remarque :

S'il n'y avait pas de dépendance entre les symptômes c -à-d $W_{22} = 0$, alors w_1 sera le vecteur propre de $(W_{12}W_{21})$ et w_2 sera le vecteur propre de $(W_{21}W_{12})$. w_1 et w_2 correspondent, dans le contexte du théorème de Bayes aux probabilités à priori et aux probabilités marginales des symptômes respectivement, ce qui est conforme avec les Théorèmes 1 et 2.

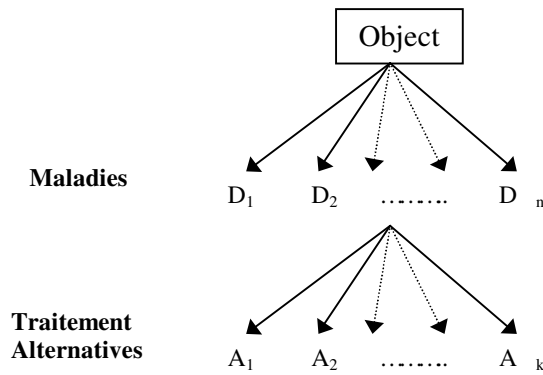
7. Au-delà du théorème de Bayes :

Le modèle de la figure 2, nous permet d'établir les priorités limites des maladies et des symptômes, permettant nous ainsi de déterminer la maladie la plus probable parmi les maladies suspectées.

On suppose maintenant qu'on a plusieurs traitements et on cherche celui le plus adéquat selon les priorités des maladies.

On décrira pour cela deux méthodes équivalentes.

La **première méthode** consiste à établir une hiérarchie comprenant les maladies et les traitements.



Il faut établir des comparaisons par paires afin de déterminer l'importance relative des traitements pour chaque maladie.

Supposons que $W(A / D)$ est la matrice des priorités locales des traitements par rapports aux maladies, alors les priorités globales des traitements sont données par : $W(A / D) \cdot (w_1 / \|w_1\|)$ avec $\|w_1\|$ est la norme du vecteur w_1 : $\|w_1\| = e^T w_1$ avec $e^T = (1, 1, \dots, 1)$

On note ici que, selon le médecin, cette hiérarchie peut contenir plus de niveau si le médecin doit définir des sous critères pour choisir le meilleur traitement à considérer selon la maladie.

La **deuxième méthode** consiste à inclure le groupe traitements dans le réseau de la figure 2, ce réseau devient :

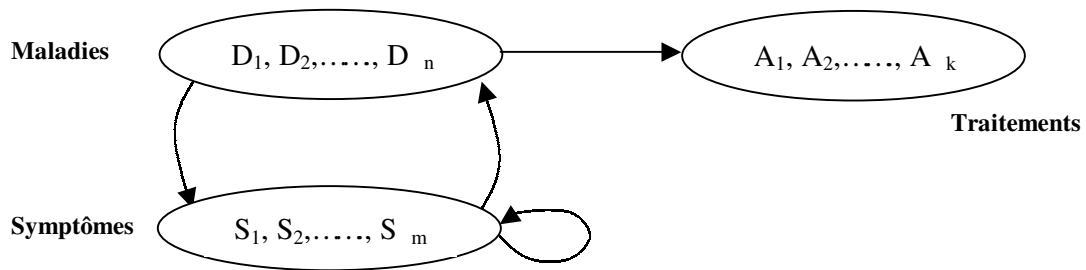


Figure 3

Donc, on a ajouté le groupe traitement, qui est influencé seulement par le groupe maladie
La super matrice correspondante est :

$$W = \begin{matrix} & \begin{matrix} Malad & Symp & Trait \end{matrix} \\ \begin{matrix} maladies \\ Syntomes \\ Traitements \end{matrix} & \begin{pmatrix} 0 & \alpha_1 W_{12} & 0 \\ \beta_1 W_{21} & \alpha_2 W_{22} & 0 \\ \beta_2 W_{31} & 0 & I \end{pmatrix} \end{matrix}$$

Cette matrice n'est plus irréductible (le graphe n'est pas fortement connexe) et donc pour déterminer les priorités limites, on cherche tout d'abord les priorités des éléments des composantes qui influent et qui sont influencées par d'autres composantes.

Donc on calcul Q^∞ avec $Q = \begin{pmatrix} 0 & \alpha_1 W_{12} \\ W_{21} & \alpha_2 W_{22} \end{pmatrix}$

D'après (*) on obtient w_1 et w_2 .

Les priorités limites des traitements s'obtiennent par : $(W_{31}, 0)Q^\infty$

Comme Q^∞ est de colonnes toutes identiques et égales à $(w_1, w_2)^T$. et W_{31} représente l'impact des maladies sur les traitements (c'est la même matrice que $W(A / D)$ de la première méthode) alors, on remarque bien que les deux méthodes permettent de trouver le même résultat.

La raison principale de ceci est que, dans la figure 3, le groupe traitements n'influe aucun autre groupe du réseau et est influencé seulement par le groupe maladies, donc la hiérarchie de la première méthode prend en considération les influences relatives au groupe traitements seulement.

S'il y avait une dépendance entre les traitements, ou le groupe des traitements influe un autre groupe, alors la première méthode ne serait plus applicable et seule, un réseau et une super matrice tel que celle de la figure 3, permet d'obtenir les priorités limites, d'où la généralité de la deuxième méthode.

8. Conclusion :

On a vu que si on utilise la même information que dans le théorème de Bayes, les priorités limites obtenues par la super matrice sont les mêmes informations que ceux dont on a commencé avec, dans le théorème de Bayes, c-à-d les probabilités à priori et les probabilités des symptômes $P(\theta)$ et $P(X)$, donc si on estime les matrices W_{21} et W_{12} (les vraisemblances et les probabilités à posteriori) alors les probabilités a priori correspondent au vecteur propre principal de $W_{12}W_{21}$, et les probabilités des symptômes sont le vecteur propre principal de $W_{21}W_{12}$.

Si de plus, le modèle inclus une dépendance entre les symptômes, alors le résultat final est différent et plus général que celui obtenu par l'approche Bayésienne.

Il existe d'autres techniques pour combiner l'information à priori avec les connaissances passées (expériences, données et fréquences statistiques ..etc.) pour modéliser une situation d'incertitude, le théorème de Bayes résout ce problème en établissant des probabilités à posteriori, le modèle AHP permet d'obtenir un résultat plus général sous une certaine condition suffisante mais pas nécessaire (satisfaite dans le cas d'une approche Bayésienne), l'AHP permet ainsi d'obtenir plusieurs matrices relatives à plusieurs interprétations (probabilités, préférences ..etc.)

Le diagnostic avec des symptômes dépendants nécessite des jugements d'expert, des informations statistiques et des expériences sur les relations entre les symptômes et des combinaisons de symptômes, le théorème de Bayes dans ce cas n'apparaît pas l'arbitre final du diagnostic médical.

L'approche de la super matrice permet d'associer des symptômes avec d'autres symptômes connus d'une certaine maladie, en utilisant l'expérience et les jugements d'un expert plutôt d'utiliser des données statistiques qui peuvent ne pas être disponible.

L'approche de la super matrice permet aussi de dépasser la difficulté combinatoire du problème qui en résulte si on veut étudier des combinaisons de symptômes (car le nombre de combinaisons peut être très grand même pour un nombre réduit de symptômes), de plus les jugements (caractérisés par des comparaisons par paires) permettent à l'expert de se concentrer sur une situation spécifique, plutôt que de prendre tous les symptômes ou toutes les maladies à la fois.

Le modèle de l'AHP a permis de retrouver les probabilités à priori du théorème de Bayes sans utiliser les probabilités conditionnelles sur lesquelles ce théorème est basé, car les comparaisons par paires de deux éléments d'un niveau de la hiérarchie remplacent les probabilités conditionnelles. Si ces probabilités sont aussi disponibles, elles peuvent remplacer les jugements ou au moins, aider à les établir.

Chapitre V

Application

1. Introduction

Dans ce chapitre, on appliquera notre modèle sur un cas pratique, au niveau du Centre Hospitalo Universitaire de Bab El Oued, service de médecine interne.

Un homme de 26 ans consulte pour une dyspnée d'apparition progressive, l'interrogatoire retrouve la notion d'amaigrissement, fièvre, sueurs et prurit.

L'examen clinique retrouve une discrète pâleur, une température à 38.5°C, des lésions de grattage, des aphtes buccaux, ainsi qu'un syndrome d'épanchement pleural liquidien dont la ponction révèle un exsudat avec 100 lymphocytes/CC.

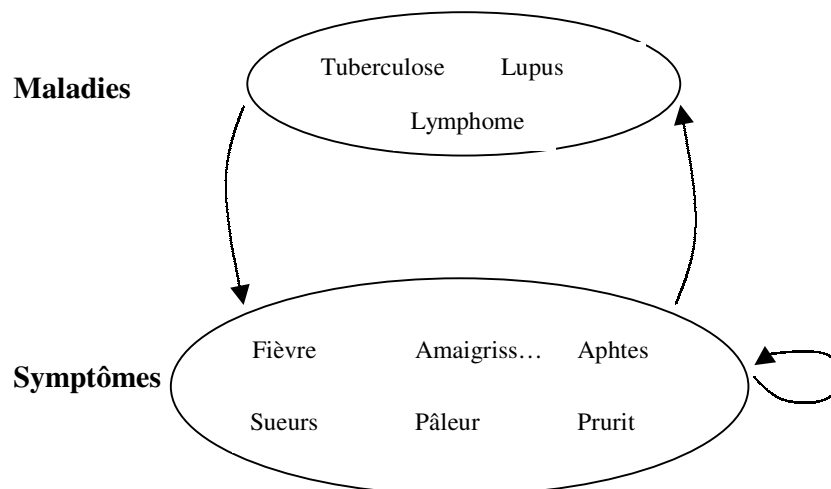
Le médecin traitant pense à trois maladies :

1. Tuberculose.
2. Lymphome.
3. Lupus.

Les symptômes observés sont :

1. Amaigrissement.
2. Fièvre.
3. Sueur.
4. Pâleur.
5. Aphtes.
6. Prurit.

On note que les symptômes ci-dessus ne sont pas tous les symptômes qui caractérisent les maladies suspectées mais sont seulement les symptômes observés.



Modèle de diagnostic médical avec feedback

La super matrice de cette figure est : $W = \begin{pmatrix} 0 & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$

Notre objectif est de déterminer les priorités des maladies, ce qui permettra de savoir la maladie la plus suspectée.

Les jugements qui suivent, peuvent être remplacés par des fréquences statistiques obtenues par l'observation des symptômes les plus dominant pour chaque maladie, comparées avec les autres symptômes. Vu que les fréquences statistiques ne caractérisent pas les individus, il est essentiel d'ajuster ces fréquences aux situations spécifiques en tenant compte des données spécifiques du patient, les jugements peuvent prendre en considération les perceptions du médecin à propos des changements des données statistiques d'un malade à un autre.

Dans notre application, et à cause de l'absence des fréquences statistiques, on n'a utilisé que les jugements du médecin.

2. Evaluation des priorités des symptômes par rapport aux maladies :

Dans la première étape, on cherche à déterminer les priorités des symptômes par rapport à chaque maladie.

On note que les comparaisons par paires ne font pas intervenir tous les symptômes à la fois, mais seulement les symptômes qui caractérisent la maladie considérée.

La question posée au médecin à chaque comparaison est " Pour une maladie donnée, et deux symptômes, quel symptôme est plus caractérisant de cette maladie, et de combien ? "

La réponse sera verbale parmi les réponses : également, légèrement plus, fortement plus, très fortement plus, absolument plus caractérisant. Ces jugements seront transformés en les nombre 1, 3, 5, 7 et 9 respectivement, les nombre 2, 4, 6 et 8 seront utilisés pour réaliser un compromis entre les jugements précédents. De plus, si le symptôme 1, par exemple, est fortement plus caractérisant la maladie que le symptôme 2, c-à-d exprimé par la valeur 5, alors le jugement 2 est 1/5 plus caractérisant la maladie que le symptôme 1.

On rappelle l'échelle fondamentale de 1 à 9 utilisé :

Valeurs numériques	Définition
1	Egalement important (aucune préférence)
3	Légèrement plus important
5	Fortement plus important
7	Très fortement plus important
9	Absolument plus important (une préférence absolue)
2, 4, 6 et 8	Valeurs intermédiaires pour mettre en évidence des compromis
Valeurs inverses	Valeurs Inverses utilisées pour montrer la dominance du second élément par rapport au premier

Par exemple : par rapport à la tuberculose, l'amaigrissement est légèrement plus caractérisant que la fièvre d'où la valeur 3.

TUBERCULOSE	Amaigriss...	Fièvre	Sueurs	Pâleur	Priorités
Amaigriss...	1	3	3	7	0.5281
Fièvre	1/3	1	1	5	0.21
Sueurs	1/3	1	1	5	0.21
Pâleur	1/7	1/5	1/5	1	0.0519

CR = 0.027 (Consistency Ratio : rapport de cohérence)

on note l'absence des symptômes Aphtes et Prurit car il ne caractérise plus la tuberculose.

Les priorités locales des symptômes sont données à droite du tableau, elles sont obtenues par le calcul du vecteur propre principale de cette matrice.

Le CR est très faible (inférieur à 0.1, valeur suggérée par Saaty) indiquant que ces jugements sont acceptables.

On remarque que, par rapport à la tuberculose, l'amaigrissement est de priorité beaucoup plus importante que celle de la fièvre ou de la sueur (plus que le double) qui eux même dépassent de loin celle de la pâleur.

On trouve ci-dessous, les comparaisons par paires des symptômes par rapport à lymphome et lupus, ainsi qu'un tableau récapitulatif qui représente la matrice W_{21} dans la super matrice W (si un symptôme est absent par rapport à une maladie alors la valeur zéro sera mise dans la case correspondante).

LYMPHOME	Amaigriss...	Fièvre	Sueurs	Pâleur	Prurit	Priorités
Amaigriss...	1	1/9	1/5	1/2	1/3	0.0454
Fièvre	9	1	3	7	3	0.5118
Sueurs	5	1/3	1	3	2	0.2229
Pâleur	2	1/7	1/3	1	2	0.1086
Prurit	3	1/3	1/2	1/2	1	0.1113

CR = 0.015

LUPUS	Amaigriss...	Fièvre	Sueurs	Pâleur	Aphtes	Priorités
Amaigriss...	1	1/3	2	1/5	1/7	0.0603
Fièvre	3	1	3	1/3	1/5	0.1209
Sueurs	1/2	1/3	1	1/7	1/9	0.0403
Pâleur	5	3	7	1	2	0.4021
Aphtes	7	5	9	1/2	1	0.3764

CR = 0.05

Tableau récapitulatif

	Tuberculose	Lymphome	Lupus
Amaigriss...	0.5281	0.0454	0.0603
Fièvre	0.21	0.5118	0.1209
Sueurs	0.21	0.2229	0.0403
Pâleur	0.0519	0.1086	0.4021
Aphtes	0	0	0.3764
Prurit	0	0.1113	0

3. Evaluation des priorités des maladies par rapport aux symptômes :

D'une manière analogue, on détermine les priorités des maladies par rapport à chaque symptôme.

La question posée au médecin est : "Etant donné un symptôme et deux maladies, quelle maladie engendre le plus ce symptôme, et de combien ?"

Par exemple : la tuberculose engendre très fortement la fièvre que lymphome, d'où la valeur 7 est donnée à la tuberculose contre lymphome par rapport à la fièvre.

Les rapports de cohérence CR sont généralement acceptables sauf pour la pâleur et le prurit. Dans une tentative de révision des jugements par rapport à la pâleur et surtout par rapport au prurit, le médecin a confirmé encore ses jugements, ce qui nous indique que ce cas particulier contient une certaine incohérence, et pas les jugements du médecin, ce qui confirme ce qu'on a dit sur l'incohérence et la transitivité des jugements dans le chapitre 2.

Amaigriss...	Tuberculose	Lymphome	Lupus	Priorités
Tuberculose	1	7	5	0.7306
Lymphome	1/7	1	1/3	0.081
Lupus	1/5	3	1	0.1884

CR = 0.0620

Fièvre	Tuberculose	Lymphome	Lupus	Priorités
Tuberculose	1	7	5	0.7306
Lymphome	1/7	1	1/3	0.081
Lupus	1/5	3	1	0.1884

CR = 0.0620

Sueurs	Tuberculose	Lymphome	Lupus	Priorités
Tuberculose	1	2	5	0.5695
Lymphome	1/2	1	4	0.3331
Lupus	1/5	1/4	1	0.0974

CR = 0.0234

Pâleur	Tuberculose	Lymphome	Lupus	Priorités
Tuberculose	1	1/4	1/7	0.0727
Lymphome	4	1	1/5	0.205
Lupus	7	5	1	0.7223

CR = 0.1181

Aphtes	Tuberculose	Lymphome	Lupus	Priorités
Tuberculose	1	1/2	1/9	0.0738
Lymphome	2	1	1/8	0.1218
Lupus	9	8	1	0.8044

CR = 0.0351

Prurit	Tuberculose	Lymphome	Lupus	Priorités
Tuberculose	1	1/9	1/7	0.0510
Lymphome	9	1	5	0.7219
Lupus	7	1/5	1	0.2271

CR = 0.1987

Voici le tableau récapitulatif des priorités des maladies par rapport aux symptômes, qui représente la matrice W_{12} dans W .

	Amaigriss...	Fièvre	Sueurs	Pâleur	Aphtes	Prurit
Tuberculose	0.7306	0.7306	0.5695	0.0727	0.0738	0.0510
Lymphome	0.081	0.081	0.3331	0.205	0.1218	0.7219
Lupus	0.1884	0.1884	0.0974	0.7223	0.8044	0.2271

4. Evaluation des priorités des symptômes par rapport aux symptômes :

On traite dans cette étape la dépendance entre les symptômes (feedback).

Dans ce cas la question adressée au médecin est : " Etant donné un symptôme, et deux autres symptômes reliés avec lui, quel est le symptôme le plus associé ? "

Par exemple : la sueur est très fortement plus associée avec la fièvre que les aphtes, d'où par rapport à la fièvre, la sueur est très fortement plus importante que les aphtes (la valeur 7), on remarque aussi que les symptômes pâleur et prurit sont absents de ces comparaisons car ils ne sont pas reliés avec la fièvre.

De plus seule l'amaigrissement est relié à la pâleur avec une importance équivalente d'où les priorités 0.5 ; 0.5. Le prurit n'est relié à aucun autre symptôme.

Les CR sont tous acceptables.

Amaigriss...	Fièvre	Sueurs	Pâleur	Aphtes	Priorités
Fièvre	1	1	4	5	0.4027
Sueurs	1	1	4	5	0.4027
Pâleur	1/4	1/4	1	3	0.1288
Aphtes	1/5	1/5	1/3	1	0.0658

CR = 0.0367

Fièvre	Amaigriss...	Sueurs	Aphtes	Priorités
Amaigriss...	1	1	7	0.4667
Sueurs	1	1	7	0.4667
Aphtes	1/7	1/7	1	0.0666

CR = 0.0002

Sueurs	Amaigriss...	Fièvre	Prurit	Priorités
Amaigriss...	1	1	9	0.4737
Fièvre	1	1	9	0.4737
Prurit	1/9	1/9	1	0.0526

CR = 0.0001

Aphtes	Amaigriss...	Fièvre	Sueurs	Pâleur	Priorités
Amaigriss...	1	5	4	7	0.6192
Fièvre	1/5	1	2	3	0.1873
Sueurs	1/4	1/2	1	2	0.1247
Pâleur	1/7	1/3	1/2	1	0.0688

CR = 0.0364

Tableau récapitulatif

	Amaigriss...	Fièvre	Sueurs	Pâleur	Aphtes	Prurit
Amaigriss...	0	0.4667	0.4737	0.5	0.6192	0
Fièvre	0.4027	0	0.4737	0	0.1873	0
Sueurs	0.4027	0.4667	0	0	0.1247	0
Pâleur	0.1288	0	0	0.5	0.0688	0
Aphtes	0.0658	0.0666	0	0	0	0
Prurit	0	0	0.0526	0	0	1

Ce tableau correspond à W_{22} dans W .

Possédant les matrices W_{12} , W_{21} et W_{22} ; la super matrice W est la suivante :

	Tuberculose	Lymphome	Lupus	Amaigriss	Fièvre	Sueurs	Pâleur	Aphtes	Prurit
Tuberculose	0	0	0	0.7306	0.7306	0.5695	0.0727	0.0738	0.0510
Lymphome	0	0	0	0.081	0.081	0.3331	0.205	0.1218	0.7219
Lupus	0	0	0	0.1884	0.1884	0.0974	0.7223	0.8044	0.2271
Amaigriss...	0.5281	0.0454	0.0603	0	0.4667	0.4737	0.5	0.6192	0
Fièvre	0.21	0.5118	0.1209	0.4027	0	0.4737	0	0.1873	0
Sueurs	0.21	0.2229	0.0403	0.4027	0.4667	0	0	0.1247	0
Pâleur	0.0519	0.1086	0.4021	0.1288	0	0	0.5	0.0688	0
Aphtes	0	0	0.3764	0.0658	0.0666	0	0	0	0
Prurit	0	0.1113	0	0	0	0.0526	0	0	1

Il reste maintenant à rendre W stochastique, pour cela, on pondère les deux blocks

W_{12} et W_{22} par α_1 et α_2 avec $\alpha_1 + \alpha_2 = 1$.

On pose au médecin la question suivante : " Est-ce-que la connaissance d'un symptôme est le résultat direct de la connaissance des maladies ou c'est dû à la connaissance des autres symptômes " ou autrement " Est-ce-que la présence d'un symptôme est dû aux maladies ou aux autres symptômes "

si les deux participent avec le même degré alors :

$$\alpha_1 = \alpha_2 = 0.5.$$

la réponse du médecin était :

$$\alpha_1 = 0.3$$

$$\alpha_2 = 0.7$$

Finalement les priorités globales des maladies et des symptômes s'obtiennent en élevant la super matrice W à des puissances infinies.

5. Priorités globales :

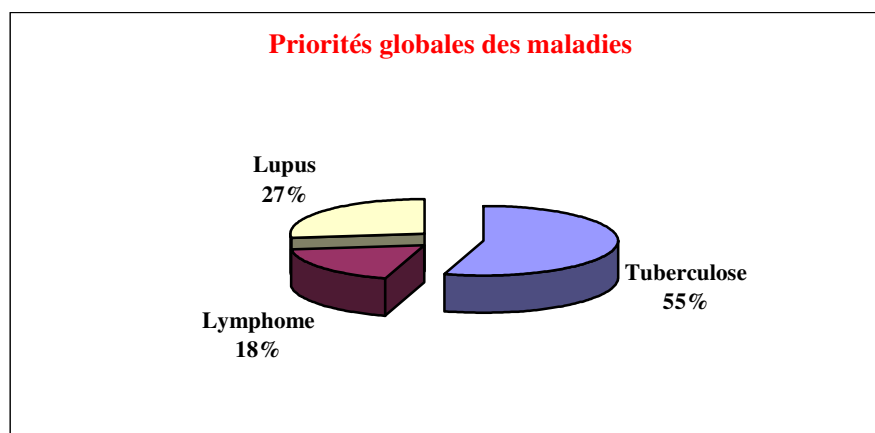
Tuberculose	Lymphome	Lupus	Amaigriss...	Fièvre	Sueurs	Pâleur	Aphtes	Prurit
0,1262	0,0426	0,0619	0,2406	0,1859	0,1708	0,0921	0,0431	0,0368

Priorités des symptômes

Amaigriss...	Fièvre	Sueurs	Pâleur	Aphtes	Prurit
0,3128	0,2416	0,2220	0,1197	0,0560	0,0478
31.28 %	24.16 %	22.2 %	11.97 %	5.6 %	4.78 %

Priorités des maladies

Tuberculose	Lymphome	Lupus
0.5470	0.1847	0.2683
54.7 %	18.47 %	26.83 %



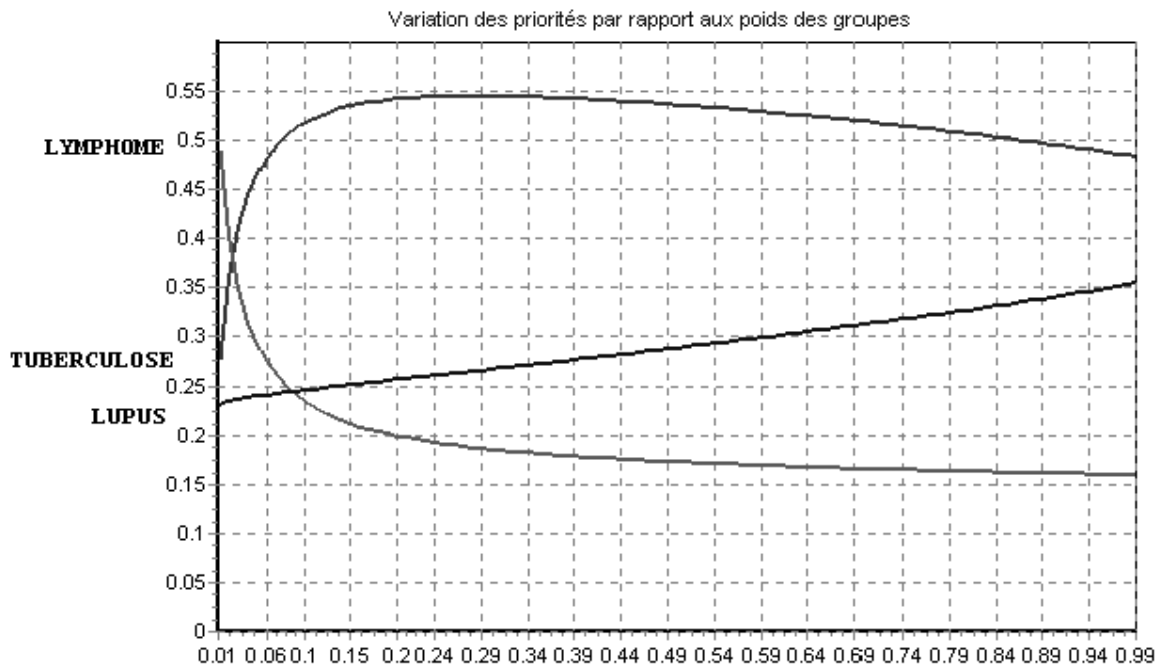
6. Analyse des résultats :

La maladie ayant la plus grande priorité est la Tuberculose (54.7 %) qui est plus que le double de la priorité juste après de Lupus, Lymphome est en troisième position avec 18.47 %. Donc la tuberculose est la maladie la plus suspecté.

Cet ordre de priorité est conforme avec ce que le médecin a pensé sur la situation du Patient, et donc le médecin valide les résultats du modèle.

Dans le but d'analyser la stabilité des résultats par rapport à la pondération des blocks W_{12} et W_{22} , on a fait varier le paramètre α_1 de 0 à 1 (et donc α_2 de 1 à 0) pour voir la variation des priorités des maladies.

Voici les courbes des variations des priorités des maladies pour différentes valeurs de α_1 et α_2 .



D'après ces résultats, on remarque que la tuberculose est la plus dominante presque par tout (sauf pour des valeurs de $\alpha_1 < 0.03$) et sa priorité est toujours beaucoup plus grande que les deux autres.

Une augmentation presque linéaire mais lente de la priorité de Lupus.

Une diminution rapide de la priorité de Lymphome suivie d' une stabilisation relative pendant le reste de l'intervalle $[0,1]$.

La tuberculose augmente rapidement, et puis diminue lentement.

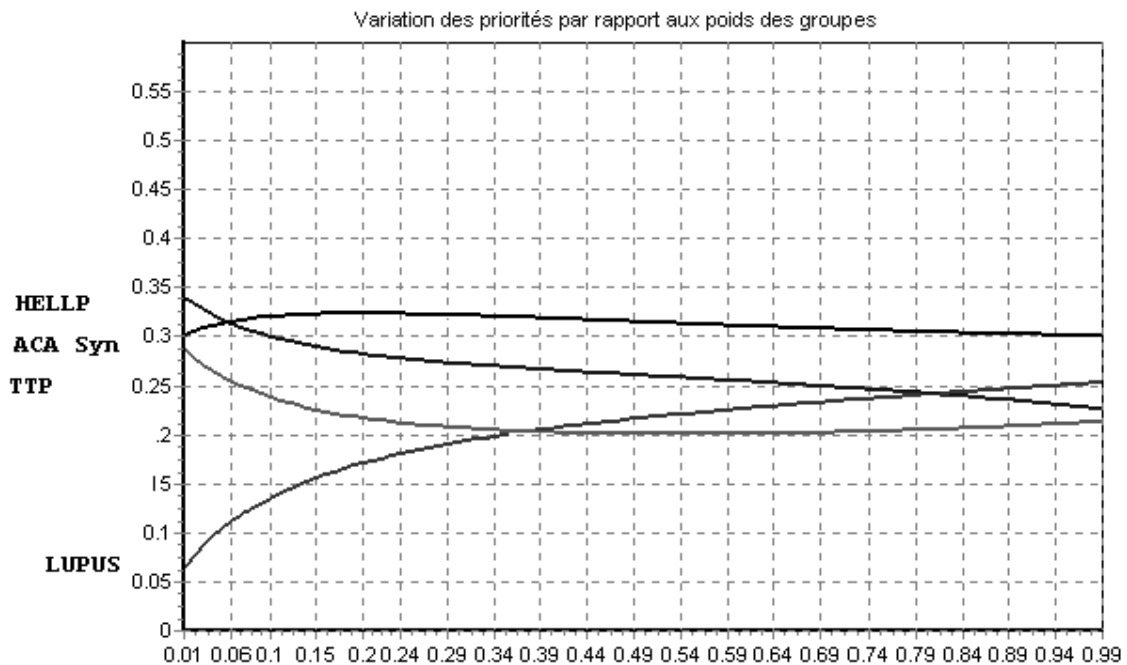
Il est utile de noter que la priorité maximale de la tuberculose est obtenue à la valeur donnée par le médecin ($\alpha_1 = 0.3$).

On remarque que la pondération très faible du groupe des symptômes ($\alpha_2 = 1 - \alpha_1$ très proche de zéro) ne perturbe pas fortement les résultats finals ce qui signifie que la dépendance entre les symptômes ne joue pas un rôle très déterminant dans ce diagnostic médical.

A titre de comparaison, on a fait varier les poids de la super matrice de l'application réalisée par Saaty et Vargas.[SAT98]

Les 4 maladies sont :

1. LUPUS
2. TTP : Thrombotic Thrombocytopenic Purpura
3. HELLP : Hemolysis, Elevated Liver function and Low Platelets
4. ACA Syn : AntiCardiolipin Antibody Syndrome



Cette figure montre que la maladie Aca Syn est généralement plus dominante que les autres maladies sur tout l'intervalle $[0,1]$, mais sa priorité reste proche des priorités des autres maladies, contrairement à notre cas.

D'autre part, même dans cette application la dépendance entre les symptômes n'influe pas fortement sur les priorités globales des maladies.

Conclusion

Dans le diagnostic médical, les jugements sont nécessaires pour déterminer les tests à réaliser en se basant sur certains symptômes.

Pour plusieurs maladies, les informations devant être récoltées sur les symptômes et les combinaisons des symptômes qui mènent à ces maladies ne sont pas bien connues.

Il y a un besoin, dans le diagnostic médical d' un modèle qui permet de déterminer les préférences du médecin et donner des priorités aux maladies suspectées tout en exploitant au maximum les informations disponibles telles que les données statistiques, les taux d'apparition des maladies, l'expérience antérieure du médecin..etc.

L'approche Bayésienne, connue de sa capacité d'exploiter les informations à priori, résout ce problème en établissant des probabilités à posteriori des maladies. Le modèle développé ici, basé sur l'AHP avec feedback permet de structurer les jugements du médecin et fournir des priorités globales aux maladies. Si des données statistiques sont disponibles alors les probabilités à posteriori obtenues par l'approche Bayésienne peuvent être utilisées dans ce modèle,

On a montré que, les mêmes priorités globales peuvent être obtenues par d'autres priorités autres que les probabilités à posteriori si elles vérifient une certaine condition, ce qui rend le modèle exploitable dans des situations de manque de données statistiques.

Notre application sur un cas pratique a permis au médecin de voir en détail l'impact des maladies sur les symptômes et l'impact des symptômes sur les maladies suspectés, ainsi que la sensibilité des priorités de ces derniers par rapport à la dépendance entre les symptômes.

On propose une perspective d'étude de cette sensibilité en détail et un essai d'exprimer cette variation en fonction des priorités globales des maladies. Il convient d'appliquer le modèle en utilisant des probabilités à priori des maladies suspectées et des données statistiques pour analyser davantage le rôle du théorème de Bayes dans notre modèle.

Annexe

L'échelle fondamentale [SAT] :

Au lieu d'attribuer deux valeurs w_i et w_j à deux alternatives i et j et considérer le rapport w_i / w_j pour mesurer l'importance de i par rapport à j , on utilise simplement un nombre de l'échelle fondamentale de 1 à 9, c'est une approximation entière de w_i / w_j .

En 1846, Weber a trouvé que, si une personne prend deux poids dans ses deux mains, s'il peut distinguer entre 20g et 21g mais pas avec 20.5g, alors il pourra entre 40g et 42g et pas avec 41g, et ainsi de suite. Donc on doit ajouter à une quantité s , une quantité minimum Δs pour atteindre un seuil au-delà duquel nos sensations peuvent distinguer entre s et $s + \Delta s$. Ce seuil est appelé la différence juste notable (The just noticeable difference JND).

Le rapport $r = \Delta s / s$ ne dépend pas de s , et alors la loi de Weber énonce que la sensation est notée quand le stimulus est changé d'une proportion constante du stimulus lui-même. A condition que le stimulus ne change pas très fortement.

En 1860, Fechner a considéré une séquence de stimulus juste notable croissante,

Soit s_0 le premier stimulus :

$$s_1 = s_0 + \Delta s_0 = s_0 + \frac{\Delta s_0}{s_0} \cdot s_0 = s_0(1 + r) = s_0 \alpha$$

$$s_2 = \alpha \cdot s_1 = \alpha^2 \cdot s_0$$

$$s_n = \alpha^n \cdot s_0$$

donc les stimulus des différences notables croient d'une manière géométrique.

$$s_n = \alpha^n s_0 \Rightarrow n = \frac{\log s_n - \log s_0}{\log \alpha}$$

la sensation n est une fonction linéaire du logarithme du stimulus.

D'où la loi psychophysique de Weber-Fechner est donnée par : $M = a \cdot \log s + b$ ($a \neq 0$) où s est le stimulus et M la sensation.

Dans la comparaison par paires des alternatives, on cherche un rapport donc $b = 0$, d'où $s_0 = 1$ (cela signifie de prendre un stimulus s_0 unitaire)

On obtient : $s_1 = s_0 \alpha = \alpha$ et la sensation correspondante est $a \log \alpha$

$$s_2 = s_0 \alpha^2 = \alpha^2 \text{ et la sensation correspondante est } 2a \log \alpha$$

par conséquent :

$$M_0 = a \log s_0 = 0$$

$$M_1 = a \log \alpha$$

$$M_2 = 2a \log \alpha$$

...

$$M_n = n.a. \log \alpha$$

On remarque que le stimulus juste notable croit géométriquement, et les sensations correspondantes croient arithmétiquement.

On divisant M_i par M_1 , on obtient la suite 1,2,3,... de l'échelle fondamentale.

Une question qui se pose maintenant est : pourquoi l'échelle s'arrête à 9 ?

Pour une personne donnée, plus le nombre d'éléments à comparer est réduit, plus la comparaison est meilleure.

D'autre part, généralement, on peut diviser nos réactions en trois catégories : élevé, moyen et bas, on peut aussi diviser chacune d'eux en 3 sous catégories : élevé, moyen et bas d'où 9 catégories.

Bibliographie :

[ADM86] Adams I.D. 1986 : Computer-Aided Diagnosis of Acute Abdominal Pain ; A Multicentre Study. *British Medical J.* 293, 800-804

[ANT94] Anthony O' hagan Kendall' s advanced theory of statistics. Volume 2B Bayesian inference. professor of statistics, University of Nottingham, UK. Published in 1994 by Arnold, a member of the Hodder Headline group, 338 Euston Road, London NW1 3BH.

[ARL] : History of Decision Making, research department of Arlington Software Corporation.
http://benli.bcc.bilkent.edu.tr/~omer/downloads/dec_analy/history.html

[BOZ01] Bozóki 2001 : An Analysis of the Analytic Hierarchy Process, Sándor Bozóki. Eötvös University, Budapest (ELTE), Faculty of sciences, Department of Operations Research 2001.
http://www.cs.elte.hu/alkmatdiploma/index_eng.html

[DED72] DeDombal, F.T. 1972 : Computer-Aided Diagnosis Of Acute Abdominal Pain. *British Medical J.* 2, 9-13.

[DEG] Degoulet & Fieschi : Traitement de l' information médicale, Méthodes et applications hospitalières.
<http://www.cybermed.jussieu.fr/Broussais/InforMed/LIVRES/TraitInfo/Fic/Chapitre7/Chap7.html>

[FOR] Forman : How additional information can lead to inferior decisions – a paradox, Ernest H. Forman, George Washington University
http://www.decisionsciences.org/DecisionLine/Vol24/24_4/ed24_4.htm

[FOS] Forman & Selly : Decision By Objectives, (How to convince others that you are right). Ernest Forman, DSc. Professor of Management Science, George Washington University, - Mary Ann Selly, Expert Choice Inc.
<http://www.expertchoice.com/dbo/>

[GAM91] Gammerman & Thatcher 1991 : Bayesian Diagnostic probabilities without assuming independence of symptoms. *Methods Info. Medicine* 30, 15-22.

[GAN66] Gantmacher 1966 : Théorie des matrices (Tome I et II) F.R Gantmacher, Collection universitaire de mathématiques DUNOD Paris 1966.

[JAB] Jablonský : Simulation analysis of the interval Analytic Hierarchic Process, Josef Jablonský, University of Economics, Department of Econometrics, 130 67 Praha 3, Czech Republic.

[JON91] Jonson, N. E. G. 1991 : Everyday Diagnostics- A Critique of the Bayesian Model. *Medical Hypotheses* 34, 289-295.

[MAR] Martel : L' aide multicritère à la décision: méthodes et applications. JeanMarc Martel, Faculté des sciences de l' administration, Université Laval, Québec, Canada G1K 7P4

http://www.cors.ca/bulletin/v33n1_1f.pdf

[NET] Adresses Internet

Evaluation de l' intérêt diagnostique des informations médicales,
www.chups.jussieu.fr/poly/biostats/poly/index.html

Systèmes d'aide à l'action médicale
http://www.paris-ouest.univ-paris5.fr/pedagogie/bime/im_aide.htm

Nouvelles Méthodes de Traitement de l' Information Médicale
www.hbroussais.fr/Broussais/InforMed/InforSante/Volume5/Index5.html

[OUN99] OUNNAR F. 1999 : Prise en compte des aspects décision dans la modélisation par réseaux de petri des systèmes flexibles de production. Institut national polytechnique de Grenoble, Laboratoire d' automatique de Grenoble.

[PEW01] Pewsner & al 2001 : Sur la voie de l'intuition ? Théorème de Bayes et diagnostic en médecine générale, D. Pewsner, J. P. Bleuer, H. C. Bucher, M. Battaglia, P. Jüni, M. Egger. Forum Med Suisse Nr3, 17 Janvier 2001
http://www.medicalforum.ch/pdf_f/2001/2001-03/2001-03-293.pdf

[SAT77] Saaty T.L. 1977 : A Scaling Method for Priorities in Hierarchical Structures, Thomas L. Saaty. University of Pennsylvania, Wharton School, Philadelphia, Pennsylvania. Journal of mathematical psychology 15, 234 – 281 (1977)

[SAT86] Saaty T.L. 1986 : Axiomatic Foundation of the Analytic Hierarchy Process, Thomas L. Saaty. Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania 15260. Management science vol 32. N° 7, July 1986

[SAT90] Saaty T.L. 1990 : The Analytic Hierarchy Process, Thomas L. Saaty, University of Pittsburgh 1990.

[SAT96] Saaty T.L. 1996 : Decision Making with Dependence and feedback, The Analytic Network Process Thomas Saaty, RWS Publications 4922 Ellsworth Avenue, Pittsburgh, PA 15213 USA. First Edition 1996.

[SAT98] Saaty T.L. & Vargas L. 1998 : Diagnosis with dependent symptoms : Bayes theorem and the Analytic Hierarchy Process, Thomas L. Saaty, and Luis G. Vargas. University of Pittsburgh, Pittsburgh, Pennsylvania, Operations research Vol 46, 491 – 502.

[SAT99] Saaty T.L. 1999 : Fundamentals of the Analytic Network Process, Thomas L. Saaty. 322 Mervis Hall, University of Pittsburgh, Pittsburgh, PA 15260, USA, ISAHP 1999

[SAT01] Saaty T.L. Decision-Making with the AHP: Why is The Principal Eigenvector Necessary. ISAHP 2001, Berne, Switzerland, August 2-4, 2001.

[SAT] Saaty T.J. : The seven pillars of the Analytic Hierarchy Process, Thomas L. Saaty, 322 Mervis Hall, University of Pittsburgh, Pittsburgh, PA 15260 USA.
http://www.creativedecisions.net/papers/papers_etc/SevenPillars.doc

[SPI84] Spiegelhalter, DJ. & R.P. Knill-Jones 1984 : Statistical and Knowledge-based approaches to clinical decision support systems with an Application in Gastro-Enterology. *J. Royal Statis. Soc., Series A*, 147, 35-77

[TAJ01] TAJI & Sagayama : A group Analytic Network Process (ANP) for Incomplete Information, Kouichi TAJI & Yousuke Sagayama Graduate School of Engineering Science, OSAKA UNIVERSITY, 13 Machikaneyama, Toyonaka, Osaka 560-8531, JAPAN.
July 24, 2001
<http://www-analab.sys.es.osaka-u.ac.jp/%7Etaji/file/ganp-e.ps.gz>

[YAN02] Yannou & Limayem : Les méthodes de comparaison par paires, Bernard Yannou, Frej Limayem. Ecole Centrale Paris, Laboratoire Productique – Logistique 92295 Châtenay – Malabry. 12/04/2002
<http://www.afav.asso.fr/download/>