



THESE

présenté à

l'Ecole Nationale d'Ingénieurs de Sfax

en vue de l'obtention du

DOCTORAT

Dans la discipline *Informatique*
Ingénierie des Systèmes Informatiques

Par

Ali BEN MRAD

(Mastère en Informatique)

Observations Probabilistes dans les Réseaux Bayésiens

soutenu le 20 juin 2015, devant le jury composé de :

M. Afif MASMOUDI	Professeur à l'Université de Sfax, Tunisie	Président
M. Philippe LERAY	Professeur à l'Université de Nantes, France	Rapporteur
Mme. Nahla BEN AMOR	Professeur à l'Université de Tunis, Tunisie	Rapporteur
M. Mounir BEN AYED	Maître de Conférences HDR à l'Université de Sfax, Tunisie	Examineur
M. Pierre-Henri WUILLEMIN	Maître de Conférences à l'Université Pierre et Marie Curie, France	Examineur
M. Sylvain PIECHOWIAK	Professeur à l'Université de Valenciennes, France	Co-directeur
M. Mohamed ABID	Professeur à l'Université de Sfax, Tunisie	Co-directeur
Mme. Véronique DELCROIX	Maître de Conférences à l'Université de Valenciennes, France	Co-encadrant

Thèse de doctorat

**Pour obtenir le grade de Docteur de l'Université de
VALENCIENNES ET DU HAINAUT-CAMBRESIS**

et de l'Ecole Nationale d'Ingénieurs de Sfax
(mention informatique)

Présentée et soutenue par Ali BEN MRAD

Le 20/06/2015, à L'Ecole Nationale d'Ingénieurs de Sfax

Ecole doctorale :

Sciences Pour l'Ingénieur (SPI)

Equipe de recherche, Laboratoire :

Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines (LAMIH)

Observations Probabilistes dans les Réseaux Bayésiens

JURY

Président :

M. Afif MASMOUDI, Professeur à l'Université de Sfax, Tunisie

Rapporteurs :

M. Philippe LERAY, Professeur à l'Université de Nantes, France

Mme. Nahla BEN AMOR, Professeur à l'Université de Tunis, Tunisie

Examineurs :

M. Mounir BEN AYED, Maître de Conférences HDR à l'Université de Sfax, Tunisie

M. Pierre-Henri WUILLEMIN, Maître de Conférences à l'Université Pierre et Marie Curie, France

Co-directeur de thèse : M. Sylvain PIECHOWIAK, Professeur à l'Université de Valenciennes, France

Co-directeur de thèse : M. Mohamed ABID, Professeur à l'Université de Sfax, Tunisie

Co-encadrant :

Mme. Véronique DELCROIX, Maître de Conférences à l'Université de Valenciennes, France

Remerciements

C'est avec un grand plaisir que je réserve cette page en signe de gratitude et de reconnaissance à tous ceux qui ont assisté ce travail.

Je tiens à remercier, en premier lieu et tout particulièrement, mon encadrant Véronique Delcroix. Je ne remercierai jamais assez Véronique pour toutes les discussions intéressantes et fructueuses que nous avons eues, mais aussi, pour la patience dont elle a fait part à mon égard.

Je voudrais exprimer mes plus profonds remerciements à mes deux directeurs de thèse Sylvain Piechowiak et Mohamed Abid pour leurs investissements scientifique et humain, leurs précieux conseils et leurs encouragements.

Je remercie aussi Mohamed Amine Maalej pour sa gentillesse, son écoute et ses conseils.

Je remercie également Nahla Ben Amor, Philippe Leray, Afif Masmoudi, Mounir Ben Ayed et Pierre Henri Wullemmin qui m'ont fait l'honneur de juger ce travail.

Je salue tous les chercheurs que j'ai pu rencontrer lors des conférences, ceux-ci étant trop nombreux, je ne peux les citer tous.

Et bien sûr, ceux sans qui je ne serais rien : mes parents, mes sœurs, ma famille et tous mes amis qui me supportent et soutiennent depuis toujours.

Enfin, mes remerciements vont également à tous ceux qui ont participé plus ou moins indirectement au bon déroulement de ma thèse.

Table des matières

1	Introduction	1
1.1	Le raisonnement dans l'incertain	1
1.2	Présentation de la thèse	3
1.2.1	But de la thèse	4
1.2.2	Organisation de la thèse	4
2	Modèles graphiques probabilistes et observations incertaines : état des lieux	7
2.1	Les modèles graphiques probabilistes	8
2.1.1	Notions de probabilité	8
2.1.2	Les réseaux bayésiens	10
2.1.2.1	Définition	10
2.1.2.2	Exemple simple	11
2.1.2.3	Indépendance conditionnelle et d-séparation	12
2.1.3	Construction d'un réseau bayésien	14
2.1.3.1	Apprentissage de la structure	14
2.1.3.2	Apprentissage des paramètres	15
2.1.4	Inférence dans les réseaux bayésiens	15
2.1.4.1	Inférence exacte	16
2.1.4.2	Inférence approchée	17
2.1.5	Extensions des réseaux bayésiens	18
2.1.5.1	Réseaux bayésiens dynamiques	18
2.1.5.2	Réseaux bayésiens orientés objet	18
2.1.5.3	Réseaux bayésiens distribués	19
2.1.5.4	Diagramme d'influence	19

2.1.5.5	Réseaux bayésiens flous	20
2.2	Observations incertaines dans les réseaux bayésiens	20
2.2.1	Observation certaine	21
2.2.2	Observations incertaines	22
2.2.3	Propagation de l'observation de vraisemblance : méthode de Pearl	23
2.2.3.1	Présentation et exemple	23
2.2.3.2	Propagation de l'observation de vraisemblance dans les logiciels	24
2.2.4	"Soft evidence"	26
2.2.5	Règle de Jeffrey	27
2.2.6	Bilan sur les observations incertaines spécifiées par une distribution locale de probabilités	28
2.2.7	Influence de l'ordre de propagation d'observations incertaines spécifiées par une distribution locale de probabilités	29
2.3	Conclusion	30
3	Observations probabilistes dans les réseaux bayésiens	33
3.1	Observations incertaines dans les réseaux bayésiens : proposition d'une terminologie	34
3.2	Observations de vraisemblance	36
3.2.1	Présentation	36
3.2.2	Exemple d'observation de vraisemblance	38
3.2.3	Observation négative (ou disjonctive)	39
3.3	Observations probabilistes fixes ou non fixes	40
3.3.1	Présentation	40
3.3.2	Propriétés des observations probabilistes	42
3.4	Observations probabilistes non fixes	43
3.4.1	Exemples d'observations probabilistes non fixes	44
3.4.2	Propagation	45
3.5	Observations probabilistes fixes	47
3.5.1	Exemples d'observations probabilistes fixes	47
3.5.2	Propagation des observations probabilistes fixes	50

3.5.2.1	Les algorithmes de propagation d'observations probabilistes fixes (multiples)	51
3.5.2.2	Algorithme BN-IPFP-1	54
3.5.2.3	Trace de l'algorithme BN-IPFP-1 en propageant deux observations probabilistes fixes.	55
3.5.3	Extension de la définition d'observations probabilistes fixes . .	57
3.5.4	Propagation de l'observation probabiliste versus révision du modèle	59
3.5.4.1	Révision et propagation : différence entre observation ponctuelle et connaissance	59
3.5.4.2	Modifier la distribution de probabilité <i>a priori</i> ou propager des observations probabilistes fixes?	61
3.6	Synthèse des différents types d'observations incertaines	62
3.7	Conclusion	64
4	Applications des observations probabilistes fixes	67
4.1	Applications aux systèmes d'information géographique	67
4.2	Observations probabilistes dans les systèmes à base d'agents équipés d'un réseau bayésien local	70
4.2.1	Modèle à base d'agents équipés d'un réseau bayésien local . .	71
4.2.1.1	Concept et définition	71
4.2.1.2	Communication entre agents	72
4.2.2	Application à un problème de décision récurrent multicritères : l'aide au choix d'un vélo	72
4.2.2.1	Problèmes récurrents d'aide à la décision multicritères	73
4.2.2.2	Graphe des agents pour un système de recommandations pour le choix d'un vélo	75
4.2.2.3	Fonctionnement du système de recommandations à l'aide d'agents équipés d'un réseau bayésien local . .	76
4.3	Observations sur des variables continues dans un réseau bayésien discret	82
4.3.1	Principe	83
4.3.2	Problème du cancer de la prostate	84
4.3.3	Description du réseau bayésien	85
4.3.4	Méthodologie et tests	86

4.3.5	Autres algorithmes de classification pour le diagnostic du cancer de la prostate	89
4.3.5.1	Arbres de décision	89
4.3.5.2	K Plus Proches voisins	90
4.3.5.3	Machines à Vecteur de Support	90
4.3.6	Résultats	91
4.4	Conclusion	94
5	Conclusion générale et perspectives	95
5.1	Conclusion	95
5.2	Limites et Perspectives	97
	Bibliographie	101

Table des figures

2.1	(a) Graphe causal (b) Graphe non causal.	11
2.2	Exemple simple de réseau bayésien.	12
2.3	Connexion en série (a), divergente (b), ou convergente (c).	13
2.4	Observation de vraisemblance sur D	24
3.1	Observation de vraisemblance sur X	37
3.2	Traitement d'une observation de vraisemblance sur X (b) fournie par un système de reconnaissance de caractères à partir de l'image d'une lettre (a).	39
3.3	Exemple Asia ; observations probabilistes fixes et non fixes.	42
3.4	Observation probabiliste non fixe sur X	44
3.5	Observation probabiliste fixe sur X : l'observation d'une variable conti- nue est transformée à l'aide d'une discrétisation floue.	47
3.6	Fonctions d'appartenance de "Age".	48
3.7	Observation probabiliste fixe sur X : observation d'une sous-population.	48
3.8	Exemple Asia avec observations probabilistes fixes.	49
3.9	Observation probabiliste fixe sur X : un "agent" expert sur X envoie son évaluation de X à un autre agent.	50
3.10	Réseau Bayésien modélisant le cancer de la prostate.	55
3.11	Un réseau bayésien reçoit une observation probabiliste fixe sur un sous ensemble de variables (X_1, \dots, X_P)	58
4.1	Intégration conceptuelle de réseau bayésien (RB) et de système d'in- formation géographique (SIG).	68
4.2	Observations probabilistes fixes sur une zone géographique d'intérêt.	69

4.3	Graphe des agents d'un système de recommandations pour le choix d'un vélo.	75
4.4	Agent d'un critère de décision et agent vélo, avec leur réseau bayésien local et leurs liens avec les autres agents.	77
4.5	Une partie du graphe du réseau bayésien local des agents représentant (a) le cycliste et (b) l'utilisation du vélo.	79
4.6	Graphe du réseau bayésien local des agents associés (a) au critère de stabilité, (b) au critère de confort de chargement du vélo dans un véhicule ou en train.	80
4.7	Extrait du réseau bayésien local de l'agent "expert du vélo".	81
4.8	Réseau Bayésien modélisant le cancer de la prostate	86
4.9	Fonction d'appartenance associée à l'état flou "PSA en faveur d'un cancer de la prostate"	88

Liste des tableaux

2.1	Terminologie utilisée dans les logiciels dédiés aux réseaux bayésiens pour désigner les observations de vraisemblance (avril 2014).	25
3.1	Terminologie proposée sur les observations incertaines dans les réseaux bayésiens	35
3.2	Propagations des observations probabilistes non fixes : exemple Asia.	46
3.3	Algorithmes de propagation d'observations probabilistes fixes dans un réseau bayésien (soft evidential update).	52
3.4	Trace de l'algorithme BN-IPFP1 ; chaque case contient les valeurs correspondant aux deux états de la variable avant normalisation.	57
3.5	Algorithmes de révision d'un réseau bayésien à partir de distribution locale de probabilités.	61
3.6	Synthèse des propriétés des différents types d'observations dans un réseau bayésien.	64
4.1	Les trois réseaux bayésiens utilisés dans les tests (nombre d'intervalles de discrétisation et type d'observations)	87
4.2	Vecteur d'observation obtenu après discrétisation classique et floue pour deux valeurs observées de la variable <i>PSA</i>	89
4.3	Matrices de confusion obtenues en appliquant uniquement des observations certaines dans chaque réseau bayésien	92
4.4	Matrice de confusion obtenues avec le réseau bayésien RB1 en appliquant deux observations probabilistes (*)	92
4.5	Résultats de classification avec des observations classiques dans différents réseaux bayésiens	92
4.6	Pourcentage de bonne classification et test de concordance (Kappa)	93

Chapitre 1

Introduction

Ce mémoire de thèse en informatique se situe dans le domaine de l'intelligence artificielle. Le travail présenté concerne les modèles graphiques probabilistes, qui sont d'une part des modèles de représentation de la connaissance et de l'incertitude et d'autre part des modèles de raisonnement permettant de propager des observations. Nous commençons par introduire le problème du raisonnement incertain puis nous présentons la problématique de cette thèse et le plan de ce mémoire.

1.1 Le raisonnement dans l'incertain

L'une des caractéristiques de l'intelligence humaine est la capacité de raisonner avec une connaissance incertaine. Très souvent, il est impossible ou difficile d'obtenir une connaissance certaine. Dans la vie réelle, on se trouve fréquemment dans des situations de prise de décisions dans lesquelles on ne dispose que d'informations incomplètes ou incertaines. Un des défis de l'intelligence artificielle est de proposer des modèles de raisonnement permettant de représenter et de manipuler des informations incertaines. Dans cet objectif, les chercheurs ont développé des modèles et des méthodes durant les dernières décennies. Les principales sont la théorie des ensembles flous ([Zad65](#)), la théorie de l'évidence ([S⁺76](#)), la théorie des possibilités ([DP88](#)), la théorie des probabilités sur laquelle repose les réseaux bayésiens ([Pea88b](#)) et la théorie des probabilités imprécises ([Wal91](#)).

On peut diviser l'incertitude en deux grandes catégories : l'incertitude aléatoire et l'incertitude épistémique.

L'incertitude aléatoire, également appelée variabilité, incertitude objective ou incertitude de type A, provient du caractère aléatoire de l'information dû à une variabilité naturelle résultant de phénomènes stochastiques par exemple (1) le temps de défaillance d'un composant ou (2) la durée de réparation d'un composant ou (3) le $n^{\text{ème}}$ lancer d'une pièce ou (4) le choix au hasard d'un individu au sein d'une population connue. L'incertitude aléatoire est généralement considérée comme une propriété du phénomène représenté, telle que l'issue d'une expérience menée de nombreuses fois dans le passé. Elle peut ainsi être représentée par les réalisations d'une variable aléatoire dont la distribution de probabilités est connue.

L'incertitude épistémique, ou encore ignorance, incertitude subjective ou incertitude de type B, est liée au caractère imprécis de l'information dû à un manque de connaissance et qui résulte par exemple (5) d'erreurs lors de mesures ou (6) d'avis d'experts. Elle est une propriété de l'observateur, mesurant son degré de connaissance du phénomène représenté. Elle porte donc sur la caractérisation d'un paramètre ou d'un modèle inconnus, mais que l'on pourrait supposer déterminés.

Si l'observateur n'a pas de connaissance sur les expériences passées concernant des événements de même type que l'événement actuel, par exemple le temps de réparation d'un composant par un prestataire qu'il ne connaît pas, il va estimer ce temps avec son expérience personnelle, sa sensibilité, ses connaissances *a priori*, et la on peut parler d'incertitude subjective.

Par exemple, prenons le cas d'un capteur de température dont on ne connaît pas la précision, ni la justesse. Ce capteur donne une information sur laquelle pèse une incertitude épistémique. L'ignorance vient du fait que le modèle du capteur n'est pas connu, mais il pourrait l'être (au moins dans l'absolu). Si le modèle du capteur était connu, l'incertitude sur l'information donnée par le capteur serait de type aléatoire, par exemple résultant d'un grand nombre d'expériences passées avec ce type de capteur.

Ces deux types d'incertitude sont présents dans les réseaux bayésiens, au niveau du modèle des connaissances. Nous nous intéressons dans cette thèse aux cas où l'incertitude concerne les observations, qu'elle soit de type aléatoire ou épistémique.

1.2 Présentation de la thèse

Parmi les différents modèles de l'intelligence artificielle proposés pour représenter et manipuler des connaissances incertaines, la théorie des probabilités est la plus ancienne. Elle permet de représenter l'incertitude sur un ensemble de variables aléatoires. La combinaison du conditionnement en probabilité et d'une représentation graphique de l'ensemble des variables aléatoires donne lieu au concept de modèle graphique probabiliste, et en particulier au concept de réseau bayésien (Pea88b) lorsque le graphe est orienté.

Un réseau bayésien permet d'exprimer une distribution de probabilité sur un ensemble de variables aléatoires à l'aide d'un produit de distributions locales de probabilités. Ce passage du raisonnement probabiliste à un niveau local a des implications profondes aussi bien la conception du modèle que sur les méthodes de propagation de nouvelles informations dans le modèle.

Un réseau bayésien est un modèle de représentation des connaissances et de l'incertitude, ainsi qu'un modèle de raisonnement. Il fonctionne schématiquement comme suit :

- construction du modèle à l'aide de connaissances (données et / ou experts),
- utilisation du modèle pour le raisonnement : mise à jour des croyances sur les variables du modèle de façon à prendre en compte de nouvelles informations. Cette étape que l'on appelle inférence est basée sur l'arrivée de nouvelles informations, en général locales et ponctuelles, que l'on appelle observations.

Dans un réseau bayésien, une observation sur une variable signifie en général que cette variable est instanciée. Ce qui signifie que l'observateur peut affirmer avec certitude que la variable est dans l'état signalé.

Cette thèse porte sur d'autres types d'observations, souvent appelées observations incertaines, qui ne peuvent pas être représentées par la simple affectation de la variable.

De nombreux ouvrages sur les réseaux bayésiens éludent totalement la question des observations incertaines et il en est de même pour les logiciels sur les réseaux bayésiens. Lorsque la question des observations incertaines est traitée :

- les articles sont essentiellement centrés sur les méthodes de propagation et non pas sur les notions, définitions, caractéristiques et comparaison des différents types d'observations incertaines dans les réseaux bayésiens ;

- divers points concernant la définition précise de ces observations restent incomplets et donnent lieu à des confusions ;
- aucun vocabulaire concernant les observations incertaines dans les réseaux bayésiens n'est généralement accepté ni en français, ni en anglais, ce qui donne lieu à une multitude de termes dont certains sont utilisés avec des significations différentes.

1.2.1 But de la thèse

L'objectif de ce travail est de clarifier et d'étudier les différents concepts d'observation incertaines portant sur une variable dans un réseau bayésien et menant à réviser les croyances sur d'autres variables du réseau bayésien.

Il s'agit de proposer une terminologie et des définitions claires et cohérentes des différents types d'observations, mais aussi de fournir de nombreuses illustrations de chacun des concepts afin de faciliter leur appropriation, ainsi que leur intégration dans les logiciels sur les réseaux bayésiens. Cette thèse propose également différentes applications et utilisations des observations incertaines dans les réseaux bayésiens.

1.2.2 Organisation de la thèse

Le présent mémoire est organisé de la manière suivante :

Le chapitre 2 illustre les différentes définitions et notations liées aux modèles graphiques probabilistes, qui seront utilisées tout au long du présent mémoire. La suite de ce chapitre est dédiée à un état de l'art sur les observations incertaines dans les réseaux bayésiens dans la littérature et au niveau de l'offre logicielle sur les réseaux bayésiens.

Le chapitre 3 permet de répondre aux limitations de l'état de l'art existant sur les observations incertaines décrit dans le second chapitre.

Ce chapitre identifie trois types d'observations incertaines dans les réseaux bayésiens : les observations de vraisemblance, les observations probabilistes fixes et non fixes. Ce chapitre donne les définitions, les propriétés et de nombreux exemples, et présente les algorithmes de propagation.

Le chapitre 4 présente trois applications des observations probabilistes fixes. La première application concerne l'intégration des réseaux bayésiens avec des systèmes

d'information géographique (SIG). La deuxième application des observations probabilistes fixes concerne la communication entre agents dans un système à base d'agents munis d'un réseau bayésien local. La troisième application, s'intéresse à la propagation d'une observation sur une variable continue dans un réseau bayésien discret.

Chapitre 2

Modèles graphiques probabilistes et observations incertaines : état des lieux

Ce chapitre est consacré à la présentation de la problématique de notre travail et à l'étude des différents enjeux qui ont conduit à l'élaboration de cette thèse. Il est basé sur un état des lieux concernant les observations incertaines dans les réseaux bayésiens. Les applications qui relèvent des réseaux bayésiens intègrent un certain degré d'incertitude. Ces modèles graphiques probabilistes permettent de stocker de la connaissance sur un ensemble de variables et de mettre à jour l'état des croyances sur certaines variables cibles, en fonction des informations disponibles sur d'autres variables du modèle. L'état des croyances à un moment donné est exprimé sous la forme de distributions de probabilités *a posteriori*, obtenues par la propagation des observations sur certaines variables du modèle grâce à des algorithmes d'inférence. La connaissance enregistrée dans un réseau bayésien inclut différentes formes d'incertitude. En revanche, les observations sont le plus souvent considérées comme certaines. Cependant, plusieurs questions se posent : Est-il possible de prendre en compte des observations "incertaines" dans un réseau bayésien ? Qu'est-ce que cela signifie ? Comment ces observations sont-elles spécifiées ? Comment sont-elles propagées ?

Dans ce chapitre, nous présentons les modèles graphiques probabilistes, puis nous faisons le point sur ces questions à l'aide d'une revue de la littérature sur les observations incertaines dans les réseaux bayésiens.

Grâce à cet état de l'art sur les observations incertaines, nous verrons que le vocabulaire, tout comme les définitions et les concepts relatifs aux observations incertaines dans les réseaux bayésiens présentent des lacunes, des contradictions et des confusions.

2.1 Les modèles graphiques probabilistes

Les modèles graphiques probabilistes sont des outils de représentation des connaissances et du raisonnement sous incertitude. Ces modèles, très utilisés en pratique, sont dotés d'un mode de raisonnement à la fois simple et puissant. Ils permettent de modéliser des systèmes¹ simples ou complexes. Parmi les modèles graphiques probabilistes, on peut distinguer les modèles orientés et les modèles non orientés. Parmi les modèles graphiques probabilistes orientés, le réseau bayésien est sans doute le plus utilisé. La construction du modèle peut se faire manuellement ou par apprentissage à partir de données. Une fois le modèle construit, l'inférence consiste à propager dans le réseau bayésien un ensemble de valeurs dites observées pour obtenir la probabilité *a posteriori* des variables cibles. L'inférence permet de combiner à la fois la connaissance et les observations. Dans cette section nous présentons les réseaux bayésiens qui font l'objet d'une grande partie de nos contributions dans cette thèse.

La section suivante donne les éléments principaux concernant les probabilités et probabilités conditionnelles pour présenter ensuite les réseaux bayésiens.

2.1.1 Notions de probabilité

Dans cette partie, nous présentons les principales notions sur le calcul probabiliste qui servent au traitement de l'incertain.

Le triplet (Ω, Z, P) , appelé espace probabilisé, est formé de l'ensemble Ω appelé univers, d'une tribu Z sur Ω et d'une mesure de probabilité P sur Z dans $[0,1]$. Les

1. Le terme système désigne ici la partie du monde à laquelle on s'intéresse pour la modéliser en vue de répondre à un objectif précis. Le système peut représenter un patient, ses symptômes et sa maladie, ou bien un système de chauffage, ou encore l'ensemble des paramètres, attributs et variables qui interviennent dans un problème d'aide à la décision.

éléments de Z sont appelés événements. Le nombre $P(A)$ est appelé la probabilité de l'événement $A \in Z$.

Si deux événements A et B sont incompatibles ou mutuellement exclusifs, alors $P(A \cup B) = P(A) + P(B)$ et $P(A \cap B) = 0$. Par contre, si A et B sont deux événements quelconques, alors $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

On dit que deux événements A et B sont indépendants si et seulement si $P(A \cap B) = P(A).P(B)$.

Si B est une conséquence logique de A i.e. $A \subset B$ ou $B \Leftrightarrow A$, alors $P(A) \leq P(B)$ et $P(B \setminus A) = P(B) - P(A)$.

Si A et B sont deux événements quelconques avec $P(B) > 0$, la probabilité de A conditionnellement à B est noté $P(A|B)$ et vaut :

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

Dans le cas où les événements A et B sont indépendants et $P(B) > 0$, alors $P(A|B) = P(A)$. D'après la définition de la probabilité conditionnelle on a :

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

Et donc

$$P(A|B) = \frac{P(B \cap A)}{P(B)} \quad (2.1)$$

Le théorème de Bayes (1763) découle de la généralisation de l'équation 2.1 à des ensembles d'événements A et B :

$$P(A|B) = \frac{P(B|A).P(A)}{\sum_a P(B|A=a) * P(A=a)} \quad (2.2)$$

Dans la suite, on note $P(A, B)$ au lieu de $P(A \cap B)$.

2.1.2 Les réseaux bayésiens

2.1.2.1 Définition

La représentation des connaissances et le raisonnement à partir de ces représentations a donné naissance à de nombreux modèles. Les modèles graphiques probabilistes, et plus précisément les réseaux bayésiens, initiés par Judea Pearl dans les années 80, se sont révélés des outils très pratiques pour la représentation des connaissances incertaines. Les réseaux bayésiens figurent parmi les modèles d'analyse probabiliste. Ils proposent un formalisme mathématique et des bases théoriques solides pour la modélisation des systèmes complexes.

Un réseau bayésien est constitué de deux composantes :

- Une composante graphique qui consiste en un graphe orienté sans circuit. Les nœuds représentent les variables pertinentes du domaine et les arcs représentent les relations de dépendance entre les variables.
- Une composante numérique qui consiste en un ensemble de distributions de probabilités conditionnelles de chaque nœud dans le contexte de ses parents.

Ces modèles ont été présentés dans (Pea88b; LS88a; Jen96; Jor98; KP87).

Définition 2.1.1 (Réseau Bayésien). *Un réseau bayésien $B = (G, P)$ est défini par :*

- *un graphe dirigé sans circuit $G = (\mathbf{X}, E)$ où \mathbf{X} est l'ensemble des nœuds et E est l'ensemble des arcs,*
- *un espace probabilisé (Ω, P) ,*
- *un ensemble de variables aléatoires $\mathbf{X} = \{X_1, \dots, X_n\}$ associées aux nœuds du graphe et définies sur (Ω, P) telles que*

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i))$$

où $Pa(X_i)$ est l'ensemble des parents du nœud X_i dans G .

Dans ce qui suit, les lettres en majuscule sont utilisées pour représenter des variables aléatoires, et les lettres en minuscules représentent leurs valeurs. Les lettres en majuscules gras correspondent à des ensembles de variables. Voici quelques notations utilisées dans le reste du manuscrit : X désigne un nœud du réseau bayésien, X appartient à \mathbf{X} ayant ses états (ou valeurs) dans $D_X = \{x_1, \dots, x_m\}$, $P(\mathbf{X}) = P(X_1, \dots, X_n)$ représente la distribution de probabilité jointe qui définit un réseau

bayésien sur l'ensemble \mathbf{X} . $P(x)$ représente $P(X = x)$, $P(X)$ est la distribution de probabilité ($P(X = x_1), \dots, P(X = x_m)$). $R(X_i)$ et $R(X_j, X_k)$ sont des distributions locales de probabilités utilisés pour décrire les observations incertaines sur X_i et (X_j, X_k) .

La distribution de probabilités $P(X_i | Pa(X_i))$ représente la connaissance *a priori* sur la variable X_i et $P(\mathbf{X})$ représente la distribution de probabilités jointe sur l'ensemble des variables du réseau.

Parmi l'ensemble des réseaux bayésiens, les réseaux bayésiens causaux sont ceux dans lesquels les arcs vont des causes vers les conséquences.

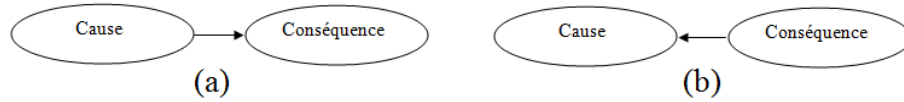


FIGURE 2.1 – (a) Graphe causal (b) Graphe non causal.

Dans les deux réseaux bayésiens de la figure 2.1, toute information sur le nœud "Cause" peut modifier les croyances que nous avons sur le nœud "Conséquence", et, réciproquement, toute information sur le nœud "Conséquence" peut modifier la connaissance que nous avons sur "Cause". Dans un graphe causal, le sens des flèche indique le sens de la causalité. D'une façon générale, dans un réseau bayésien, le sens des flèches n'indique pas le sens de la propagation et différents graphes peuvent être associés à la même distribution de probabilité.

La construction d'un réseau bayésien passe par trois étapes principales. La première est l'étape qualitative qui correspond à l'identification des variables et de leur domaine. La deuxième étape consiste à définir les arcs du graphe. Finalement, l'étape quantitative consiste à spécifier numériquement les distributions de probabilités conditionnelles pour chaque variable.

2.1.2.2 Exemple simple

Exemple 2.1.1 (Arroseur). La figure 2.2 présente un exemple classique dans la littérature sur les réseaux bayésiens. Il a été proposé par Pearl dans (Pea88a). On y voit les variables et la structure du graphe, ainsi que les distributions de probabilités conditionnelles associées à chaque nœud. Dans cet exemple, les nœuds J et W sont déterministes (probabilités 0 ou 1).

Ce matin-là, alors que le temps est clair et sec, M. Holmes sort de sa maison. Il s'aperçoit que la pelouse de son jardin est humide. Il se demande alors s'il a plu pendant la nuit, ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette un coup d'oeil sur la pelouse de son voisin, M. Watson, et s'aperçoit qu'elle est également humide. Il en déduit qu'il a probablement plu et décide de partir au travail sans vérifier son arroseur automatique. La représentation graphique du modèle causal utilisé par M. Holmes est composé de 4 nœuds : Arroseur (A), Pluie (P), Jardin humide (J), Jardin du voisin humide (W).

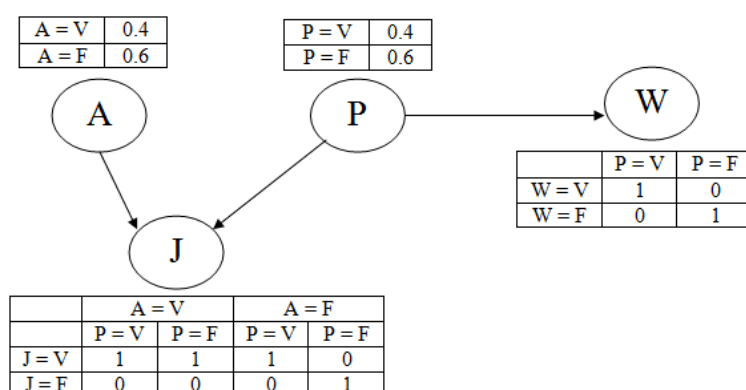


FIGURE 2.2 – Exemple simple de réseau bayésien.

2.1.2.3 Indépendance conditionnelle et d-séparation

Les connexions entre les nœuds définissent des lois de circulation de l'information dans le graphe. On distingue trois types de connexions :

- connexion en série (Figure 2.3 a) : l'information ne peut circuler entre X et Y que si la valeur de Z n'est pas connue, si non c'est directement la connaissance sur le nœud Z qui influe ;
- connexion divergente (Figure 2.3 b) : comme précédemment, l'information ne peut circuler entre X et Y que si la valeur de Z n'est pas connu ;
- connexion convergente ou connexion en V (Figure 2.3 c) : l'information ne peut circuler entre X et Y que si la valeur de Z est connue.

Ainsi, la circulation de l'information à l'intérieur d'un graphe dépend du type des connexions, plutôt que du sens des flèches.

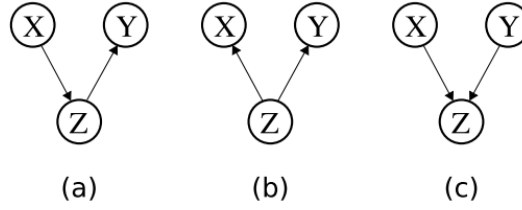


FIGURE 2.3 – Connexion en série (a), divergente (b), ou convergente (c).

La notion de *d-séparation* est essentielle pour le calcul des probabilités *a posteriori* car elle permet de définir l'indépendance conditionnelle entre certains nœuds. Soient X , Y et Z trois nœuds du graphe, on dit que X et Y sont *d-séparés* par Z (et on note $\langle X|Z|Y \rangle$) si, pour tous les chemins entre X et Y , l'une au moins des deux conditions suivantes est vérifiée :

- le chemin converge en un nœud W , tel que $W \neq Z$, et Z n'est pas un descendant de W .
- Le chemin passe par Z avec une connexion divergente ou en série.

Dans la figure 2.3 (a) et (b), X et Y sont *d-séparés* par Z , alors que dans la figure 2.3 (c) Z empêche la *d-séparation*. En d'autres termes, lorsque Z est connu (on dit observé), l'information circule entre X et Y dans les deux sens. La définition de *d-séparation* s'étend à des sous-ensemble de variables.

Le théorème fondamental des réseaux bayésiens est défini comme suit :

« si X et Y sont *d-séparés* par Z , alors X et Y sont indépendants sachant Z » :

$$\langle X|Z|Y \rangle \Rightarrow P(X|Y, Z) = P(X|Z)$$

Ainsi la structure du graphe d'un réseau bayésien encode un certain nombre d'indépendances conditionnelles. Cette propriété est exploitée dans les algorithmes d'inférence pour réduire la partie du graphe à considérer.

En revanche, la réciproque est fautive : certaines indépendances conditionnelles, présentées dans la distribution de probabilité, peuvent ne pas se traduire dans le graphe. Certains graphes sont "meilleurs" que d'autres pour une même distribution de probabilité jointe sur un ensemble de variables \mathbf{X} , car ils représentent plus d'indépendances conditionnelles. Par ailleurs, différents graphes peuvent encoder les

mêmes indépendances conditionnelles, puisque seules les V-structures se différencient des deux autres types de connexion.

2.1.3 Construction d'un réseau bayésien

La construction d'un réseau bayésien peut se faire selon les trois méthodes suivantes :

- manuelle : avec l'aide d'experts humains. Les spécialistes en ingénierie de la connaissance interrogent les experts et ajoutent les nœuds, les liens et les probabilités conditionnelles au réseau bayésien sur la base de la connaissance recueillie. Dans ce type de construction, il est plus fréquent de définir un graphe causal.
- automatique : par application d'un algorithme d'apprentissage à une base de données. Les algorithmes d'apprentissage peuvent identifier à la fois la structure du graphe et les paramètres (les distributions de probabilités conditionnelles). Les données peuvent être complètes ou incomplètes. On peut aussi avoir des variables non observables ;
- hybride : cette approche combine les deux précédentes. Les données peuvent être exploitées pour améliorer un premier modèle élaboré à partir des connaissances disponibles. Elles peuvent permettre d'élaborer ou de modifier la structure du graphe du réseau bayésien ou les probabilités conditionnelles sur les nœuds du graphe.

Les algorithmes d'apprentissage de réseau bayésien permettent d'exploiter des données pour proposer un modèle qui reflète le mieux les données utilisées. On distingue deux types d'apprentissage : l'un permet d'obtenir la structure du graphe associé au réseau bayésien et l'autre permet d'obtenir la distribution de probabilité.

2.1.3.1 Apprentissage de la structure

L'objectif de l'apprentissage de la structure est de trouver une structure du graphe à partir des données disponibles et qui représente le mieux un problème.

Une solution naïve pour trouver la meilleure structure d'un réseau bayésien, est de parcourir tous les graphes possibles, de leur associer un score, puis de choisir le graphe qui a le score le plus élevé. Cependant, le nombre de structures différentes pour un réseau bayésien de n nœuds est super exponentiel. Il est donc impossible

d'effectuer un parcours exhaustif en un temps raisonnable. Pour cette raison, la plupart des méthodes d'apprentissage de structure utilisent une heuristique de recherche dans l'espace des graphes acycliques dirigés.

Les méthodes d'apprentissage de structure qui se basent sur un calcul de score maximisent le score de la structure G qui décrit le mieux les données D . Par exemple,

$$\text{Score}(G, D) = P(G|D) = \frac{P(D|G).P(G)}{P(D)}$$

Plusieurs méthodes et critères de calculs des scores ont été présentés dans la littérature (NWL⁺07; Mar03; Jor98; FL03).

2.1.3.2 Apprentissage des paramètres

L'apprentissage des paramètres pour une structure particulière consiste à estimer les distributions de probabilités *a priori* ou les paramètres des lois de probabilités à partir des données disponibles. Dans le cas où l'on dispose de données complètes, on peut utiliser le maximum de vraisemblance qui utilise la fréquence d'apparition d'un événement dans les données (FGG97; GD04). Au contraire, si les données sont incomplètes, l'algorithme Expectation-Maximisation peut être utilisé (Fri98). Cet algorithme itératif part du modèle, infère le modèle pour calculer la distribution de probabilité. Puis, sur la base de cette distribution, l'algorithme construit un modèle meilleur. Le processus se répète jusqu'à obtenir le modèle le plus vraisemblable.

2.1.4 Inférence dans les réseaux bayésiens

L'inférence, ou la mise à jour des croyances, consiste à calculer la probabilité *a posteriori* au niveau de certaines variables en prenant en compte les informations sur d'autres variables appelées observations². Les réseaux bayésiens permettent de mettre à jour les distributions de probabilités au niveau des variables : on commence par fixer des observations, c'est à dire affecter des valeurs aux variables d'observation, puis on utilise un algorithme d'inférence pour calculer les probabilités *a posteriori* des autres variables.

2. en anglais, on utilise les termes observations ou findings, ainsi que le terme indénombrable "evidence" qui désigne un ensemble d'observations. En français on utilise : observation ou nouvelle information

Pour un réseau bayésien dont l'ensemble des nœuds est $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$, les algorithmes d'inférence permettent de calculer les probabilités *a posteriori* $P(X_i|\mathbf{Y})$ d'une variable X_i sachant les observations \mathbf{Y} où $\mathbf{Y} \subset \mathbf{X}$ et $X_i \notin \mathbf{Y}$.

Aucun algorithme d'inférence n'applique directement le théorème de Bayes car cela conduit à des calculs très coûteux en temps. Il existe une multitude de travaux dans le domaine de l'inférence dans les réseaux bayésiens. Nous en présentons ici un échantillon représentatif. Deux grandes familles d'algorithmes d'inférence se distinguent : les algorithmes d'inférence exacte et les algorithmes d'inférence approchée.

2.1.4.1 Inférence exacte

Parmi les algorithmes d'inférence exacte, on peut citer l'algorithme de propagation de messages, ainsi que les algorithmes de coupe ou de conditionnement qui exploitent la structure du réseau. Judea Pearl, le père de l'inférence dans les réseaux bayésiens, a développé un algorithme d'envoi de messages (message-passing) (Pea88c). Le principe consiste, pour chaque nœud du réseau, à mettre à jour la distribution de probabilité de proche en proche avec ses voisins pour prendre en compte les observations sur un ou plusieurs nœuds. Cet algorithme s'applique aux réseaux bayésiens dont le graphe est un arbre ou un poly-arbre. Grâce à un algorithme coupe-cycle, il est possible d'étendre cette méthode à n'importe quelle structure de réseau bayésien.

D'autres algorithmes utilisent le regroupement des nœuds (LS88b; JA13) pour se ramener à un arbre de façon à pouvoir appliquer ensuite l'algorithme de Pearl. Ces algorithmes se basent sur une modification du graphe. Une première étape, la moralisation, revient à remplacer les arcs par des arrêtes et à lier les parents communs de chaque nœud afin d'obtenir des cliques. Une deuxième étape, la triangulation, consiste à réduire tout cycle du graphe moral de plus de trois nœuds en cycles de trois nœuds maximum. Ceci permet d'obtenir une structure appelée arbre de jonction. Chaque nœud représente une clique du réseau bayésien ce qui permet d'appliquer l'algorithme initial de propagation des messages dans un arbre. Il reste à définir la distribution de probabilité associée à chaque clique : $\prod_{X_i \in H} P(X_i | \text{Parents}(X_i))$ où H est l'ensemble des nœuds de la clique qui n'appartiennent pas à ses cliques parentes.

L'inférence exacte est un problème NP complet. En pratique, la complexité des algorithmes d'inférence dépend de plusieurs facteurs dont notamment : le nombre de nœuds du réseau bayésien, la structure du réseau bayésien (nombre de chemins multiples entre deux nœuds), la taille des domaines des variables et le nombre de parents d'un nœud. Ce constat peut amener à utiliser des algorithmes d'inférence approchées.

2.1.4.2 Inférence approchée

Les algorithmes d'inférence approchée sont plus souvent utilisés dans les réseaux bayésiens complexes, avec plusieurs cycles. Ces méthodes donnent une estimation des probabilités *a posteriori* et sont non déterministes. Ils sont principalement de deux types (NWL⁺07) :

- les algorithmes d'inférence approchée du premier type utilisent des méthodes exactes mais opèrent seulement sur une partie du graphe. L'idée est d'inhiber l'effet des arcs jugés insignifiants ce qui permet d'éliminer des liens et de réduire la taille des tables de probabilités conditionnelles. Ceci accélère les calculs de façon significative avec des erreurs minimales sur les probabilités calculées (Kjæ94). D'autres algorithmes propagent partiellement les informations dans le graphe d'origine en ne considérant qu'une partie des tables de probabilités (HSC13). Dans (D'A93), les nœuds dont les probabilités ont les valeurs les plus importantes sont considérés comme des *observations* (probabilité égale 1). Cette méthode est inadaptée au cas des réseaux bayésiens dont certains nœuds ont une valeur extrêmement probable, car ils risqueraient d'être considérés à tort comme des observations.
- Les algorithmes d'inférence approchée du deuxième type utilisent des méthodes de simulation stochastique. Les plus connus sont les algorithmes de Monte-Carlo ou probabilistic logic sampling (Hen86), likelihood sampling (SP13) et backward sampling (FDF94). Ces algorithmes donnent une estimation des probabilités moyennant la génération d'échantillons d'instanciation du réseau. Le problème soulevé par ces méthodes est dû au nombre important d'instanciations nécessaires pour calculer une approximation des probabilités *a posteriori*. Pour des réseaux bayésiens de grande taille ou lorsque certaines variables ont

des valeurs avec une probabilité proche de 0 ou 1, ces méthodes requièrent un très grand nombre d'échantillons.

2.1.5 Extensions des réseaux bayésiens

Il existe plusieurs extensions des réseaux bayésiens. Ces extensions couvrent (de façon non exhaustives) :

- l'ajout de l'aspect temporel (réseaux bayésiens dynamiques),
- l'ajout de nœuds de décision et des nœuds d'utilité (diagrammes d'influence),
- la combinaison avec la notion d'objet (réseaux bayésiens orientés objets),
- la combinaison avec la théorie des ensembles flous (réseaux bayésiens flous),
- la division du réseau en plusieurs sous-réseaux qui seront gérés plus facilement localement (réseaux bayésiens distribués).

D'autres modèles ont été proposés récemment, comme les réseaux bayésiens relationnels, ou les MEBN, qui ne sont pas présentés ici. Nous présentons brièvement quelques uns de ces modèles dans la suite de ce mémoire. Tous ces modèles utilisent la notion d'observation et d'inférence.

2.1.5.1 Réseaux bayésiens dynamiques

Dans de nombreux problèmes réels, le temps peut être un paramètre qu'il est important de prendre en compte. Les réseaux bayésiens dynamiques sont une extension des réseaux bayésiens dans laquelle l'évolution temporelle des variables est représentée (DK89; Mur02). Lorsque les réseaux bayésiens dynamiques ne font intervenir que des variables discrètes, ils sont appelés modèles de Markov cachés : ils sont inspirés des chaînes de Markov cachés (Sah99). Il est possible de se référer à (Mur02) pour voir différents types de réseaux bayésiens dynamiques remarquables. Un grand nombre d'applications a utilisé ce type de modèles, par exemple en reconnaissance de la parole dans (BJ05). Des réseaux bayésiens partiellement dynamiques ont été proposés dans (TDW02) et contiennent à la fois des nœuds statiques et des nœuds dynamiques.

2.1.5.2 Réseaux bayésiens orientés objet

Lorsque, dans un réseau bayésien, une sous-structure apparaît de manière répétée, il est possible de le représenter à l'aide d'un réseau bayésien dit orienté

objet (KP97). Il existe plusieurs types d'implémentations pour ces modèles. Ces modèles sont particulièrement bien adaptés pour représenter les réseaux bayésiens dynamiques. Ils permettent également de modéliser des systèmes complexes pour lesquels le même mode de raisonnement apparaît dans différents sous-systèmes.

2.1.5.3 Réseaux bayésiens distribués

Pour maîtriser le problème de la complexité temporelle et de la complexité spatiale, une solution consiste à organiser le réseau bayésien en plusieurs réseaux de plus petite taille : on parle alors de *Réseaux Bayésiens Distribués*. Deux modèles ont été proposés : *Agent Encapsulated Bayesian Network* (AEBN) (Blo98) et *Multiply Sectioned Bayesian Networks* (MSBN) (XL00). Nous présentons le modèle AEBN en détail dans le chapitre 4, section 4.2.

Ainsi, on peut décomposer les tâches de modélisation et d'inférence en plusieurs sous problèmes.

2.1.5.4 Diagramme d'influence

Les réseaux bayésiens constituent des outils efficaces pour la prise de décision dans le sens où ils permettent d'évaluer différentes probabilités en fonction de l'état connu du système. Cependant, ils ne permettent pas de modéliser directement un système sur lequel l'opérateur peut agir en particulier dans le cas d'une séquence de décisions. Il est possible d'étendre leur formalisme en introduisant de nouveaux types de nœuds : les nœuds d'utilité et les nœuds de décision. Ils permettent d'associer une valeur d'utilité à chaque décision afin de choisir la meilleure d'entre elles. L'*utilité* représente la qualité ou encore le coût lié à ces décisions.

Les diagrammes d'influence (DI) sont des modèles graphiques dirigés qui permettent la prise de décisions séquentielles en situation d'incertitude (NJ09; JN13; KM12; Ter92).

Les diagrammes d'influence incluent trois types de nœuds : les nœuds de hasard correspondent aux variables du problème et sont munis d'une distribution de probabilités conditionnellement à l'ensemble des variables parents. Les nœuds de décisions, représentés par des rectangles, correspondent aux variables de décisions. Ils ne sont associés à aucune distribution de probabilités. Une flèche vers un nœud de décision signifie que la variable qui précède a été observée avant la décision. Le (ou

les) nœud d'utilité modélise les préférences du décideur. Il est muni d'une fonction d'utilité en fonction des variables parents.

2.1.5.5 Réseaux bayésiens flous

Différents travaux de recherche ont récemment permis l'utilisation de la théorie des ensembles flous dans le cadre des réseaux bayésiens. Différentes approches ont été proposées, mais à ce jour aucune ne fait consensus. L'approche adoptée par Baldwin et Di Tomaso (BDT03) est basée sur la discrétisation floue des variables continues. Fogelberg et al. (FPA08) ont cherché à généraliser les réseaux bayésiens aux réseaux bayésiens flous qui remplacent les états discrets par états flous dans tout le réseau. Ceci contribue au renforcement de la robustesse du système. Ils proposent d'étendre la propagation des croyances dans le réseau sur des états flous. Pan et Liu (PL00) et Tang et Liu (TL07) mettent également l'accent sur l'élaboration d'un cadre qui intègre des composants flous à l'intérieur d'un réseau bayésien.

Ces différentes extensions des réseaux bayésiens permettent de propager des observations afin de mettre à jour ses croyances. A ce titre, il est permis de poser la question des observations incertaines dans ces modèles, comme il en est question dans ce mémoire pour les réseaux bayésiens.

2.2 Observations incertaines dans les réseaux bayésiens

Quelle que soit la raison pour laquelle on utilise un réseau bayésien, il est toujours question d'inférence, c'est à dire qu'on utilise un réseau bayésien pour calculer une ou des probabilités *a posteriori*, sachant un ensemble d'observations (ou d'information). Après avoir présenté le cas des observations certaines, la suite de ce chapitre est une revue de la littérature en ce qui concerne les observations incertaines dans les réseaux bayésiens. Nous introduisons brièvement le concept d'observation de vraisemblance³ et nous présentons la méthode de Pearl appelée en anglais "Pearl's method of virtual evidence". Nous présentons ensuite le concept de "soft evidence" ainsi que la règle de Jeffrey, qui constituent deux approches dans lesquelles l'observation est spécifiée par

3. Nous reviendrons plus en détail sur la notion d'observation de vraisemblance dans le chapitre suivant, avec une présentation détaillée de cette notion et de ses propriétés

une distribution locale de probabilités. Enfin nous faisons un état de l'art d'une discussion existant dans la littérature concernant l'influence de l'ordre de propagation d'observations incertaines spécifiées par une distribution de probabilités. A partir de cet état de l'art sur les observations incertaines, nous verrons que le vocabulaire, tout comme les définitions et les concepts relatifs aux observations incertaines dans les réseaux bayésiens présentent des lacunes, des contradictions et des confusions.

2.2.1 Observation certaine

Un réseau bayésien constitue un modèle d'un "système" représenté par un ensemble de variables. Un observateur du système peut en général obtenir certaines informations sur l'état du système grâce au fait qu'il peut constater l'état ou la valeur de certaines variables du système ou les obtenir directement. Il est fréquent que toutes les variables du système ne soient pas observables. La valeur (ou l'état) d'une variable constatée par un observateur, à un instant donné, est appelée observation.

Nous utilisons ici le terme "observation certaine", par opposition à "observation incertaine", mais le simple terme "observation" est le plus couramment utilisé dans la littérature sur les réseaux bayésiens. Ce type d'observation est nommé *information déterministe* dans (NWL⁺07).

Définition 2.2.1 (Observation certaine). *Dans un réseau bayésien, une observation certaine e sur une variable X à valeur dans \mathcal{D}_X est définie par un vecteur d'observations de taille $m = |\mathcal{D}_X|$ contenant un unique 1, à la position correspondant à l'état $x \in \mathcal{D}_X$ et des 0 aux positions des autres états de X . Cette observation représente l'instanciation de X avec la valeur x et est caractérisée par $P(X = x \mid e) = 1$.*

Exemple 2.2.1 (Observation certaine). *Considérons un réseau bayésien contenant les variables A et S représentant l'âge et le sexe d'une personne, avec les ensembles de définition $\mathcal{D}_A = \{[0, 11], [12, 25], [26, 65], [66, 130]\}$ et $\mathcal{D}_S = \{M, F\}$. Le cas d'une jeune fille de 20 ans se traduit par les deux observations $A = [12, 25]$ et $S = F$ et par le vecteur d'observations $(0, 1, 0, 0)$ pour A et $(0, 1)$ pour S .*

Bien que les réseaux bayésiens soient un modèle de représentation et de raisonnement sous incertitude, les observations sont le plus souvent considérées comme certaines. Dans un réseau bayésien, les tables de probabilités conditionnelles reflètent l'incertitude sur la façon dont un ensemble de variables influence une autre variable

(nœud fils). En outre, la probabilité marginale d'une variable, $P(X)$, reflète l'état des croyances sur les valeurs de X lorsqu'aucune observation n'est rapportée.

Suite à l'arrivée d'une ou plusieurs informations certaines, le calcul des probabilités *a posteriori* permet de mettre à jour les croyances sur d'autres variables en fonction de ces observations. Ces nouvelles croyances, modélisées par $P(Y \mid obs)$, précisent l'incertitude sur les valeurs de Y sachant l'ensemble des valeurs observées pour les autres variables (*obs*).

La tâche de recueil de l'observation peut être soumise aux imprécisions et incertitudes que peuvent générer l'observateur ou les instruments de mesure. Il est par exemple possible que celui-ci rencontre des difficultés à déterminer si un condensateur présente un défaut naissant ou un défaut grave. L'information fournie par l'observateur sur le mode de fonctionnement d'un condensateur peut être imprécise. Il est également possible que l'observateur soit incertain quant à l'observation qu'il a fournie. La section suivante de ce chapitre rapporte la distinction dans la littérature de deux grande classes d'observations incertaines.

2.2.2 Observations incertaines

Beaucoup de problèmes du monde réel nous obligent à raisonner avec des informations incertaines. Il est intéressant de noter que l'information incertaine a été traitée dans un cadre de la révision des croyances, qui peut être fait dans un cadre probabiliste ou non. Par exemple, l'information incertaine a été étudiée dans un cadre très proche des réseaux bayésiens : celui des réseaux possibilistes dans (BT12). La mise à jour de croyances basées sur des informations incertaines peut être différente de la mise à jour de croyances basées sur des informations certaines. Voici les deux interprétations possibles du terme "information incertaine" d'après (DMP98) :

- L'incertitude concerne l'information elle-même, en raison, par exemple, du manque de fiabilité de la source qui fournit cette information (par exemple l'instrument de mesure).
- L'information est une description partielle d'une mesure de probabilité ; l'incertitude concerne une partie de l'information et elle est prise comme une contrainte sur l'état cognitif final. L'information est alors une correction de l'état cognitif *a priori*.

Dans la suite de ce chapitre, il apparaît clairement que les différents types d'observations incertaines décrits dans la littérature relèvent tous de ces deux grandes catégories.

2.2.3 Propagation de l'observation de vraisemblance : méthode de Pearl

2.2.3.1 Présentation et exemple

La première présentation de la notion d'*observation de vraisemblance* se trouve dans (Pea88c) : il y est question de "*uncertain evidence*", de "*virtual evidence*" et de "*likelihood evidence*". La traduction littérale de *likelihood evidence* qui donne *observation de vraisemblance* n'est pas utilisée dans la littérature en français sur les réseaux bayésiens, mais comme l'indique le titre de cette partie c'est cette traduction qui nous semble la plus appropriée. Nous verrons par la suite que le terme d'*observation incertaine*, utilisé par Pearl lui même, ne peut pas être utilisé pour désigner ce type d'observation car ces termes sont souvent employés dans la littérature pour désigner de façon générique différents types d'observations non certaines.

Les observations de vraisemblance concernent les cas où l'observation d'une variable *est* incertaine. Il existe une incertitude sur l'observation elle même, par exemple parce que l'information est fournie par une source qui n'est pas parfaitement fiable ou précise. Pour spécifier cette observation, il est nécessaire d'avoir une évaluation, au moins partielle, du lien entre les différentes valeurs réelles possibles de la variable et la valeur observée.

La méthode de Pearl pour propager une observation incertaine sur X consiste à ajouter un nœud "virtuel" O au réseau bayésien pour transformer l'observation incertaine sur X en une observation classique sur O . Le nouveau nœud est ajouté comme fils du nœud X . Il comporte deux états que l'on peut noter o et \bar{o} , où la valeur o désigne l'observation. Ainsi, X représente le nœud observé par le biais d'un observateur ou un capteur imprécis et O représente l'observation, cad le "biais" introduit par l'observateur.

Dans ce réseau bayésien augmenté, l'observation incertaine sur X se traduit simplement l'instanciation $O = o$, et peut être propagée à l'aide des algorithmes classiques d'inférence.

Exemple 2.2.2. Considérons un réseau bayésien contenant une variable D représentant la valeur d'un dé à six faces (voir figure 2.4).

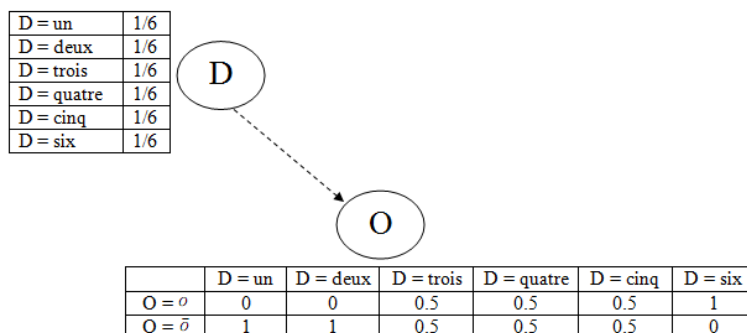


FIGURE 2.4 – Observation de vraisemblance sur D .

L'observateur est un jeune enfant qui confond encore certains chiffres, et il déclare "c'est un six". Connaissant le niveau d'acquisition des chiffres de l'enfant, on estime qu'il y a deux fois plus de chances que l'enfant déclare avoir vu un six si le dé est effectivement tombé sur la face numéro six que s'il est tombé sur la face trois, quatre ou cinq; et on estime que l'enfant n'affirmera pas cela si le dé est tombé sur la face un ou deux car il connaît ces chiffres. Cette information (déclaration de l'enfant + notre connaissance de l'enfant) indique entre autre que $P(O \mid D = \text{six}) = 2 \times P(O \mid D = \text{quatre})$. La figure 2.4 illustre l'ajout temporaire du nœud O pour prendre en compte cette observation incertaine sur la variable D . La table de probabilité conditionnelle associée au nœud est donnée par la vraisemblance de l'observation sachant X : $P(O = o \mid X)$.

Il apparait ainsi que la propagation d'une observation de vraisemblance ne soulève pas de réel problème technique. La compréhension précise de sa signification et sa spécification ne sont pas triviale. L'acceptation de cette notion d'observation incertaine par les utilisateurs passe par la disponibilité de cette fonctionnalité dans les logiciels sur les réseaux bayésiens. La section suivante illustre la propagation de l'observation de vraisemblance dans les logiciels.

2.2.3.2 Propagation de l'observation de vraisemblance dans les logiciels

La notion d'observation de vraisemblance, introduite par (Pea88c), correspond à une des deux notions présentées 10 ans plus tard par (DMP98). Cette notion

Logiciel	Termes utilisés pour désigner les observations de vraisemblance
Netica (Nor14)	likelihood finding
Hugin (Lau14)	likelihood evidence
Genie (Dru14)	virtual evidence
Infer.NET (MW14)	virtual evidence
gRain (Høj14)	virtual / likelihood evidence
BayesiaLab (JM14)	likelihoods (soft evidence in the user's manual)
BNT (Mur14)	soft evidence
Bayes Server (San14)	soft evidence
AgenaRisk (BFNT14)	soft evidence

TABLE 2.1 – Terminologie utilisée dans les logiciels dédiés aux réseaux bayésiens pour désigner les observations de vraisemblance (avril 2014).

d'observation incertaine n'est ni triviale ni intuitive à utiliser. En effet, l'observation de vraisemblance ne traduit pas directement une information du genre : "j'observe que $X = x_i$ avec une certitude de 0.95". Le rapport de vraisemblance utilisé pour spécifier une information avec incertitude correspond à une phrase de type : "il est r fois plus probable d'avoir cette observation (O) sur X si $X = x_i$ que si $X = x_j$, où r est le rapport entre $P(O|X = x_i)$ et $P(O|X = x_j)$ ". Or ce type d'expression n'est pas très naturel.

La propagation de l'observation de vraisemblance est proposée dans certains logiciels de réseaux bayésiens mais sous différentes appellations (voir table 2.1). En revanche, d'autres logiciels sur les réseaux bayésiens ne disposent d'aucune fonctionnalité concernant la propagation d'observation incertaine (par exemple Elvira(Elv14), Analytica(Hen14), Samlam(Dar14)).

La table 2.1 illustre le fait que le vocabulaire concernant l'observation de vraisemblance n'est pas normalisé et presque inexistant en français. La traduction littérale en français fournit le terme *observation virtuelle*, cependant, cette traduction n'est pas satisfaisante, car dans l'observation il n'y a rien de virtuel. En résumé, Pearl introduit dès 1988, une méthode de propagation qui permet de prendre en compte un premier type d'observation incertaine. L'accent est mis sur la méthode de propagation plutôt que sur le concept lui même.

On constate jusqu'à aujourd'hui une absence de vocabulaire spécifique pour désigner ce premier concept d'observation incertaine. Nous retenons en anglais les termes de "likelihood evidence" et "virtual evidence" et en français le terme d'observation de vraisemblance. La méthode de propagation des observations de vraisemblance est simple en terme d'implémentation et elle est présente dans de nombreux logiciels.

2.2.4 "Soft evidence"

Un second concept lié aux observations incertaines dans les réseaux bayésiens est celui de *soft evidence*. Il est apparu en 1998. Le terme *soft evidence* a été introduit dans le cadre des agents équipés d'un réseau bayésien local (Agent Encapsulated Bayesian Network, ou AEBN) (Blo98; VKV02). Aucun terme en français n'existe dans la littérature pour traduire l'expression en anglais, et à ce stade du manuscrit, nous utilisons le terme anglais. Ce concept sera présenté plus en détail dans le chapitre suivant, dans le cadre d'une présentation unifiée des différents types d'observations incertaines. Le concept sous jacent a été identifié mais non nommé dans la thèse de Bloemeke (Blo98). Puis il est repris et nommé "soft evidence" dans (VKV02). Il s'agit d'observations spécifiées par des distributions locales de probabilités qui définissent des contraintes sur la distribution de probabilité *a posteriori* et qui ne peuvent pas être améliorées par d'autres informations. Cette signification du terme *soft evidence* est utilisé par Valtorta (VKV02; KVV04; LV08; LVB10) et d'autres auteurs (BMDP⁺13; PPD06; PZP10; PDZP12; TB08; KN09).

Dans ces articles, plusieurs méthodes de propagation ont été proposées pour propager un ensemble d'observations incertaines de ce type.

En revanche, de nombreux articles utilisent le terme *soft evidence* pour désigner un autre concept. Dans (KM13; Kri01; Bil04; DTD00; CD04; Cha05), ce terme est employé pour désigner des observations de vraisemblance⁴. Dans un autre article (BF05), *soft evidence* signifie qu'une variable X ne peut pas prendre une valeur spécifique x , c'est-à-dire $X \neq x$.

Cette confusion sur le terme *soft evidence* est présente aussi dans les logiciels sur les réseaux bayésiens. Parmi les logiciels qui implémentent la méthode de Pearl pour la propagation des observations de vraisemblance, trois d'entre eux utilisent

4. L'usage des termes *soft evidence* est abandonné dans (CD05)

les termes *soft evidence* pour désigner des observations de vraisemblance (voir table 2.1).

En conclusion, le terme *soft evidence* porte à confusion du fait qu'il désigne différents types d'observations incertaines aussi bien dans la littérature que dans les logiciels. Plus généralement, le concept *soft evidence* reste peu connu et mal maîtrisé par les utilisateurs des réseaux bayésiens.

2.2.5 Règle de Jeffrey

Une autre méthode de propagation concernant la propagation des observations incertaines est issue de la règle de Jeffrey (Jef90). Cette règle concerne le cadre de la révision d'une distribution de probabilité jointe par une distribution locale de probabilités. Elle n'a pas été pensée dans le cadre Bayésien, et ne s'adapte pas facilement aux probabilités conditionnelles. La règle de Jeffrey est basée sur le principe de la "cinématique des probabilités" (probability kinematics). Elle peut être vue comme un principe pour minimiser les changements de croyances c'est à dire les changements sur les distributions de probabilités.

Considérons la révision d'une distribution de probabilité $P(\mathbf{X}) = P(X_1, \dots, X_n)$ par une distribution locale $R(\mathbf{Y}) = R(Y_1, \dots, Y_p)$ où $\mathbf{Y} \subset \mathbf{X}$ ($p < n$). On appelle Q la distribution de probabilité révisée $Q(\mathbf{X})$.

Voilà l'idée de la règle de Jeffrey : adoptons la distribution $R(\mathbf{Y})$ comme notre état des croyances finale sur \mathbf{Y} , donc R et P ne sont pas d'accord sur la distribution de \mathbf{Y} . Etendons la définition de R en une distribution jointe Q sur l'ensemble des variables \mathbf{X} de sorte que l'influence de \mathbf{Y} sur les autres variables soit identique dans Q et dans P : $\forall Z \in \mathbf{X} \setminus \{\mathbf{Y}\}, Q(Z \mid \mathbf{Y}) = P(Z \mid \mathbf{Y})$.

La règle de Jeffrey est la suivante : $Q(Z) = \sum_{\mathbf{y}} P(Z \mid \mathbf{Y} = \mathbf{y}) * R(\mathbf{Y} = \mathbf{y})$

La règle de Jeffrey ne concerne que les distributions jointes et non pas les distributions de probabilités conditionnelles. Ainsi, cette règle est applicable sur des réseaux bayésiens de petite taille, en considérant la révision de la distribution jointe de toutes les variables du réseau bayésien, mais elle n'est pas utilisable sur les réseaux bayésiens de grande taille.

Chan et Darwiche (CD05) ont montré que la propagation d'une observation de vraisemblance avec la méthode de Pearl fournit les mêmes résultats que la règle de Jeffrey après avoir converti l'observation spécifiée par une distribution locale de

probabilités en un rapport de vraisemblance. Cette équivalence des deux méthodes est présentée et expliquée dans la chapitre 3, ainsi que les raisons pour lesquelles cette transformation est limitée à une seule observation.

2.2.6 Bilan sur les observations incertaines spécifiées par une distribution locale de probabilités

Le terme observation incertaine est souvent employé de manière générique pour nommer les différents types d’observations non certaines (BMMD⁺11; BMDM⁺12; MDPL14; Vom04; CD05; PPD06; Dar09; PZP10). Ces articles concernent à la fois la définition de l’observation incertaine et sa propagation. Dans ces articles, on distingue clairement deux types d’observations incertaines dans les réseaux bayésiens qui correspondent aux deux types proposés dans (DMP98) et présenté dans la section 2.2.2.

Le premier type concerne l’observation de vraisemblance, initialement proposée par Pearl dans (Pea88c). Le deuxième type d’observation incertaine concerne les observations incertaines spécifiées par une distribution locale de probabilités. Ce type d’observation n’est cependant pas clairement défini.

Nous faisons le point sur la littérature concernant les observations incertaines spécifiées par une distribution locale de probabilités. D’un côté Chan et Darwiche, à la suite de Pearl, ont étudié comment la règle de Jeffrey pouvait être utilisée dans le cadre des réseaux bayésiens pour propager une information spécifiée par une distribution de probabilité sur un sous ensemble de variables (CD05). Les auteurs étudient et comparent la spécification et la propagation des observations de vraisemblance avec les observations incertaines basées sur la règle de Jeffrey, sans proposer de terme spécifique pour désigner ce type d’observation incertaine.

Par ailleurs, Valtorta et son équipe, et à sa suite Peng, ont travaillé sur le concept de *soft evidence*, et en particulier sur les algorithmes de propagation (VKV02; Vom04; KVV04; PPD06; LV08; LVB10; PZP10; PDZP12). Dans leurs travaux, la distinction entre *soft evidence* et observation incertaine basée sur la règle de Jeffrey n’est pas clairement établie, bien que la règle de Jeffrey y soit présentée.

Actuellement, il n’y a pas de terminologie ou définition proposés afin de considérer une distribution locale de probabilités comme un type particulier d’observation incertaine dans un réseau bayésien. Il apparait clairement qu’il est nécessaire de

définir les observations incertaines spécifiées par des distributions locales de probabilités et les propriétés attendues. La section suivante expose une divergence de points de vue entre différents auteurs dans le cas où plusieurs observations sont spécifiées par des distributions locales de probabilités. Concernant les logiciels des réseaux bayésiens, il n’y a que peu de produits proposant la propagation de l’observation incertaine spécifiée par une distribution locale de probabilités. Cela sera présenté dans le chapitre suivant.

2.2.7 Influence de l’ordre de propagation d’observations incertaines spécifiées par une distribution locale de probabilités

Les auteurs (PZP10) font état d’une divergence de point de vue concernant la question de la commutation suite à la propagation de plusieurs observations spécifiées par une distribution locale de probabilités dans un réseau bayésien. Dans (CD05), les auteurs examinent la question de la commutation après la propagation de plusieurs observations incertaines dans un réseau bayésien. Avec la méthode de Pearl, il y a commutation : plusieurs observations de vraisemblance peuvent être propagées successivement sans que l’ordre de leur propagation modifie le résultat sur l’état final des croyances (la distribution de probabilité *a posteriori*). En revanche, ce n’est pas le cas avec la règle de Jeffrey : lorsqu’il y a plusieurs observations spécifiées par une distribution locale de probabilités, la propagation de chacune conduit à des résultats différents suivant l’ordre de prise en compte des observations. En effet, les distributions locales de probabilités utilisées pour spécifier chaque observation ne sont pas toutes préservées par les propagations multiples. Certains auteurs (CD05) affirment que la propagation de plusieurs observations spécifiées ”tout bien considéré” (comme c’est le cas pour la règle de Jeffrey) n’est pas et ne doit pas être commutative. D’autres auteurs (PZP10) expliquent que lorsque plusieurs observations spécifiées par une distribution locale de probabilités représentent chacune une véritable observation de la distribution de certains événements, alors chacune de ces distributions locales de probabilités doit être préservée dans la distribution de probabilité *a posteriori* quel que soit l’ordre de prise en compte des observations. Dans le premier cas, une information sur une variable spécifiée par une distribution locale de probabilités est susceptible d’être modifiée par de nouvelles informations,

alors que dans le second cas, l'information probabiliste se comporte comme une observation certaine et n'est pas modifiée par de nouvelles informations. Il en résulte que selon les auteurs, les concepts sous jacents à la notion d'observation spécifiée par une distribution locale de probabilités sont différents, en particulier dans le cas d'observations multiples.

Cette revue de la littérature conduit au constat que plusieurs types d'observations incertaines ont été identifiés et définis, mais que ni la terminologie ni les concepts ne sont complètement et clairement définis dans la communauté des réseaux bayésiens.

2.3 Conclusion

Le début de ce chapitre a brièvement présenté les modèles graphiques probabilistes et en particulier les réseaux bayésiens. Tous ces modèles fonctionnent sur le même principe : une phase d'élaboration du modèle, puis une phase d'utilisation du modèle qui utilise l'inférence pour calculer des probabilités *a posteriori* à partir d'observations.

Nous avons étudié la notion d'observation comme déclencheur de la révision des croyances dans un réseau bayésien. Nous avons présenté le cas usuel des observations classiques où une ou plusieurs variables sont instanciées. Nous avons ensuite examiné le cas où l'on cherche à prendre en compte une information incertaine spécifiée à l'aide d'un rapport de vraisemblance et nous avons finalement exposé le cas des observations spécifiées par une distribution locale de probabilités.

Les observations de vraisemblance sont spécifiées par un rapport de vraisemblance et non par une distribution de probabilité. Elles sont connues depuis longtemps et de nombreux logiciels de réseau bayésien proposent la propagation de telles observations par la méthode de l'observation virtuelle de Pearl.

En revanche, les observations spécifiées par une distribution locale de probabilités restent mal connues des utilisateurs de réseau bayésien. En particulier, le concept de *soft evidence* n'est pas connu et les termes sont régulièrement employés dans une acceptation différente de celle proposée par Valtorta, et acceptée au moins par Vomlel, Langevin, Bloemeke, Peng, Pan, Zhang, Ding, Di Tomaso, Baldwin (Blo98; PDZP12; LVB10; TB08).

C'est sur la base de ces constats que nous présentons, dans le troisième chapitre, une présentation unifiée sur les différents types d'observations incertaines. Nous y proposons une distinction des observations incertaines spécifiées par une distribution locale de probabilités, selon que la distribution doit rester fixe ou non après l'arrivée de nouvelles informations. Cette présentation permettra de situer correctement et de comparer les concepts proposés par l'ensemble des auteurs qui fait référence dans ce domaine.

Définition[section] Exemple[section] Propriété[section]

Chapitre 3

Observations probabilistes dans les réseaux bayésiens

L'état des lieux du chapitre précédant sur les observations incertaines dans les réseaux bayésiens a permis d'établir les constats suivants :

1. Le concept classique d'observation (certaine) dans un réseau bayésien est largement partagé, mais il est désigné sous différentes appellations. Une observation d'une variable du réseau bayésien correspond à l'instanciation de la variable par la valeur ou l'état observé.
2. Une définition plus large du concept d'observation (*finding*) proposée dans (JN07) correspond au cas où le vecteur d'observation inclut plusieurs fois la valeur 1. Nous ne retenons pas cette définition du fait qu'elle correspond à une observation certaine lorsque le vecteur d'observation contient un seul 1, et à une observation incertaine sinon.
3. Le concept d'observation de vraisemblance (likelihood evidence ou virtual evidence) proposé par Pearl (Pea88c) est clairement identifié et reconnu par de nombreux auteurs, même s'il est parfois mal nommé. La propagation de ce type d'observation est disponible dans de nombreux logiciels sur les réseaux bayésiens.
4. Un autre type d'observations appelé *soft evidence* en anglais (sans équivalent en français) est identifié dans la littérature, mais il y a parfois confusion dans les termes. C'est un type d'observation incertaine spécifié par une distribution

locale de probabilités. Il y a confusion sur le terme *soft evidence* puisqu'il est parfois employé pour désigner des observations de vraisemblance.

5. Un dernier concept d'observation correspond à l'information considérée dans la règle de Jeffrey pour réviser une distribution de probabilités jointe à partir d'une distribution locale de probabilités. Ce concept est mal identifié dans le cadre des réseaux bayésiens et rarement distingué du concept de *soft evidence*, excepté dans (BT12), où les trois types d'observations incertaines sont identifiés dans le même article.

L'objectif de ce chapitre est de clarifier les termes et les concepts concernant les différents types d'observations incertaines dans un réseau bayésien. Dans ce chapitre, nous commençons par proposer une terminologie pour les différents types d'observations dans les réseaux bayésiens. Nous proposons le terme d'*observations probabilistes* pour désigner les observations incertaines spécifiées par une distribution locale de probabilités dans un réseau bayésien. Dans cette catégorie, nous identifions deux types d'observations incertaines : les observations probabilistes fixes et les observations probabilistes non fixes.

3.1 Observations incertaines dans les réseaux bayésiens : proposition d'une terminologie

Lorsqu'on observe un système qui peut être modélisé par un réseau bayésien, toute nouvelle information concernant certaines variables du réseau bayésien est susceptible de modifier les croyances sur les valeurs d'autres variables du réseau bayésien. Lorsque cette nouvelle information ne peut se traduire par l'instanciation de variables du réseau bayésien, bien qu'elle apporte une information, alors on parle d'observation incertaine.

Cette dénomination regroupe trois types d'observations qui sont présentés dans la table 3.1 : d'une part les cas où l'observation *est* incertaine, que l'on appelle observation de vraisemblance, et d'autre part les observations spécifiées par une distribution locale de probabilités, qui se décomposent en deux sous types.

Afin de remédier à l'absence d'une terminologie non ambiguë dans la théorie et la pratique des réseaux bayésiens, nous proposons l'utilisation du terme *observation probabiliste* pour désigner les observations incertaines spécifiées par une distribution

TABLE 3.1 – Terminologie proposée sur les observations incertaines dans les réseaux bayésiens

Nom	Observation certaine	Observations incertaines		
		Observation de vraisemblance	Observations probabilistes fixes ou non fixes	
Spécification	instanciation d'une variable	rapport de vraisemblance	distribution locale de probabilités	
Caractéristiques (selon (DMP98))	pas d'incertitude	l'observation est incertaine	définit une contrainte sur les croyances <i>a posteriori</i>	
La probabilité <i>a posteriori</i> peut être modifiée par d'autres informations ?	non	oui	oui ↓ observations probabilistes non fixes	non ↓ observations probabilistes fixes
Références	(Pea88c)	(Pea88c)	(CD05)	(VKV02) (PZP10)

locale de probabilités. Afin de faire une distinction claire entre les deux sous-types d'observation probabiliste, nous proposons les termes *observation probabiliste fixe* en référence aux *soft evidence* comme définis par Valtorta (VKV02; PZP10), et *observation probabiliste non fixe* pour désigner le concept utilisé dans la règle de Jeffrey et discuté dans (CD05; PZP10).

Le terme *probabiliste* désigne la manière dont l'observation est spécifiée. Les adjectifs *fixe* et *non fixe* désignent le comportement attendu de la distribution de probabilité *a posteriori* après qu'une observation supplémentaire soit obtenue.

La table 3.1 présente d'une part la terminologie proposée, mais aussi les principales caractéristiques qui découlent de l'état des lieux réalisé précédemment :

- Une observation certaine correspond à l'instanciation de la variable observée. L'observation étant certaine, la probabilité *a posteriori* sur la variable vaut 1 pour la valeur observée et aucune autre information ne peut modifier cela, pour l'état observé du système.
- Une observation de vraisemblance est spécifiée par un rapport de vraisemblance qui reflète l'incertitude sur l'observation. Les croyances *a posteriori* sur la variable observée peut être modifiée par d'autres informations.

- Les observations probabilistes sont spécifiées par une distribution locale de probabilités qui définit une contrainte sur les croyances *a posteriori*. Pour ce type d'observations incertaines, on définit deux sous-types selon que la distribution de probabilités *a posteriori* doit rester fixe ou non après l'arrivée de nouvelles informations. Les observations probabilistes fixes correspondent à la définition de *soft evidence* telle que proposée par Valtorta et al.

La suite de ce chapitre présente en détail chacun de ces types d'observation incertaine, ainsi que les méthodes de propagation.

3.2 Observations de vraisemblance

Dans le chapitre précédant, nous avons présenté la méthode de Pearl pour la propagation des observations de vraisemblance et l'équivalence entre la propagation de ce type d'observation par la méthode de Pearl et par la règle de Jeffrey. Dans cette section nous proposons une définition formelle et nous présentons les propriétés du concept de l'observation de vraisemblance.

3.2.1 Présentation

Les observations de vraisemblance (en anglais *likelihood evidence* ou *virtual evidence*) concernent les cas où l'observation d'une variable est incertaine. Il existe une incertitude sur l'observation elle-même, par exemple parce que l'information est fournie par une source qui n'est pas parfaitement fiable ni précise. Pour spécifier cette observation, il est nécessaire d'avoir une évaluation, au moins partielle, du lien entre les différentes valeurs possibles de la variable réelle et la valeur observée.

Les observations de vraisemblance permettent de prendre en compte l'incertitude qui porte sur l'observation lorsqu'elle est de nature épistémique : l'incertitude peut être considérée comme une propriété de l'observateur, ou de la source de l'information (capteur imparfait, observateur partiellement ignorant, conditions d'observation imparfaites, etc.). La spécification d'une observation de vraisemblance sur une variable est une mesure de la connaissance de l'observateur concernant cette variable.

Définition 3.2.1 (Observations de vraisemblance). *Une observation de vraisemblance sur une variable X d'un réseau bayésien est une observation sur laquelle porte une incertitude. Elle est spécifiée par un rapport de vraisemblance $L(X) =$*

$(L(X = x_1) : \dots : L(X = x_m))$, où les $L(X = x_i)$ sont des quantités relatives et le rapport de vraisemblance¹ est défini par :

$$L(X) = (P(\text{Obs} = \text{obs} \mid x_1) : \dots : P(\text{Obs} = \text{obs} \mid x_n))$$

$P(\text{Obs} = \text{obs} \mid x_i)$ est interprété comme la probabilité de l'évènement observé sachant que X est dans l'état x_i .

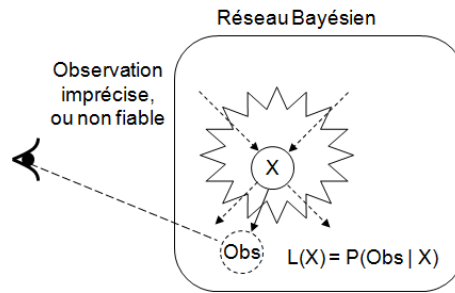


FIGURE 3.1 – Observation de vraisemblance sur X .

La figure 3.1 illustre une observation de vraisemblance sur une variable X d'un réseau bayésien. La variable X n'est pas observée directement, avec certitude. L'observation permet néanmoins d'apprendre quelque chose sur la valeur réelle de X . Cette information sur X apportée par l'observation est spécifiée par un rapport de vraisemblance de l'observation par rapport à chaque valeur de X .

Les deux propriétés suivantes décrivent comment l'observation de vraisemblance interagit avec les croyances avant et après sa propagation.

Propriété 3.2.1. *L'observation de vraisemblance est spécifiée indépendamment des croyances a priori ; par conséquent, la propagation d'une observation de vraisemblance tient compte de l'état des croyances sur la variable avant l'arrivée de l'observation.*

Propriété 3.2.2. *L'état des croyances sur une variable après la propagation d'une observation de vraisemblance sur cette variable n'est pas fixe. En effet, la probabilité a posteriori sur la variable observée peut être modifiée par de nouvelles observations sur d'autres variables.*

1. Il n'y a pas de condition sur la somme des termes d'un rapport de vraisemblance. Les deux points indiquent que seul le rapport entre les termes qui compte.

Notons Q_1 et Q_2 deux mesures de probabilité telles que $Q_1(\mathbf{X})$ représente les croyances sur l'ensemble \mathbf{X} après la propagation d'une observation de vraisemblance sur la variable $X_1 \in \mathbf{X}$ et $Q_2(\mathbf{X})$ représente les croyances sur l'ensemble \mathbf{X} après la propagation d'une deuxième observation (de n'importe quel type) sur la variable $X_2 \in \mathbf{X}$.

3.2.2 Exemple d'observation de vraisemblance

Une observation de vraisemblance concerne une ou plusieurs valeurs de la variable observée avec incertitude. L'exemple suivant illustre un cas d'observation de vraisemblance.

Exemple 3.2.1 (Reconnaissance de caractères). *Considérons un modèle où l'observation d'une variable X est fournie par un système de reconnaissance de caractères manuscrits. Ce système prend en entrée l'image d'un caractère et calcule un vecteur donnant pour chaque lettre de l'alphabet un indice de ressemblance de l'image avec cette lettre. Dans cet exemple (voir figure 3.2), le réseau bayésien contient une variable X qui représente la lettre que l'auteur a voulu tracer. Notons O ce que l'on voit sur l'image, mais O n'est pas une variable du réseau bayésien. Considérons le cas où l'image contient une lettre mal tracée qui peut faire penser à un 'v' ou à un 'u' (voir figure 3.2(a)) ; le système de reconnaissance de caractères fournit les indices 0.8 pour la lettre 'v', 0.4 pour la lettre 'u' et 0 pour les autres lettres. Ces indices, peuvent être interprétés comme $P(O = o \mid X = 'v') = 0.8$ et $P(O = o \mid X = 'u') = 0.4$, ce qui signifie qu'il y a deux fois plus de chances d'observer ce tracé si l'auteur a effectivement voulu tracer la lettre 'v' que si l'auteur avait voulu tracer la lettre 'u'. L'observation de vraisemblance sur X est spécifiée par le rapport de vraisemblance $L(X) = (0 : \dots : 0 : 0.4 : 0.8 : 0 : 0 : 0 : 0)$.*

Ce système de reconnaissance de caractères n'intègre pas de connaissance a priori sur la distribution des lettres dans la langue dont est issu le caractère. En revanche, cette connaissance est intégrée dans le réseau bayésien par la distribution de probabilités a priori $P(X)$. Autrement dit, l'information donnée par ce système de reconnaissance de caractères (le vecteur d'indices) doit être combinée avec l'information sur la fréquence des lettres dans la langue du texte pour mettre à jour les croyances sur les valeurs de X . Par ailleurs, une fois mise à jour les croyances sur la lettre représentée par l'image, ces croyances ne sont pas figées : elles peuvent encore être

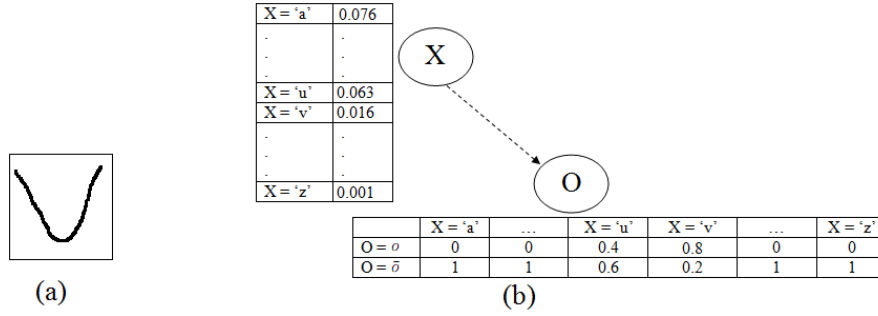


FIGURE 3.2 – Traitement d’une observation de vraisemblance sur X (b) fournie par un système de reconnaissance de caractères à partir de l’image d’une lettre (a).

modifiées par d’autres informations. Par exemple, une information concernant les caractères situés juste avant et après le caractère concerné peut être prise en compte et modifier les croyances sur X .

Le rapport de vraisemblance qui spécifie l’observation de vraisemblance est défini à une constante multiplicative près : $L'(X) = c * L(X)$. Ce rapport ne change rien au niveau du résultat de la propagation dans le réseau bayésien augmenté. En effet, comme on considère uniquement l’instanciation $O = o$, qui représente l’observation e , on a $P(O = o|e) = 1$, et de ce fait, la probabilité de cette observation est indifférente.

3.2.3 Observation négative (ou disjonctive)

Une observation de vraisemblance sur X est appelée observation *négative* ou observation *disjonctive* lorsqu’elle est spécifiée par un rapport de vraisemblance ne contenant que des 0 et des 1.

Ce type d’observation apporte les deux informations suivantes :

- les états correspondant à la valeur 0 sont des états exclus (et on ne sait rien dire de plus sur les autres états),
- la variable se trouve dans l’un ou l’autre des états correspondant aux valeurs 1.

Chacune de ces deux informations implique implicitement l’autre.

Exemple 3.2.2 (Observations disjonctive). *Considérons de nouveau l’exemple du dé à six faces. L’information ”le dé est tombé sur un nombre pair” se traduit par le*

vecteur d'observation $L(D) = (0 : 1 : 0 : 1 : 0 : 1)$ qui traduit le fait que le dé est tombé sur la face deux, quatre ou six, et que les faces un, trois et cinq sont exclues.

3.3 Observations probabilistes fixes ou non fixes

3.3.1 Présentation

Les observations probabilistes sont des observations spécifiées par une distribution locale de probabilités qui décrit l'état des croyances *a posteriori*. Ce type d'observation englobe d'une part la proposition de "soft evidence" de Valtorta (VKV02) et d'autre part la notion d'observation incertaine prise en compte par la règle de Jeffrey.

Il n'existe à l'heure actuelle aucune traduction en français du terme "soft evidence" et il est parfois utilisé à tort dans la littérature anglaise comme un synonyme de "likelihood evidence". Nous proposons le terme français "observation probabiliste (fixe ou non fixe)" qui est inspiré du vocabulaire employé dans le logiciel Bayesialab : "probability distribution (fixed or not)".

Le concept d'observation probabiliste (fixe ou non fixe) est parfaitement décrit dans (DMP98) : l'information est une description partielle d'une mesure de probabilité et constitue une contrainte sur l'état cognitif final. L'information vient donc corriger (remplacer) l'état cognitif initial. L'observation probabiliste comprend à la fois l'observation et l'état des croyances avant l'observation. On dit qu'une observation probabiliste est spécifiée "tout bien considéré", ce qui signifie que l'observation probabiliste prend en compte toutes les informations et connaissances disponibles sur la variable considérée.

Définition 3.3.1 (Observation probabiliste fixe ou non fixe). *Une observation probabiliste sur une variable $X \in \mathbf{X}$ dans un réseau bayésien est spécifiée par une distribution locale de probabilités $R(X)$ qui définit l'état des croyances sur la variable X après que cette information ait été propagée. Une observation probabiliste est fixe lorsque la distribution de probabilités $R(X)$ ne peut pas être modifiée par la propagation d'autres observations. Elle est dite non fixe dans le cas contraire.*

La différence entre les observations probabilistes fixes et non fixes apparaît lorsqu'on propage successivement plusieurs observations.

Exemple 3.3.1 (Asia). Considérons le réseau bayésien Asia ([LS88b](#)), qui contient huit nœuds binaires, dont un nœud racine S (Fumeur) et un nœud feuille D (Difficultés respiratoires). Prenons le cas de M. Flipo qui travaille dans l'usine FunT dont un récent sondage a révélé que la moitié de ses travailleurs souffre de difficultés respiratoires et que un sur dix est fumeur (voir Figure 3.3). Sans autres informations concernant M. Flipo, les distributions de probabilité $R(D) = (0.5, 0.5)$ et $R(S) = (0.1, 0.9)$ représentent les croyances a posteriori sur les variables D et S pour M. Flipo. Les croyances sur D et S avant l'arrivée de la nouvelle information, représentées par $P(D)$ et $P(S)$ sont remplacées par $R(D)$ et $R(S)$ qui décrivent l'état des croyances sur D et S "tout bien considéré". Étant donné que la propagation de $R(D)$ doit être conservée lors de la propagation de $R(S)$ (ou inversement suivant l'ordre de propagation). La première observation probabiliste, par exemple $R(S) = (0.1, 0.9)$, doit rester fixe durant la propagation de la deuxième observation. En revanche, si de nouvelles informations concernant M. Flipo sont reçues, comme une récente visite en Asie, elles peuvent modifier les croyances sur les variables S et D . Ainsi, $R(D)$ et $R(S)$ deviennent des observations probabilistes non fixes (l'observation probabiliste sur S n'est plus maintenue fixe).

Considérons maintenant le cas où la première information concernant M. Flipo est sa récente visite en Asie ($A = \text{vrai}$), et la deuxième information est que M. Flipo travaille dans l'usine FunT (dans lequel la moitié des travailleurs souffrent de difficultés respiratoires et un dixième d'entre eux sont des fumeurs). Les distributions de probabilité $R(D) = (0.5, 0.5)$ et $R(S) = (0.1, 0.9)$ ne sont pas données "tout bien considéré" puisqu'elles ne comprennent pas l'information initiale de la visite en Asie. Elles ne peuvent pas être utilisées pour spécifier des observations probabilistes.

Dans cet exemple, un ensemble d'observations probabilistes non fixes est déduit de la même information. Puisque chacune de ces observations définit une contrainte sur la distribution de probabilité a posteriori, elle doit être conservée fixe jusqu'à ce qu'elles soient toutes propagées. Ensuite, les observations probabilistes peuvent être à nouveau modifiées par l'apparition d'autres informations.

Nous avons défini une observation probabiliste sur une variable, mais cette définition peut-être étendue à un sous ensemble de variables. Dans ce cas, l'observation probabiliste peut être spécifiée par une distribution de probabilité jointe sur le sous ensemble de variable, ou par une distribution de probabilités conditionnelle sur le sous ensemble de variables, ou encore par d'autres expressions probabilistes sur le

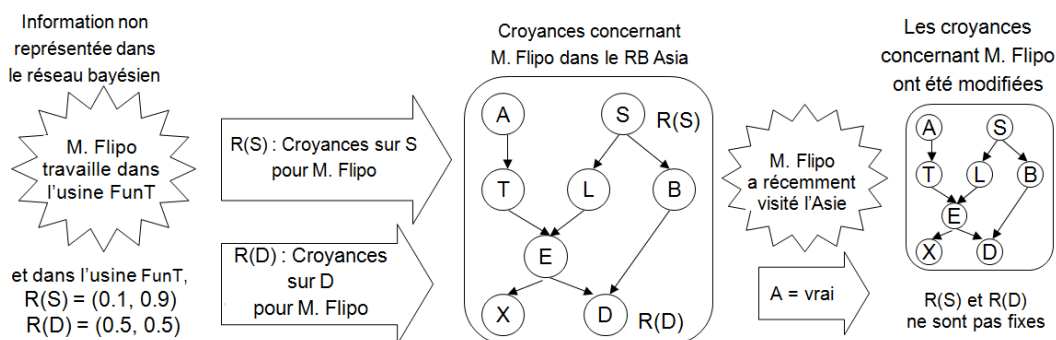


FIGURE 3.3 – Exemple Asia ; observations probabilistes fixes et non fixes.

sous ensemble de variables. Cette définition étendue a été proposée par Valtorta (VKV02) et nous la présentons plus bas dans la section 3.5.3.

3.3.2 Propriétés des observations probabilistes

Propager une observation probabiliste sur X nécessite de réviser la distribution de probabilités sur l'ensemble de variables \mathbf{X} par la distribution locale de probabilités $R(X)$. La difficulté réside dans le fait que la règle de Bayes ne peut être appliquée car $R(X)$ n'est pas un événement (PZP10). Une observation probabiliste $R(X)$ impose par sa nature une remise en cause de la distribution de probabilité jointe P car elle remplace la probabilité *a priori* de la variable X . La propagation de ce type d'observation nécessite de remplacer (temporairement) la distribution P par une distribution Q qui reflète les croyances sur les variables du modèle après avoir accepté l'observation probabiliste. La distribution Q doit vérifier $Q(X) = R(X)$ et réduire au maximum les changements de croyances sur les autres variables.

Propriété 3.3.1. Une observation probabiliste $R(X)$ sur une variable X remplace toutes les croyances *a priori* ou connaissance sur X . Par conséquent, la probabilité *a priori* $P(X)$ n'est pas utilisée dans la propagation de $R(X)$ et toute observation précédente sur X sera ignorée.

Propriété 3.3.2. La distribution locale de probabilités $R(X)$ qui spécifie une observation probabiliste sur une variable X est préservée. Après avoir propagé l'observation probabiliste sur X , les croyances sont représentées par une distribution de probabilité Q sur \mathbf{X} telle que $Q(X) = R(X)$.

Propriété 3.3.3. *Une observation probabiliste fixe sur X ne sera pas modifiée par d'autres observations sur d'autres variables du modèle. De plus, aucune autre observation sur X n'est possible, aussi longtemps que l'observation probabiliste fixe est valide. Lors de l'arrivée d'une nouvelle observation sur une autre variable après la propagation d'une observation probabiliste fixe sur X , il faut re-propager l'observation probabiliste fixe sur X en même temps que la nouvelle observation, afin de maintenir fixe la distribution de probabilité sur X . De ce fait, la propagation de plusieurs observations probabilistes fixes commute, c'est à dire que le résultat de la propagation est indépendant de l'ordre de l'arrivée des observations probabilistes fixes.*

Une observation probabiliste fixe se comporte comme une observation certaine dans le sens où l'observation reste inchangée après sa propagation, et aussi après l'arrivée d'autres informations sur d'autres variables.

Propriété 3.3.4. *Une observation probabiliste non fixe sur X peut être modifiée par d'autres observations sur une variable du réseau, y compris des observations de vraisemblance sur X . En conséquence, la propagation de plusieurs observations probabilistes non-fixes ne commute pas.*

3.4 Observations probabilistes non fixes

Une observation probabiliste non fixe sur une variable X est une observation incertaine spécifiée par une distribution locale de probabilités qui décrit l'état des croyances sur X "tout bien considéré". Les croyances sur X sont susceptibles d'être modifiées par de nouvelles informations.

Par exemple, lorsque le cas observé appartient à une sous-population dont les propriétés spécifiques sont connues et qu'elles ne sont pas modélisées dans le graphe. Dans l'exemple 3.3.1 ci-dessus, le réseau bayésien ne modélise pas le fait d'appartenir à la sous population des travailleurs de l'usine FunT, ni la façon dont cela influe les variables S et D . Cette information définit les nouvelles croyances sur S .

La figure 3.4 illustre de façon schématique une observation probabiliste non fixe sur une variable X d'un réseau bayésien. Le réseau bayésien ne modélise pas l'information concernant l'appartenance à une sous population spécifique, ni le comportement de cette sous population pour X . La nouvelle information inclut l'observation

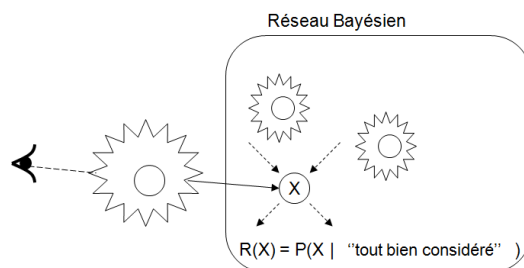


FIGURE 3.4 – Observation probabiliste non fixe sur X

d'une variable qui ne fait pas partie du réseau bayésien, ainsi que la connaissance de la façon dont cette observation influe sur X . L'observation probabiliste non fixe sur X est spécifiée par une distribution de probabilité $R(X)$ qui intègre l'influence de toutes les croyances *a priori*. Cette distribution locale de probabilités remplace la distribution de probabilité *a priori* $P(X)$ aussi longtemps que l'information est valide.

3.4.1 Exemples d'observations probabilistes non fixes

Exemple 3.4.1. (suite de l'exemple du dé : exemple 2.2.2). Le réseau bayésien contient une variable représentant le résultat d'un dé à jouer, mais ne contient pas de variable pour représenter la nature du dé. Dans ce cas, comment traduire l'information "il s'agit d'un dé truqué qui favorise l'apparition du 6". Cette information concerne l'observation d'une variable extérieure au modèle qui modifie nos croyances sur les valeurs du dé. Cette information peut être représentée par une observation probabiliste non fixe sur la variable D , à condition d'y associer une information supplémentaire sur la façon dont le dé truqué favorise l'apparition du six. Cette information se traduit par la distribution locale de probabilités $R(D) = P(D \mid \text{déTruqué} = \text{vrai})$. L'information que le dé est truqué peut être combinée avec une autre information sur X (observation de vraisemblance ou autre observation probabiliste non fixe). Une fois que l'on sait que le dé est truqué, les croyances sur la valeur du dé peuvent être modifiées par des informations sur d'autres variables, qui influencent D .

Exemple 3.4.2. (Reconnaissance de caractères, exemple 3.2.1 modifié) Le réseau bayésien contient une variable X qui représente un caractère, mais ne contient pas de variable pour représenter la langue du mot dont le caractère est extrait. Supposons

un cas où cette information est connue, ainsi que la fréquence des lettres dans cette langue. Cette information peut être introduite comme observation probabiliste non fixe sur la variable X , spécifiée par : $R(X) = (R(X = 'a'), R(X = 'b'), \dots, R(X = 'z'))$. Si rien d'autre n'influence les croyances sur X , alors $R(X)$ est donné par la fréquence des lettres dans cette langue et représente les croyances sur X sachant cette information. Si d'autres éléments influencent les croyances sur X , ils doivent être combinés avec la fréquence des lettres dans la langue du texte pour définir $R(X = x)$, de sorte que $R(X)$ représente les croyances sur X "tout bien considéré". Ainsi les croyances a priori sur X sont remplacées par $R(X)$. Puisque l'information sur X pourrait être améliorée par d'autres observations supplémentaires, alors c'est une observation probabiliste non fixe sur X .

3.4.2 Propagation

Une méthode pour propager des observations probabilistes non fixes consiste à convertir la distribution locale de probabilités en un rapport de vraisemblance, puis à utiliser la méthode de l'observation virtuelle de Pearl pour la propagation.

La distribution de probabilité $R(X)$ est convertie en un rapport de vraisemblance

$$L(X) = \frac{R(x_1)}{P(x_1)} : \dots : \frac{R(x_m)}{P(x_m)} \quad (3.1)$$

où $\mathcal{D}_X = \{x_1, \dots, x_m\}$.

Chan et Darwiche discutent cette méthode dans (CD05) et expliquent qu'il est normal que différentes révisions ne commutent pas lorsqu'on propage des observations incertaines spécifiées par une distributions locale de probabilités. Les observations incertaines dont il est question dans (CD05) sont donc des observations probabilistes non fixes. Lorsqu'on spécifie une information probabiliste non fixe sur une variable, on prend en compte le degré de certitude de l'observation (*strength of observation*), mais aussi les croyances sur les variables du modèle avant l'arrivée de l'information.

Rappelons que d'autres auteurs (VKV02; PZP10) estiment que les mises à jour successives après plusieurs observations probabilistes doivent être indépendantes de l'ordre des observations. C'est précisément cette exigence qui nous a conduit à distinguer deux types d'observations probabilistes : fixes et non fixes. Le fait d'imposer que l'observation probabiliste reste fixe même après la propagation de nouvelles

observations garantit que les croyances finales ne dépendent pas de l'ordre de propagation des observations.

Que se passe-t-il lors de la propagation de deux observations probabilistes non fixes ? Reprenons l'exemple Asia (voir exemple 3.3.1). La table 3.2 détaille les probabilités *a posteriori* des deux nœuds observés. La méthode de propagation est celle décrite ci-dessus : conversion des observations probabilistes non fixes en rapport de vraisemblance puis propagation avec la méthode de Pearl. Les lignes 2 et 4 montrent que dans le cas de la propagation d'une seule observation probabiliste, la probabilité *a posteriori* du nœud observé reste fixe. En revanche, les lignes 3 et 5 montrent que dans le cas de la propagation de deux observations probabilistes, seule la deuxième observation reste maintenue fixe.

Cet exemple illustre le fait que la méthode de propagation d'une observation probabiliste non fixe qui consiste à convertir la distribution locale de probabilités en un rapport de vraisemblance ne permet pas de maintenir fixe l'observation probabiliste dès lors que plusieurs observations doivent être considérées.

Valeur des deux Obs. probabilistes non fixes	$R(S = \text{oui}) = 0.4$	$R(D = \text{oui}) = 0.3$
	Croyances sur "S = oui"	Croyances sur "D = oui"
Croyances avant toute Obs.	0.5	0.436
Croyances après une seule Obs. probabiliste non fixe sur "S"	0.4	0.413
Croyances après une Obs. probabiliste non fixe sur "S" puis une Obs. probabiliste non fixe sur "D"	0.374	0.3
Croyances après une seule Obs. probabiliste non fixe sur "D"	0.468	0.3
Croyances après une Obs. probabiliste non fixe sur "D" puis une Obs. probabiliste non fixe sur "S"	0.4	0.286

TABLE 3.2 – Propagations des observations probabilistes non fixes : exemple Asia.

3.5 Observations probabilistes fixes

Une *observation probabiliste fixe* représente une information qui ne peut pas être remise en cause. De ce fait, elle se comporte comme une observation certaine. Une observation probabiliste fixe sur une variable X définit une contrainte sur la distribution de probabilité de X qui ne peut être modifiée par aucune autre information. Elle ne peut être combinée avec une autre information sur la même variable, elle remplace toute probabilité *a priori* sur X . Enfin, la propagation de plusieurs observations probabilistes fixes sur différentes variables du modèle aboutit au même résultat quel que soit l'ordre de leur propagation : il y a commutation. Ce dernier point constitue une différence essentielle entre les observations probabilistes fixes et non fixes. C'est le caractère "certain" des observations probabilistes fixes qui les différencie essentiellement des observations probabilistes non fixes. Au contraire, une observation probabiliste non fixe sur une variable peut être modifiée par l'arrivée d'autres informations.

3.5.1 Exemples d'observations probabilistes fixes

Les figures 3.5, 3.7 et 3.9 illustrent trois types de cas d'observations probabilistes fixes sur une variable X . Ces trois types d'exemples sont présentés ici de façon schématique avec un exemple illustratif rapide. Chacun de ces trois types d'application des observations probabilistes fixes est repris dans le chapitre 4 où il fait l'objet d'une présentation détaillée avec des applications réelles.

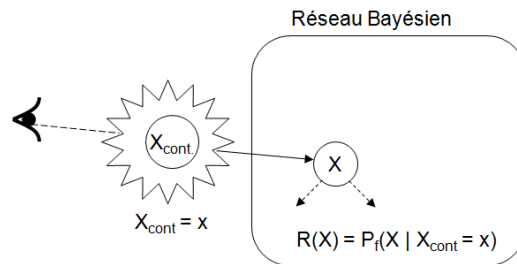


FIGURE 3.5 – Observation probabiliste fixe sur X : l'observation d'une variable continue est transformée à l'aide d'une discrétisation floue.

Dans le premier cas (figure 3.5), l'information vient de l'observation d'une variable continue ; la variable discrète X du réseau bayésien est associée à la variable

continue par une partition floue. L'observation probabiliste fixe sur X est spécifiée par une distribution de probabilités $R(X)$ qui projette la valeur observée sur la partition floue du domaine de définition de la variable X . L'exemple 3.5.1 illustre ce cas sur la variable "Age".

Exemple 3.5.1 (Observation probabiliste fixe : discrétisation floue). *Considérons un réseau bayésien contenant une variable A (âge), à valeur dans $\mathcal{D}_A = \{\text{enfant, adulte, personne âgée}\}$. Cette variable est associée à une variable continue "Age" qui ne fait pas partie du réseau bayésien. L'information "le sujet a 15 ans" peut se traduire par une observation probabiliste fixe sur la variable A de notre réseau bayésien, si on y ajoute un élément de connaissance concernant la fonction floue qui associe à chaque valeur réelle de l'âge un degré d'appartenance aux intervalles flous de \mathcal{D}_A (figure 3.6). On a par exemple $R(A) = (0.7, 0.3, 0)$. Cette observation probabiliste sur A est fixe car aucune autre observation ne doit modifier cette distribution.*

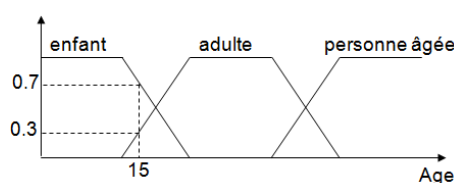


FIGURE 3.6 – Fonctions d'appartenance de "Age".

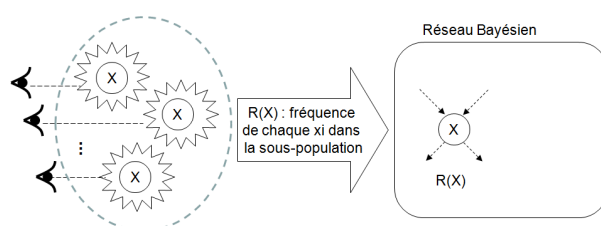


FIGURE 3.7 – Observation probabiliste fixe sur X : observation d'une sous-population.

Le deuxième type d'exemples d'observation probabiliste fixe concerne l'observation d'une sous-population, et non plus d'un seul cas (figure 3.7). L'information vient d'un ensemble d'observations d'une variable X , issu de l'observation de chaque individu de la sous-population. L'observation probabiliste fixe sur X est spécifiée par une distribution de probabilité $R(X)$ qui représente la fréquence d'apparition de chaque

valeur dans la sous-population observée. L'exemple 3.5.2 illustre ce type d'exemple sur le réseau bayésien Asia, où l'on considère les travailleurs d'une usine (au lieu d'un individu particulier).

Exemple 3.5.2 (Observations probabilistes fixes : observation d'une sous-population). *Considérons de nouveau le réseau bayésien Asia (voir figure 3.8). L'information concerne les travailleurs de l'usine FunT et indique que la moitié d'entre eux souffre de difficultés respiratoires, ce qui peut se traduire par une observation probabiliste fixe sur D : $R(D) = (0.5, 0.5)$. Une deuxième information indique que seul un dixième d'entre eux sont des fumeurs : $R(S) = (0.1, 0.9)$. La première information doit être préservée par la propagation de la deuxième information car chacune des observations probabilistes fixes doit rester inchangée. Si l'ensemble des couples d'observations (D, S) est disponible pour chaque individu de la sous-population, on peut considérer une seule observation probabiliste fixe $R(D, S)$ grâce à la définition étendue des observations probabilistes fixes (voir définition 3.5.1) : cette distribution locale de probabilités jointe sur D et S représente la variabilité du couple (D, S) parmi les travailleurs de l'usine observée.*

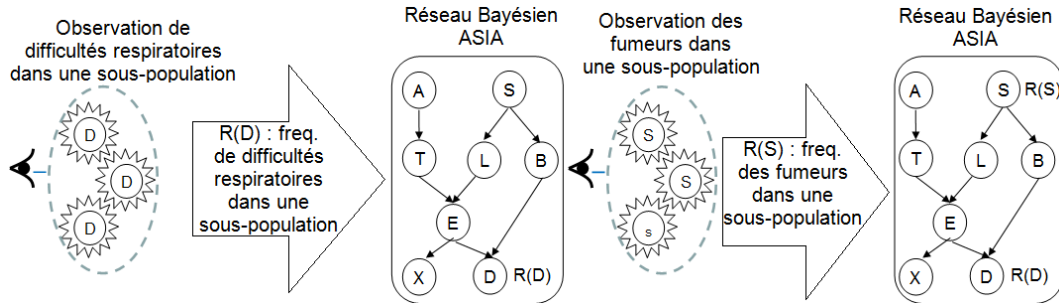


FIGURE 3.8 – Exemple Asia avec observations probabilistes fixes.

Le troisième type de cas d'observation probabiliste fixe se situe dans le cadre des réseaux bayésiens distribués initialement proposé par Bloemeke (Blo98). Il s'agit d'un modèle à base d'agents équipés d'un réseau bayésien local (Agent Encapsulated Bayesian Network (AEBN)). Dans ce cadre, on considère un ensemble "d'agents" dont chacun est muni d'un réseau bayésien local qui représente son modèle de raisonnement. Une observation probabiliste sur une variable X est une information qu'un agent expert (RB1) sur X envoie à un autre agent (RB2) qui la lui demande.

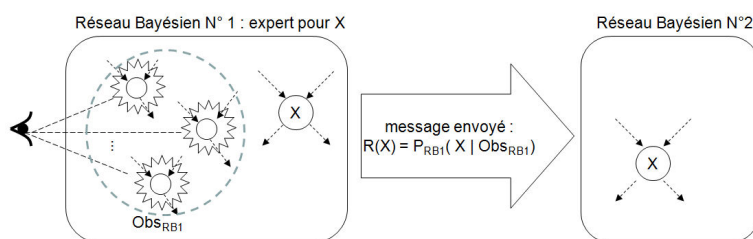


FIGURE 3.9 – Observation probabiliste fixe sur X : un "agent" expert sur X envoie son évaluation de X à un autre agent.

Le premier agent (RB1) évalue la valeur de X grâce à un ensemble d'observations locales et envoie son évaluation $R(X)$ au second agent (RB2). Ce dernier met à jour son réseau bayésien local (RB2) à partir de l'observation probabiliste fixe sur X spécifiée par $R(X)$ (figure 3.9). L'exemple 3.5.3 illustre cette utilisation des observations probabilistes fixes dans le cadre de l'aide au choix d'un fauteuil roulant (DSL13).

Exemple 3.5.3 (Observation probabiliste fixe : agents équipés d'un réseau bayésien local.). *Considérons deux agents, chacun doté d'un réseau bayésien local comme modèle de connaissance et de raisonnement. Le premier agent est spécialiste de l'évaluation des capacités motrices d'une personne handicapée. Le second agent a pour objectif de fournir des recommandations en vue du choix d'un fauteuil roulant (DSL13). Il fait appel au premier pour obtenir une évaluation de la capacité de marche (CM) de la personne qui est une variable partagée par ces deux agents. Le premier agent actualise les distributions de probabilité de la variable CM en fonction des informations recueillies sur la personne, et la transmet au second agent qui la reçoit, et la propage "chez lui" sans la modifier.*

3.5.2 Propagation des observations probabilistes fixes

Une observation probabiliste fixe impose que la distribution de probabilité qui la définit ne soit modifiée par aucune autre observation (aussi longtemps que l'observation est valide). Plusieurs observations probabilistes peuvent donc être propagées dans n'importe quel ordre sans que cela ne modifie le résultat final. En particulier, toutes les distributions locales de probabilités qui spécifient des observations probabilistes fixes doivent rester inchangées. Rappelons que la propagation d'une seule

observation probabiliste (fixe ou non fixe) peut être obtenue par sa transformation en une observation de vraisemblance comme indiqué dans l'équation 3.1 mais que cette méthode ne permet pas de maintenir fixes plusieurs observations probabilistes.

Cette section concerne la propagation de plusieurs observations probabilistes fixes. La synthèse ci-dessous sur les algorithmes de propagation des observations probabilistes fixes est inspirée de (PZP10; PDZP12).

3.5.2.1 Les algorithmes de propagation d'observations probabilistes fixes (multiples)

Plusieurs algorithmes ont récemment été proposés pour propager un ensemble d'observations probabilistes fixes dans un réseau bayésien. La plupart d'entre eux sont basés sur l'algorithme IPFP (*Iterative Proportional Fitting Procedure*) (Kru37; DS40). Cet algorithme itératif permet de réviser une distribution de probabilité jointe P à partir d'un ensemble de contraintes probabilistes locales tout en minimisant la distance de Kullback-Leibler (appelée aussi I-divergence) (KL51; Vom99) avec la distribution de probabilité P . Cet algorithme a fait l'objet de plusieurs adaptations pour des distributions de probabilités conditionnelles (VKV02; PPD06; PZP10). Cependant, l'algorithme IPFP fonctionne sur les distributions jointes complètes et, de ce fait, il n'est pas directement applicable pour la mise à jour des croyances dans les réseaux bayésiens. Cet algorithme ne peut être appliqué que pour des réseaux bayésiens de petite taille, car il nécessite de modifier l'entrée de la table de distribution de probabilité jointe à chaque itération.

La table 3.3 présente la liste des algorithmes qui ont été proposés pour propager un ensemble d'observations probabilistes fixes. L'algorithme *Big clique* (Blo98; VKV02) intègre les observations probabilistes en combinant les algorithmes *IPFP* et arbre de jonction. Lors de la construction de l'*arbre de jonction*, toutes les variables concernées par une observation probabiliste fixe sont entièrement reliées les unes aux autres par des liens supplémentaires non orientés. Après l'étape de triangulation, ces nœuds apparaissent dans une seule clique (*Big clique*). La mise à jour des croyances se fait en deux temps : la mise à jour de la distribution de probabilités de cette clique en exécutant *IPFP* jusqu'à la convergence, puis par la propagation de la distribution de probabilité résultant de cette clique au reste de l'arbre de jonction.

Les algorithmes *BN-IPFP1* et *BN-IPFP2* (PZP10) ne modifient pas l'arbre de jonction et peuvent fonctionner avec n'importe quel algorithme d'inférence de réseau bayésien. Ces deux algorithmes utilisent *IPFP* mais de manière différente.

L'algorithme *BN-IPFP-1* convertit d'abord toutes les observations probabilistes en observations de vraisemblance. Ensuite, il fait la mise à jour des croyances dans le réseau bayésien à l'aide de *IPFP* jusqu'à avoir une distribution qui satisfait toutes les observations probabilistes fixes propagées.

L'algorithme *BN-IPFP-2*, quant à lui, ajoute temporairement un nœud supplémentaire au réseau bayésien. Ses parents sont les nœuds qui ont une observation probabiliste fixe. La distribution locale sur ce nœud ajouté est mise à jour en utilisant *IPFP*. Puis elle est propagée simplement comme une unique observation probabiliste.

Algorithme	Réf.	Caractéristiques
Big-Clique	(VKV02) (Blo98)	variation de l'algorithme de l'arbre de jonction basé sur IPFP, avec recodage de l'arbre de jonction
BN-IPFP 1 et 2	(PPD06)	gère des observations non contradictoires, indépendant de l'algorithme d'inférence, combine IPFP et la conversion en observation de vraisemblance, (éq. 3.1)
BN-IPFP1	(PPD06)	convertit séparément chaque observation probabiliste, très coûteux si le réseau bayésien est grand.
BN-IPFP2	(PPD06)	utilise IPFP pour calculer $R(X_1, \dots, X_p)$ à partir de $R(X_1), \dots, R(X_p)$ puis propage en transformant en une observation de vraisemblance
Lazy big clique	(LV08)	modifie l'algorithme de propagation (MJ99)
SMOOTH	(ZPW08) (PZP10)	propagation d'observations probabilistes contradictoires (preuve de convergence pour deux obs.)

TABLE 3.3 – Algorithmes de propagation d'observations probabilistes fixes dans un réseau bayésien (soft evidential update).

L'algorithme *SMOOTH* a été développé en modifiant l'algorithme *IPFP* pour supporter la mise à jour des croyances sur un ensemble d'observations probabilistes fixes contradictoire. Dans la suite nous présentons les détails de l'algorithme *SMOOTH*.

Algorithme 1 SMOOTH

Entrée : un réseau bayésien (G, P) avec $G = (\mathbf{X}, \mathbf{E})$ et une distribution de probabilité $P(\mathbf{X})$,

m observations probabilistes,

$\alpha \in (0, 1)$ est le facteur smooth,

$\mathbf{R} = R(Y^1), \dots, R(Y^m)$.

Sortie : la distribution de probabilité *a posteriori* $Q(\mathbf{X})$

Phase 1 :

Appliquer l'algorithme IPFP standard en utilisant toutes les observations probabilistes dans \mathbf{R} jusqu'à convergence ou commencer un cycle ;

Si la convergence est atteinte, l'algorithme s'arrête.

Phase 2 :

1. Pour j de 1 à m

$R_0(y^j) \leftarrow R(y^j)$;

Fin Pour

2. $k \leftarrow 1$ [k : indice d'iterations]

3. Répéter

3.1. $j \leftarrow 1 + (k - 1) \bmod m$;

$l \leftarrow 1 + \lfloor (k - 1)/m \rfloor$;

3.2. $R_l(y^j) \leftarrow \alpha R_{l-1}(y^j) + (1 - \alpha) Q_{k-1}(y^j)$;

3.3.

$$Q_k(x) \leftarrow Q_{k-1}(x) \frac{R_l(y^j)}{Q_{k-1}(y^j)} ;$$

3.4. $k \leftarrow k + 1$

Jusqu'à (convergence)

L'algorithme SMOOTH se compose de deux phases. La première phase effectue l'algorithme IPFP standard en utilisant toutes les observations probabilistes fixes d'origine. Il s'arrête lorsqu'on obtient la convergence pour le cas des observations probabilistes fixes non contradictoires ou commence un cycle pour le cas des observations probabilistes fixes contradictoires.

La deuxième phase est exécutée uniquement lorsqu'un cycle est détecté à la fin de la phase 1. Cette deuxième phase diffère de la phase 1 par le fait que, à chaque itération, non seulement la distribution courante $Q_{k-1}(X)$ est modifiée par la contrainte choisie $R(Y^j)$, mais $R(Y^j)$ elle-même sera modifié par $Q_{k-1}(X)$.

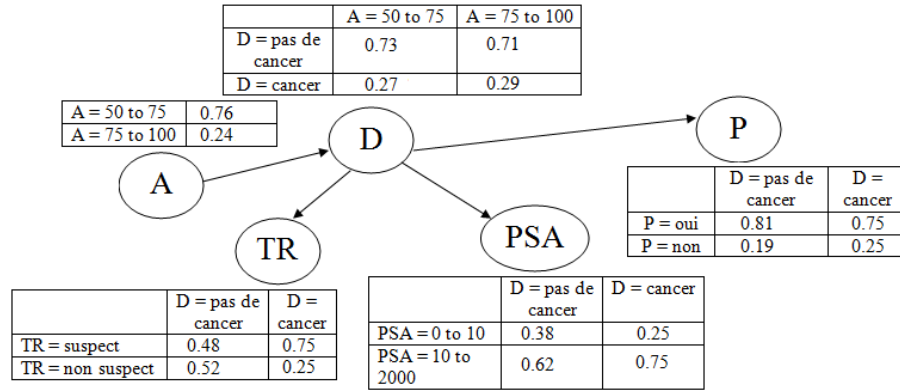
L'algorithme SMOOTH est similaire à l'algorithme IPFP, sauf que, dans la phase 2, il utilise des observations probabilistes modifiées plutôt que les observations d'origine pour mettre à jours Q_k . Cela rend SMOOTH directement applicable pour la mise à jour des croyances dans les réseaux bayésiens. Lorsque l'ensemble des observations probabilistes est non contradictoire, l'algorithme SMOOTH est similaire à l'algorithme IPFP standard, et il converge dans la phase 1.

La performance en termes de temps pour l'algorithme SMOOTH, comme pour tous les algorithmes basés sur IPFP, dépend du nombre d'itérations nécessaire pour atteindre la convergence. Les expériences fournis par (PZP10) montrent que l'algorithme SMOOTH atteint la convergence plus rapidement lorsque est très proche de 1.

3.5.2.2 Algorithme BN-IPFP-1

Nous présentons ici les détails de l'algorithme *BN-IPFP-1*. C'est l'algorithme que nous avons choisi d'implémenter pour nos tests, sur la base des résultats de la comparaison des trois algorithmes (*BN-IPFP-1*, *BN-IPFP-2* et *Big-clique*) (LV08; PZP10). L'un des avantages de l'algorithme *BN-IPFP-1* est qu'il peut fonctionner facilement avec n'importe quel algorithme d'inférence. Soit un réseau bayésien (G, P) et \mathbf{X} l'ensemble des variables associées aux nœuds du graphe G . L'algorithme *BN-IPFP-1* permet de propager un ensemble d'observations probabilistes fixes non contradictoires sur m variables $\{Y_1, \dots, Y_m\}$ inclus dans \mathbf{X} et spécifiées par $\{R(Y_f), f = 1, \dots, m\}$, tel que chaque distribution locale de probabilités est dominée² par P . L'algorithme BN-IPFP-1 est indépendant de l'algorithme d'inférence. Il combine

2. Une mesure R est *dominée* par une autre mesure P si pour tout événement A , $P(A) = 0$ implique $R(A) = 0$; on note alors $R \ll P$.



l'algorithme IPFP et la conversion des observations probabilistes en observations de vraisemblance. L'algorithme BN-IPFP-1 convertit séparément chaque observation probabiliste en une observation de vraisemblance. Ensuite, il met à jour les croyances à l'aide d'un algorithme de propagation quelconque dans le réseau bayésien jusqu'à avoir une distribution qui satisfait toutes les observations probabilistes propagées. A chaque itération, un nouveau rapport de vraisemblance est obtenu pour l'une des observations probabilistes fixes (à tour de rôle), en divisant la distribution de probabilité de l'observation probabiliste par la probabilité marginale sur cette variable obtenue à l'étape précédente. Ce nouveau rapport de vraisemblance est ensuite combiné avec tous les rapports de vraisemblance antérieurs sur la même variable qui sont obtenus dans les itérations précédentes (un pour m itérations).

La preuve de la convergence de l'algorithme BN-IPFP-1 (PZP10) est basée sur la convergence de IPFP.

3.5.2.3 Trace de l'algorithme BN-IPFP-1 en propageant deux observations probabilistes fixes.

Nous avons implémenté l'algorithme BN-IPFP-1 en utilisant "Bayes Net Toolbox" (BNT) en Matlab. La table 3.4 illustre la trace de l'algorithme BN-IPFP-1 pour le réseau bayésien "cancer de la prostate" présenté dans la figure 3.10. Ce réseau bayésien contient 5 noeuds : AGE (A), Diagnostic (D), TR, Prostatisme (P), et PSA.

Algorithme 2 BN-IPFP-1

Entrée : un réseau bayésien (G, P) avec $G = (\mathbf{X}, \mathbf{E})$ et une distribution de probabilité $P(\mathbf{X})$,

m observations probabilistes non contradictoires

$\{Y_f, f = 1, \dots, m\} \subset \mathbf{X}$, spécifiées par $R(Y_1), \dots, R(Y_m)$ telles que $R(Y_f) \ll P(Y_f)$.

Sortie : la distribution de probabilité *a posteriori* $Q(\mathbf{X})$

$Q_0(\mathbf{X}) = P(\mathbf{X})$

$k \leftarrow 1$ [k : indice d'itérations]

Répéter

$f \leftarrow 1 + (k - 1) \bmod m$ [f : indice de l'observation actuelle]

$l \leftarrow 1 + \lfloor (k - 1)/m \rfloor$ [l : indice d'itérations par observation]

Construire le rapport de vraisemblance

$$L_{f,l}(Y_f) \leftarrow \frac{R(y_f^1)}{Q_{k-1}(y_f^1)} : \dots : \frac{R(y_f^{f_s})}{Q_{k-1}(y_f^{f_s})}$$

avec $y_f^1, \dots, y_f^{f_s}$ sont les états de Y_f

Combiner les rapports de vraisemblance antérieurs sur Y_f

$$L_k(Y_f) \leftarrow \prod_l L_{f,l}(Y_f)$$

Obtenir $Q_k(\mathbf{X})$ en mettant à jour $Q_{k-1}(\mathbf{X})$ avec $L_k(Y_f)$ en utilisant la méthode de Pearl de l'observation virtuelle

$k \leftarrow k + 1$

Jusqu'à (convergence)

Les résultats présentés dans la table 3.4 sont obtenus en propageant deux observations probabilistes fixes sur les variables AGE (Y_1) et PSA (Y_2). Les observations probabilistes sont définies comme suit : $R(\text{AGE}) = (0.52, 0.48)$ et $R(\text{PSA}) = (0.487, 0.513)$. La table 3.4 illustre les rapports de vraisemblance construits à chaque étape, et le résultat de la probabilité *a posteriori* obtenu à chaque étape. Les valeurs des distributions de probabilité *a posteriori* de la dernière itération sont égales aux distributions de probabilité des observations probabilistes fixes.

TABLE 3.4 – Trace de l’algorithme BN-IPFP1 ; chaque case contient les valeurs correspondant aux deux états de la variable avant normalisation.

k	1	2	3	4	5
f	1	2	1	2	1
l	1	1	2	2	3
$L_{f,l}(Y_f)$	0.7469 1.5801	1.5487 0.7483	0.9613 1.0456	1.0053 0.9950	0.9994 1.0007
$L_k(Y_f)$	0.7469 1.5801	1.5487 0.7483	0.7180 1.6522	1.5569 0.7446	0.7175 1.6533
$Q_k(\text{AGE})$	0.5200 0.4800	0.5409 0.4591	0.5200 0.4800	0.5203 0.4797	0.5200 0.4800
$Q_k(\text{PSA})$	0.3145 0.6855	0.4870 0.5130	0.4844 0.5156	0.4870 0.5130	0.4870 0.5130

3.5.3 Extension de la définition d’observations probabilistes fixes

La définition 3.3.1 a été étendue par Valtorta (VKV02) afin de considérer des informations sur une ou plusieurs variables du modèle, spécifiées par différentes formes d’observations probabilistes fixes.

Définition 3.5.1 (Extension de l’observation probabiliste fixe). *Une observation probabiliste fixe sur un sous-ensemble de variables $\mathbf{Y} \subset \mathbf{X}$ est spécifiée par une description partielle d’une mesure de probabilité. La spécification peut être donnée par :*

- (a) *une distribution de probabilité jointe $R(\mathbf{Y})$,*
- (b) *une distribution de probabilité conditionnelle $R(\mathbf{Y} \mid \mathbf{Z})$ avec $\mathbf{Z} \subset \mathbf{X} \setminus \mathbf{Y}$,*

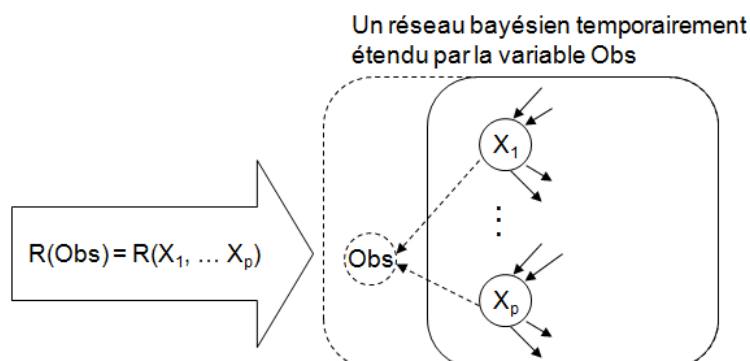


FIGURE 3.11 – Un réseau bayésien reçoit une observation probabiliste fixe sur un sous ensemble de variables (X_1, \dots, X_P)

- (c) affectation de probabilité sur des événements arbitraires sur des variables de \mathbf{Y} ,
- (d) affectation de probabilité sur des formules logiques arbitraires sur des variables de \mathbf{Y} .

La procédure suivante permet de traiter la propagation pour l'extension des observations probabilistes fixes. Ceci est réalisé en ajoutant une variable d'observation qui est créé comme suit :

- Premièrement, on crée une variable d'observation Obs pour chaque observation probabiliste fixe concernant un sous ensemble de variables. Chaque état de la variable ajoutée correspond aux résultats possibles de l'observation probabiliste.
- Deuxièmement, on ajoute des arcs dirigés vers la variable Obs à partir de toutes les variables du réseau bayésien qui ont une influence directe sur l'observation, c'est à dire les variables impliquées dans l'observation probabiliste (Figure 3.11).
- Troisièmement, on modélise la dépendance de la variable ajoutée en spécifiant la table de probabilité conditionnelle $P(Obs \mid pa(Obs))$.

L'observation probabiliste sur la variable d'observation ajoutée se propage dans le réseau bayésien augmenté grâce à l'un des algorithmes de la table 3.3.

Dans le cas d'un ensemble d'observations sur des variables indépendantes $\{E_1, \dots, E_p\}$, la propagation peut se faire en considérant une seule observation probabiliste $R(E_1, \dots, E_p) = R(E_1) \times \dots \times R(E_p)$.

3.5.4 Propagation de l'observation probabiliste versus révision du modèle

3.5.4.1 Révision et propagation : différence entre observation ponctuelle et connaissance

La révision des connaissances représentées par le réseau bayésien peut concerner les paramètres ou la structure. La révision des paramètres consiste à remplacer la distribution de probabilité P qui définit le réseau bayésien par une autre, de façon définitive. Ici nous discutons la révision des paramètres du réseau bayésien versus la propagation d'observations probabilistes fixes qui consiste aussi à remplacer la distribution de probabilité initiale par une autre, mais de façon temporaire : aussi longtemps que les observations probabilistes fixes sont valables.

Par définition, une observation probabiliste définit une contrainte sur la distribution de probabilité *a posteriori*, de ce fait elle remplace la distribution de probabilité marginale sur les variables concernées et met en question la distribution de probabilité jointe.

Doit-on considérer une observation probabiliste comme un nouvel élément de connaissance ? Autrement dit ne doit-on pas changer le modèle qui représente la connaissance plutôt que de propager une observation probabiliste ?

Dans la suite, nous présentons deux scénarios qui montrent clairement la différence entre la mise à jour de croyance par les observations probabilistes et la révision du modèle. Tout d'abord, considérons une observation qui a une validité temporaire. Cette observation peut venir de l'observation partielle d'un état particulier du système modélisé. Ainsi, l'observation probabiliste peut être considérée comme une observation certaine temporaire, donc, ce n'est pas pertinent de réviser le modèle. En revanche, la distribution de probabilité P d'un réseau bayésien (G, P) , dans lequel l'observation probabiliste temporaire est appliquée, représente une connaissance permanente qui n'a pas besoin d'être révisé par des observations à court terme.

D'autre part, un modèle est souvent un compromis entre l'efficacité relative à son traitement et la précision des données qu'elles peuvent le représenter. Ceci signifie que le modèle ne comprend jamais tous les paramètres du système modélisé. Dans un réseau bayésien, la taille de l'ensemble des variables et le nombre de leurs valeurs

possibles sont souvent limités, notamment, en raison de la quantité de données d'apprentissage ou pour maîtriser les coûts liés aux traitements.

Considérons l'exemple 3.4.2 où on peut supposer que le modèle a été conçu, par défaut, avec la langue française. Ainsi, le modèle ne comprend pas de variable représentative de la langue du texte à partir de laquelle le caractère est extrait. Cette limite du modèle peut être partiellement surmontée en appliquant une observation probabiliste afin d'envisager des informations sur une variable qui ne sont pas incluses dans le modèle. Lorsque l'observateur est capable de traduire la caractéristique observée (ici la langue) sur une variable du modèle (ici la lettre), il peut appliquer une observation probabiliste sur cette variable. Bien que la connaissance intégrée dans la distribution de probabilité est permanente (ici la distribution des lettres de l'alphabet en français), cette information n'est pertinente que lorsqu'on est dans le même contexte c'est-à-dire la lettre observée est issue d'un texte de la même langue considérée dans le modèle.

Ainsi, on peut conclure que l'information sur un état spécifique du système modélisé peut être prise en compte via la propagation d'observation certaine, de vraisemblance ou probabiliste, afin de prendre en compte l'influence des variables qui ne sont pas dans le modèle. La révision du modèle se produit lorsque des hypothèses qui lui sont associés sont changées.

La différence entre la révision des paramètres du réseau bayésien et la propagation d'observations probabilistes fixes est essentiellement basée sur la nature des éléments nouveaux à prendre en compte. En termes algorithmiques, ces opérations sont proches.

Certains des algorithmes listés dans la table 3.3 ont été adaptés afin de réviser le réseau bayésien en remplaçant la distribution de probabilité initiale par la distribution de probabilité révisée (voir table 3.5). La table 3.5 liste des algorithmes qui permettent de réviser (définitivement) la distribution de probabilité d'un réseau bayésien pour intégrer de nouvelles connaissances.

L'algorithme E-IPFP intègre les contraintes en changeant seulement les tables de probabilités conditionnelles du réseau bayésien tout en conservant la structure

Algorithme	Réf.	Caractéristiques
E-IPFP	(PDZP12)	révise les TPC, mais conserve la structure du réseau bayésien.
E-IPFP-SMOOTH	(PDZP12)	idem, observation probabilistes contradictoires
D-IPFP	(PDZP12)	idem

TABLE 3.5 – Algorithmes de révision d’un réseau bayésien à partir de distribution locale de probabilités.

du réseau. Les algorithmes E-IPFP-SMOOTH et D-IPFP sont obtenus par des modifications de l’algorithme E-IPFP. Le premier est utilisé dans la situation où les contraintes probabilistes sont contradictoires entre elles ou avec la structure du réseau bayésien. Le second réduit le coût de calcul par décomposition d’un E-IPFP globale en un ensemble de petits problèmes E-IPFP locaux.

3.5.4.2 Modifier la distribution de probabilité *a priori* ou propager des observations probabilistes fixes ?

Une observation probabiliste fixe sur une variable remplace les connaissances *a priori* sur cette variable. Il est donc légitime de se demander si une telle observation probabiliste ne peut pas être prise en compte en remplaçant la distribution de probabilité du nœud observé par la distribution de probabilité qui spécifie l’observation probabiliste fixe. Cette opération est facile à réaliser lorsqu’il s’agit d’un nœud sans parent. L’exemple suivant illustre que même dans ce cas simple, une observation probabiliste sur un nœud sans parent ne peut pas être maintenue fixe lorsqu’on remplace la probabilité *a priori* du nœud observé par la distribution de probabilité qui spécifie l’observation probabiliste.

Reprenons l’exemple Asia (voir exemple 3.3.1). Notons $R(X)$: la distribution de probabilité d’une observation probabiliste sur le nœud X et $Q(X)$: la distribution de probabilité *a posteriori* du nœud X . On considère deux cas différents.

Dans le premier cas³ on souhaite propager deux observations probabilistes fixes sur deux nœuds "sources" (sans parents) et conditionnellement indépendants. Prenons par exemple une observation probabiliste fixe sur le nœud A avec $R(A) = (0.2, 0.8)$ et une deuxième observation probabiliste fixe sur le nœud S avec $R(S) = (0.4, 0.6)$. Dans ce cas, en remplaçant les tables de probabilité *a priori* des nœuds A et S par $R(A)$ et $R(S)$, on obtient les mêmes distributions de probabilité *a posteriori* que lorsqu'on propage des observations probabilistes fixes. Les probabilités *a posteriori* des nœuds A et S restent inchangées et on a $Q(A) = R(A) = (0.2, 0.8)$ et $Q(S) = R(S) = (0.4, 0.6)$.

Dans le deuxième cas, on souhaite propager deux observations probabilistes fixes sur deux nœuds conditionnellement dépendants. Prenons par exemple une observation probabiliste fixe sur le nœud S avec $R(S) = (0.2, 0.8)$ et une deuxième observation probabiliste fixe sur le nœud D avec $R(D) = (0.4, 0.6)$. Dans cette expérience, on a remplacé la table de probabilité *a priori* du nœud (S) par $R(S)$ puis on a propagé l'observation probabiliste $R(D)$. La distribution de probabilités *a posteriori* du nœud (S) est modifiée : $Q(S) = (0.374, 0.626)$.

Ce second cas illustre le fait qu'une observation probabiliste fixe ne peut pas être prise en compte en remplaçant localement la distribution de probabilité du réseau bayésien par la distribution de probabilité de cette observation probabiliste fixe.

3.6 Synthèse des différents types d'observations incertaines

Dans ce chapitre nous avons identifié trois types d'observations incertaines (BMDP⁺15) : observation de vraisemblance, observation probabiliste fixe et observation probabiliste non-fixe.

Une observation de vraisemblance est une observation imprécise ou fournie par une source non fiable, elle est spécifiée par un rapport de vraisemblance et propagée par la méthode de Pearl de l'observation virtuelle. Cette méthode consiste à trans-

3. Ce procédé est utilisé dans plusieurs articles qui proposent de combiner des réseaux bayésiens avec des SIGs : le SIG fourni des informations pour un point (ou une zone) géographique, concernant certaines variables du réseau bayésien. Pour exploiter ces informations, un nouveau réseau bayésien est généré pour chaque zone géographique concernée, avec une structure toujours identique, et une distribution de probabilité issue des observations locales.

former une observation de vraisemblance sur une variable X en une observation certaine sur un nouveau nœud virtuel, ajouté en tant que fils du nœud X .

Une observation probabiliste fixe ou non-fixe exprime une contrainte sur l'état de certaines variables après que cette information soit propagée dans le réseau bayésien. Une observation probabiliste sur X est spécifiée par une distribution de probabilité $R(X)$ définie "tout bien considéré", ce que signifie qu'elle remplace toutes croyances précédentes et toutes les connaissances qu'on a sur X .

Une observation probabiliste non-fixe sur X peut être modifiée, par la suite, par d'autres observations sur n'importe quels nœuds du réseau bayésien, y compris le nœud X . Elle peut être propagée par la conversion en observation de vraisemblance.

Une observation probabiliste fixe sur une variable X ne peut être modifiée par aucune information supplémentaire sur les variables du modèle. Ainsi, la propagation de plusieurs observations probabilistes fixes est commutative. Les observations probabilistes fixes doivent être propagées par des algorithmes spécifiques tels que big clique ou BN-IPFP (Table 3.3).

Une observation certaine sur une variable signifie que cette variable est instanciée. Ceci est représenté par un vecteur de zéros contenant la valeur un à la position correspondant à la valeur observée.

Une observation négative ou disjonctive sur une variable X est un cas particulier d'observation de vraisemblance. Elle est défini par un vecteur d'observation de zéros et de uns. Il représente l'information que X peut être seulement dans l'un des états qui correspond aux uns et que les autres états sont impossibles.

Une observation de vraisemblance se distingue clairement d'une observation probabiliste en termes de spécification et de propagation. L'observation probabiliste comprend deux sous-groupes : les observations probabilistes fixes et non-fixes qui se différencient par l'impact de nouvelles informations sur la contrainte définie par l'observation probabiliste.

Le tableau 3.6 présente une synthèse des caractéristiques des différents types d'observations dans un réseau bayésien présentés dans cette partie.

	Obs. certaine	Obs. incertaine		
		Obs. de vraisemblance	Obs. probabiliste non fixe	Obs. probabiliste fixe
L'observation définit une contrainte sur la distribution de probabilité <i>a posteriori</i> . Elle est spécifiée "tout bien considéré"	Oui	Non	Oui	Oui
Plusieurs observations peuvent être combinées sur la même variable	Non	Oui	Oui	Non
La distribution de probabilité <i>a posteriori</i> peut être modifiée par la propagation d'observations sur d' autres variables	Non	Oui	Oui	Non
La distribution de probabilité <i>a priori</i> sur la variable est utilisée lors de la propagation	Non	Oui	Non	Non
La propagation de plusieurs observations commute	Oui	Oui	Non	Oui

TABLE 3.6 – Synthèse des propriétés des différents types d'observations dans un réseau bayésien.

3.7 Conclusion

Dans ce chapitre, nous avons présenté trois types d'observations incertaines. Quelle est la nature de l'incertitude qui concerne ces différents types d'observations ? Les deux grands types d'observations sont présents :

- incertitude épistémique : par exemple l'information fournie par l'observateur est peu fiable (observations de vraisemblance).
- incertitude aléatoire (ou stochastique) : l'information fournie par l'observateur provient d'un ensemble d'observations sur une même variable et la variabilité de ces observations se traduit par une distribution de probabilité sur la ou les variables observées (distribution locale de probabilités). L'ensemble des observations provient :

- soit de l’observation d’une sous population : on observe une classe en tant que groupe d’individus, et non pas un individu unique (observation probabiliste fixe),
- soit d’un individu unique (un cas précis), dont on sait l’appartenance à une sous-population pour laquelle la variable a été observée (observation probabiliste non fixe).

Actuellement, très peu de logiciels permettant de propager des observations probabilistes. Pourtant, cette fonctionnalité présente un réel intérêt et permet d’élargir les champs d’utilisations des réseaux bayésiens. Un indicateur de la demande des utilisateurs pour propager différents types d’observations incertaines est fourni par le nombre important d’articles relatifs aux réseaux bayésiens flous ([BDT03](#); [FPA08](#); [TL07](#); [CYLH07](#); [DPC⁺14a](#); [PYC06](#); [PL00](#); [HOA12](#); [IY10](#); [FB12](#); [DTB08](#)). A ce jour, il n’y a pas de définition du concept de réseau bayésien flou qui soit généralement accepté par la communauté académique des chercheurs sur les réseaux bayésiens. Cependant, ces articles ont souvent pour objectif de propager des observations sur lesquelles existe une incertitude. Le manque de clarification de la terminologie et des propriétés des différents types d’observations probabilistes est un facteur d’explication de l’absence des fonctionnalités pour la propagation des observations probabilistes dans les logiciels sur les réseaux bayésiens. Nous espérons avoir remédié à ces insuffisances dans ce chapitre.

Le chapitre suivant propose trois types d’exemples d’utilisation des observations probabilistes dans les réseaux bayésiens.

Chapitre 4

Applications des observations probabilistes fixes

Bien que la notion d’observation probabiliste dans les réseaux bayésiens a été introduite en 1998 (Blo98; DMP98), elle est restée très peu utilisée par les utilisateurs des réseaux bayésiens. Dans ce chapitre, nous proposons trois types d’exemples d’utilisation des observations probabilistes. Le premier concerne l’intégration des réseaux bayésiens avec des systèmes d’information géographique (SIG). Le second traite des échanges d’informations dans le cadre des réseaux bayésiens distribués. Le troisième type d’applications s’intéresse à la propagation d’observations sur des variables continues.

4.1 Applications aux systèmes d’information géographique

Un type d’applications des observations probabilistes concerne l’intégration d’un système d’information géographique (SIG) avec un réseau bayésien. Les SIGs sont basés sur le principe du découpage d’un territoire en petites zones afin de fournir un ensemble d’informations relatif à chaque zone. Plusieurs propositions ont été faites dans le but de combiner réseau bayésien et SIG (AMD08; JLCM12; GDA⁺13; GRS06; LBD⁺13b; LWdC10b; PP07). De façon schématique, les réseaux bayésiens sont utilisés en amont et/ou en aval des SIGs. Par exemple, un premier SIG fournit un ensemble d’informations par ”point géographique” qui sont utilisées pour définir

les paramètres d'un réseau bayésien. Celui-ci grâce à l'inférence, permet de déduire de nouvelles informations qui peuvent alimenter un second SIG (voir figure 4.1).

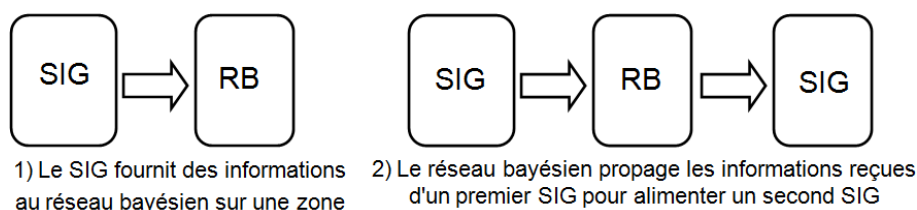


FIGURE 4.1 – Intégration conceptuelle de réseau bayésien (RB) et de système d'information géographique (SIG).

Dans (LBD⁺13a), les auteurs affirment que dans la plupart des études qui intègrent un réseau bayésien à un SIG, le SIG est utilisé comme entrée du réseau bayésien. C'est à dire que pour une variable donnée, et pour chaque zone géographique, le SIG fournit des distributions de probabilités qui reflètent la variabilité des variables X_i dans la zone considérée. Ces distributions de probabilités sont utilisées pour déterminer les paramètres d'un réseau bayésien qui par ce fait devient spécifique à une zone géographique.

Dans cette association des SIGs et des réseaux bayésiens, les distributions de probabilités fournies par le SIG pour chaque zone géographique constituent des observations probabilistes à propager. L'usage des observations probabilistes dans ce type d'application évite de modifier les paramètres du réseau bayésien pour chaque zone et permet d'utiliser le même modèle partout. La révision du réseau bayésien est remplacée par une simple propagation d'observations probabilistes fixes.

La combinaison (SIG + RB) a aussi été utilisée dans (LGR12; LGR13) pour évaluer les impacts sociaux, économiques et environnementaux de l'utilisation des énergies renouvelables. La figure 4.2 illustre comment utiliser des observations probabilistes fixes dans un réseau bayésien pour intégrer des observations concernant une zone géographique. Une zone géographique est composée de plusieurs unités, chacune d'entre elles ayant une valeur spécifique pour chaque variable X_i . L'unité peut être un pixel, une maison, etc. Chaque observation d'une variable X_i dans une région A_k est obtenue à partir d'une source S_i qui permet de calculer la répartition des X_i dans la région A_k . La source S_i peut être une base de données ou un SIG. Cette information est spécifiée par une liste d'observations probabilistes plutôt qu'une unique distribution de probabilités jointe sur les variables X_1, \dots, X_p .

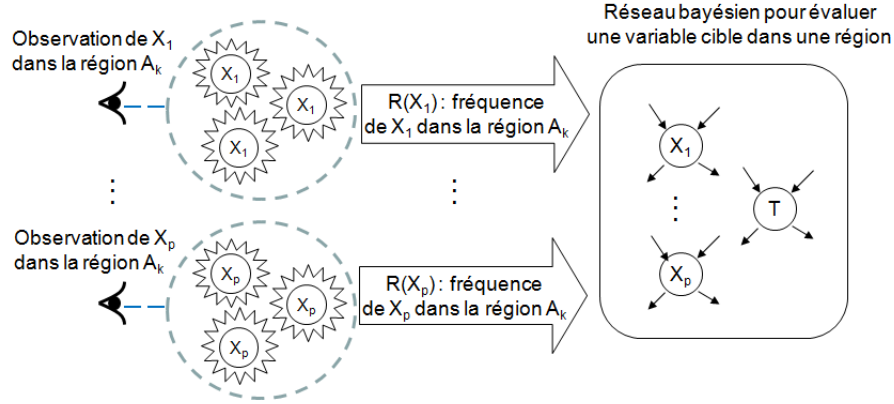


FIGURE 4.2 – Observations probabilistes fixes sur une zone géographique d'intérêt.

La question de la représentation et la propagation de l'incertitude pour l'information géospatiale a été discutée par Laskey et Wright ([LWdC10a](#)). En particulier, les auteurs proposent de représenter des régularités statistiques et des observations incertaines avec des ontologies probabilistes. Il s'agit d'un modèle générique basé sur Multi-Entity Bayesian Network (MEBN) qui permet la génération d'un nouveau réseau bayésien pour chaque zone géographique dans la base de données.

En conclusion, les SIGs réunissent tous les ingrédients pour justifier l'intérêt d'utiliser des observations probabilistes fixes dans les réseaux bayésiens :

- La multitude des zones géographiques qui composent un territoire augmente l'intérêt de définir un modèle de raisonnement unique et réutilisable pour chaque zone géographique, au lieu de définir un nouveau modèle pour chaque zone.
- Les liens causaux entre les différentes informations considérées dans un SIG et les différentes formes d'incertitudes concernant ces informations conduisent vers l'emploi de réseau bayésien.
- Le fait de considérer une zone géographique comme un point unique, conduit à rassembler les informations de cette zone pour la considérer comme une seule entité. La variabilité des observations au sein de la zone définit une distribution de probabilités, ce qui conduit à utiliser des observations probabilistes au lieu des observations classiques.

En définitive, l'utilisation des observations probabilistes fixes dans les réseaux bayésiens en interaction avec les SIGs est un sujet prometteur à explorer.

4.2 Observations probabilistes dans les systèmes à base d'agents équipés d'un réseau bayésien local

La modélisation de grands systèmes du monde réel par des modèles centralisés est souvent trop complexe et coûteuse. Plusieurs travaux ont été réalisés pour résoudre ce problème dans le contexte des modèles graphiques probabilistes à savoir : Multi-agents Influence Diagrams (MAID) (KM03; GP08), Multiply Sectioned Influence Diagrams (MSID) (Zen06), Multiply Sectioned Bayesian Network (MSBN) (XL00; Xia02), et Agent Encapsulated Bayesian Network (AEBN) (Blo98; Lan11). Les deux premiers concernent les diagrammes d'influence, qui sont appliqués à des problèmes ayant une séquence de décisions, alors que les modèles MSBN et AEBN sont basés sur les réseaux bayésiens.

La proposition de Xiang (MSBN, (Xia02)) considère une unique distribution de probabilité sur l'ensemble des variables regroupant les variables de chaque réseau bayésien local. Ce point de vue amène à de fortes contraintes sur la façon dont le réseau bayésien initial est sectionné. Au contraire, dans un modèle à base d'agents équipés de réseaux bayésiens locaux (Blo98; Lan11), chaque réseau bayésien local possède sa propre distribution de probabilité, et il n'y a pas de raison de considérer une distribution de probabilité générale sur l'union de toutes les variables des différents réseaux bayésiens. Cette caractéristique permet aux modèles basés sur des agents équipés d'un réseau bayésien local de traiter les problèmes réels où les acteurs et les informations sont naturellement distribués.

Dans cette section, nous présentons le concept des AEBNs que nous appelons agents équipés d'un réseau bayésien local. Ensuite, nous expliquons l'utilisation des agents équipés d'un réseau bayésien local dans le cas d'un problème de décision récurrent multi-critères pour choisir un vélo.

4.2.1 Modèle à base d'agents équipés d'un réseau bayésien local

4.2.1.1 Concept et définition

Les modèles à base d'agents équipés d'un réseau bayésien local ont été d'abord proposés par Bloemeke (Blo98) et étendus par Langevin (Lan11). Ce cadre des agents équipés d'un réseau bayésien local a été la première motivation de l'apparition et l'utilisation des observations probabilistes fixes dans un réseau bayésien. En effet, dans un modèle à base d'agents équipés d'un réseau bayésien local, un agent "demandeur d'informations pour un ensemble de variables" met à jour ses croyances suite à la réception d'observations probabilistes transmises par des agents "fournisseurs d'informations pour un ensemble de variables". Dans un système basé sur des agents équipés d'un réseau bayésien, chaque agent est détenteur de connaissances sur une partie restreinte du système global considéré. Ces connaissances sont modélisées sous forme d'un réseau bayésien dont l'ensemble des variables représentent la partie du monde que cet agent "comprend" ou perçoit.

L'agent utilise ce réseau bayésien local pour son raisonnement propre. Pour cela, il est capable de faire lui même certaines observations directement et il fait appel à d'autres agents pour obtenir une estimation sur d'autres variables qu'il ne peut obtenir par lui même. A l'aide de ces observations, il met à jour ses croyances sur les autres variables de son modèle de raisonnement et il peut les transmettre à un autre agent qui les lui demande. De ce fait, chaque agent est d'une part "fournisseur d'information" pour les variables sur lesquelles il a le plus d'expertise et d'autre part "demandeur d'informations pour un ensemble de variables" pour les variables sur lesquelles il a besoin de l'expertise d'un autre agent. En fonction des objectifs, les agents sont organisés en une hiérarchie "fournisseur / demandeur".

Définition 4.2.1 (Agent équipé d'un réseau bayésien local). *Un agent équipé d'un réseau bayésien local est un agent dont la base de connaissance interne est représentée par un réseau bayésien (G, P) tel que $G = (\mathbf{X}, \mathbf{E})$ et l'ensemble des variables est divisé en trois sous ensembles $\mathbf{X} = \mathbf{I} \cup \mathbf{L} \cup \mathbf{O}$:*

- les variables d'entrée (\mathbf{I}) : les variables pour lesquelles d'autres agents ont une meilleure connaissance,

- les variables de sortie (\mathbf{O}) : les variables pour lesquelles cet agent a les meilleures connaissances et qui sont partagées avec d'autres agents,
- les variables locales ou cachés (\mathbf{L}) : les variables qui sont privées à cet agent.

L'ensemble \mathbf{E} contient les arcs du modèle qui définissent les relations de causalité entre les variables de \mathbf{X} et P est une distribution de probabilité jointe définie sur \mathbf{X} . Les variables de l'ensemble $\mathbf{I} \cup \mathbf{O}$ sont des variables partagées entre cet agent et d'autres agents, tandis que les variables de l'ensemble \mathbf{L} sont privées.

4.2.1.2 Communication entre agents

Un agent demande des informations sur une ou plusieurs de ses variables d'entrées à un autre agent pour qui ces variables sont des variables de sortie. L'information envoyée par un agent à un autre sur un ensemble de variables communes aux deux agents consiste en une ou plusieurs distributions locales de probabilités sur les variables partagées. Un agent considère les informations qu'il reçoit sur ses variables d'entrée comme des "vérités" qu'il ne remet pas en cause. Il doit donc prendre en compte ces informations sans les modifier, ce qui définit des observations probabilistes fixes. Lorsqu'une information concerne plusieurs variables partagées entre deux agents, il s'agit d'une observation probabiliste fixe au sens de la définition étendue (voir section 3.5.3).

Comme mentionné dans (VKV02), la mise à jour des croyances dans le réseau bayésien local d'un agent par la transmission d'observations probabilistes fixes assure une sorte de cohérence globale car les croyances sur chaque variable partagée, représentée par la probabilité *a posteriori* marginale, est la même pour tous les agents.

4.2.2 Application à un problème de décision récurrent multicritères : l'aide au choix d'un vélo

Sedki et Delcroix ont proposé une structure à base de réseau bayésien pour fournir un système de recommandations pour les problèmes récurrents d'aide à la décision multi-critères (ADMC) appliqué à l'aide au choix d'un fauteuil roulant (DSL13; SDL⁺10). Nous nous sommes inspirés de ces travaux pour proposer une structure à base d'agents équipés d'un réseau bayésien local, pour produire un ensemble de recommandations pour l'aide au choix d'un vélo. Cette proposition est basée sur le

système de recommandations proposé dans (DSL13) avec en plus l'utilisation d'un ensemble d'agents équipés d'un réseau bayésien local. L'exemple du choix d'un vélo que nous présentons ici permet d'illustrer l'utilisation des observations probabilistes fixes au niveau des messages échangés entre les agents.

4.2.2.1 Problèmes récurrents d'aide à la décision multicritères

Un problème de décision *récurrent* est un problème de décision qui est répété plusieurs fois dans des contextes différents et avec différents acteurs, mais toujours avec le même type d'alternatives et les mêmes critères de choix. Les problèmes de décision récurrents sont caractérisés par le fait qu'ils concernent tout type de décideur. En effet, chaque occurrence de la décision implique un ensemble d'acteurs différents, dans un contexte spécifique, avec leur contraintes propres. Ils sont aussi caractérisés par l'absence possible d'un expert, capable d'évaluer une alternative sur un critère, mais aussi l'importance des critères. Les problèmes de décision récurrents multi-critères multi-attributs sont un sous-ensemble des problèmes de décision dans lequel les alternatives sont caractérisées par un ensemble d'attributs fixes, un ensemble de critères fixes, mais où l'importance de chaque critère et l'évaluation d'une alternative en fonction d'un critère donné dépendent de chaque cas de décision.

Choisir un vélo est un problème de décision récurrent multi-critères. Il est défini comme suit :

- l'ensemble des caractéristiques du cas de décision qui peuvent avoir une influence sur le choix du vélo (âge et sexe du cycliste, type d'usage souhaité, etc.) ;
- l'ensemble des alternatives \mathbf{A} (une alternative désigne un vélo dont l'ensemble des valeurs des attributs est complètement précisé) ;
- l'ensemble $\mathbf{V} = \{V_1, \dots, V_n\}$ des attributs d'une alternative (type de cadre, de guidon, de roue et de pneus, etc.). Une alternative $\mathbf{a} \in \mathbf{A}$ est complètement décrite par l'ensemble des valeurs de ses attributs $\mathbf{a} = (v_1, \dots, v_n)$ où $v_i \in \mathcal{D}_{V_i}$ et \mathcal{D}_{V_i} est l'ensemble des valeurs possibles de V_i .
- l'ensemble des critères $\{g_1, \dots, g_m\}$ pour guider le choix. Par exemple la stabilité du vélo et le confort de chargement du vélo dans un autre véhicule sont des critères de choix.

Choisir un vélo est un problème de décision *récurrent* puisque le problème de décision se produit plusieurs fois avec le même ensemble d'alternatives (au sens où l'ensemble \mathbf{V} des attributs d'un vélo ne change pas) et la même liste de critères, mais à chaque fois avec un cas de décision spécifique : le décideur, le contexte et les préférences sont différents.

Les caractéristiques d'un cas de décision sont regroupées en trois ensembles : (1) les caractéristiques directes du cycliste (âge, poids, sexe, taille, niveau d'activité, etc.), (2) ses usages du vélo : type et fréquence d'utilisation du vélo, besoin de transporter (courses, bagage en randonnée, enfant, etc.), (3) l'environnement et le contexte de l'utilisation du vélo (type de sol des lieux d'usage (chemin, route de campagne, ville, etc.), relief du lieu d'usage, précipitation moyenne du lieu d'usage, etc.).

L'objectif est de fournir au cycliste voulant acheter un vélo un ensemble de recommandations pertinentes sur les caractéristiques du vélo qui répondra à ses besoins et à ses projets. Un tel système de recommandations doit intégrer des connaissances d'experts sur les vélos et leur utilisation alors que ces experts ne sont pas présents lors du choix du vélo. Ces recommandations se basent sur la description du cycliste, de ses besoins et objectifs, et la description du contexte. Aucune préselection de vélos n'est requise. Au contraire, les recommandations fournies ont pour but d'aider le décideur à sélectionner un ensemble d'alternatives satisfaisantes, avant d'en choisir finalement une.

Dans le problème du choix d'un vélo, l'incertitude est à deux niveaux : d'une part elle vient d'une description incomplète ou imprécise du cas de décision, d'autre part, l'incertitude porte sur l'évaluation d'un vélo donné sur un critère donné, pour un cycliste donné avec ses objectifs (incertitude épistémique). Nous avons choisi un modèle à base de réseau bayésien car il intègre ces diverses formes d'incertitude et permet de propager des informations sur les variables observées vers les autres variables. Le choix d'un modèle distribué est motivé par deux points. Premièrement, les paramètres qui peuvent influencer sur le choix sont naturellement répartis sur plusieurs entités, telles que le cycliste, l'environnement d'utilisation, le vélo, etc. Le choix d'une organisation basée sur des agents équipés d'un réseau bayésien local permet de représenter chaque entité par un agent spécifique, et de construire séparément le réseau bayésien local de chaque agent.

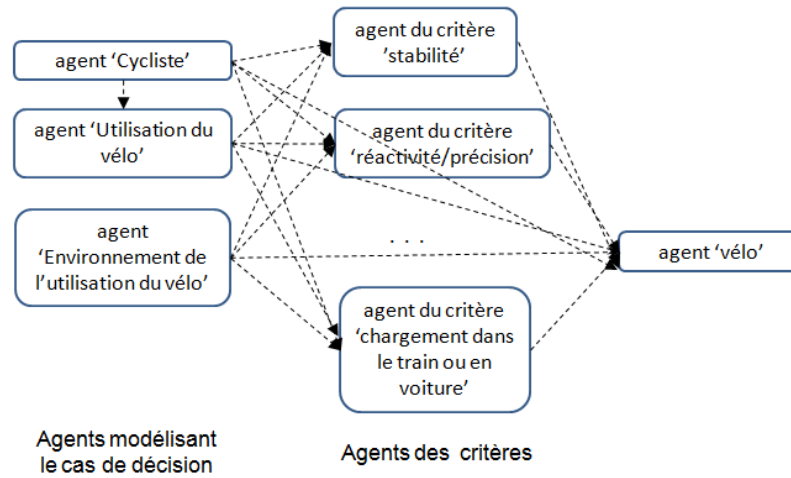


FIGURE 4.3 – Graphe des agents d'un système de recommandations pour le choix d'un vélo.

Deuxièmement, le nombre total de paramètres est très élevé et la construction d'un seul réseau bayésien conduirait à une structure de graphe avec de nombreux chemins multiples entre deux nœuds ce qui augmenterait le risque d'obtenir un modèle inutilisable.

4.2.2.2 Graphe des agents pour un système de recommandations pour le choix d'un vélo

Dans cette partie, nous décrivons la structure du modèle proposé à base d'agents équipés d'un réseau bayésien local pour fournir un ensemble de recommandations sur le choix d'un vélo, en fonction des particularités de chaque situation. Nous commençons par décrire le graphe des agents, avec le rôle de chaque agent, puis nous décrivons le fonctionnement du système, à l'aide du graphe du réseau bayésien local d'un agent expert d'un critère, en illustrant ces explications avec le problème de l'aide au choix d'un vélo.

La figure 4.3 montre le graphe des agents impliqués dans le système chargé de fournir les recommandations. Les agents sont de trois catégories :

Sur la gauche, trois agents représentent les caractéristiques d'un cas de décision. L'agent "cycliste" est un expert de la personne pour laquelle il est question de choisir un vélo. Les deux autres agents représentent l'utilisation souhaitée du vélo et l'environnement dans lequel le vélo sera utilisé (climat, relief, réseau routier, etc.).

Ces trois agents tirent leurs informations de questionnaires, d'entretiens directs ou indirects avec le cycliste. Ils les complètent avec l'expertise incluse dans leur réseau bayésien local et grâce au mécanisme d'inférence. Ces agents sont fournisseurs d'informations pour la plupart de leur variables.

Au centre de la figure 4.3 se trouvent les agents associés aux critères de décision. Pour l'aide au choix d'un vélo, trois critères sont représentés : la stabilité du vélo, sa réactivité et le confort de chargement du vélo dans un véhicule. Chaque critère est modélisé par un agent qui a pour mission d'évaluer l'importance de ce critère en fonction du cas de décision et qui l'utilise pour fournir une distribution de probabilités sur le niveau de qualité attendue d'une alternative sur ce critère. Ce niveau de qualité est représenté par un indice entre 0 et 1. L'agent d'un critère fournit une distribution de probabilité sur les différents niveaux de qualité qui reflète l'attente sur ce critère en fonction des caractéristiques du cas de décision.

Sur la droite, l'agent "vélo" est un expert du vélo. Son réseau bayésien local contient d'une part l'ensemble des caractéristiques directes d'un vélo et leurs liens, ainsi qu'un ensemble de variables complexes qui combinent les caractéristiques simples du vélo et si besoin les caractéristiques du cycliste pour évaluer un vélo sur chacun des critères. Chaque noeud d'évaluation d'une alternative sur un critère est partagé avec l'agent responsable de ce critère.

4.2.2.3 Fonctionnement du système de recommandations à l'aide d'agents équipés d'un réseau bayésien local

Le système de recommandations fonctionne suivant les trois étapes ci-dessous :

- Etape 1 : les agents qui représentent le cas de décision collectent leurs observations et fournissent aux agents des critères les informations sur les caractéristiques du cycliste concerné, son environnement et ses objectifs en termes d'utilisation du vélo. Ces informations sont des observations probabilistes fixes.
- Etape 2 : chaque agent d'un critère évalue l'importance du critère dans ce cas de décision à l'aide des informations reçues et envoie à l'agent "vélo" le niveau de qualité attendue pour ce critère. Ce niveau se présente sous la forme d'une observation probabiliste fixe.
- Etape 3 : l'agent vélo propage l'ensemble des observations probabilistes fixes qu'il a reçues concernant le niveau de qualité attendu des critères et obtient

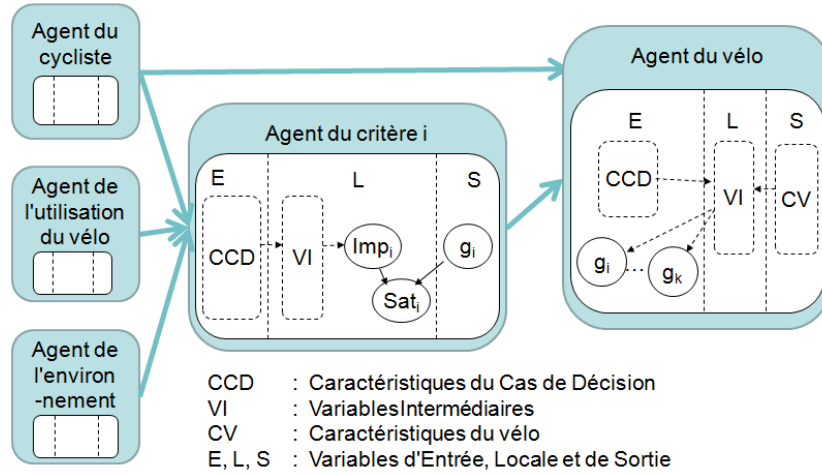


FIGURE 4.4 – Agent d'un critère de décision et agent vélo, avec leur réseau bayésien local et leurs liens avec les autres agents.

ainsi une recommandation pour chaque caractéristique directe du vélo sous la forme d'une distribution de probabilité.

Examinons plus en détail ce fonctionnement à l'aide de la figure 4.4 qui montre le détail du graphe des réseaux bayésiens locaux d'un agent chargé d'un critère et de l'agent "vélo" ainsi que leurs liens avec les autres agents.

La figure 4.4 montre les liens entre un agent responsable d'un critère et les autres agents. Sur la gauche, les flèches représentent les informations demandées par l'agent d'un critère aux agents décrivant le cas de décision en vue d'évaluer l'importance du critère pour ce cas de décision. Ces informations sont échangées sous la forme d'observations probabilistes fixes (qui peuvent être aussi des observations certaines). Sur la droite, la flèche de l'agent du critère i vers l'agent "vélo" représente l'information envoyée par l'agent chargé d'un critère à l'agent vélo. Cette information concerne le niveau de qualité attendue sur le critère concerné pour le cas de décision en cours. Il s'agit d'une distribution de probabilités sur l'ensemble des valeurs du noeud g_i d'évaluation du critère i qui résulte de l'inférence dans le réseau bayésien local de l'agent de ce critère. Cette évaluation est une observation probabiliste fixe que l'agent vélo reçoit, en même temps que les évaluations des autres critères.

L'information échangée entre deux agents concerne une ou plusieurs variables partagées entre les deux agents. Du côté de l'agent fournisseur d'informations pour un ensemble de variables, il s'agit de variables de sortie, alors que du côté de l'agent

demandeur d'informations pour un ensemble de variables, il s'agit d'une variable d'entrée.

Examinons maintenant la structure du graphe du réseau bayésien local de l'agent chargé du $i^{\text{ème}}$ critère de décision (voir figure 4.4). Ce graphe contient une V-structure incluant trois nœuds directement associés au critère i :

- Imp_i : indice d'importance du critère (évalué en fonction du cas de décision). Les nœuds parents de Imp_i sont des caractéristiques du cas de décision.
- g_i : indice d'évaluation d'une alternative donnée pour ce critère (évalué en fonction des caractéristiques de l'alternative et du cas de décision). Les nœuds parents de g_i sont des caractéristiques de l'alternative et du cas de décision.
- Sat_i : indice de satisfaction du critère fournie par une alternative donnée en fonction du cas de décision ($sat = g^{Imp}$). Cette définition permet d'exiger le même niveau de satisfaction pour tout critère, quel que soit son niveau d'importance.

Lorsque le niveau de satisfaction attendu est fixé, alors le chemin entre les nœuds Imp et g est passant. Ce principe est utilisé pour évaluer l'indice g pour chaque critère en fonction de l'importance de ce critère : le niveau de satisfaction est fixé dans un intervalle de valeurs correspondant à une satisfaction élevée pour tous les critères, et la propagation du nœud Imp_i vers le nœud g_i permet d'obtenir le niveau de qualité souhaité sur le critère i dans ce cas de décision.

La définition de l'indice de satisfaction et la V-structure assurent le comportement suivant : lorsqu'un critère est peu ou pas important (indice proche de zéro) selon le cas de décision, le niveau de qualité attendu sur ce critère est quelconque (distribution de probabilité uniforme sur le niveau de qualité attendu). A l'inverse, pour un critère très important (indice proche de 1), le niveau de qualité attendu est élevé (proche de 1).

La figure 4.5 illustre une partie des réseaux bayésiens locaux des agents représentant un cas de décision : (a) agent "Cycliste" et (b) agent chargé de décrire l'utilisation du vélo. La plupart des nœuds de ces réseaux bayésiens sont partagés avec les réseaux bayésiens locaux d'autres agents.

Deux exemples concrets des graphes d'agents de critères sont donnés dans la figure 4.6 pour les critères de stabilité (a) et de confort de chargement du vélo dans un train ou un véhicule (b). Par exemple, l'agent du critère de stabilité du vélo commence par demander à ses fournisseurs les informations sur le cycliste, pourquoi,

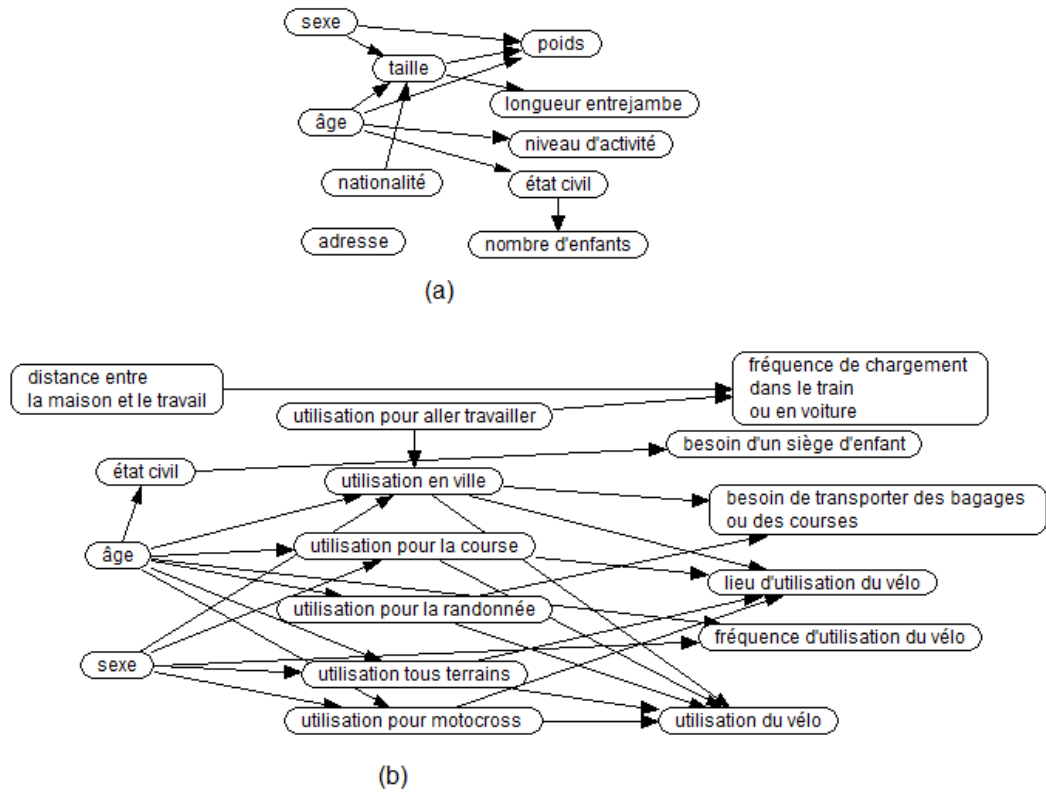


FIGURE 4.5 – Une partie du graphe du réseau bayésien local des agents représentant (a) le cycliste et (b) l'utilisation du vélo.

comment et où il veut faire du vélo. A l'aide de ces informations, l'agent du critère stabilité évalue l'importance d'avoir un vélo stable et en déduit le niveau de stabilité du vélo attendu pour ce cas de décision.

Examinons maintenant le réseau bayésien local de l'agent représentant le vélo (figure 4.7). Les caractéristiques directes du vélo (sur le type de cadre, de pneu, de guidon, etc.) sont ses variables de sortie car l'objectif de cet agent est de fournir des recommandations sur ces caractéristiques. Les variables d'entrées de l'agent vélo sont les niveaux de qualité attendus sur chacun des critères (variables g_i). Le réseau bayésien local de l'agent "vélo" combine l'ensemble des exigences concernant chaque critère de décision et les propage sur les caractéristiques directes du vélo.

Dans ce modèle, les niveaux de qualités attendus pour chaque critère ne peuvent pas être des observations certaines puisqu'ils résultent d'un raisonnement incertain. Le niveau de qualité attendu pour un critère dépend de l'importance de ce critère

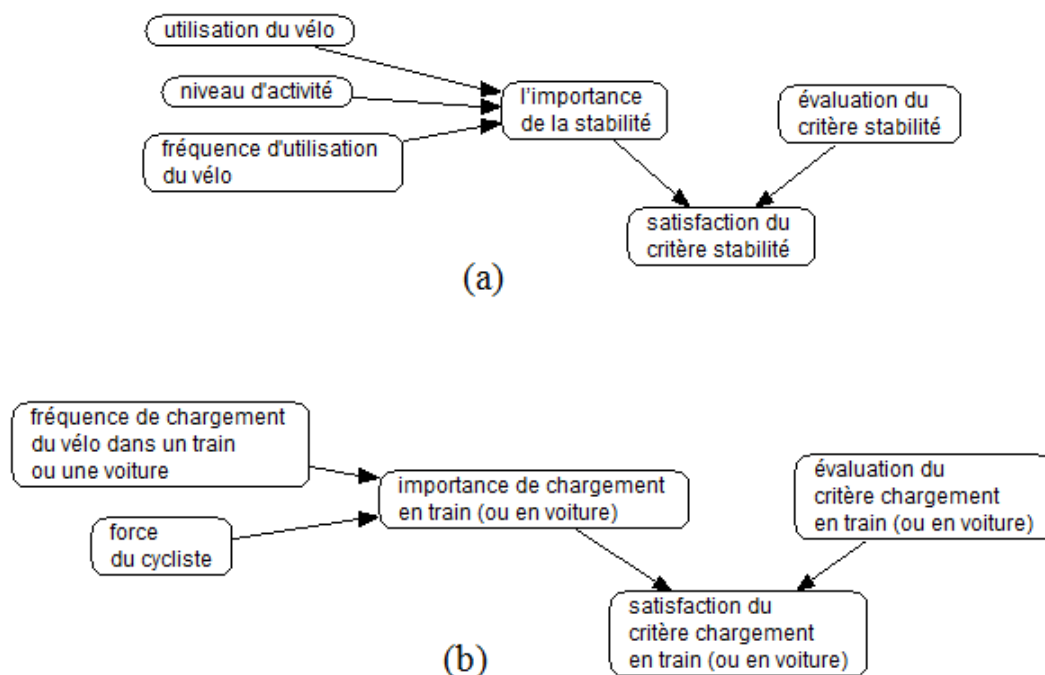


FIGURE 4.6 – Graphe du réseau bayésien local des agents associés (a) au critère de stabilité, (b) au critère de confort de chargement du vélo dans un véhicule ou en train.

qui est par nature une information incertaine. Une première source d'incertitude dans l'évaluation de l'importance des critères est dû au fait que les informations qui décrivent le cas de décision sont souvent incomplètes. La deuxième raison est que l'incertitude de l'évaluation de l'importance des critères est basée sur une synthèse des connaissances par les experts pour le choix d'un vélo.

Finalement, expliquons en détail pourquoi les distributions de probabilité échangées entre les agents doivent rester fixes. Considérons les informations envoyées par deux agents critères à l'agent expert du vélo. Dans le réseau bayésien local de l'agent vélo, les variables d'entrée (g_1 et g_2 par exemple) qui reçoivent des informations représentent les niveaux de *qualité souhaités* pour les critères 1 et 2. Ces distributions de probabilités ne dépendent que des caractéristiques du cas de décision, et ne doivent pas être modifiées par la nécessité du compromis entre les critères. Or dans le réseau bayésien local de l'agent spécialiste du vélo, les variables g_1 et g_2 ne sont en

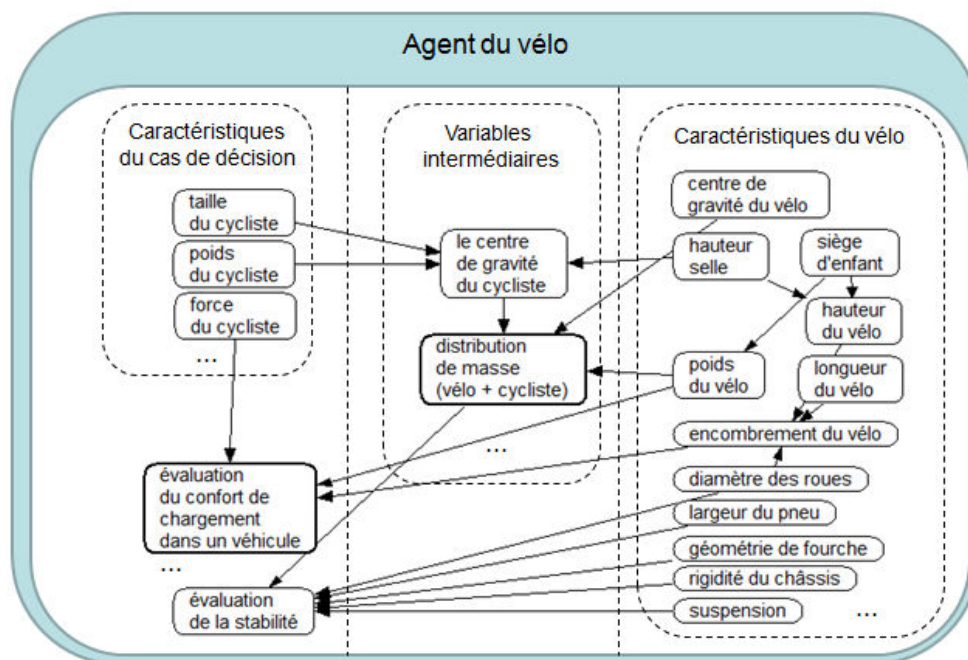


FIGURE 4.7 – Extrait du réseau bayésien local de l'agent "expert du vélo".

général pas d-séparées par les autres variables observées¹. Donc si les distributions de probabilités reçues sur g_1 et g_2 ne sont pas maintenues fixes, le niveau de qualité souhaité pour le critère 2 pourrait modifier le niveau de qualité souhaité pour le critère 1. Ce qui n'est pas correct vu que les variables g_i représentent les niveaux de qualité attendus pour chaque critère, et non pas l'évaluation d'un vélo précis.

Le système de recommandation mis en oeuvre par l'agent vélo consiste donc à propager l'ensemble des attentes sur le niveau de qualité de chaque critère, sous la forme d'un ensemble d'observations probabilistes fixes (une par critère). Une explication détaillée du fonctionnement du système de recommandations se trouve dans (DSL13).

Dans cette section, nous avons présenté comment les observations probabilistes sont utilisées dans les structures composées d'agents équipés d'un réseau bayésien local, pour permettre à un agent de prendre en compte dans son modèle de raisonnement une information fournie par un autre agent expert sur une variable partagée. Ensuite, nous avons montré comment utiliser un tel modèle à base d'agents équipés

1. il n'est pas possible de montrer que tous les chemins entre g_1 et g_2 sont bloqués.

de réseaux bayésiens locaux pour un problème récurrent d'aide à la décision multi-critères et nous avons illustré cette problématique avec le problème du choix d'un vélo. Nous avons expliqué pourquoi les observations probabilistes fournies par un agent à un autre qui les lui demandent doivent rester fixes.

En perspective, il nous semble intéressant d'implémenter complètement cette application avec des agents équipés d'un réseau bayésien local

4.3 Observations sur des variables continues dans un réseau bayésien discret

Dans cette partie, nous présentons un cas d'utilisation d'observations probabilistes fixes pour prendre en compte une observation certaine sur une variable continue. Lorsque cette variable est discrétisée grossièrement (c'est à dire avec peu d'intervalles) dans le réseau bayésien, le fait de traduire l'observation certaine en une distribution de probabilité sur l'ensemble des intervalles peut permettre de compenser partiellement la perte d'information liée à la discrétisation grossière.

La discrétisation des variables continues dans un réseau bayésien est un compromis entre trois critères qui n'ont pas toujours la même importance :

- (1) La qualité de l'information : la discrétisation doit éviter ou minimiser toute perte d'information du modèle. Les intervalles doivent être définis de telle sorte que, si deux valeurs différentes pour un nœud observé fournissent des résultats différents pour les nœuds cibles, ils doivent être contenus dans deux intervalles différents. Ce critère conduit à préférer un nombre d'intervalles de discrétisation important.
- (2) La qualité statistique : la discrétisation doit assurer qu'il ya suffisamment d'échantillons dans chaque intervalle ; Ce critère conduit à choisir un petit nombre d'intervalles de discrétisation, surtout lorsque la base de données contient un petit nombre d'échantillons.
- (3) L'efficacité des calculs : cet aspect est d'autant plus important que la variable discrétisée a plusieurs nœuds parents et / ou des nœuds fils, et que la taille globale du réseau bayésien est grande. Ce critère conduit à préférer des petits nombres d'intervalles de discrétisation.

Dans l'exemple que nous avons traité, c'est le deuxième critère qui est prépondérant, et conduit au choix d'une discrétisation grossière.

Dans la suite, nous expliquons le principe de la discrétisation floue des variables continues. Puis, nous présentons en détail une application concrète de ce principe sur un réseau bayésien simple associé au problème du diagnostic du cancer de la prostate. L'objectif des tests réalisés est de comparer d'une part les résultats obtenus par des réseaux bayésiens avec et sans observation probabiliste, et d'autre part de comparer les résultats obtenus avec d'autres classifieurs classiques. Après avoir présenté le problème du cancer de la prostate sur lequel nous avons expérimenté notre approche, nous décrivons les variables et la structure du réseau bayésien construit pour ce problème à l'aide des données réelles de patients de l'hôpital Habib Bourguiba de Sfax ; nous expliquons ensuite la méthodologie suivie pour réaliser les tests et nous présentons les autres classifieurs utilisés pour la comparaison. Enfin, nous présentons les résultats et le bilan de ces tests.

Toutes les expériences ont été effectuées sur une machine d'un processeur *Intel^R CoreTM i3*, 2.27 GHz avec une RAM de 3 Go. Les programmes ont été implémentés en matlab en utilisant "Bayes Net Toolbox" (BNT).

4.3.1 Principe

La discrétisation d'une variable continue nécessite de déterminer le nombre d'intervalles et l'ensemble des seuils. Ces choix sont parfois un compromis difficile. En effet, dans le cas d'un réseau bayésien discret, le nombre d'états des variables, ainsi que le nombre des parents d'un nœud détermine la taille de la table de probabilités conditionnelles de ce nœud, et donc la taille du modèle. Or, la taille du modèle est un facteur important pour l'apprentissage et l'inférence dans le réseau bayésien. A l'opposé, une discrétisation grossière peut conduire à une perte de précision.

La solution proposée dans (TB08) pour propager des observations sur les variables continues dans un réseau bayésien discret consiste à choisir une discrétisation avec un petit nombre d'intervalles, et à propager des observations probabilistes fixes sur ces états pour représenter des observations certaines sur les variables continues. Ainsi, les observations certaines sur les variables continues seront remplacées par des observations probabilistes sur les mêmes variables après discrétisation.

Les observations probabilistes sont obtenues par discrétisation floue.

Une *partition floue* sur un univers Ω est un ensemble d'états flous $\{f_1, \dots, f_p\}$ tel que $\forall x \in \Omega, \sum_{i=1}^p X_{f_i}(x) = 1$ où X_{f_i} est la fonction d'appartenance de f_i , c'est à dire une fonction $X_{f_i} : \Omega \longrightarrow [0, 1]$.

L'ensemble des intervalles de la discrétisation de la variable continue est remplacé par un ensemble d'états flous $\{f_1, \dots, f_p\}$ tel que défini ci-dessus. L'observation initiale sur la variable continue : $X = x$ est transformée en une distribution de probabilités $R(X_d) = (X_{f_1}(x), \dots, X_{f_p}(x))$.

L'utilisation d'une observation probabiliste fixe garantit que la distribution de probabilité de la variable observée ne sera pas modifiée ensuite par des observations sur d'autres variables. Ceci est nécessaire car l'observation certaine initiale sur la variable continue signifie que la valeur observée est connue de manière certaine.

Dans la suite, nous présentons un exemple d'application de la discrétisation floue des variables continues sur un réseau bayésien concernant le problème du diagnostic du cancer de la prostate.

4.3.2 Problème du cancer de la prostate

Le cancer de la prostate, ou adénocarcinome ([ASBW70](#)) est actuellement le cancer le plus fréquent chez l'homme et constitue la seconde cause de mortalité par cancer chez l'homme. En 1970, le dosage de PSA (prostate spécifique antigène) a permis une avancée considérable dans le diagnostic du cancer de la prostate. Pour suivre l'évolution des patients traités pour ce cancer, les médecins utilisent le dosage sanguin du PSA. Ainsi, une augmentation de son taux dans le temps signale une progression de la maladie. Néanmoins, aucun chiffre de taux de PSA ne peut donner la certitude que la prostate est normale et qu'il n'y a pas de cancer. Il peut s'agir d'une prostatite (infection aiguë ou chronique de la prostate), ou d'un adénome (tumeur bénigne).

A partir d'un ensemble de symptômes, le médecin qui craint que son patient soit atteint de ce cancer, doit décider de faire ou non des examens complémentaires comme une biopsie de la prostate. Or il arrive que ce type d'examen soit demandé de manière abusive, car la démarche mise en jeu pour le diagnostic du cancer de la prostate reste mal connue du clinicien ([Hak08](#)).

Un diagnostic précoce de ce cancer est très intéressant pour l'éradiquer dans un stade primaire, car les traitements sont d'autant plus efficaces que la maladie est

peu développée. Une meilleure évaluation du risque de cancer de la prostate permet d'éviter de prescrire des tests supplémentaires tels que la biopsie de la prostate lorsque le risque est considéré comme assez faible.

4.3.3 Description du réseau bayésien

Dans le cadre de notre travail, nous avons travaillé avec les professeurs en urologie à l'hôpital Habib Bourguiba de Sfax (Tunisie). Les entretiens avec les médecins et la littérature disponible sur ce sujet nous ont permis d'extraire les variables et la structure du modèle. Nous utilisons un réseau bayésien causal car il est plus facile à comprendre et à valider par les experts médicaux. Les variables concernent les données que les médecins rassemblent pour donner un premier diagnostic. La première question du médecin concerne l'âge du patient. En effet, la probabilité d'avoir une tumeur de la prostate augmente avec la progression de l'âge. A l'inverse, plus le patient est jeune plus le médecin pense à une prostatite. Le *toucher rectal* donne les informations sur le volume de la prostate et son aspect. Le médecin note aussi les plaintes du patient : le *prostatisme* est un ensemble des troubles urinaires dus à l'hypertrophie de la prostate. La valeur de PSA est un élément clé du diagnostic. Voici les variables du modèle et leur domaine de définition :

- Diagnostic : {pas de cancer, cancer}
- Toucher rectal (TR) : {suspect, non-suspect}
- Prostatisme : {oui, non}
- Age : [50,100]
- PSA : [0, 2000]

Les deux variables Age et PSA sont des variables continues. Différents choix de discrétisation de ces deux variables sont décrits et comparés dans la suite, avec et sans observations probabilistes.

Avec ces cinq variables, nous avons défini la structure du réseau bayésien avec l'aide de spécialistes (voir figure 4.8).

Afin de définir les paramètres du modèle, nous avons rassemblé une collection rétrospective de 61 fiches provenant de 180 patients de l'hôpital Habib Bourguiba de Sfax, ayant une prostatite, un adénome ou un adénocarcinome. Seules les fiches suffisamment complètes par rapport aux variables retenues ont été conservées. Deux

groupes de patients ont été sélectionnés : 39 patients n’ayant pas de cancer (soit prostatite ou adénome), et 22 patients ayant un cancer (adénocarcinome).

Nous avons défini les probabilités conditionnelles des nœuds du réseau bayésien par un apprentissage automatique des paramètres avec le logiciel Netica (Nor14), avec l’algorithme EM (N⁺04). Vu le petit nombre de données disponibles, nous avons utilisé la méthode de la validation croisée avec 10 paquets.

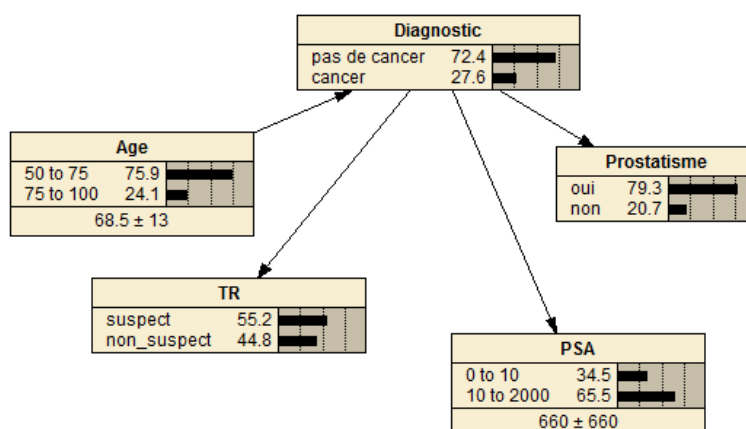


FIGURE 4.8 – Réseau Bayésien modélisant le cancer de la prostate

4.3.4 Méthodologie et tests

L’objectif des tests suivants est d’étudier l’utilisation des observations probabilistes pour propager des observations sur des variables continues dans un réseau bayésien discret. Nous avons défini trois réseaux bayésiens avec la même structure (voir figure 4.8) et correspondant chacun à une discrétisation différente pour les variables continues. Les variables continues *Age* et *PSA* sont discrétisées grossièrement dans le premier réseau bayésien (RB1), et un peu plus finement dans les RB2 et RB3. Les trois variables discrètes sont identiques dans les trois réseaux bayésiens. La table 4.1 présente les trois réseaux bayésiens utilisés dans les tests avec le nombre d’intervalles de discrétisation des variables continues et les types d’observations utilisées sur ces variables : observations certaines et / ou observation(s) probabiliste(s) fixe(s).

La variable *Age* est discrétisée avec la méthode de *equal width discretization* (DKS⁺95) dans le réseau bayésien RB1 en deux intervalles et dans les réseaux bayé-

TABLE 4.1 – Les trois réseaux bayésiens utilisés dans les tests (nombre d’intervalles de discrétisation et type d’observations)

	Discrétisation <i>Age</i>	Discrétisation <i>PSA</i>	nombre de paramètres	Type d’observation pour les variables <i>Age</i> et <i>PSA</i>
RB1	2 intervalles	2 intervalles	18	observations certaines et observations probabilistes (deux tests)
RB2	4 intervalles	2 intervalles	24	observations certaines
RB3	4 intervalles	4 intervalles	28	observations certaines

siens RB2 et RB3 en 4 intervalles. Dans les réseaux bayésiens RB1 et RB2, la variable *PSA* est discrétisée en deux intervalles suivant le seuil de 10 généralement admis comme seuil du risque de cancer de la prostate. Dans le réseau bayésien RB3 la variable *PSA* est discrétisée en quatre intervalles suivant les seuils de 4, 10 et 100. Ce choix des seuils est basé sur l’expérience des urologues : la valeur de *PSA* est généralement inférieure à 4 ng/ml pour une personne en bonne santé. La valeur 10 ng/ml est généralement acceptée par les médecins spécialistes comme un seuil au-delà duquel le risque de cancer de la prostate devient élevé, lorsqu’il est combiné avec d’autres symptômes. Un taux de *PSA* supérieure à 100 ng/ml est certainement du à un cancer de la prostate. Le choix de la discrétisation a un impact direct sur la taille du modèle. Le nombre de paramètres dans un réseau bayésien discret est la taille totale des tables de probabilités conditionnelles. Dans notre étude, le réseau bayésien RB1 possède 18 paramètres alors que le réseau bayésien RB3 possède 28 paramètres : 4 (taille de la TPC *Age*) + 2×4 (taille de la TPC Diagnostic) + 2×4 (taille de la TPC *PSA*) + 2×2 (taille de la TPC *TR*) + 2×2 (taille de la TPC *Prostatisme*) = 28.

Par la suite, nous comparons les résultats de classification des réseaux bayésiens RB1, RB2 et RB3, en propageant uniquement des observations certaines (dans les réseaux bayésiens RB1, RB2 et RB3) ou des observations probabilistes sur les variables *Age* et *PSA* et des observations certaines sur les variables *TR* et *Prostatisme* (dans le réseau bayésien RB1).

Nous présentons maintenant la discrétisation floue utilisée pour obtenir des observations probabilistes sur les variables *Age* et *PSA*. Pour la variable *PSA*, on

considère deux états flous qui correspondent aux taux de PSA en faveur ou non du cancer de la prostate. La figure 4.9 donne la fonction d'appartenance à l'ensemble flou correspondant à un taux de PSA en faveur du cancer de la prostate. Cette fonction est issue des données statistiques présentes dans la littérature (Hak08). A partir du taux de PSA observé chez un patient, cette fonction permet de définir la distribution de probabilité sur les deux états flous de la variable PSA . Cette distribution constitue une observation probabiliste fixe.

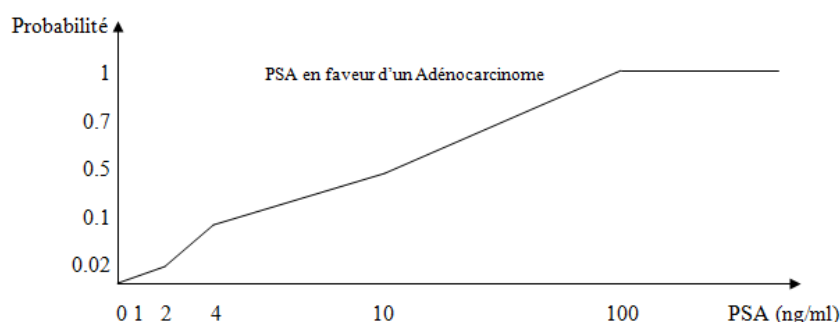


FIGURE 4.9 – Fonction d'appartenance associée à l'état flou "PSA en faveur d'un cancer de la prostate"

Du fait que la discrétisation floue de la variable PSA ne contient que deux états flous, la fonction d'appartenance associée au 2^{ème} état flou peut être déduite de la fonction d'appartenance associée au premier état flou. On peut remarquer que la valeur de $PSA = 10$ correspond à un degré d'appartenance de 0.5, ce qui détermine le point de croisement des courbes des fonctions d'appartenance des deux états flous.

Le tableau 4.2 donne le vecteur d'observation obtenu après chaque type de discrétisation de la variables PSA pour deux valeurs observées ($PSA = 10.3$ et $PSA = 6.8$). La distribution de probabilité obtenue à partir de la discrétisation floue est calculée grâce à la fonction d'appartenance représentée sur la figure 4.9.

Nous avons utilisé les trois réseaux bayésiens décrits ci-dessus, avec et sans observations probabilistes pour prédire le cancer de la prostate. Nous avons comparé les résultats obtenus avec ceux obtenus par d'autres types de classifieurs classiques : les arbres de décision, les K plus proches voisins, et les machines à vecteur de support. Avant de présenter les résultats obtenus, nous présentons brièvement ces différents classifieurs.

TABLE 4.2 – Vecteur d’observation obtenu après discrétisation classique et floue pour deux valeurs observées de la variable *PSA*

	Observation certaine		Observation probabiliste (discrétisation floue)
	(RB1)	(RB2)	(RB1)
Domaine de PSA	{”<10”, ”>=10”}	{”<4”, [4,10[, [10,100[, ”>=100”}	{”PSA en faveur du cancer”, ”PSA en défaveur du cancer”}
$PSA = 10.3$	(0, 1)	(0, 0, 1, 0)	(0.498, 0.502)
$PSA = 6.8$	(1, 0)	(0, 1, 0, 0)	(0.712, 0.288)

4.3.5 Autres algorithmes de classification pour le diagnostic du cancer de la prostate

Les classifieurs que nous présentons ci-dessous nécessitent de choisir les valeurs de divers paramètres. Pour cela, nous avons testé différentes valeurs des paramètres et retenu celles qui donnent les meilleurs résultats.

4.3.5.1 Arbres de décision

Les arbres de décision, font partie des méthodes de classification dites non paramétriques. Aucune hypothèse n’est faite sur la distribution conditionnelle de probabilité du modèle. C’est à partir d’une approche structurelle des données que les frontières de décisions dans l’espace des attributs sont déterminées.

La procédure de classification est apprise de manière hiérarchique, sous la forme d’un arbre où les nœuds correspondent aux attributs, et les feuilles portent les différentes classes du problème. La méthodologie employée pour construire l’arbre de décision consiste à séparer un nœud en plusieurs branches par rapport à un attribut donné, de manière à maximiser la proportion d’objets appartenant à une même classe dans chacune des branches formées. Ce procédé est répété de manière récursive sur chacun des nœuds fils obtenus, jusqu’à obtention de sous-régions de l’espace des attributs dans lesquelles les exemples appartiennent tous à une même classe. Ce nœud est alors remplacé par une feuille étiquetée par la classe à laquelle appartiennent les objets y figurant. La division d’un nœud en plusieurs branches s’effectue généralement de façon binaire. Dans notre travail, nous avons utilisé l’algorithme CART (BFSO84) pour apprendre l’arbre de décision.

4.3.5.2 K Plus Proches voisins

L'algorithme des K Plus Proches voisins (KPP) (Das91) utilise une mesure de distance pour définir la notion de proximité entre deux cas (ou individus). Cet algorithme permet de classer un échantillon en identifiant les individus présentant le plus de similitudes. Dans les tests qui suivent, nous avons utilisé l'algorithme KPP avec la distance euclidienne et un nombre de voisins $K = 14$ car cette valeur offre les meilleures performances du classifieur avec le jeu de données que nous avons utilisé.

4.3.5.3 Machines à Vecteur de Support

Les Machines à Vecteurs de Support (MVS ou SVM en anglais) (CV95) constituent une famille d'algorithmes d'apprentissage permettant de contrôler la capacité de généralisation des modèles qu'ils produisent. En s'appuyant sur le principe de minimisation du risque structurel, cette famille d'algorithmes formule l'apprentissage sous la forme d'un problème d'optimisation. Les MVS constituent une approche intéressante au problème de classification, et ont été utilisées avec succès à des applications issues du monde réel.

Les choix de la fonction noyau, de ses paramètres et de la valeur du coût de dépassement des contraintes "C" sont fixés avant de résoudre le problème d'optimisation. Les noyaux polynomiaux peuvent donner des temps de résolution excessivement longs. On préférera en général utiliser un noyau Gaussien pour obtenir plus rapidement la solution optimale du problème. Une faible valeur pour "C" facilite la résolution du problème d'optimisation au détriment de l'optimalité de la solution obtenue. À l'inverse, la complexité du problème étant fonction du nombre de contraintes, une valeur trop importante pour "C" augmente de manière significative le temps de calcul.

Dans notre travail, les paramètres définis pour le MVS sont "la fonction noyau" avec la valeur "fonction noyau Gaussienne" et "l'implémentation multi-classe" avec la valeur "one-versus-one". Nous avons fixé la valeur du coût de dépassement des contraintes "C" à 1 et la valeur du paramètre noyau à 10^{-3} .

4.3.6 Résultats

Les résultats des quatre séries de tests sur les réseaux bayésiens décrits ci-dessus sont présentés maintenant. Chacun des 61 échantillons a été testé en utilisant la validation croisée avec 10 paquets en insérant les observations certaines ou probabilistes sur *Age*, *PSA*, *Prostatisme* et *TR* suivant la table 4.1. Pour chaque test, le diagnostic retenu est celui qui a la probabilité *a posteriori* la plus élevée.

La table 4.3 présente les résultats de classification des réseaux bayésiens (RB1, RB2 et RB3) utilisés avec uniquement des observations certaines. Les résultats sont présentés sous forme de matrice de confusion² avec dans chaque case le nombre de cas bien ou mal classés ; la somme des deux nombres sur chaque diagonale (cases grisées) donne le nombre de cas bien classés par ce réseau bayésien, sur un total de 61 cas.

Rappelons que la seule différence entre ces trois réseaux bayésiens concerne la discrétisation des deux variables continues *Age* et *PSA*, avec la discrétisation la plus grossière pour le réseau bayésien RB1 et la plus fine pour le réseau bayésien RB3. Ces résultats indiquent des performances très proches pour ces trois classifieurs, avec un cas de plus bien classé entre les réseaux bayésiens RB1 et RB2 et un cas de plus de bien classé entre les réseaux bayésiens RB2 et RB3. Il n'est cependant pas possible d'améliorer ces résultats en affinant la discrétisation des variables à cause du trop petit nombre de cas disponibles pour l'apprentissage des paramètres du réseau bayésien.

La table 4.4 illustre les résultats issus du réseau bayésien RB1 en appliquant deux observations probabilistes (sur *PSA* et *Age*) et des observations certaines (sur *TR* et *Prostatisme*).

Les résultats du réseau bayésien RB1 en utilisant des observations probabilistes pour les variables continues *Age* et *PSA* sont meilleurs que ceux obtenus avec des observations classiques, même sur des réseaux bayésiens avec une discrétisation plus fine. La matrice de confusion du réseau bayésien RB1 utilisé avec des observations probabilistes affiche 3 cas bien classés de plus que le même réseau bayésien avec des observations classiques, (5% de plus) et 1 cas de plus que le réseau bayésien RB3.

2. Il s'agit d'un tableau de contingence confrontant les états obtenus (colonnes) et les états désirés (lignes) pour l'échantillon. Sur la diagonale, on trouve donc les valeurs bien classées, hors de la diagonale les éléments mal classés ; la somme des valeurs sur une ligne donne le nombre d'exemplaires théoriques de la catégorie.

TABLE 4.3 – Matrices de confusion obtenues en appliquant uniquement des observations certaines dans chaque réseau bayésien

	RB1		RB2		RB3	
Cas réel	pas de cancer	cancer	pas de cancer	cancer	pas de cancer	cancer
pas de cancer	29	4	29	3	30	3
cancer	7	21	7	22	6	22

TABLE 4.4 – Matrice de confusion obtenues avec le réseau bayésien RB1 en appliquant deux observations probabilistes (*)

	RB1 (*)	
Cas réel	pas de cancer	cancer
pas de cancer	31	3
cancer	5	22

Afin de comparer ces résultats avec ceux d'autres classifieurs, nous avons étendu les tests en utilisant les algorithmes de classification suivants : KPP, arbre de décision et machine à vecteurs de support. Comme pour les réseaux bayésiens, chacun des 61 cas a été classé à l'aide d'un classifieur appris à partir d'un ensemble de données d'apprentissage constitué de 55 cas (90% des données). Avec nos données, tous ces algorithmes fournissent les mêmes résultats en termes de classification. La matrice de confusion commune est présenté dans la table 4.5.

TABLE 4.5 – Résultats de classification avec des observations classiques dans différents réseaux bayésiens

	KPP, arbre de décision et MVS (identique)	
Cas réel	pas de cancer	cancer
pas de cancer	29	2
cancer	7	23

Dans cette étude, les résultats fournis par les classifieurs classiques sont très similaires à ceux fournis par les réseaux bayésiens avec des observations classiques, et moins bons que les résultats du réseau bayésien RB1 avec des discrétisations floues

pour les variables continues. Afin de donner une vue synthétique des résultats, nous avons rassemblé dans la table 4.6 le pourcentage de bonne classification de chaque classifieur, et le coefficient kappa qui évalue la qualité de la concordance pour chaque test.

TABLE 4.6 – Pourcentage de bonne classification et test de concordance (Kappa)

Test	RB1	RB2	RB3	RB1 *	KPP, arbre de décision, MVS
Pourcentage de bonne classification (%)	82	84	85	87	85
coefficient de concordance (κ)	0.63	0.67	0.7	0.73	0.7

Le meilleur pourcentage de bonne classification est obtenu avec le réseau bayésien RB1, où on a appliqué une discrétisation grossière avec deux états flous pour les variables continues et propagé des observations probabilistes. Ce même réseau bayésien utilisé avec seulement des observations certaines fournit des résultats moins bons que tous les autres classifieurs considérés ici. D'une part, l'utilisation de seulement deux intervalles pour la discrétisation des variables continues (RB1) conduit à une perte de précision, comparativement aux réseaux bayésiens RB2 et RB3 où la discrétisation se fait avec plus d'intervalles. D'autre part, l'augmentation du nombre d'intervalles dans la discrétisation conduit à un nombre de paramètres du réseau bayésien plus élevé, ce qui nécessite des données supplémentaires pour l'apprentissage. Un plus petit nombre de paramètres du réseau bayésien permet un meilleur apprentissage des paramètres, mais peut conduire à une perte de précision, qui peut être partiellement compensée à l'aide des observations probabilistes fixes.

Le test du kappa permet d'évaluer la concordance entre le classifieur et le cas réel. Le coefficient de kappa de Cohen est défini par : $\kappa = \frac{P_o - P_e}{1 - P_e}$ tel que P_o est l'accord observé (la concordance observée) et P_e est l'accord de chances prévu (la concordance théorique). Il est calculé à partir des données des tables 4.3, 4.4 et 4.5. Dans notre exemple, pour le test sur le RB1(*) (table 4.4), on a :

$$P_o = (31 + 22)/61 = 0.8668$$

$$P_e = (34 \times 36 + 27 \times 25)/61^2 = 0.5103$$

$$\kappa = (0.8668 - 0.5103)/(1 - 0.5103) \approx 0.73$$

Situé dans l'intervalle $[0.6, 0.8]$, ce coefficient kappa prouve une très bonne concordance entre les deux classifications. D'après ces résultats, l'utilisation des observations probabilistes pour propager des observations sur des variables continues dans un réseau bayésien discret semble intéressant, en particulier dans les cas où la quantité limitée de données ne permet pas d'utiliser une discrétisation assez fine pour les variables continues et conduit à une perte de précision.

La transformation d'une observation certaine sur une variable continue en une observation probabiliste fixe permet de garantir que cette observation ne sera pas modifiée par d'autres informations, ce qui est nécessaire, de la même façon qu'une observation certaine n'est pas remise en cause par d'autres informations. Ni les observations de vraisemblance ni les observations probabilistes non fixes ne garantissent cette propriété.

4.4 Conclusion

Dans ce chapitre, nous avons proposé trois types d'exemples d'utilisation des observations probabilistes dans les réseaux bayésiens. Dans la première section, nous avons présenté une application qui concerne l'intégration des réseaux bayésiens avec des systèmes d'information géographique et nous avons montré l'intérêt d'utiliser des observations probabilistes fixes pour ce type d'applications. Dans la deuxième section, nous avons présenté comment les observations probabilistes sont utilisées dans les structures composées d'agents équipés d'un réseau bayésien local, pour permettre à un agent de prendre en compte dans son modèle de raisonnement une information fournie par un autre agent expert sur une variable partagée. Nous avons illustré un exemple d'utilisation d'un modèle à base d'agent équipés de réseaux bayésiens locaux pour un problème récurrent d'ADMC. Dans la troisième section, nous avons présenté un cas d'utilisation d'observation probabiliste pour prendre en compte une observation certaine sur une variable continue. Lorsque cette variable est discrétisée grossièrement dans le réseau bayésien et le fait de traduire l'observation certaine en une distribution de probabilité sur l'ensemble des intervalles peut permettre de compenser partiellement la perte d'information liée à la discrétisation grossière.

Chapitre 5

Conclusion générale et perspectives

Ce dernier chapitre propose une synthèse des contributions de ce travail, puis souligne certaines limites et propose diverses perspectives pour l'avenir.

5.1 Conclusion

Cette thèse porte sur les observations incertaines dans les réseaux bayésiens, c'est à dire les observations qui ne peuvent pas être représentées simplement par l'instanciation de certaines variables du réseau bayésien, comme c'est le cas lorsque l'observateur peut affirmer avec certitude que la variable est dans l'état observé. Cette thèse clarifie et étudie les différents concepts d'observations incertaines portant sur une variable dans un réseau bayésien et menant à réviser les croyances sur d'autres variables du réseau bayésien. Elle propose également différents exemples et applications des observations incertaines dans les réseaux bayésiens.

De nombreux ouvrages sur les réseaux bayésiens éludent totalement la question des observations incertaines et il en est de même pour les logiciels sur les réseaux bayésiens. Le chapitre 2 dresse un état des lieux sur les observations incertaines dans les réseaux bayésiens dans la littérature et dans les logiciels, en termes de terminologie, de définition, de spécification et de propagation. Il en ressort que les articles sont essentiellement centrés sur les méthodes de propagation et non pas sur les notions, définitions, caractéristiques et comparaison des différents types

d'observations incertaines dans les réseaux bayésiens. Divers points concernant la définition précise et le vocabulaire de ces observations restent incomplets et donnent lieu à des confusions.

Trois types d'observations incertaines dans les réseaux bayésiens sont identifiés et une terminologie est définie. Le chapitre 3 présente les concepts, définitions, propriétés et exemples de chacun des types d'observations incertaines puis les méthodes de propagation. L'observation de vraisemblance est une observation imprécise ou fournie par une source non fiable, elle est spécifiée par un rapport de vraisemblance et propagée par la méthode de Pearl de l'observation virtuelle. Cette méthode consiste à transformer une observation de vraisemblance sur une variable X en une observation certaine sur un nouveau nœud virtuel, ajouté en tant que fils du nœud X . L'observation probabiliste fixe ou non-fixe exprime une contrainte sur l'état de certaines variables après que cette information soit propagée dans le réseau bayésien. Une observation probabiliste sur X est spécifiée par une distribution de probabilité $R(X)$ définie "tout bien considéré", ce qui signifie qu'elle remplace toutes croyances précédentes et qu'elle intègre toutes les connaissances qu'on a sur X . Une observation probabiliste non-fixe peut être modifiée, par la suite, par d'autres observations sur n'importe quels nœuds du réseau bayésien, y compris le nœud X . Elle peut être propagée par la conversion en observation de vraisemblance. Une observation probabiliste fixe sur une variable X ne peut être modifiée par aucune information supplémentaire sur les variables du modèle. Ainsi, la propagation de plusieurs observations probabilistes fixes est commutative.

Le chapitre 4 propose trois types d'exemples d'utilisation des observations probabilistes fixe dans les réseaux bayésiens. Premièrement, nous avons présenté une application des observations probabilistes fixes qui consiste à représenter des observations faites sur une sous population et non sur un individu. C'est le cas avec les systèmes d'information géographique et nous avons montré l'intérêt d'utiliser des observations probabilistes fixes pour ce type d'applications. Deuxièmement, nous avons présenté comment les observations probabilistes sont utilisées dans les structures distribuées composées d'agents équipés d'un réseau bayésien local. Les observations probabilistes fixes représentent l'information envoyée par un agent "expert" sur ces variables à un agent "demandeur" sur ces mêmes variables. Elles permettent à l'agent demandeur de mettre à jour son modèle local de raisonnement en propageant l'information reçue. Nous avons illustré l'utilisation d'un modèle à base

d'agents équipés de réseaux bayésiens locaux pour produire un système de recommandations dans un problème récurrent d'aide à la décision multi-critères. Troisièmement, nous avons présenté un cas d'utilisation d'observations probabilistes fixes pour prendre en compte une observation certaine sur une variable continue dans un réseau bayésien discret. Lorsque cette variable est discrétisée grossièrement dans le réseau bayésien, le fait de traduire l'observation certaine en une distribution de probabilités sur l'ensemble des intervalles peut permettre de compenser partiellement la perte d'information liée à la discrétisation grossière. Nous avons implémenté l'algorithme BN-IPFP-1 qui permet de propager des observations probabilistes fixes dans les réseaux bayésiens. Nous avons étudié un exemple médical qui concerne le diagnostic du cancer de la prostate. Les tests réalisés permettent de comparer d'une part les résultats obtenus par des réseaux bayésiens avec et sans observations probabilistes fixes pour représenter les observations sur les variables continues du réseau bayésien, et d'autre part les résultats obtenus avec des réseaux bayésiens et avec d'autres classifieurs classiques. Les résultats obtenus en utilisant des observations probabilistes fixes pour les variables continues sont légèrement meilleurs que ceux obtenus avec des observations classiques ou par d'autre classifieurs.

5.2 Limites et Perspectives

En perspective à court terme et en ce qui concerne les tests réalisés pour la propagation des observations sur des variables continues avec une discrétisation floue, nous projetons d'étendre les tests avec d'autres jeux de données de taille plus importante et avec des réseaux bayésiens de plus grandes taille, et de structures plus complexes, afin de mieux étudier dans quels cas cette proposition présente le plus de bénéfices.

Nous envisageons d'implémenter l'application qui concerne l'intégration des réseaux bayésiens avec des systèmes d'information géographique (SIG). Nous continuons la collaboration avec Philip Leicester, de l'université de Loughborough, pour l'aide à la décision concernant les politiques d'énergies renouvelables, à l'aide de réseau bayésien intégré dans un système d'information géographique. Cette étude vise à montrer l'intérêt des observations probabilistes dans ce cadre en comparaison avec d'autres méthodes intégrant des réseaux bayésiens dans des systèmes d'information géographique, mais n'utilisant que des observations classiques. Ces travaux

concernent aussi les problèmes d'aide à la décision multi-critères, car le choix concernant les politiques sur les énergies renouvelables repose sur des critères sociaux économiques, environnementaux, etc. Nous souhaitons aussi développer le modèle à base d'agents équipés de réseaux bayésiens locaux pour tester le modèle de recommandation proposé.

Il est à noter que l'étude des algorithmes de propagation des observations probabilistes fixes est traitée dans plusieurs articles. Cependant, plusieurs logiciels sur les réseaux bayésiens ne disposent d'aucune fonctionnalité concernant la propagation d'observations incertaines. Une perspective consiste à ajouter, dans les logiciels sur les réseaux bayésiens, la possibilité de propager les trois types d'observations incertaines, avec un effort pédagogique particulier pour permettre aux utilisateurs de comprendre le plus facilement possible le sens de chaque type d'observation¹. Un autre aspect pratique concerne la possibilité de cumuler plusieurs observations incertaines sur une même variable lorsque cela est possible².

Les perspectives d'implémentation des observations probabilistes dans les logiciels présentent un deuxième niveau de fonctionnalités pour permettre aux utilisateurs des réseaux bayésiens d'exploiter facilement les trois usages possibles des observations probabilistes fixes à savoir :

- rendre possible le fait d'insérer des observations sur une sous population au lieu d'un cas unique,
- proposer à l'utilisateur la transformation en observations probabilistes fixes lorsqu'il s'agit d'insérer des observations sur des variables continues discrétisées,
- implémenter l'utilisation d'agents équipés d'un réseau bayésien local et leurs échanges.

Une autre perspective concerne l'étude des indépendances conditionnelles et d-séparation avec des observations probabilistes.

Une dernière perspective consiste à étudier les observations probabilistes dans les différentes extensions des réseaux bayésiens. En particulier, la comparaison des différentes propositions des réseaux bayésiens flous avec les observations incertaines ; comme nous l'avons détaillé dans l'article ([BMDP⁺15](#)) (partie 7.2), certains articles

1. à ce propos remarquons les travaux réalisés dans le logiciel NETICA pour la saisie des observations de vraisemblance et les observations probabilistes non fixes (appelées calibration).

2. à ce propos, à notre connaissance, seul le logiciel NETICA qui permet de combiner plusieurs observations de vraisemblance sur la même variable.

traitant les réseaux bayésiens flous visent explicitement à prendre en compte des observations incertaines (DPC⁺14b).

Bibliographie

- [AMD08] Steven C Amstrup, Bruce G Marcot, and David C Douglas. A bayesian network modeling approach to forecasting the 21st century worldwide status of polar bears. *Arctic sea ice decline : observations, projections, mechanisms, and implications*, pages 213–268, 2008.
- [ASBW70] R J Ablin, W A Soanes, P Bronson, and E Witebsky. Precipitating antigens of the normal human prostate. *J Reprod Fertil*, 22(3) :573–4, 1970.
- [BDT03] Jim F Baldwin and Enza Di Tomaso. Inference and learning in fuzzy bayesian networks. In *Fuzzy Systems, 2003. FUZZ’03. The 12th IEEE International Conference on*, volume 1, pages 630–635. IEEE, 2003.
- [BF05] C. J. Butz and F. Fang. Incorporating evidence in Bayesian networks with the select operator. In *In Proceedings of the 18 th Canadian Conference on Artificial Intelligence, Springer-Verlag, Victoria British-Columbia*, pages 297–301, 2005.
- [BFNT14] Nic Birtles, Norman Fenton, Martin Neil, and Ed Tranham. *Agenda-Risk manual (Version 6.1) [Computer software]*. Cambridge, United Kingdom, 2014.
- [BFSO84] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.
- [Bil04] Jeff Bilmes. On soft evidence in Bayesian networks. Technical Report UWEETR-2004-00016, Department of Electrical Engineering University of Washington, Seattle, 2004.

- [BJ05] Francis R Bach and Michael I Jordan. Discriminative training of hidden markov models for multiple pitch tracking [speech processing examples]. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 5, pages v–489. IEEE, 2005.
- [Blo98] Mark Bloemeke. *Agent encapsulated Bayesian networks*. Ph.d. thesis, Department of Computer Science, University of South Carolina, 1998.
- [BMDM⁺12] A. Ben Mrad, V. Delcroix, M. A. Maalej, S. Piechowiak, and M. Abid. Uncertain evidence in Bayesian networks : Presentation and comparison on a simple example. In *Proceedings of the 14th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2012*, pages 39–48, Catania, Italy, July 2012.
- [BMDP⁺13] Ali Ben Mrad, Veronique Delcroix, Sylvain Piechowiak, Mohamed Amine Maalej, and Mohamed Abid. Understanding soft evidence as probabilistic evidence : Illustration with several use cases. In *Modeling, Simulation and Applied Optimization (ICMSAO), 2013 5th International Conference on*, pages 1–6, 2013.
- [BMDP⁺15] Ali Ben Mrad, Veronique Delcroix, Sylvain Piechowiak, Philip Leicester, and Mohamed Abid. A unified presentation of uncertain evidence in bayesian networks : Likelihood evidence and probabilistic evidence. *Applied Intelligence*, 2015.
- [BMMD⁺11] A. Ben Mrad, M. A. Maalej, V. Delcroix, S. Piechowiak, and M. Abid. Fuzzy evidence in Bayesian networks. In *Proc. of Soft Computing and Pattern Recognition, Dalian, China*, 2011.
- [BT12] Salem Benferhat and Karim Tabia. Inference in possibilistic network classifiers under uncertain observations. *Annals of Mathematics and Artificial Intelligence*, 64(2-3) :269–309, 2012.
- [CD04] Hei Chan and Adnan Darwiche. Sensitivity analysis in Bayesian networks : From single to multiple parameters. In *UAI*, pages 67–75, 2004.

- [CD05] Hei Chan and Adnan Darwiche. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163(1) :67–90, 2005.
- [Cha05] Hei Chan. *Sensitivity Analysis of Probabilistic Graphical Models*. Ph.d. thesis, University of California, Los Angeles, 2005.
- [CV95] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3) :273–297, 1995.
- [CYLH07] Chiu Chui-Yu, Chih-Chung Lo, and Yan-Xin Hsu. Integrating bayesian theory and fuzzy logics with case-based reasoning for car-diagnosing problems. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, volume 4, pages 344–348. IEEE, 2007.
- [D'A93] Bruce D'Ambrosio. Incremental probabilistic inference. In *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*, pages 301–308. Morgan Kaufmann Publishers Inc., 1993.
- [Dar09] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.
- [Dar14] Adnan Darwiche. *SamIam [Computer software]*. University of California Los Angeles, 2014.
- [Das91] Belur V Dasarathy. Nearest neighbor ({NN}) norms :{NN} pattern classification techniques. 1991.
- [DK89] Thomas Dean and Keiji Kanazawa. A model for reasoning about persistence and causation. *Computational intelligence*, 5(2) :142–150, 1989.
- [DKS⁺95] James Dougherty, Ron Kohavi, Mehran Sahami, et al. Supervised and unsupervised discretization of continuous features. In *ICML*, pages 194–202, 1995.
- [DMP98] Didier Dubois, Serafin Moral, and Henri Prade. Belief change rules in ordinal and numerical uncertainty theories. In D.M. Gabbay and

- Ph. Smets, editors, *Belief Change*, (D. Dubois, H. Prade, eds.), Vol. 3 of the *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 311–392. Kluwer Academic Publishers, Dordrecht, 1998.
- [DP88] Didier Dubois and Henri Prade. *Possibility theory*. Springer, 1988.
- [DPC⁺14a] Marcos F. S. V. D’Angelo, Reinaldo M. Palhares, Luciana B. Cosme, Lucas A. Aguiar, Felipe S. Fonseca, and Walmir M. Caminhas. Fault detection in dynamic systems by a fuzzy/bayesian network formulation. *Appl. Soft Comput.*, 21 :647–653, 2014.
- [DPC⁺14b] Marcos FSV D’Angelo, Reinaldo M Palhares, Luciana B Cosme, Lucas A Aguiar, Felipe S Fonseca, and Walmir M Caminhas. Fault detection in dynamic systems by a fuzzy/bayesian network formulation. *Applied Soft Computing*, 21 :647–653, 2014.
- [Dru14] Marek J. Druzdzel. *Genie Smile, (Version 20.) [Computer software]*. Decision Systems Laboratory, University of Pittsburgh, Pennsylvania, USA, 2014.
- [DS40] W. E. Deming and F. F. Stephan. On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11 :427–444, 1940.
- [DSL13] Veronique Delcroix, Karima Sedki, and François Xavier Lepoutre. A Bayesian network for recurrent multi-criteria and multi-attribute decision problems : Choosing a manual wheelchair. *Expert Systems with Applications*, 40(7) :2541–2551, 2013.
- [DTB08] E Di Tomaso and JF Baldwin. An approach to hybrid probabilistic models. *International Journal of Approximate Reasoning*, 47(2) :202–218, 2008.
- [DTD00] B. D’Ambrosio, M. Takikawa, and Upper D. Representation for dynamic situation modeling. Technical Report Technical report, Information Extraction and Transport, Inc., 2000.

- [Elv14] Elvira. Elvira project, 2014.
- [FB12] Luciano Ferreira and Denis Borenstein. A fuzzy-bayesian model for supplier selection. *Expert Systems with Applications*, 39(9) :7834–7844, 2012.
- [FDF94] Robert Fung and Brendan Del Favero. Backward simulation in bayesian networks. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, pages 227–234. Morgan Kaufmann Publishers Inc., 1994.
- [FGG97] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3) :131–163, 1997.
- [FL03] Olivier François and Philippe Leray. Etude comparative d’algorithmes d’apprentissage de structure dans les réseaux bayésiens. *Rencontres des Jeunes Chercheurs en IA*, 2003.
- [FPA08] Christopher Fogelberg, Vasile Palade, and Phil Assheton. Belief propagation in fuzzy bayesian networks. In *1st International Workshop on Combinations of Intelligent Methods and Applications (CIMA) at ECAI’08*, pages 19–24, 2008.
- [Fri98] Nir Friedman. The bayesian structural em algorithm. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 129–138. Morgan Kaufmann Publishers Inc., 1998.
- [GD04] Daniel Grossman and Pedro Domingos. Learning bayesian network classifiers by maximizing conditional likelihood. In *Proceedings of the twenty-first international conference on Machine learning*, page 46. ACM, 2004.
- [GDA⁺13] Raffaele Giordano, Daniela D’Agostino, Ciro Apollonio, Nicola Lamaddalena, and Michele Vurro. Bayesian belief network to support conflict analysis for groundwater protection : The case of the apulia region. *Journal of environmental management*, 115 :136–146, 2013.

- [GP08] Ya'akov Gal and Avi Pfeffer. Networks of influence diagrams : A formalism for representing agents' beliefs and decision-making processes. *Journal of Artificial Intelligence Research*, 33(1) :109–147, 2008.
- [GRS06] A Grêt-Regamey and D Straub. Spatially explicit avalanche risk assessment linking bayesian networks to a gis. *Natural Hazards and Earth System Science*, 6(6) :911–926, 2006.
- [Hak08] Asma Kacem Hakim. *Intérêt du PSA et du f-PSA dans le diagnostic précoce du cancer prostatique*. Thèse de doctorat en médecine, Faculté de Médecine de Sfax, 2008.
- [Hen86] Max Henrion. Propagating uncertainty in bayesian networks by probabilistic logic sampling. In *UAI*, pages 149–164, 1986.
- [Hen14] Max Henrion. *Analytica, (Version) [Computer software]*. Lumina decision systems, Los Gatos, CA, USA, 2014.
- [HOA12] Abdelkader Heni, Mohamed Nazih Omri, and Adel Alimi. Certain bayesian network based on fuzzy knowledge bases. *arXiv preprint arXiv :1206.1319*, 2012.
- [Høj14] Søren Højsgaard. *gRain, (Version 1.2-3) [Computer software]*. Aalborg University, Denmark, 2014.
- [HSC13] Eric J Horvitz, Jaap Suermondt, and Gregory F Cooper. Bounded conditioning : Flexible inference for decisions under scarce resources. *arXiv preprint arXiv :1304.1512*, 2013.
- [IY10] Rolly Intan and Oviliani Yenty Yuliana. Fuzzy bayesian belief network for analyzing medical track record. In *Advances in Intelligent Information and Database Systems*, pages 279–290. Springer, 2010.
- [JA13] Frank Jensen and SK Anderson. Approximations in bayesian belief universe for knowledge based systems. *arXiv preprint arXiv :1304.1101*, 2013.
- [Jef90] Richard C. Jeffrey. *The Logic of Decision*. (2nd edition) University of Chicago Press, 1990. 246 pages.

- [Jen96] Finn V Jensen. *An introduction to Bayesian networks*, volume 210. UCL press London, 1996.
- [JLCM12] Sandra Johnson, Sama Low-Choy, and Kerrie Mengersen. Integrating bayesian networks and geographic information systems : Good practice examples. *Integrated environmental assessment and management*, 8(3) :473–479, 2012.
- [JM14] Lionel Jouffe and Paul Munteanu. *BayesiaLab*. Laval, France, 2014.
- [JN07] Finn Verner Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer Publishing Company, Incorporated, 2nd edition, 2007.
- [JN13] Finn V Jensen and Thomas Dyhre Nielsen. Probabilistic decision graphs for optimization under uncertainty. *Annals of Operations Research*, 204(1) :223–248, 2013.
- [Jor98] Michael Irwin Jordan. *Learning in Graphical Models :[proceedings of the NATO Advanced Study Institute... : Ettore Majorana Center, Erice, Italy, September 27-October 7, 1996]*, volume 89. Springer, 1998.
- [Kjæ94] Uffe Kjærulff. Reduction of computational complexity in bayesian networksthrough removal of weak dependences. In *Proceedings of the tenth international conference on uncertainty in artificial intelligence*, pages 374–382. Morgan Kaufmann Publishers Inc., 1994.
- [KL51] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22 :79–86, 1951.
- [KM03] Daphne Koller and Brian Milch. Multi-agent influence diagrams for representing and solving games. *Games and Economic Behavior*, 45(1) :181–221, 2003.
- [KM12] Uffe B Kjaerulff and Anders L Madsen. *Bayesian Networks and Influence Diagrams : A Guide to Construction and Analysis : A Guide to Construction and Analysis*, volume 22. Springer, 2012.

- [KM13] U.B. Kjaerulff and A.L. Madsen. *Bayesian Networks and Influence Diagrams : A Guide to Construction and Analysis*, volume 22 of *Information science and statistics*. Springer, 2 edition, 2013.
- [KN09] T. Koski and J. Noble. *Bayesian Networks : An Introduction*. Wiley Series in Probability and Statistics. Wiley, 2009.
- [KP87] Jin H Kim and Judea Pearl. Convince : A conversational inference consolidation engine. *Systems, Man and Cybernetics, IEEE Transactions on*, 17(2) :120–132, 1987.
- [KP97] Daphne Koller and Avi Pfeffer. Object-oriented bayesian networks. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 302–313. Morgan Kaufmann Publishers Inc., 1997.
- [Kri01] Mark L. Krieg. A tutorial on Bayesian belief networks. Technical Report DSTO-TN-0403, Surveillance Systems Division, Electronics and Surveillance Research Laboratory, Defense science and technology organisation, Edinburgh, South Australia, Australia, 2001.
- [Kru37] R. Kruithof. Telefoonverkeersrekening. *De Ingenieur*, 52 :15–25, 1937.
- [KVV04] Young-Gyun Kim, Marco Valtorta, and Jiri Vomlel. A prototypical system for soft evidential update. *Applied Intelligence*, 21(1) :81–97, 2004.
- [Lan11] Scott Langevin. *Knowledge representation, communication, and update in probability-based multiagent systems*. PhD thesis, University of South Carolina, Columbia, SC, USA, 2011. AAI3454755.
- [Lau14] Steffen L. Lauritzen. *Hugin, (Version 8.0) [Computer software]*. Aalborg, Denmark, 2014.
- [LBD⁺13a] Dries Landuyt, Steven Broekx, Rob D’hondt, Guy Engelen, Joris Aertsens, and Peter Goethals. A review of Bayesian belief networks in ecosystem service modelling. *Environmental Modelling and software*, 46 :1–11, 2013.

- [LBD⁺13b] Dries Landuyt, Steven Broekx, Rob D’hondt, Guy Engelen, Joris Aertsens, and Peter LM Goethals. A review of bayesian belief networks in ecosystem service modelling. *Environmental Modelling & Software*, 46 :1–11, 2013.
- [LGR12] Philip A. Leicester, Chris I. Goodier, and Paul Rowley. Community energy delivers megawatts, pounds, carbon reductions, etcetera. In *Midlands Energy Graduate School, (MEGS)*, 2012.
- [LGR13] Philip A. Leicester, Chris I. Goodier, and Paul Rowley. Using a Bayesian network to evaluate the social, economic and environmental impacts of community renewable energy. In *Clean Technology for Smart Cities and Buildings (CISBAT)*, 2013.
- [LS88a] Steffen L Lauritzen and David J Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 157–224, 1988.
- [LS88b] Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50 :157–224, 1988.
- [LV08] Scott Langevin and Marco Valtorta. Performance evaluation of algorithms for soft evidential update in Bayesian networks : First results. In *SUM*, pages 284–297, 2008.
- [LVB10] Scott Langevin, Marco Valtorta, and Mark Bloemeke. Agent-encapsulated Bayesian networks and the rumor problem. In *AAMAS ’10 Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, volume 1, pages 1553–1554, 2010.
- [LWdC10a] Kathryn B. Laskey, Edward J. Wright, and Paulo Cesar G. da Costa. Envisioning uncertainty in geospatial information. *International Journal of Approximate Reasoning*, 51(2) :209–223, 2010.

- [LWdC10b] Kathryn Blackmond Laskey, Edward J Wright, and Paulo CG da Costa. Envisioning uncertainty in geospatial information. *International journal of approximate reasoning*, 51(2) :209–223, 2010.
- [Mar03] Dimitris Margaritis. *Learning Bayesian network model structure from data*. PhD thesis, University of Pittsburgh, 2003.
- [MDPL14] Ali Ben Mrad, Véronique Delcroix, Sylvain Piechowiak, and Philip Leicester. From information to evidence in a bayesian network. In *Probabilistic Graphical Models*, pages 33–48. Springer, 2014.
- [MJ99] Anders L. Madsen and Finn Verner Jensen. Lazy propagation : A junction tree inference algorithm based on lazy evaluation. *Artificial Intelligence*, 113(1-2) :203–245, 1999.
- [Mur02] Kevin Patrick Murphy. *Dynamic bayesian networks : representation, inference and learning*. PhD thesis, University of California, Berkeley, 2002.
- [Mur14] Kevin Murphy. *Bayesian Network Toolbox (BNT), (Version 1.0.7) [Computer software]*. MIT AI lab, Cambridge, MA, USA, 2014.
- [MW14] Tom Minka and John Winn. *Infer.NET, (Version 205) [Computer software]*. Microsoft Research Cambridge, MA, USA, 2014.
- [N⁺04] Richard E Neapolitan et al. *Learning bayesian networks*, volume 1. Prentice Hall Upper Saddle River, 2004.
- [NJ09] Thomas Dyhre Nielsen and Finn Verner Jensen. *Bayesian networks and decision graphs*. Springer, 2009.
- [Nor14] Norsys. *Netica Application, (Version 5.12) [Computer software]*. Norsys Software Corp., Vancouver, BC, CANADA, 2014.
- [NWL⁺07] Patrick Naïm, Pierre-Henri Willemin, Philippe Leray, Olivier Pourret, and Anna Becker. *Réseaux bayésiens*. Eyrolles, 3 edition, 2007.
- [PDZP12] Yun Peng, Zhongli Ding, Shenyong Zhang, and Rong Pan. Bayesian network revision with probabilistic constraints. *International Journal*

of *Uncertainty, Fuzziness and Knowledge-Based Systems*, 20(3) :317–337, 2012.

- [Pea88a] J Pearl. Probabilistic reasoning in intelligent systems. 1988.
- [Pea88b] Judea Pearl. *Probabilistic reasoning in intelligent systems : networks of plausible inference*. Morgan Kaufmann, 1988.
- [Pea88c] Judea Pearl. *Probabilistic reasoning in intelligent systems : networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.
- [PL00] Heping Pan and Lin Liu. Fuzzy bayesian networks—a general formalism for representation, inference and learning with hybrid bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(07) :941–962, 2000.
- [PP07] DV Pullar and TH Phan. Using a bayesian network in a gis to model relationships and threats to koala populations close to urban environments. In *Les Oxley and Don Kulasiri, MODSIM 2007 : International Congress on Modelling and Simulation, University of Canterbury, Christchurch, New Zealand*, pages 1370–1375, 2007.
- [PPD06] Rong Pan, Yun Peng, and Zhongli Ding. Belief update in Bayesian networks using uncertain evidence. In *ICTAI*, pages 441–444, 2006.
- [PYC06] Han-Saem Park, Ji-Oh Yoo, and Sung-Bae Cho. A context-aware music recommendation system using fuzzy bayesian networks with utility theory. In *Fuzzy systems and knowledge discovery*, pages 970–979. Springer, 2006.
- [PZP10] Y. Peng, S. Zhang, and R. Pan. Bayesian network reasoning with uncertain evidences. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 18(5) :539–564, 2010.
- [S⁺76] Glenn Shafer et al. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.

- [Sah99] Maneesh Sahani. *Latent variable models for neural data analysis*. PhD thesis, California Institute of Technology, 1999.
- [San14] John Sandiford. *Bayes Server, (Version 5.5) [Computer software]*. East Preston, United Kingdom, 2014.
- [SDL⁺10] K. Sedki, V. Delcroix, F.-X. Lepoutre, E. Adam, A.-P. Maquinghen-Godillon, and I. Ville. Bayesian network model for decision problems. In M.a. Klopotek, M. Marciniak, A. Mykowiecka, W. Penczek, and S.t. Wierzchon (Ed.), editors, *Intelligent Information Systems, new approaches*, pages 285–298, Publishing House of University of Podlasie, Siedlce, Poland, June 2010.
- [SP13] Ross D Shachter and Mark Alan Peot. Simulation approaches to general probabilistic inference on belief networks. *arXiv preprint arXiv :1304.1526*, 2013.
- [TB08] E. Di Tomaso and J. F. Baldwin. An approach to hybrid probabilistic models. *International Journal of Approximate Reasoning*, 47(2) :202–218, 2008.
- [TDW02] Masami Takikawa, Bruce D’Ambrosio, and Ed Wright. Real-time inference with large-scale temporal bayes nets. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 477–484. Morgan Kaufmann Publishers Inc., 2002.
- [Ter92] George R Terre11. Influence diagrams, belief nets, and decision analysis. *Technometrics*, 34(1) :114–115, 1992.
- [TL07] Hao Tang and Shi Liu. Basic theory of fuzzy bayesian networks and its application in machinery fault diagnosis. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, volume 4, pages 132–137. IEEE, 2007.
- [VKV02] Marco Valtorta, Young-Gyun Kim, and Jirí Vomlel. Soft evidential update for probabilistic multiagent systems. *International Journal of Approximate Reasoning*, 29(1) :71–106, 2002.

- [Vom99] Jirí Vomlel. *Methods of probabilistic knowledge integration*. Ph.d. thesis, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, 1999.
- [Vom04] Jirí Vomlel. Probabilistic reasoning with uncertain evidence. *Neural Network World, International Journal on Neural and Mass-Parallel Computing and Information Systems*, 14(5) :453–465, 2004.
- [Wal91] Peter Walley. *Statistical reasoning with imprecise probabilities*, volume 42. Peter Walley, 1991.
- [Xia02] Yang Xiang. *Probabilistic Reasoning in Multiagent Systems : A graphical models approach*. Cambridge University Press, 2002.
- [XL00] Yang Xiang and Victor Lesser. Justifying multiply sectioned Bayesian networks. In *MultiAgent Systems, 2000. Proceedings. Fourth International Conference on*, pages 349–356. IEEE, 2000.
- [Zad65] LA Zadeh. nxz. *Information and Control*, 8 :338–353, 1965.
- [Zen06] Yifeng Zeng. *Probabilistic modeling and reasoning in multiagent decision systems*. Ph.d. thesis, National University of Singapore, 2006.
- [ZPW08] Shenyong Zhang, Yun Peng, and Xiaopu Wang. An Efficient Method for Probabilistic Knowledge Integration. In *Proceedings of The 20th IEEE International Conference on Tools with Artificial Intelligence*. IEEE Computer Society, November 2008.

Liste de publications

Revue internationale avec comité de lecture

- **Ben Mrad A.** , Delcroix, V., Piechowiak, S., Leicester, P., & Abid, M. (2015). An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. **Applied Intelligence**, 1-23.

Conférence internationale avec actes et comité de lecture

- **Ben Mrad A.**, Delcroix V., Piechowiak S., Leicester P. (2014). From information to evidence in a Bayesian network. L.C. van der Gaag, A. J. Feelders, Probabilistic Graphical Models, 7th European Workshop, **PGM 2014**, Utrecht, The Netherlands, September 17-19, 2014, Proceedings, **Springer**, LNCS 8754, pp. 33-48, septembre.
- **Ben Mrad A.**, Delcroix V., Maalej M.-A., Piechowiak S., Abid M. (2013). Prostate cancer diagnosis based on probabilistic evidence for propagating observations on continuous variables in Bayesian networks. International Conference for Engineering Sciences for Biology and Medicine, **ESBM'2013** (November 29-December 1st), Mahdia, Tunisia, novembre .
- **Ben Mrad A.**, Delcroix V., Piechowiak S., Maalej M.-A., Abid M. (2013). Understanding soft evidence as probabilistic evidence : illustration with several use cases. 5th International Conference on Modeling, Simulation and Applied Optimization, **ICMSAO'2013** (April 28-30), **IEEE**, Hammamet, Tunisia, avril.
- **Ben Mrad A.**, Delcroix V., Maalej M.-A., Piechowiak S., Abid M. (2012). Uncertain Evidence in Bayesian Network: presentation and comparison on a simple example. 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, **IPMU'2012** (July 9-13), Lecture Notes in Computer Science, **Springer**, Catania, Italy, pp. 39-48, juillet.
- **Ben Mrad A.**, Maalej M.-A., Delcroix V., Piechowiak S., Abid M. (2011). Fuzzy Evidence in Bayesian Network. Third International Conference on Soft Computing and Pattern Recognition, **SoCPaR 2011** (October 14-16), **IEEE Press**, Dalian, China, pp. 486-491, octobre.

Conférence nationale avec actes et comité de lecture

- Delcroix V., **Ben Mrad A.**, Piechowiak S. (2014). De l'information à l'observation dans un réseau bayésien. Huitièmes Journées de l'Intelligence Artificielle Fondamentale, **JIAF'2014** (11-13 juin 2014), Angers, France, juin.
- **Ben Mrad A.**, Delcroix V., Maalej M.-A., Piechowiak S. (2013). Observations probabilistes dans un réseau Bayésien : diagnostic du cancer de la prostate. 11èmes Rencontres des Jeunes Chercheurs en Intelligence Artificielle, **RJCIA'2013** (1-2 Juillet), juillet.
- Delcroix V., **Ben Mrad A.** (2013). Observations incertaines et observations probabilistes dans un réseau bayésien. Septièmes Journées de l'Intelligence Artificielle Fondamentale, **JIAF'2013** (12-14 juin), Aix-en-Provence, France, juin.
- **Ben Mrad A.**, Delcroix V., Maalej M.-A., Piechowiak S., Abid M. (2012). Different types of evidence in bayesian network. First Tunisian Workshop on Embedded Systems Design, **TWESD'2012** (May 25-27), Monastir, Tunisia, mai.

Séminaire et autres communications

- **Ben Mrad A.**, Delcroix V., Piechowiak S., Maalej M.-A., Abid M. (2014). Evidences incertaines dans les Réseaux Bayésiens. 7èmes Journée Francophone des Réseaux Bayésiens, **JFRB'2014** (25-27 Juin), Paris, France, juin.
- **Ben Mrad A.**, Delcroix V., Piechowiak S., Maalej M.-A., Abid M. (2014). Observations non-déterministes dans un réseau Bayésien. Third Tunisian Workshop on Embedded Systems Design, **TWESD'2014** (09-11 Mai), Mahdia, Tunisia, mai .
- **Ben Mrad A.** (2013). Uncertain Evidence in Bayesian Network. Workshop "Healthcare systems under uncertainty: research and applications", TEMPO, UVHC, Valenciennes, juin.
- **Ben Mrad A.**, Delcroix V., Maalej M.-A., Piechowiak S., Abid M. (2013). Illustration of the notion of soft evidence in Bayesian Network. Second Tunisian Workshop on Embedded Systems Design, **TWESD'2013** (March 25-28), Sousse, Tunisia, mars .
- **Ben Mrad A.**, Delcroix V., Maalej M.-A., Piechowiak S., Abid M. (2012). Evidences incertaines dans les Réseaux Bayésiens. 6 ème Journée Francophone des Réseaux Bayésiens, **JFRB'2012** (11-13 mai), Kerkennah, Tunisie, mai.
- **Ben Mrad A.**, Maalej M.-A., Delcroix V., Piechowiak S., Abid M. (2011). Fuzzy Evidence in Bayesian Network. 2èmes journées sur les Réseaux Bayésiens et leurs Applications (30 septembre - 2 octobre), Sfax, Tunisie, janvier.
- **Ben Mrad A.**, Maalej M.-A., Abid M. (2010). Réseau bayésien flou. 1ères journées sur les Réseaux Bayésiens et leurs Applications (**JRBA'2010**), Sfax, Tunisie, janvier.

RESUME :

Dans un réseau bayésien, une observation sur une variable signifie en général que cette variable est instanciée. Ceci signifie que l'observateur peut affirmer avec certitude que la variable est dans l'état signalé. Cette thèse porte sur d'autres types d'observations, souvent appelées observations incertaines, qui ne peuvent pas être représentées par la simple affectation de la variable. Cette thèse clarifie et étudie les différents concepts d'observations incertaines et propose différentes applications des observations incertaines dans les réseaux bayésiens.

Nous commençons par dresser un état des lieux sur les observations incertaines dans les réseaux bayésiens dans la littérature et dans les logiciels, en termes de terminologie, de définition, de spécification et de propagation. Il en ressort que le vocabulaire n'est pas clairement établi et que les définitions proposées couvrent parfois des notions différentes.

Nous identifions trois types d'observations incertaines dans les réseaux bayésiens et nous proposons la terminologie suivante : observation de vraisemblance, observation probabiliste fixe et observation probabiliste non-fixe. Nous exposons ensuite la façon dont ces observations peuvent être traitées et propagées.

Enfin, nous donnons plusieurs exemples d'utilisation des observations probabilistes fixes dans les réseaux bayésiens. Le premier exemple concerne la propagation d'observations sur une sous-population, appliquée aux systèmes d'information géographique. Le second exemple concerne une organisation de plusieurs agents équipés d'un réseau bayésien local et qui doivent collaborer pour résoudre un problème. Le troisième exemple concerne la prise en compte d'observations sur des variables continues dans un RB discret. Pour cela, l'algorithme BN-IPFP-1 a été implémenté et utilisé sur des données médicales de l'hôpital Bourguiba de Sfax.

Mots clefs : intelligence artificielle, incertitude, modèle graphique probabiliste, réseau bayésien, observation, observation incertaine, observation probabiliste, observation de vraisemblance.

ABSTRACT:

In a Bayesian network, evidence on a variable usually signifies that this variable is instantiated, meaning that the observer can affirm with certainty that the variable is in the signaled state. This thesis focuses on other types of evidence, often called uncertain evidence, which cannot be represented by the simple assignment of the variables. This thesis clarifies and studies different concepts of uncertain evidence in a Bayesian network and offers various applications of uncertain evidence in Bayesian networks.

Firstly, we present a review of uncertain evidence in Bayesian networks in terms of terminology, definition, specification and propagation. It shows that the vocabulary is not clear and that some terms are used to represent different concepts.

We identify three types of uncertain evidence in Bayesian networks and we propose the following terminology: likelihood evidence, fixed probabilistic evidence and not-fixed probabilistic evidence. We define them and describe updating algorithms for the propagation of uncertain evidence. Finally, we propose several examples of the use of fixed probabilistic evidence in Bayesian networks. The first example concerns evidence on a subpopulation applied in the context of a geographical information system. The second example is an organization of agent encapsulated Bayesian networks that have to collaborate together to solve a problem. The third example concerns the transformation of evidence on continuous variables into fixed probabilistic evidence. The algorithm BN-IPFP-1 has been implemented and used on medical data from CHU Habib Bourguiba in Sfax.

Keywords: artificial intelligence, uncertainty, probabilistic graphical models, Bayesian network, evidence, uncertain evidence, probabilistic evidence, likelihood finding, soft evidence, virtual evidence.