Transformations probabilistes-possibilistes: conditionnement, inférence et modèles graphiques

Salem Benferhat   Amélie Levray   Karim Tabia

Univ Lille Nord de France, F-59000 Lille, France
UArtois, CRIL UMR CNRS 8188, F-62300 Lens, France
{benferhat,levray,tabia}@cril.univ-artois.fr

Résumé

Le problème de la représentation et du raisonnement dans l'incertain est un sujet important en Intelligence artificielle. Dans ce papier, nous étudions les transformations probabilistes-possibilistes (dans le sens des probabilités vers les possibilités) à travers plusieurs questions telles que l'inférence et les modèles graphiques. Les travaux qui traitent des transformations sont principalement centrés sur les propriétés intéressantes à satisfaire et proposent des transformations les satisfaisant. Peu de travaux ont été développés en ce qui concerne l'analyse du comportement de ces transformations face à des opérations comme le conditionnement et la marginalisation. Ce papier s'intéresse à la commutativité des transformations lors de tâches de raisonnement telles que marginaliser et conditionner. Un autre problème abordé dans ce papier est celui des transformations dans les modèles graphiques, plus précisément l'indépendance des événements et des variables.

Abstract

Representing and reasoning with uncertain information is a common topic in Artificial Intelligence. In this paper, we focus on probability-possibility transformations (from probability to possibility) in the context of changing operations and graphical models. Existing works mainly propose probability-possibility transformations satisfying some desirable properties. Regarding the analysis of the behavior of these transformations with respect to changing operations (such as conditioning and marginalization), only few works addressed such issues. This paper concerns the commutativity of transformations with respect to some reasoning tasks such as marginalization and conditioning. Another crucial issue addressed in this paper is the one of probability-possibility transformations in the context of graphical models, especially the independence of events and variables.

1 Introduction

De nombreux cadres existent pour représenter l'information en présence d'imprécision, d'incertitude et de contradictions, tels que la théorie des probabilités, la théorie des possibilités, les probabilités imprécises ou encore la théorie de Dempster-Shafer. La théorie des probabilités est la plus connue et par conséquent la plus développée mais lorsqu'on fait face à des cas d'ignorance totale ou partielle, ou à des informations incomplètes en général, cette théorie n'est peut-être pas la plus adaptée. C'est pour cela que des cadres de représentation alternatifs tels que la théorie des possibilités ont été proposés.

Les premiers travaux mêlant la théorie des probabilités et la théorie des possibilités ont essayé de trouver les connections entre ces théories en définissant des transformations pour passer d'un cadre à un autre. On peut voir plusieurs raisons pour lesquelles les transformations seraient utiles. Par exemple, pour fusionner des informations multi-sources, plusieurs agents peuvent fournir des informations mais dans différents formalismes. La question est de savoir comment fusionner ces informations. Une solution serait de transformer toutes les informations dans un seul cadre. Un autre exemple serait celui où on est en présence d'informations statistiquement pauvres. En effet, les probabilités sont appropriées dans un cadre empirique, mais cela requiert une quantité d'informations importante. Et quand ces informations ne sont pas disponibles en quantité suffisante, les possibilités peuvent combler ce manque comme dans [13]. On trouve dans [1] un exemple de propagation probabiliste (stochastique) et d'information possibiliste en analyse de risques. Une motivation supplémentaire aux transformations est de pouvoir utiliser des outils développés dans un cadre (comme les algorithmes d'inférence, les logiciels) sans avoir à tout réimplémenter ou reprogrammer pour le cadre que l'on utilise. Par exemple, de nombreux outils existent pour faire de l'inférence dans les réseaux Bayésiens (comme Javabayes). Dans ce cas, il serait utile de pouvoir passer des possibilités aux probabilités, avec un minimum de perte d'information, pour utiliser ces outils.

Les transformations probabilistes-possibilistes ont fait l'objet de plusieurs travaux [20, 12, 15], mais ces derniers se sont plus concentrés à trouver de bonnes propriétés à satisfaire et les transformations les satisfaisant. Un exemple de telles propriétés est le principe de consistance dont le but est de préserver autant d'information que possible. Dans [20, 12, 15], plusieurs principes de consistance ont été définis. Parmi les travaux sur les transformations, l'auteur du papier [18] s'est intéressé, précisément, à la commutativité des tâches de raisonnement. Un autre papier [16] a étudié les transformations dans le cadre des modèles graphiques en transformant un réseau Bayésien en réseau possibiliste. Dans [5], les auteurs proposent des transformations des probabilités imprécises vers d'autres formalismes incertains. Un travail plus récent [19] utilise les transformations probabilistes-possibilistes pour passer d'un réseau Bayésien à une carte cognitive floue avec une application en analyse causale.

Dans le présent travail, nous étudions les transformations probabilistes-possibilistes et leur comportement lors de tâches de raisonnement et dans les modèles graphiques. Il est à noter que nous nous sommes intéressés uniquement aux transformations du cadre probabiliste vers le cadre possibiliste. Étant donnée une distribution encodant des informations incertaines, qu'elles soient probabilistes ou possibilistes, nous sommes supposés pouvoir raisonner à propos d'événements d'intérêt. Ce travail s'intéresse à des problématiques complémentaires telles que préserver la marginalisation, le conditionnement et les relations d'indépendance. Nous étudions ces questions lorsque les informations sont représentées à l'aide de distributions ou de modèles graphiques. Nous montrons qu'il n'existe pas de transformation, du cadre probabiliste vers le cadre possibiliste, qui préserve la plupart des tâches de raisonnement présentées dans ce travail. Par exemple, pour la préservation de la marginalisation, nous montrons qu'aucune transformation ne peut préserver l'ordre relatif des événements même si l'ordre relatif des interprétations est préservé. De même, l'ordre des interprétations ne peut être préservé lorsqu'on transforme les modèles graphiques probabilistes en modèles possibilistes. Rappelons, avant de présenter nos résultats, les principaux concepts des transformations.

2 Théorie des probabilités, théorie des possibilités et modèles graphiques

Les probabilités sont le cadre le plus connu et le plus ancien pour représenter et raisonner avec des informations incertaines. Le principal concept des probabilités est la notion de distribution de probabilités p qui associe à chaque état élémentaire (ou interprétation) ω_i du monde Ω, un degré de probabilité. La théorie des probabilités est régie par les axiomes de Kolmogorov (non-négativité, normalisation et additivité) et peut s'interpréter de deux manières (de manière fréquentiste ou de manière subjective).

Parmi les alternatives à la théorie des probabilités, il y a la théorie des possibilités [20, 8]. Cette théorie, très répandue, est basée sur la notion de distribution de possibilités π qui pour chaque état, de l'univers du discours, associe un degré dans l'intervalle $[0, 1]$ exprimant une connaissance partielle sur l'état du monde. Elle est aussi régie par trois axiomes qui sont la normalisation (il doit exister un état du monde qui soit totalement possible, c'est à dire avec un degré de 1), la non-négativité et la maxitivité. Par convention, $\pi(\omega_i)=1$ correspond à un état complètement possible alors que $\pi(\omega_i)=0$ correspond à un état impossible. Notons que nous pouvons interpréter les possibilités soit i) *qualitativement* (de manière ordinale) où seul l'ordre des interprétations a de l'importance, soit ii) *quantitativement* où les degrés prennent un sens sur l'intervalle comme en probabilités.

Une notion est présente dans ces deux théories, celle de mesure de probabilités (resp. de possibilités), elle est définie comme étant le degré de probabilités (resp. de possibilités) d'un ensemble d'interprétations (états du monde). Cette mesure, notée P en probabilités et Π en possibilités se calcule à l'aide de l'axiome d'additivité (en probabilités, où l'on somme les degrés des interprétations appartenant à l'ensemble) et avec l'axiome de maxitivité (en possibilités, où le degré de l'ensemble est le degré maximal des interprétations de l'ensemble). La différence principale entre la théorie des probabilités et la théorie des possibilités se trouve donc dans la façon de calculer le degré d'une disjonction de deux événements. En effet, $\forall \phi, \psi \subseteq \Omega$ avec $\phi \cap \psi = \emptyset$, $P(\phi \cup \psi) = P(\phi) + P(\psi)$ alors que $\Pi(\phi \cup \psi) = \max(\Pi(\phi), \Pi(\psi))$.

Le conditionnement est un important opérateur de changement de croyances qui consiste à mettre à jour les connaissances actuelles, encodées soit par une distribution de probabilités soit par une distribution de possibilités, lorsqu'un nouvel événement sûr est observé. Alors que le conditionnement quantitatif en possibilités et le conditionnement probabiliste se ressemblent (en effet, ils sont définis de manière similaire), le conditionnement qualitatif, quant à lui, diffère de manière significative. Remarquons que les deux définitions de conditionnement possibiliste suivantes satisfont la condition : $\forall \omega \in \phi$, $\pi(\omega)=\pi(\omega|\phi)\otimes\Pi(\phi)$ où $\otimes$ est soit le produit soit l'opérateur min. Dans le cadre quantitatif, la règle de conditionnement (aussi connue comme la règle de Dempster) est définie comme suit :

$$\pi(w_i|_p\phi) = \begin{cases} \frac{\pi(w_i)}{\Pi(\phi)} & \text{si } w_i \in \phi\,; \\ 0 & \text{sinon.} \end{cases} \tag{1}$$

Le conditionnement qualitatif est défini de la manière suivante [11] :

$$\pi(w_i|_m\phi) = \begin{cases} 1 & \text{si } \pi(w_i)=\Pi(\phi) \text{ et } w_i \in \phi\,; \\ \pi(w_i) & \text{si } \pi(w_i)< \Pi(\phi) \text{ et } w_i \in \phi\,; \\ 0 & \text{sinon.} \end{cases} \tag{2}$$

Travailler directement sur les distributions (de probabilités ou de possibilités) n'est pas pratique en terme de complexité (spatiale et temporelle). En effet, la taille de la distribution peut vite devenir trop volumineuse, nous avons donc besoin de représentations compactes. Les modèles graphiques [4] nous fournissent un outil efficace pour la représentation d'informations. Et parmi eux, on trouve les réseaux Bayésiens qui sont l'une des représentations compactes les plus utilisées.

Réseaux Bayésiens

Un réseau Bayésien [4] est défini par :

— **Une composante graphique** : Les variables et les arcs forment un graphe orienté acyclique (DAG). Chaque sommet du DAG représente une variable du problème modélisé et les arcs encodent les indépendances entre les variables.

— **Une composante numérique** : Chaque lien dans le DAG est quantifié par une distribution conditionnelle $p(A_i|par(A_i))$ pour chaque noeud A_i dans le contexte de ses parents $par(A_i)$. Pour les variables sans parents, on associe aussi une table locale a priori (sans parents).

Un réseau Bayésien code une distribution jointe qui se calcule avec la règle de chaînage.

$$p(A_1, ..., A_n) = \prod_{i=1}^{n} p(A_i|par(A_i)) \tag{3}$$

Les réseaux Bayésiens ne servent pas uniquement à la représentation mais aussi au raisonnement (calculer des probabilités d'intérêt). Les modèles graphiques sont dotés de mécanismes d'inférence [4] pour répondre à des questions d'intérêt afin de réaliser certaines tâches comme l'explication, le diagnostic, la classification, etc.

Réseaux possibilistes

Un réseau possibiliste [3] est aussi spécifié par une composante graphique et une autre numérique mais où les tables locales sont des distributions de possibilités. La règle de chaînage est définie comme suit :

$$\pi(A_1, ..., A_n) = \otimes_{i=1..n}\pi(A_i \mid par(A_i)) \tag{4}$$

où $\otimes$ est, selon le cadre possibiliste choisi, soit le produit soit l'opérateur min (autrement dit, $\otimes{=}*$ ou $\otimes{=}min$).

A moins qu'il en soit précisé autrement, par la suite les résultats valent dans les deux cadres, quantitatif et qualitatif.

3 Transformations probabilistes - possibilistes

Dans cette section, nous présentons en premier les principaux principes des transformations probabilistes-possibilistes. En particulier, puisque la théorie des probabilités et la théorie des possibilités représentent différentes sortes d'incertitude, il faut se focaliser sur le principe de consistance défini par Zadeh [20] et redéfini par plusieurs auteurs comme Dubois et Prade [7].

3.1 Principes des transformations

Principe de consistance de Zadeh. Zadeh [20] décrit le principe de consistance entre probabilités et possibilités comme *"a high degree of possibility does not imply a high degree of probability, and a low degree of probability does not imply a low degree of possibility"*. Le degré de consistance entre une distribution de probabilités p et une distribution de possibilités π est défini par :

$$C_z(\pi, p) = \sum_{i=1..n} \pi_i * p_i \tag{5}$$

Cependant, Zadeh signale que $C_z(\pi, p)$ n'est pas une loi précise ou une relation entre les distributions de probabilités et de possibilités. *"Rather it is an approximate formalization of the heuristic observation that a lessening of the possibility of a event tends to lessen its probability, but not vice-versa"*.

En effet, le degré de consistance calculé est discutable [7, 12] puisqu'il est d'autant plus proche de 1 que ω_i a une probabilité non nulle ($\forall \omega_i \in \Omega$, si $p(\omega_i) > 0$ alors $\pi(\omega_i) = 1$). Nous pouvons donc avoir plusieurs distributions (différentes) mais *ayant le même degré de consistance.*

Principe de consistance de Dubois et Prade. Dubois et Prade ont défini le principe de consistance, différemment, utilisant ces trois postulats : [7]

— La *condition de consistance* : $P(\phi)\leq\Pi(\phi)$, $\forall\phi\subseteq\Omega$.

— La *préservation de la préférence* : $\forall(\omega_1, \omega_2){\in}\Omega^2$, si $p(\omega_1){>}p(\omega_2)$ alors $\pi(\omega_1){>}\pi(\omega_2)$ et si $p(\omega_1){=}p(\omega_2)$ alors $\pi(\omega_1){=}\pi(\omega_2)$.

— Le principe du *maximum de spécificité* : Il faut chercher la distribution la plus spécifique qui satisfait les deux postulats précédents.

La notion de spécificité permet de sélectionner une distribution de possibilités en fonction de la quantité d'informations contenue dans cette distribution. Pour rappel, soient π_1 et π_2 deux distributions de possibilités, π_1 est dite plus spécifique que π_2 si $\forall\omega_i \in \Omega$, $\pi_1(\omega_i){\leq}\pi_2(\omega_i)$.

3.2 Règles de transformations

Il existe plusieurs transformations dans la litterature. On peut citer : *Optimal transformation* (*OT*) [7], *Klir transformation* (*KT*) [12], *Symmetric transformation* (*ST*) [10] et *Variable transformation* (*VT*) [14]. La transformation dite "optimale" a été définie par Dubois et Prade [7] et garantit que la distribution de possibilités résultante est la plus spécifique. *OT* est définie par :

$$\pi_i = \sum_{j/p_j \leq p_i} p_j \tag{6}$$

Il est à noter qu'il existe des transformations du cadre possibiliste vers le cadre probabiliste [10, 12]. Il existe aussi des transformations vers d'autres cadres incertains [5], en particulier des transformations des probabilités imprécises vers d'autres théories, le but étant de diminuer la complexité des calculs. Dans cet article, nous ne nous focalisons pas sur une transformation en particulier mais sur les transformations des probabilités vers les possibilités en général. On parlera de transformations probabilistes-possibilistes pour désigner des probabilités aux possibilités.

4 Transformations et raisonnement

Le but de ce papier est d'étudier la commutativité des transformations sur les tâches de raisonnement. Un premier papier [18] a étudié cette question mais cherchait seulement à savoir si les distributions résultantes étaient *identiques*. Il a été montré qu'il n'existe pas de transformations qui donnent les mêmes distributions par rapport au conditionnement et à la marginalisation. Nous utilisons $\triangleright(p){=}\pi$ pour désigner la transformation de la distribution de probabilités p en distribution de possibilités π satisfaisant le principe de préservation de préférence de Dubois et Prade. Dans la suite de ce papier, nous étudions la commutativité des transformations par rapport à i) l'ordre des événements quelconques et ii) deux opérations de raisonnement qui sont la marginalisation et le conditionnement. Nous étudions ces questions particulières pour des raisons pratiques. En effet parmi les requêtes les plus utilisées dans les modèles probabilistes, on trouve les requêtes MPE (la recherche de l'explication (instanciation de toutes les variables) la plus probable étant donnée une observation) et les requêtes MAP (où étant donné une ou plusieurs observations, l'objectif est de trouver l'hypothèse la plus probable sur certaines variables) [4]. Pour être plus explicite, développons deux petits exemples.

Exemple 1. *Soit $p(ABC)$ une distribution de probabilités sur trois variables binaires A, B et C. Ω, l'ensemble des états possibles, est le produit carthésien des domaines D_A, D_B, D_C. Etant donnée une observation, par exemple, $C{=}\neg c$, une requête MPE serait "quelle est l'instanciation des variables la plus probable étant donnée l'observation $C{=}\neg c$?". Ceci s'écrit en terme de probabilités : $argmax_{\omega_i\in\Omega}(p(\omega_i, \neg c))$. En d'autres termes, parmi les états de Ω où $C{=}\neg c$, ie. l'ensemble $\{ab\neg c, a\neg b\neg c, \neg ab\neg c, \neg a\neg b\neg c\}$, quelle est l'interprétation qui a le plus haut degré ?*

Exemple 2. *Soit $p(ABC)$ une distribution de probabilités sur trois variables binaires A, B et C (l'ensemble des variables est noté $\mathcal{A}$). Etant donné un sous-ensemble de variables $E{\subset}\mathcal{A}$ qui représente l'observation, et e une instanciation de E. Etant donné un sous-ensemble de variables $Q{\subset}\mathcal{A}\backslash E$ qui représente les variables de requêtes. Alors, si $e{=}\neg c$, une requête MAP serait "quelle est l'instanciation (l'hypothèse) des variables de Q la plus probable étant donnée l'observation $e{=}\neg c$?". Ceci s'écrit en terme de probabilités : $argmax_{q\in D_Q}(p(q|e))$. En d'autres termes, parmi les états du domaine de Q après conditionnement par l'évidence e, quelle est l'interprétation qui a le plus haut degré ?*

Par conséquent, pour répondre à ces questions en utilisant les transformations probabilistes-possibilistes, il est nécessaire d'étudier la commutativité des transformations par rapport à la marginalisation et le conditionnement. En particulier, il est nécessaire d'analyser l'impact des transformations sur l'ordre des événements.

Nous considérons les opérations de changement sur les distributions comme décrit par la Figure 1. La première distribution de possibilités π'' est obtenue en appliquant l'opération (marginalisation, conditionnement...) puis en transformant par $\triangleright$. Et nous obtenons la seconde distribution de possibilités π' en transformant puis en appliquant l'opération. Notre objectif est de comparer ces deux distributions de possibilités et vérifier si elles encodent le même ordre entre les événements.

$$
\begin{tikzcd}
p \arrow[rr, "\text{Opération de changement}"] \arrow[d] & & p' \arrow[d] \\
\pi = \rhd(p) \arrow[r, "\text{Opération de changement}"] & \pi' \arrow[r, leftrightarrow, "?"'] & \pi'' = \rhd(p')
\end{tikzcd}
$$

FIGURE 1 – Commutativité des tâches de raisonnement.

Considérons en premier l'opération de marginalisation qui consiste à construire les distributions marginales à partir d'une distribution jointe.

4.1 Transformations et marginalisation : préservation de l'ordre des événements

Dans la section précédente, nous avons vu que l'un des principes de Dubois et Prade requiert que l'ordre des interprétations doit être préservé, mais qu'en est-il de l'ordre des événements (ensemble d'interprétations) ? Par exemple, est-ce que le principe de préservation de préférence suffit-il pour que la transformation préserve l'ordre d'événements arbitraires ? Pour rappel, la marginalisation consiste à calculer le degré de probabilité ou de possibilité, d'un sous-ensemble de variables. Ainsi, soit A=$\{A_1..A_k\}$ un sous-ensemble de variables de $\mathcal{A}$, la marginalisation en probabilité se calcule, grâce à l'axiome d'additivité, par : $P(A_1,..,A_k)$=$\sum_{A_{k+1}..A_n}(p(A_1A_2..A_n))$. Quant à la marginalisation en possibilités, elle se calcule via l'axiome de maxitivité. Ainsi, $\Pi(A_1,..,A_k)$=$\max_{A_{k+1}..A_n}(\pi(A_1A_2..A_n))$. Il s'agit de sommer[1] (ou maximiser) sur les variables restantes.

La Proposition 1 établit qu'il n'y a pas de transformation qui préserve l'ordre des événements.

Proposition 1. *Soit $\rhd$ une transformation probabiliste-possibiliste*[2]*. Alors il existe p une distribution de probabilités et $\rhd(p)$=π où il existe $\phi \subseteq \Omega$, $\psi \subseteq \Omega$, avec $\phi \neq \psi$ tel que :*

$$P(\phi) < P(\psi) \not\Rightarrow \Pi(\phi) < \Pi(\psi)$$

La perte de l'ordre est due à la différence de comportement de l'axiome d'additivité du cadre probabiliste et de l'axiome de maxitivité du cadre possibiliste.

1. Si $p(ABC)$ est une distribution de probabilités conjointes aux trois variables binaires A, B et C, alors la distribution marginale $p(A)$ s'obtient depuis $p(ABC)$ en sommant pour obtenir $P(A{=}\neg a) = p(\neg abc) + p(\neg ab\neg c) + p(\neg a\neg bc) + p(\neg a\neg b\neg c)$. Et de manière similaire on obtient $P(A{=}a)$.

2. $\rhd$ est toujours supposée satisfaire les principes de préservation de préférence et de consistance de Dubois et Prade.

En conséquence de la Proposition 1, si l'univers du discours Ω est le produit cartésien de l'ensemble des domaines des variables, alors la marginalisation sur un sous-ensemble de variables ne préservera pas l'ordre relatif des événements après la transformation. Prenons un exemple afin d'illustre notre propos.

Exemple 3. *Soient deux variables binaires A et B. On considère la distribution de probabilités telle que représentée par la Table 1a. Nous pouvons marginali-*

A	B	$p(A,B)$
a	b	0.4
a	$\neg b$	0.1
$\neg a$	b	0.3
$\neg a$	$\neg b$	0.2

(a)

A	B	$\pi(A,B)$
a	b	1
a	$\neg b$	x_1
$\neg a$	b	x_2
$\neg a$	$\neg b$	x_3

(b)

TABLE 1 – Distribution de probabilités et sa transformée par $\rhd$ pour l'illustration de la Proposition 1

ser sur A ou B, ainsi $p(a)$=0.5 et $p(\neg a)$=0.5 et pour B, $p(b)$=0.7 et $p(\neg b)$=0.3. En transformant les marginales avec $\rhd$, nous obtenons $\pi(a)$=1 et $\pi(\neg a)$=1 et pour B, $\pi(b)$=1 et $\pi(\neg b)$=$y < 1$. Notons x_1, x_2, x_3 les degrés de possibilités (avec $x_2{>}x_3{>}x_1$) qui varient selon la transformation utilisée.

Maintenant, considérons la construction des distributions de possibilités marginales à partir de la transformation de la Table 1a, que l'on trouve dans la Table 1b. Les x_i sont ordonnées telles que : $x_1{<}x_3{<}x_2{<}1$. En marginalisant sur B, $\pi''(b)$=1 et $\pi''(\neg b)$=$max(x_1, x_3)$. En marginalisant sur A, $\pi''(a)$=1 et $\pi''(\neg a)$=$max(x_2, x_3)$. Par conséquent, en transformant les marginales, on a $\pi''(a)$=$\pi''(\neg a)$=1 alors qu'en marginalisant la transformée, on a $\pi'(a)$>$\pi'(\neg a)$. Donc aucune transformation $\rhd$ préservant l'ordre des interprétations ne peut préserver la marginalisation.

4.2 Transformations et conditionnement

Dans cette partie, nous essayons de répondre à la question suivante : *Étant donnée une distribution de probabilités p, une distribution de possibilités π (obtenue par transformation) et une observation, trouvons nous le même ordre après conditionnement en probabilités et en possibilités ?* La Proposition 2 montre que l'ordre des interprétations élémentaires est préservé après le conditionnement si la transformation utilisée préserve l'ordre des interprétations.

Proposition 2. *Soit $\phi{\subseteq}\Omega$ une évidence, et $\rhd$ une transformation probabiliste-possibiliste. On suppose que p' est la distribution de probabilités obtenue en* *conditionnant p par ϕ, $\pi'' = \triangleright(p')$ et π' est la distribution de possibilités obtenue en conditionnant $\pi = \triangleright(p)$ par ϕ. Alors, $\forall \omega_i, \omega_j \in \Omega$,*

$$\pi'(\omega_i) < \pi'(\omega_j) \Leftrightarrow \pi''(\omega_i) < \pi''(\omega_j)$$

et

$$\pi'(\omega_i) = \pi'(\omega_j) \Leftrightarrow \pi''(\omega_i) = \pi''(\omega_j)$$

La Proposition 2 est valide dans les deux cadres possibilistes, quantitatif et qualitatif.

Idée de la preuve. Dans le cadre quantitatif comme dans le cadre qualitatif, la preuve est simple. Il suffit de considérer deux interprétations appartenant à l'ensemble des modèles de ϕ. Ainsi, soient deux interprétations $\omega_1, \omega_2 \in \phi$ (puisque si $\omega_i \notin \phi$ alors $p(\omega_i)$=0, de même en possibilités) avec $p(\omega_1)<p(\omega_2)$, nous regardons après conditionnement l'ordre des deux interprétations, qui est préservé et puisque la transformation probabiliste-possibiliste $\triangleright$ préserve l'ordre des interprétations, $\pi(\omega_1)<\pi(\omega_2)$. □

Voyons-en un exemple dans le cadre quantitatif.

Exemple 4. *Soit la distribution de probabilités de la Table 2, décrivant les croyances sur 3 variables A, B et C. Nous souhaitons conditionner par ϕ=$\neg a$.*

A	B	C	$p(A,B,C)$	$\pi(A,B,C)$
$\neg a$	$\neg b$	$\neg c$	0.216	α_1
$\neg a$	$\neg b$	c	0.27	1
$\neg a$	b	$\neg c$	0.144	α_2
$\neg a$	b	c	0.27	1
a	$\neg b$	$\neg c$	0.012	α_5
a	$\neg b$	c	0.048	α_3
a	b	$\neg c$	0.028	α_4
a	b	c	0.012	α_5

TABLE 2 – Distribution de probabilités et sa transformée par $\triangleright$ où $1>\alpha_1>\alpha_2>\alpha_3>\alpha_4>\alpha_5>0$.

On utilise la définition du conditionnement donnée dans la Section 2. Les résultats sont décrits par la Table 3, où on trouve dans la Table 3a, la distribution résultant du conditionnement puis de la transformation (ie. π''); Et dans la Table 3b, on trouve la distribution résultante de la transformation puis du conditionnement, autrement dit π'. Maintenant si l'on regarde la Table 3, nous pouvons voir que l'ordre est respecté puisqu'on a : $\alpha_1 < \alpha_2$ et $\beta_1 < \beta_2$.

Comme conséquence immédiate de la préservation de l'ordre des interprétations par l'opération de conditionnement, nous pouvons déduire de la Proposition 2 que pour les requêtes MPE, les réponses aux requêtes

$A\ B\ C$	$\pi''(ABC\|\neg a)$
$\neg a \neg b \neg c$	β_1
$\neg a \neg b\ c$	1
$\neg a\ b \neg c$	β_2
$\neg a\ b\ c$	1
$a \neg b \neg c$	0
$a \neg b\ c$	0
$a\ b \neg c$	0
$a\ b\ c$	0

(a)

$A\ B\ C$	$\pi'(ABC\|\neg a)$
$\neg a \neg b \neg c$	α_1
$\neg a \neg b\ c$	1
$\neg a\ b \neg c$	α_2
$\neg a\ b\ c$	1
$a \neg b \neg c$	0
$a \neg b\ c$	0
$a\ b \neg c$	0
$a\ b\ c$	0

(b)

TABLE 3 – Conditionnement quantitatif de la Table 2.

seront les mêmes en probabilités et en possibilités après la transformation, comme le stipule la proposition suivante :

Proposition 3. *Soit p une distribution de probabilités et $\triangleright$ une transformation préservant l'ordre des interprétations alors $\forall \phi \subseteq \Omega$,*

$$argmax_{\omega_i \in \Omega}(p(\omega_i, \phi)) = \ argmax_{\omega_i \in \Omega}(\pi(\omega_i, \phi))$$

Idée de la preuve. L'idée de la preuve tient au fait que la réponse aux requêtes MPE dépend uniquement de l'ordre des interprétations qui est préservé par la transformation par hypothèse. □

La Proposition 3 affirme que les requêtes MPE ont les mêmes réponses dans les deux cadres. Toutefois, lorsqu'on s'intéresse à des requêtes de type MAP, nous n'obtenons pas les mêmes réponses, cette perte est due à la marginalisation. En effet, les requêtes MAP sont évaluées en utilisant le conditionnement et la marginalisation. La réponse aux requêtes MAP n'est pas la même à cause de la marginalisation qui ne préserve pas l'ordre des événements (voir Proposition 1).

Proposition 4. *Soit $\mathcal{A}$=$\{A_1, A_2, .., A_n\}$ un ensemble de variables discrètes. Soit $E \subset \mathcal{A}$, un sous-ensemble de variables observées et soit e la valeur observée pour E (E=e). Soit un autre sous-ensemble de variables $Q \subset \mathcal{A}$ tel que $Q \cap E$=$\emptyset$ et $Q \cup E \neq \mathcal{A}$. Il existe p une distribution de probabilités et π la distribution de possibilités résultant de la transformation $\triangleright$, tel que :*

$$argmax_{q \in D_Q}(p(q|e)) \neq argmax_{q \in D_Q}(\pi(q|e))$$

avec D_Q le domaine associé aux variables de Q (produit cartésien des domaines des variables $A_i \in Q$).

4.3 Relations d'indépendance et transformations

La nature incomplète et incertaine des informations que l'on traite nous amène à nous intéresser à la notion d'indépendance, qui est une notion très importante dans les modèles graphiques. Pour rappels, soient δ, ϕ et ψ trois événements, en probabilités (resp. en possibilités), ϕ est dit indépendant de ψ sachant δ si et seulement si $P(\phi|\psi\delta)$=$P(\phi|\delta)$ (resp. $\Pi(\phi|\psi\delta)$=$\Pi(\phi|\delta)$).

Cette section analyse si les relations d'indépendance entre événements sont préservées lors des transformations. Bien évidement, le concept d'indépendance est lié à ceux du conditionnement et de la marginalisation. La Proposition 5 montre qu'il n'existe pas de transformation $\triangleright$ préservant les relations d'indépendance.

Proposition 5. *Soient 3 événements ϕ, ψ et $\delta \subseteq \Omega$. Soit $\triangleright$ une transformation probabiliste-possibiliste, alors il existe une distribution de probabilités p et $\pi = \triangleright(p)$ tel que,*

$$P(\phi|\psi\delta) = P(\phi|\delta) \not\Rightarrow \Pi(\phi|_{\otimes}\psi\delta) = \Pi(\phi|_{\otimes}\delta)$$

Dans la Proposition 5, $|_{\otimes}$ représente soit le conditionnement quantitatif soit le conditionnement qualitatif. Les deux exemples suivants démontrent la perte des relations d'indépendance.

Exemple 5 (Cadre qualitatif)**.** *Soit p une distribution probabiliste et $\pi = \triangleright(p)$ la distribution possibiliste. La Table 4 décrit les deux distributions où :*

A et B sont indépendants : $P(A|B)$=$P(A)$ et $P(B|A)$=$P(B)$.

A	B	$p(A,B)$	$\pi(A,B)$
a	b	0.12	β_4
a	$\neg b$	0.18	β_3
$\neg a$	b	0.28	β_2
$\neg a$	$\neg b$	0.42	1

Table 4 – Distribution de probabilités et la distribution de possibilités obtenue par transformation.

Puisque la transformation $\triangleright$ préserve l'ordre des interprétations alors $1 > \beta_2 > \beta_3 > \beta_4$.
Il s'agit maintenant de vérifier si $\Pi(A|_{min}B)$=$\Pi(A)$. En particulier, pour $\Pi(a|_{min}b)$ et $\Pi(a)$ on trouve :

- $\Pi(a|_{min}b)$=β_4
- $\Pi(a)$=β_3

On peut conclure que puisqu'il n'y a pas d'égalité, l'indépendance n'est pas préservée.

Exemple 6 (Cadre quantitatif)**.** *Soit p une distribution probabiliste et $\pi = \triangleright(p)$ la distribution possibiliste. La Table 5 décrit les deux distributions où :*

- *les événements $\phi = \{\omega_3, \omega_4, \omega_5\}$ and $\psi = \{\omega_2, \omega_3, \omega_4, \omega_5\}$ sont indépendants en probabilités. En effet, $P(\phi \mid \psi) = \frac{P(\phi\cap\psi)}{P(\psi)} = 0.3 = P(\phi)$.*

ω_i	$p(\omega_i)$	$\pi(\omega_i)$
ω_1	0.42	1
ω_2	0.18	α_1
ω_3	0.12	α_2
ω_4	0.10	α_3
ω_5	0.08	α_4
ω_6	0.05	α_5
ω_7	0.05	α_5

Table 5 – Distribution de probabilités et distribution de possibilités obtenue par transformation.

Puisque la transformation $\triangleright$ préserve l'ordre des interprétations alors $1 > \alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 > \alpha_5$.
Le calcul de $\Pi(\phi \mid \psi)$ et de $\Pi(\phi)$ nous donne :

- $\Pi(\phi \mid \psi) = \frac{\Pi(\phi\cap\psi)}{\Pi(\psi)} = \frac{\alpha_2}{\alpha_1}$
- $\Pi(\phi) = \alpha_2$

Mais, $\frac{\alpha_2}{\alpha_1} \neq \alpha_2$ puisque $\alpha_1 < 1$. Alors l'équation $\Pi(\phi \mid \psi) = \Pi(\phi)$ ne tient pas, donc la préservation des événements non plus.

On peut déduire que puisque les indépendances d'événements ne sont pas préservées, alors les indépendances de variables ne sont pas préservées non plus.

La perte des relations d'indépendance entre les variables dans la distribution jointe pose un vrai problème en terme de perte d'informations, c'est pour cela que l'on s'intéresse maintenant à transformer directement un modèle graphique, le but étant de préserver ces relations d'indépendance.

5 Transformations dans les modèles graphiques

Transformer un modèle graphique probabiliste, ici un réseau Bayésien, vers un réseau possibiliste [16] consiste à transformer les tables locales et conserver la structure qui encode les relations d'indépendance conditionnelle.

Définition 1. *Soit un réseau Bayésien $\mathcal{BN}$ sur un ensemble de variables $\mathcal{A}$=$\{A_1, .., A_n\}$, le réseau possibiliste $\mathcal{PN}$ sur le même ensemble de variables A est le réseau résultant de la transformation, il est composé de :*

- *Une composante graphique qui est la même que le réseau $\mathcal{BN}$.*
- *Une composante numérique, qui est telle que l'ensemble des tables locales de $\mathcal{PN}$ sont celles de $\mathcal{BN}$ transformées par $\triangleright$.*

L'avantage de transformer directement un réseau Bayésien en réseau possibiliste utilisant la Définition 1, est que l'on préserve les indépendances, et que calculatoirement cela est moins coûteux de transformer un ensemble de tables locales que de transformer une distribution jointe. Le problème maintenant est qu'il n'y a pas de garantie que l'ordre des interprétations et des événements soient préservés dans le réseau possibiliste résultant et sa distribution jointe. La Figure 2 illustre la question de la transformation d'un réseau probabiliste (Bayésien) en réseau possibiliste.

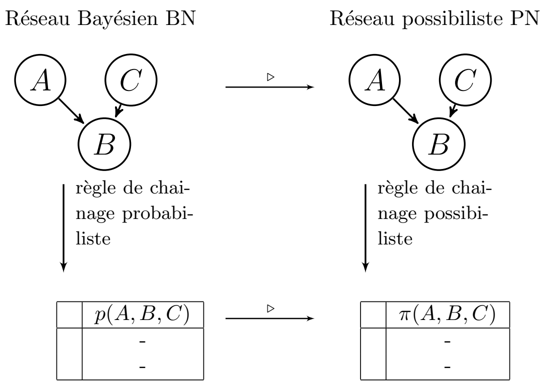

FIGURE 2 – Transformation de modèles graphiques

Nous nous intéressons donc à la différence entre l'ordre induit par la règle de chaînage probabiliste et l'ordre induit par la transformation puis le chaînage possibiliste. La Proposition 6 répond à la question : *Est-ce que l'ordre induit par la distribution $p_{\mathcal{BN}}$ (distribution jointe encodée par $\mathcal{BN}$) est préservé dans la distribution $\pi_{\mathcal{PN}}$ (la distribution jointe encodée par le réseau possibiliste $\mathcal{PN}$) ?*

Proposition 6. *Soit $\triangleright$ une transformation probabiliste-possibiliste. Alors il existe un réseau Bayésien $\mathcal{BN}$ tel que $\exists\omega_1 \in \Omega$, $\exists\omega_2 \in \Omega$ où :*

$$\pi'(\omega_1) < \pi'(\omega_2) \not\Rightarrow \pi''(\omega_1) < \pi''(\omega_2)$$

où : i) $\pi'(\omega) = \triangleright(p(\omega))$, et p est la distribution jointe induite par $\mathcal{BN}$ et ii) π'' est la distribution jointe induite par $\mathcal{PN}$ en utilisant la Définition 1.

Le contre-exemple qui suit prouve la Proposition 6 dans le cadre qualitatif.

Exemple 7. *Soit $\mathcal{BN}$ le réseau Bayésien de la Figure 3 sur les variables A et B. Notons que la distribution de probabilités $p(A)$ dans $\mathcal{BN}$ est une permutation[3] de la distribution de probabilités $p(B)$. Ainsi, la transformation de $p(A)$ et $p(B)$ par $\triangleright$ nous donne respectivement $\pi(A)$ et $\pi(B)$ où $\pi(B)$ est aussi une permutation de $\pi(A)$. Dans cet exemple, puisque $\triangleright$ est supposé préserver l'ordre des interprétations, nous avons $1{>}\alpha_1{>}\alpha_2{>}\alpha_3$.*

3. La propriété de permutation dans les transformations probabiliste-possibiliste est discutée dans [18].

A	$p(A)$	$\pi(A)$
a_1	0.4	1
a_2	0.2	α_2
a_3	0.25	α_1
a_4	0.15	α_3

B	$p(B)$	$\pi(B)$
b_1	0.15	α_3
b_2	0.2	α_2
b_3	0.25	α_1
b_4	0.4	1

FIGURE 3 – Exemple de transformation de réseaux Bayésiens-possibilistes.

Regardons de près les degrés de probabilités et de possibilités des interprétations a_1b_1 et a_2b_2 :

— $p(a_1b_1) = 0.4 * 0.15 = 0.06$

— $p(a_2b_2) = 0.2 * 0.2 = 0.04$

alors, $p(a_1b_1) > p(a_2b_2)$ $(*_1)$

— $\pi(a_1b_1) = \alpha_3$

— $\pi(a_2b_2) = \alpha_2$

alors, $\pi(a_1b_1) < \pi(a_2b_2)$ $(*_2)$

A partir de $(*_1)$ et $(*_2)$ il est clair que l'ordre de ces deux interprétations est inversé peu importe la transformation qui sera utilisée.

De la même manière, dans le cadre quantitatif, nous pouvons prouver que l'ordre ne sera pas préservé quelque soit la transformation utilisée.

Pour résumer, si on transforme une distribution de probabilités jointe induite par $\mathcal{BN}$, on préserve l'ordre des interprétations mais cela est coûteux (la taille ie. le nombre d'interprétations de la distribution jointe est exponentielle dans le nombre de variables de $\mathcal{BN}$) et on perd les relations d'indépendance. Si on transforme directement un réseau Bayésien selon la Définition 1, on préserve les indépendances, c'est moins coûteux mais on risque de perdre l'ordre des interprétations.

6 Discussions et conclusions

Ce travail préliminaire traite des problèmes des transformations probabilistes-possibilistes, plus précisément des problèmes lors du raisonnement et des problèmes des transformations dans les modèles graphiques. Nous avons montré qu'il n'existe pas de transformation préservant l'ordre des interprétations qui puisse préserver l'ordre des événements. En plus de cela, nous avons montré que pour l'opération de marginalisation, nous ne pouvons pas trouver de transformation qui nous donne le même ordre sur les événements, lorsqu'on transforme puis marginalise ou l'inverse (que l'on marginalise puis que l'on transforme). Quant à la question des relations d'indépendance entre variables ou événements, le papier montre que les transformations ne préservent pas les indépendances. Enfin, dans la dernière partie, nous nous sommes penchés sur les transformations au sein des modèles graphiques. Malgré le fait que les relations d'indépendance soient préservées, si nous regardons les distributions jointes, l'ordre des interprétations ne sera pas toujours le même.

Dans la littérature, il existe principalement deux travaux qui traitent des mêmes problèmes que dans le présent article. Dans le premier travail [16], les auteurs étudient les transformations de réseaux Bayésiens vers les réseaux possibilistes. Ils étendent la définition du principe de consistance, afin de préserver l'ordre des interprétations dans les distributions obtenues après transformation. Notons que ce travail s'intéresse plus aux transformations existantes (comme OT et ST) alors que nous avons essayé de travailler de manière générale en considérant toutes les transformations vérifiant le principe de préservation de préférences de Dubois et Prade. Le second travail [18], quant à lui, a étudié cette question des transformations à travers les tâches de raisonnement. Son objectif était de montrer que les distributions résultantes n'étaient pas identiques (au niveau des degrés de possibilités). Dans notre travail, nous nous sommes plus concentrés à montrer qu'en plus d'être différentes, les distributions n'induisent pas le même ordre entre les événements. Certaines de ces questions ont été étudiées dans le contexte de l'analyse d'intervalles flous [9].

Une question qui mérite d'être développée est de savoir s'il existe des formes particulières de distributions de probabilités qui pour n'importe quelles transformations $\triangleright$ préserveraient l'ordre des événements (après la marginalisation, par exemple). L'idée la plus triviale concerne les distributions de probabilités uniformes. En effet, toute transformation $\triangleright$ doit préserver la normalisation, ce qui donne une distribution de possibilités uniforme (où chaque état est associé à un degré de possibilité égal à 1). Par conséquent, les tâches de raisonnement appliquées aux réseaux ou aux distributions ne modifieront pas les connaissances (e.g. le sens des interprétations). Entre autres, les distributions marginales que l'on pourra calculer seront des distributions uniformes. Dans les modèles graphiques, l'ordre des interprétations dans les distributions jointe sera le même. Un autre type de distributions de probabilités est appelé : "atomic bound system" [17] ou encore "big-stepped" [6, 2] ou "lexicographic probability distributions", une telle distribution p est définie par : $\forall \omega_i \in \Omega, p(\omega_i) > \sum\{p(\omega_j) : \omega_j \in \Omega \text{ et } p(\omega_j) < p(\omega_i)\}$. Clairement, si p est une distribution "big-stepped" alors la transformation $\triangleright$ préserve l'ordre entre les interprétations après marginalisation. Notons que pour les deux types de distributions décrits ci-dessus (uniforme et big-stepped), l'ordre entre les événements non élémentaires n'est pas préservé.

Il est connu que les transformations probabilistes-possibilistes souffrent d'une perte d'informations puisqu'on passe d'un cadre additif à un cadre qualitatif ou semi-qualitatif. Mais l'impact sur les tâches de raisonnement sur les distributions n'était pas complètement étudié. Les résultats obtenus confirment cette perte d'information à plusieurs niveaux. Ce n'est pas pour autant qu'on ne peut rien faire avec les transformations. En particulier, les réponses aux requêtes MPE ne sont pas affectées par les transformations, contrairement aux requêtes MAP. Nos travaux futurs étudieront en particulier l'inférence MAP dans les réseaux crédaux (basés sur les probabilités imprécises et connus pour leur très grande complexité en comparaison avec les réseaux Bayésiens ou possibilistes) en les transformant en réseaux possibilistes. Cette piste peut apporter une bonne et efficace approximation pour l'inférence MAP avec un meilleur coût de calcul. D'autres perspectives sont envisageables, notamment concernant la commutativité des transformations avec d'autres définitions de conditionnement et d'indépendances dans le cadre possibiliste.

Références

[1] C. Baudrit, I. Couso, and D. Dubois. Joint propagation of probability and possibility in risk analysis : Towards a formal framework. *International Journal of Approximate Reasoning*, 45(1) :82 – 105, 2007.

[2] S. Benferhat, D. Dubois, and H. Prade. Possibilistic and standard probabilistic semantics of conditional knowledge bases. *J. Log. Comput.*, 9(6) :873–895, 1999.

[3] C. Borgelt, J. Gebhardt, and R. Kruse. Graphical models. In *In Proceedings of International School for the Synthesis of Expert Knowledge 98*, pages 51–68. Wiley, 2002.

[4] A. Darwiche. *Modeling and Reasoning with Bayesian Networks.* Cambridge University Press, 2009.

[5] S. Destercke, D. Dubois, and E. Chojnacki. Transforming probability intervals into other uncertainty models. In *5th EUSFLAT Conference, Ostrava, Czech Republic, September 11-14, 2007, Volume 2 : Regular Sessions*, pages 367–373, 2007.

[6] D. Dubois, H. Fargier, and H. Prade. Ordinal and probabilistic representations of acceptance. *J. Artif. Int. Res.*, 22(1) :23–56, July 2004.

[7] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10(4) :273–297, 2004.

[8] D. Dubois and H. Prade. *Possibility Theory : An Approach to Computerized Processing of Uncertainty.* Plenum Press, New York, 1988.

[9] D Dubois and H Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 42(1) :87–101, 1991. DP165.

[10] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In *Fuzzy Logic*, pages 103–112. 1993.

[11] E. Hisdal. Conditional possibilities independence and non interaction. *Fuzzy Sets and Systems*, pages 283–297, 1978.

[12] G. J. Klir and J. F. Geer. Information-preserving probability- possibility transformations : Recent developments. In *Fuzzy Logic*, pages 417–428. 1993.

[13] M.-H. Masson and T. Denoeux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157(3) :319–340, 2006.

[14] P. Billaudel M.S Mouchaweh, M.S Bouguelid and B. Riera. Variable probability-possibility transformation. pages 417–428. September 2006.

[15] H. Prade and D. Dubois. *Fuzzy Sets and Systems : Theory and Applications*. Academic Press, Inc., Orlando, FL, USA, 1997.

[16] Y. B. Slimen, R. Ayachi, and N. B. Amor. Probability-possibility transformation : - application to bayesian and possibilistic networks. In *WILF*, volume 8256 of *Lecture Notes in Computer Science*, pages 122–130. Springer, 2013.

[17] P. Snow. Standard probability distributions described by rational default entailment. 1996.

[18] T. Sudkamp. On probability-possibility transformations. *Fuzzy Sets and Systems*, pages 73–81, 1992.

[19] Y.Y. Wee, W.P. Cheah, S.C. Tan, and K. Wee. A method for root cause analysis with a bayesian belief network and fuzzy cognitive map. *Expert Syst. Appl.*, 42(1) :468–487, 2015.

[20] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100 :9–34, 1999.