

LUDWIG-MAXIMILIANS UNIVERSITÄT MÜNCHEN

MASTER-STUDIENGANG STATISTIK MIT WIRTSCHAFTS- UND  
SOZIALWISSENSCHAFTLICHER AUSRICHTUNG

MASTER-SEMINAR: (IMPRECISE) PROBABILISTIC GRAPHICAL MODELS

HAUSARBEIT

# Einführung in die diskreten Bayes- und Kredal-Netze



Patrick Schwaferts

Betreuer:

M.Sc. Eva Endres

M.Sc. Paul Fink

Datum: 04.03.2016

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Diskrete Bayes-Netze</b>	<b>1</b>
2.1	Definition eines Bayes-Netzes	1
2.1.1	Gerichtete Azyklische Graphen	2
2.1.2	Bedingte Verteilungen	2
2.1.3	Faktorisierung einer Verteilung	3
2.1.4	Bayes-Netz	3
2.2	Abhängigkeitsmodelle	5
2.3	Unabhängigkeitsaussagen eines Graphen	5
2.3.1	Lokale Markov Unabhängigkeiten	6
2.3.2	d-separation und Globale Markov Unabhängigkeiten	6
2.4	Graphoide	8
2.5	Unabhängigkeiten der gemeinsamen Verteilung	9
2.6	Ausblick	10
<b>3</b>	<b>Diskrete Kredal-Netze</b>	<b>10</b>
3.1	Kredal-Mengen	10
3.1.1	Definition, Darstellung und Extrempunkte	10
3.1.2	Bedingte und Marginale Kredal-Mengen	13
3.1.3	Komposition von Kredal-Mengen	15
3.2	Kredal-Netze	17
3.2.1	Definition	17
3.2.2	Starke Unabhängigkeit	18
3.2.3	Starke Erweiterung	19
3.2.4	Epistemische Unabhängigkeit	21
3.2.5	Epistemische Erweiterung	22
<b>4</b>	<b>Zusammenfassung</b>	<b>22</b>
	<b>Literaturverzeichnis</b>	<b>24</b>

# 1 Einleitung

Die gemeinsame Verteilung von diskreten Zufallsvariablen besteht aus einer Wahrscheinlichkeitsangabe für jede Kombination der Variablenausprägungen. Bei wachsender Anzahl an Variablen und Ausprägungen, steigt die Zahl der Einträge exponentiell an und die gemeinsame Verteilung wird somit extrem umfangreich [1]. Jedoch ist in der gemeinsamen Verteilung häufig eine Struktur enthalten: Manche Variablen sind voneinander bedingt unabhängig, so dass eine Vielzahl von Wahrscheinlichkeitsangaben identisch sind. Folglich ist es möglich diese gemeinsame Verteilung kompakter darzustellen. Eine Möglichkeit dazu stellen Bayes-Netze dar, die die Unabhängigkeitsstrukturen der Zufallsvariablen graphentheoretisch modellieren.

Des Weiteren gibt es Situationen, in denen präzise Wahrscheinlichkeitsangaben entweder nicht möglich oder mit einem enormen Aufwand verbunden sind. Ein Beispiel dazu sind Expertenmeinungen und -entscheidungen im medizinischen Kontext: Ein Arzt ist meistens nicht in der Lage eine präzise Wahrscheinlichkeitsangabe für das Auftreten eines Symptoms bei einer bestimmten Krankheit zu geben; vielmehr kann er diese Wahrscheinlichkeit mittels eines Intervalls oder Beschränkungen abschätzen. Für derartige Fälle lohnt es sich Modelle mit Mengen von Verteilungen, so genannten Kredal-Mengen, anstelle von präzisen Wahrscheinlichkeitsverteilungen aufzustellen. Auch Bayes-Netze können auf Kredal-Mengen generalisiert werden. Man spricht in diesem Fall von Kredal-Netzen [2].

Eine Einführung in diskrete Bayes- und Kredal-Netze stellt das Thema dieser Arbeit dar. Es werden in Kapitel 2 zuerst Bayes-Netze definiert und anschließend ihre Unabhängigkeiten untersucht. Dazu werden Unabhängigkeitsaussagen allgemein definiert, aufgezeigt inwiefern sich Unabhängigkeitsaussagen aus Graphen extrahieren lassen, dargestellt unter welchen Umständen sich Unabhängigkeitsaussagen graphisch darstellen lassen und schlussendlich gezeigt, dass in Bayes-Netzen (fast) alle bedingten Unabhängigkeiten graphisch erfasst werden. Kapitel 3 thematisiert Kredal-Netze: Nach einer Einführung in Kredal-Mengen können Kredal-Netze definiert werden. Es folgt die Darstellung von zwei möglichen Unabhängigkeitskonstrukten in Bezug auf Kredalmengen, nämlich die starke und die epistemsiche Unabhängigkeit, jeweils mit ihren Erweiterungen, sprich auf dem Kredal-Netz basierenden gemeinsamen Kredal-Mengen, in denen diese Unabhängigkeitskonstrukte gelten.

## 2 Diskrete Bayes-Netze

### 2.1 Definition eines Bayes-Netzes

Ausgangspunkt eines Bayes-Netzes sind im Rahmen dieser Arbeit diskrete Zufallsvariablen  $X_1, \dots, X_n$ ,  $n \in \mathbb{N}$  und es sei ihre gemeinsame Verteilung  $P = P(X_1, \dots, X_n)$  von Interesse. Diese beinhaltet in aller Regel bedingte Unabhängigkeiten, bzw. äquivalent dazu

Einflussstrukturen innerhalb der Zufallsvariablen, die durch einen gerichteten azyklischen Graphen, beispielhaft in Abbildung 1 auf Seite 4, dargestellt werden können. Folglich sind für eine komplette Beschreibung von  $P$  lediglich die bedingten Verteilungen der einzelnen Zufallsvariablen gegeben ihre auf sie Einfluss nehmenden Zufallsvariablen nötig. Dies alles basiert darauf, dass es sich bei  $P$  auch wirklich um eine Verteilung handelt, die sich auch aus eben beschriebenen bedingten Verteilungen errechnen lässt, sprich sich über den Graphen faktorisieren lässt.

Vor der formalen Definition des Bayes-Netz sollen erst drei Konzepte, nämlich gerichtete azyklische Graphen, auf diesen basierende bedingte Verteilungen und die Faktorisierung einer Verteilung, eingeleitet werden. Dieses Kapitel orientiert sich im Groben an [3, Kapitel 3]; sofern nicht anders angegeben, sei als Referenz auf diese Quelle verwiesen.

### 2.1.1 Gerichtete Azyklische Graphen

Ein *Graph* bezeichnet eine Menge von Knoten die mittels Kanten verbunden sein können. Dieser heißt *gerichtet*, falls die Kanten mit einer Orientierung versehen sind. Es sind also zwei verschiedene, in entgegengesetzte Richtungen zeigende Kanten zwischen jedem Knotenpaar möglich.

Folgt man den Orientierungen über mehrere Kanten hinweg, so ist es unter Umständen möglich, wieder den ursprünglichen Knoten zu erreichen und man bezeichnet diese Kantenfolge als Zyklus. Ein gerichteter Graph, der keine Zyklen enthält, ist ein *gerichteter azyklischer Graph* (DAG, „directed acyclic graph“). Ein Beispiel eines DAG ist in Abbildung 1 zu sehen. Besitzt ein Knotenpaar in einem gerichteten Graphen Kanten in beide Richtungen, so stellt dies natürlich auch einen Zyklus dar. Folglich sind jeweils zwei Knoten in einem DAG mit maximal einer Kante verbunden.

Im Kontext von Bayes-Netzen stellen die Knoten des Graphen die Zufallsvariablen  $X_1, \dots, X_n$  dar und gerichtete Kanten repräsentieren (häufig kausale [4]) Einflüsse zwischen diesen Variablen.

### 2.1.2 Bedingte Verteilungen

Die Zufallsvariablen, die nach einem DAG  $\mathcal{G}$  direkt auf eine Zufallsvariable  $X_i$  Einfluss nehmen, werden ihre *Eltern* genannt und mit  $Pa_{X_i}^{\mathcal{G}}$  bezeichnet, wobei  $Pa_{X_i}^{\mathcal{G}}$  deren Zustandsraum und  $pa_{X_i}^{\mathcal{G}} \in Pa_{X_i}^{\mathcal{G}}$  eine konkrete Realisierung davon darstellt.

Die bedingte Verteilung einer Zufallsvariable  $X_i$  hängt also nur von ihren Eltern  $Pa_{X_i}^{\mathcal{G}}$

ab<sup>1</sup> und beinhaltet für jede Ausprägung der Eltern eine Verteilung:

$$P(X_i | Pa_{X_i}^{\mathcal{G}}) = \{P(X_i | pa_{X_i}^{\mathcal{G}}) | pa_{X_i}^{\mathcal{G}} \in Pa_{X_i}^{\mathcal{G}}\}.$$

Dies kann beispielhaft in Abbildung 1 auf Seite 4 gesehen werden. Die Wahrscheinlichkeitsangaben für jedes  $pa_{X_i}^{\mathcal{G}}$  summieren sich hier zu 1 und stellen somit eine Verteilung dar.

### 2.1.3 Faktorisierung einer Verteilung

Eine gemeinsame Verteilung  $P$  der Zufallsvariablen  $X_1, \dots, X_n$  kann mithilfe eines DAG  $\mathcal{G}$  und der dazugehörigen bedingten Verteilungen durch folgende *Kettenregel für Bayes-Netze* errechnet werden:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa_{X_i}^{\mathcal{G}}). \quad (2.1)$$

Obwohl es, wie sich später herausstellen wird, zu jeder Verteilung  $P$  einen DAG  $\mathcal{G}$  gibt, so dass sich  $P$  über diese Formel (2.1) darstellen lässt (vgl. Satz 1 mit Satz 3), kann bei gegebenem  $\mathcal{G}$  nicht jede mögliche gemeinsame Verteilung  $P$  durch Formel (2.1) dargestellt werden, d.h. es gibt keine entsprechenden bedingten Verteilungen  $P(X_i | Pa_{X_i}^{\mathcal{G}})$ , so dass Formel (2.1) gilt.

Handelt es sich dennoch um eine Verteilung  $P$ , die sich bei gegebenem DAG  $\mathcal{G}$  über die Kettenregel für Bayes-Netze darstellen lässt, so spricht man davon, dass sich  $P$  nach  $\mathcal{G}$  *faktorisieren* lässt.

### 2.1.4 Bayes-Netz

Nach diesen Vorüberlegungen kann das Bayes-Netz folgendermaßen definiert werden.

**Definition 1.** Ein Bayes-Netz über die Zufallsvariablen  $X_1, \dots, X_n$  ist das geordnete Paar  $(\mathcal{G}, M)$ , bestehend aus einem auf  $X_1, \dots, X_n$  definierten, gerichteten azyklischen Graphen  $\mathcal{G}$  und einer Menge  $M = \{P(X_i | Pa_{X_i}^{\mathcal{G}}) | i = 1, \dots, n\}$  an bedingten Verteilungen der einzelnen Zufallsvariablen, wobei die jeweils bedingenden Zufallsvariablen als Eltern  $Pa_{X_i}^{\mathcal{G}}$  durch  $\mathcal{G}$  vorgegeben werden.

Zusätzlich wird angenommen, dass die gemeinsame Verteilung  $P$  von  $X_1, \dots, X_n$  nach  $\mathcal{G}$  *faktoriert* wird.

**Beispiel 1** (Bayes-Netz). Es werden die folgenden 4 binären Zufallsvariablen betrachtet (nach [1]):

---

<sup>1</sup>Mathematisch gesehen handelt es sich bei den Verteilungen der Zufallsvariablen ohne Eltern nicht um bedingte, sondern um marginale Verteilungen. Aus Gründen der Einfachheit werden diese beiden aber im Folgenden unter dem Begriff „bedingte Verteilungen“ zusammengefasst und es sei immer dann „bedingte und marginale Verteilungen“ gemeint, sofern Zufallsvariablen ohne Eltern mit vorkommen.

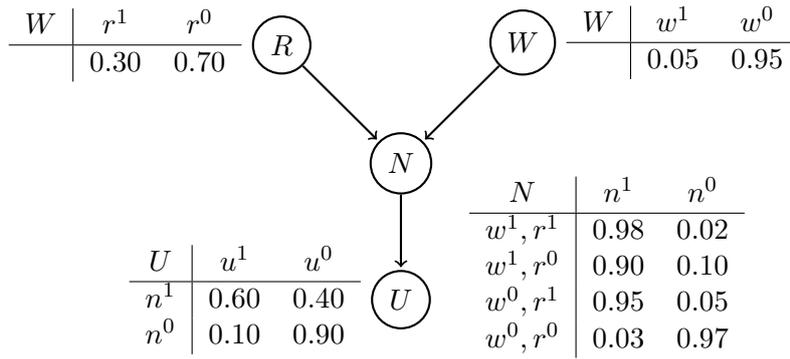


Abbildung 1: Bayes-Netz über 4 binäre Variablen.

( $R$ ) „Es regnet“

( $W$ ) „Es gibt einen Wasserschaden durch einen Rohrbruch“

( $N$ ) „Der Weg ist nass“

( $U$ ) „Auf dem Weg passiert ein Unfall“

Die in diesem Fall kausalen Einflüsse der Variablen untereinander, in Abbildung 1 dargestellt durch den Graphen  $\mathcal{G}_{\text{Regen}}$  und mit beispielhaften Wahrscheinlichkeitsangaben versehen, sind: Durch Regen und Wasserschaden folgt ein nasser Weg und dieser wiederum erhöht das Unfallrisiko. Allerdings sind diese Einflüsse nicht deterministisch in dem Sinne, dass die bedingte Wahrscheinlichkeit mancher Ereignisse gleich 1 ist. Beispielsweise folgt auf Regen ( $r^1$ ) nicht immer ein nasser Weg ( $P(n^1|w^1, r^1) = 0.98$  bzw.  $P(n^1|w^0, r^1) = 0.95$ ), denn es sind auch Situationen denkbar, in denen der Weg überdacht ist.

Die gemeinsame Verteilung  $P(W, R, N, U)$  errechnet sich für jede Kombination der Ausprägungen über die Kettenregel für Bayes-Netze (Formel (2.1)) aus den bedingten Verteilungen, die in Abbildung 1 mit arbiträr gewählten Werten dargestellt sind. Zum Beispiel ist

$$\begin{aligned} P(w^0, r^1, n^1, u^0) &= P(w^0) \cdot P(r^1) \cdot P(n^1|w^0, r^1) \cdot P(u^0|n^1) \\ &= 0.95 \cdot 0.3 \cdot 0.95 \cdot 0.4 = 0.1083. \end{aligned}$$

Obwohl in diesem Beispiel die Darstellungsweise mittels Bayes-Netz gleich viele Wahrscheinlichkeitsangaben (16) benötigt wie die gemeinsame Verteilung  $P(W, R, N, U)$ , ist leicht ersichtlich, dass bei steigender Variablen- und Ausprägungszahl des Bayes-Netz auf weniger Wahrscheinlichkeitsangaben angewiesen ist und diese Kompaktheit stark an Bedeutung gewinnt.

◆

## 2.2 Abhängigkeitsmodelle

Abhängigkeiten unter Zufallsvariablen können verschiedene Formen annehmen. Meistens wird darunter die stochastische Unabhängigkeit verstanden, die in der bedingten Version folgendermaßen definiert ist.

**Definition 2.** *Zwei Zufallsvariablen  $X$  und  $Y$  sind bedingt stochastisch unabhängig gegeben die Zufallsvariable  $Z$ , formal  $(X \perp\!\!\!\perp Y|Z)$ , wenn gilt*

$$P(x, y|z) = P(x|z) \cdot P(y|z) \quad \forall x \in X, \forall y \in Y, \forall z \in Z, \quad (2.2)$$

wobei  $X, Y, Z$  die Zustandsräume von  $X, Y, Z$  bezeichnen und letzte auch Vektoren von Zufallsvariablen sein können.

Andere Abhängigkeiten wären beispielsweise relationaler oder korrelationaler Art (siehe [1]) oder im Falle von Kredal-Mengen die in Kapitel 3.2 eingeführten Konzepte der starken Unabhängigkeit und der epistemische Unabhängigkeit.

Obwohl das Bayes-Netz in der Darstellung der gemeinsamen Verteilung das stochastische Unabhängigkeitskonzept der Zufallsvariablen verwendet, ist es möglich durch den Graphen auch andere Arten von Abhängigkeiten bzw. Unabhängigkeiten auszudrücken. Bevor dargestellt wird, inwiefern ein DAG diese repräsentiert, sollen Abhängigkeitsmodelle allgemein definiert werden (nach [1]):

**Definition 3.** *Eine Unabhängigkeitsaussage  $(X \perp\!\!\!\perp Y|Z)$  bedeutet „ $X$  ist unabhängig von  $Y$  gegeben  $Z$ “. Eine Menge  $I$  an Unabhängigkeitsaussagen innerhalb der Zufallsvariablen  $X_1, \dots, X_n$  bezeichnet man als Abhängigkeitsmodell über  $X_1, \dots, X_n$ .*

In konkreten Fällen wird anstelle der allgemeinen Unabhängigkeit  $\perp$  eine bestimmte Unabhängigkeit verwendet, so wie bei Bayes-Netzen die stochastische Unabhängigkeit  $\perp\!\!\!\perp$ .

## 2.3 Unabhängigkeitsaussagen eines Graphen

Da ein DAG die direkte Abhängigkeitsstruktur der Zufallsvariablen repräsentiert, lassen sich eine Reihe von (allgemeinen, nicht nur stochastischen) Unabhängigkeitsaussagen, d.h. ein Abhängigkeitsmodell, daraus ableiten. Manche davon sind vom Graphen direkt „ablesbar“ und werden lokale Unabhängigkeitsaussagen genannt; viele jedoch sind auf den ersten Blick nicht direkt ersichtlich. Allerdings lassen sie sich bestimmen, indem die Knoten des Graphen auf Separiertheit, die so genannte *d-separation*, untersucht werden. Alle so extrahierten Unabhängigkeitsaussagen, die die lokalen beinhalten, werden als globale Unabhängigkeitsaussagen bezeichnet und sind, wie noch beschrieben wird, eine vollständige Beschreibung der Abhängigkeitsstruktur des DAG.

### 2.3.1 Lokale Markov Unabhängigkeiten

Lokale Unabhängigkeitsaussagen lassen sich über die *Markov-Bedingung* ableiten: Jede Zufallsvariable  $X_i$  ist gegeben ihre Eltern  $Pa_{X_i}^{\mathcal{G}}$  bedingt unabhängig von allen Zufallsvariablen  $NonDes_{X_i}^{\mathcal{G}}$ , die weder ihre Eltern noch ihre Nachkömmlinge<sup>2</sup> darstellen.

**Definition 4.** Als lokale (Markov) Unabhängigkeiten  $I_l(\mathcal{G})$  eines DAG  $\mathcal{G}$  werden Unabhängigkeitsaussagen bezeichnet, die sich über die Markov-Bedingung von  $\mathcal{G}$  ableiten lassen:

$$I_l(\mathcal{G}) = \{(X_i \perp NonDes_{X_i}^{\mathcal{G}} | Pa_{X_i}^{\mathcal{G}}) | i = 1, \dots, n\} \quad (2.3)$$

[3, S. 57].

Bezogen auf ein Bayes-Netz bedeutet die Markov Bedingung, dass die so aus dem DAG entnommenen Unabhängigkeitsaussagen  $I_l(\mathcal{G})$  in der gemeinsamen Verteilung  $P$  als bedingte stochastische Unabhängigkeiten enthalten sind.

**Beispiel 2** (Lokale Markov Unabhängigkeiten). Im oben eingeführten Beispiel 1 lassen sich aus  $\mathcal{G}_{\text{Regen}}$  folgende lokale Unabhängigkeitsaussagen ableiten<sup>3</sup>

$$I_l(\mathcal{G}_{\text{Regen}}) = \{(W \perp R), (U \perp R|N), (U \perp W|N)\}.$$

Diese Unabhängigkeitsaussagen sind hier allgemein dargestellt. Wie später gezeigt wird, entsprechen im Kontext des Bayes-Netzes diese Aussagen bedingten stochastischen Unabhängigkeiten in der gemeinsamen Verteilung  $P(W, R, N, U)$ .

♦

### 2.3.2 d-separation und Globale Markov Unabhängigkeiten

Da ein DAG allerdings mehr als nur diese lokalen Unabhängigkeiten beinhaltet, bedarf es einer Methode, um weitere Unabhängigkeiten zu extrahieren. Dazu können jeweils zwei Knoten  $X$  und  $Y$  untersucht werden, ob - gegeben eine Menge an beobachteten Knoten  $Z$  - eine „Einflussübertragung“ zwischen diesen Knoten möglich ist, sprich ob die Kenntnis des einen die Wahrscheinlichkeiten des anderen beeinflusst. Ein Pfad<sup>4</sup>  $X \rightleftarrows \dots \rightleftarrows Y$  aus Knoten des DAG auf dem diese „Einflussübertragung“ möglich ist, wird aktiver Pfad zwischen  $X$  und  $Y$  genannt. Gibt es keinen aktiven Pfad zwischen den beiden Knoten, so bezeichnet man  $X$  und  $Y$  als *d-separated* gegeben  $Z$ . Mithilfe dieser Eigenschaft lassen

<sup>2</sup>Nachkömmlinge einer Zufallsvariablen  $X_i$  in einem DAG  $\mathcal{G}$  sind alle Zufallsvariablen  $NonDes_{X_i}^{\mathcal{G}}$  die von  $X_i$  ausgehend und den Kantenorientierungen folgend erreicht werden können.

<sup>3</sup>Da es sich bei Unabhängigkeit im Gegensatz zur Irrelevanz (siehe Abschnitt 3.2.4) um ein symmetrisches Konstrukt handelt, sind hier jeweils nur eine Version der Aussagen dargestellt. Beispielsweise gilt zu  $(W \perp R)$  natürlich auch  $(R \perp W)$ .

<sup>4</sup>Das Zeichen  $\rightleftarrows$  bedeutet hier das Vorhandensein einer Kante. Im Gegensatz zu den Zeichen  $\rightarrow$  und  $\leftarrow$  spielt die Orientierung der Kante dabei keine Rolle.

sich alle weiteren Unabhängigkeitsaussagen des Graphen extrahieren [5]. Diese Konzepte werden nun formal dargestellt:

**Definition 5.** Sei  $X_1 \rightleftarrows \dots \rightleftarrows X_m$  ein Pfad in einem DAG  $\mathcal{G}$  und  $Z$  eine Teilmenge von beobachteten Variablen. Der Pfad  $X_1 \rightleftarrows \dots \rightleftarrows X_m$  heißt aktiv gegeben  $Z$ , wenn folgendes gilt:

- Für jede so genannte  $v$ -Struktur  $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$  ist  $X_i$  oder eine ihrer Nachkömmlinge in  $Z$
- Jede Variable des Pfades, die nicht in dem Zentrum einer  $v$ -Struktur ist, liegt nicht in  $Z$

[3, S. 71].

**Beispiel 3** (Aktive Pfade). In  $\mathcal{G}_{\text{Regen}}$  ist der Pfad  $W \rightarrow N \leftarrow R$  nicht aktiv (gegeben  $\emptyset$ ): Wird beobachtet, dass es einen Wasserschaden gibt ( $w^1$ ), so liefert dies alleine noch keinen Hinweis darauf, ob es regnet ( $r^1$ ).

Ist allerdings zusätzlich gegeben, dass sich auf dem Weg ein Unfall ereignet ( $u^1$ ), so ändert ein beobachteter Wasserschaden ( $w^1$ ) die Wahrscheinlichkeit, dass es regnet ( $r^1$ ):  $W$  liefert nun einen zusätzlichen Hinweis auf  $R$ .  $W \rightarrow N \leftarrow R$  ist folglich aktiv gegeben  $U$ .

$W \rightarrow N \leftarrow U$  ist nicht aktiv gegeben  $N$ , da  $N$  diesen Pfad blockiert.  $W \rightarrow N \leftarrow U$  ist aber aus dem gleichen Grund auch nicht aktiv gegeben  $\{N, R\}$ .

◆

**Definition 6.** Zwei Variablen  $X$  und  $Y$  eines DAG  $\mathcal{G}$  sind  $d$ -separated ( $d$ : „directed“) gegeben eine Teilmenge an beobachteten Variablen  $Z$ , bezeichnet mit  $d\text{-sep}_{\mathcal{G}}(X; Y|Z)$ , wenn es keinen aktiven Pfad zwischen  $X$  und  $Y$  gibt [3, S. 71].

**Definition 7.** Als globale Markov Unabhängigkeiten  $I(\mathcal{G})$  eines DAG  $\mathcal{G}$  werden Unabhängigkeitsaussagen bezeichnet, die sich über die  $d$ -separation von  $\mathcal{G}$  ableiten lassen:

$$I(\mathcal{G}) = \{(X \perp Y|Z) | d\text{-sep}_{\mathcal{G}}(X; Y|Z)\} \quad (2.4)$$

[3, S. 72].

Es gilt, dass  $I_l(\mathcal{G}) \subset I(\mathcal{G})$  [5], d.h.  $d$ -separation extrahiert auch die lokalen Markov Unabhängigkeiten. Es lässt sich sogar zeigen, dass Unabhängigkeitsaussagen eines DAG durch  $d$ -separation erschöpfend beschrieben werden: Das Kriterium der  $d$ -separation ist also nicht mehr zu verbessern, so dass es zusätzliche Unabhängigkeiten aufzeigt [1].

**Beispiel 4** (Globale Markov Unabhängigkeiten). Aus den aktiven Pfaden aus Beispiel 3 folgt

$$d\text{-sep}_{\mathcal{G}}(W; R|\emptyset), d\text{-sep}_{\mathcal{G}}(U; W|N), d\text{-sep}_{\mathcal{G}}(U; W|N, R)$$

und daraus die Unabhängigkeitsaussagen

$$(W \perp R), (U \perp W|N), (U \perp W|N, R).$$

Die ersten beiden Unabhängigkeitsaussagen  $(W \perp R)$ ,  $(U \perp W|N)$  sind schon in  $I_t(\mathcal{G})$  enthalten und stellen somit keinen zusätzlichen Erkenntnisgewinn dar. Letzte  $(U \perp W|N, R)$  ist aus der Markov-Bedingung nicht direkt ersichtlich. Hier zeigt sich die Bedeutung der d-separation.

In ganz  $\mathcal{G}_{\text{Regen}}$  ergeben sich somit die folgenden Unabhängigkeitsaussagen, die über die lokalen Markov Unabhängigkeiten hinausgehen:  $(U \perp W|N, R)$  und  $(U \perp R|N, W)$

Die globalen Markov Unabhängigkeiten sind also

$$I(\mathcal{G}_{\text{Regen}}) = \{(W \perp R), (U \perp R|N), (U \perp W|N), (U \perp W|N, R), (U \perp R|N, W)\}.$$

◆

## 2.4 Graphoide

Es wurde nun gezeigt, wie sich Unabhängigkeitsaussagen aus einem DAG extrahieren lassen. Andersherum ist es auch möglich ein Abhängigkeitsmodell mithilfe eines DAG darzustellen. Dabei können entweder alle Unabhängigkeitsaussagen des Abhängigkeitsmodells mittels d-separation aus dem DAG hergeleitet werden oder aber, was in vielen Fällen ausreicht [6], nur ein Teil davon. Entsprechend kommen dem Abhängigkeitsmodell dann folgende Bezeichnungen zu:

**Definition 8.** *Seien  $I$  und  $I^*$  zwei Abhängigkeitsmodelle. Man bezeichne  $I^*$  als I-Map von  $I$ , wenn  $I^* \subset I$  gilt. Lässt sich  $I^*$  (beispielsweise über d-separation) aus einem DAG  $\mathcal{G}$  herleiten, d.h.  $I^* = I(\mathcal{G})$ , so wird  $\mathcal{G}$  auch I-Map (Independency-Map) von  $I$  genannt. Analog ist  $I^*$  bzw.  $\mathcal{G}$ , falls  $I^* = I(\mathcal{G})$ , eine Perfect-Map von  $I$ , sofern  $I^* = I$  gilt. [1, 3]*

Damit ein Abhängigkeitsmodell als DAG  $\mathcal{G}$  darstellbar ist, müssen in diesem Abhängigkeitsmodell bestimmte Eigenschaften, so genannte *Graph-Eigenschaften*, erfüllt sein. Diese Eigenschaften stellen eine Möglichkeit dar, Abhängigkeitskonzepte und Abhängigkeitsmodelle formal zu untersuchen. Je nachdem, ob  $\mathcal{G}$  eine Perfect-Map [4] oder eine I-Map [1] des Abhängigkeitsmodells sein soll, sind unterschiedlich viele dieser Graph-Eigenschaften zu erfüllen. Für den letzten Fall (I-Map) sind die folgenden vier von Relevanz:

- (A1) Symmetrie**  $(X \perp Y|Z) \Rightarrow (X \perp Y|Z)$
- (A2) Zerlegung**  $(X \perp Y, W|Z)^5 \Rightarrow (X \perp Y|Z)$
- (A3) Schwache Vereinigung**  $(X \perp Y, W|Z) \Rightarrow (X \perp Y|Z, W)$

<sup>5</sup>Diese Schreibweise bedeutet  $(X \perp Y|Z) \wedge (X \perp W|Z)$ .

**(A4) Kontraktion**  $(X \perp Y|Z) \wedge (X \perp W|Z, Y) \Rightarrow (X \perp Y, W|Z)$

**Definition 9.** *Abhängigkeitsmodelle, in denen diese vier Graph-Eigenschaften gelten, werden Semi-Graphoide genannt [6].*

**Satz 1.** *Zu jedem Semi-Graphoid ist es möglich, einen DAG  $\mathcal{G}$  zu konstruieren, so dass dieser eine I-Map<sup>6</sup> des Semi-Graphoiden ist [6].*

## 2.5 Unabhängigkeiten der gemeinsamen Verteilung

Unabhängigkeitsaussagen, die sich zunächst in allgemeiner Form aus dem DAG ableiten lassen, entsprechen in einem Bayes-Netz den bedingten stochastischen Unabhängigkeiten  $(X_i \perp\!\!\!\perp X_j|X_k)$ ,  $i, j, k \subset \{1, \dots, n\}$ , die in der gemeinsamen Verteilung  $P$  der Zufallsvariablen  $X_1, \dots, X_n$  enthalten sind. All diese bilden somit auch ein Abhängigkeitsmodell, das mit  $I(P)$  bezeichnet wird.

**Satz 2.** *Ein Abhängigkeitsmodell  $I(P)$ , das bezüglich der stochastischen Unabhängigkeit aus einer gemeinsamen Verteilung  $P$  gewonnen wird, erfüllt (A1) - (A4) und ist somit ein Semi-Graphoid [5].*

Zu jeder Verteilung  $P$  lässt sich folglich (in Zusammenhang mit Satz 1) ein DAG  $\mathcal{G}$  aufstellen, so dass dieser eine I-Map von  $P$  ist. In  $\mathcal{G}$  sind also ausschließlich jene Unabhängigkeiten  $I(\mathcal{G})$  repräsentiert, die auch in  $I(P)$  vorhanden sind; dabei können in  $I(P)$  allerdings noch weitere Unabhängigkeitsaussagen enthalten sein. Entsprechend kann  $P$  durch  $\mathcal{G}$  dargestellt werden:

**Satz 3.** *Wenn der DAG  $\mathcal{G}$  eine I-Map von  $I(P)$  ist, so lässt sich  $P$  nach  $\mathcal{G}$  faktorisieren [3, S. 62].*

Die Umkehrung dieses Zusammenhangs gilt allerdings auch und ist bei Bayes-Netzen relevant, da diese die Faktorisierung der gemeinsamen Verteilung  $P$  durch  $\mathcal{G}$  vorgeben:

**Satz 4** (Soundness der d-separation). *Wenn eine Verteilung  $P(X_1, \dots, X_n)$  über einen DAG  $\mathcal{G}$  faktorisiert, dann gilt  $I(\mathcal{G}) \subset I(P)$  [3, S. 72].*

Anders formuliert ist in diesem Fall  $\mathcal{G}$  eine I-Map von  $I(P)$ .

Nun ist es interessant zu wissen, ob  $\mathcal{G}$  auch alle Unabhängigkeitsaussagen von  $I(P)$  repräsentiert, sprich ob  $\mathcal{G}$  eine Perfect-Map von  $I(P)$  ist. Dies ist allerdings nicht immer der Fall. Denn es ist möglich, dass die bedingten Verteilungen des Bayes-Netzes gerade so beschaffen sind, dass in der sich daraus ergebenden gemeinsamen Verteilung  $P$  die Wahrscheinlichkeitswerte „zufällig“ weitere stochastische Unabhängigkeiten ausdrücken [3, S. 73]. Es gibt folglich zu einem DAG  $\mathcal{G}$  auch gemeinsame Verteilungen,

<sup>6</sup>Dieser ist sogar eine minimale I-Map, in dem Sinne, dass es nicht mehr möglich ist eine Kante des DAG wegzulassen, so dass  $\mathcal{G}$  immer noch eine I-Map des Semi-Graphoiden ist.

die mehr Unabhängigkeiten beinhalten als durch  $\mathcal{G}$  mittels d-separation ausgedrückt werden. Allerdings zeigt sich, dass diese gemeinsamen Verteilungen mit zusätzlichen Unabhängigkeiten nur die „Ausnahme“ unter allen möglichen Verteilungen sind und bei stetiger Parametrisierung des Raums aller Verteilungen mit Wahrscheinlichkeit 0 auftreten.

**Satz 5** (Completeness der d-separation). *Für fast alle Verteilungen  $P(X_1, \dots, X_n)$ , die über einen DAG  $\mathcal{G}$  faktorisieren, gilt  $I(\mathcal{G}) = I(P)$  [3, S. 73].*

Dieser Satz ist essenziell im Kontext der Bayes-Netze, da er besagt, dass der DAG  $\mathcal{G}$  in fast allen Fällen eine Perfect-Map von  $I(P)$  darstellt. Sollten dennoch in  $I(P)$  mehr Unabhängigkeiten vorhanden sein als durch  $\mathcal{G}$  repräsentiert wird, so können diese schon durch die kleinstmögliche Veränderung von  $P$  oder ihrer sie ausmachenden bedingten Verteilungen eliminiert werden [3, S. 74].

## 2.6 Ausblick

Vor diesem Hintergrund erklärt sich, dass das Bayes-Netz eine gemeinsame Verteilung in kompakterer Form darstellen kann. Auch für inferenzielle Zwecke kann damit auf die gemeinsame Verteilung, die in der Praxis für computationale Berechnungen häufig zu umfangreich ist, verzichtet werden. Stattdessen werden die Verteilungen der interessierenden Zufallsvariable unter Einbezug beobachteter Daten direkt errechnet. Für eine Beschreibung dieser als *Updating* bezeichneten Inferenz in Bayes-Netzen sei auf andere Ausarbeitungen [3, Teil II] verwiesen.

Ein anderer großer Themenbereich der Bayes-Netze bezieht sich auf das Erstellen eines Bayes-Netzes aus einem Datensatz heraus, was *Lernen* eines Bayes-Netzes genannt wird. Auch hier dienen andere Quellen, wie z. B. [3, Teil III], als Überblick.

Da sich zu jeder gegebenen Verteilung ein DAG (Sätze 1 und 2) und somit auch ein Bayes-Netz (Satz 3) herleiten lässt, bedarf es also nur noch Algorithmen, um diesen DAG als I-Map, idealerweise als Perfect-Map oder wenigstens als I-Map mit minimal möglicher Anzahl an Kanten, von  $P$  aufzustellen. Dazu sei auf [3, S. 78] verwiesen.

## 3 Diskrete Kredal-Netze

### 3.1 Kredal-Mengen

#### 3.1.1 Definition, Darstellung und Extrempunkte

Wie in der Einleitung beschrieben, kann es Situationen geben, in denen es nur unter enormem Aufwand möglich ist, präzise Wahrscheinlichkeitsangaben zu erlangen. Auch ist dies nicht immer möglich, z. B. wenn sich mehrere Experten nur über bestimmte Einschränkungen in einer Verteilung einig sind, nicht aber über ihre konkreten Werte. In

diesem Fall lohnt es sich verallgemeinerte Wahrscheinlichkeiten zu verwenden, die ein gewisses Maß an Ambiguität in den Wahrscheinlichkeitsangaben zulassen [7]. Dies kann beispielsweise mit intervallwertigen Wahrscheinlichkeiten oder mit den in dieser Arbeit vorgestellten Kredal-Mengen erreicht werden.

**Definition 10.** *Eine Kredal-Menge über eine Zufallsvariable  $X$  ist eine geschlossene, konvexe Menge  $\mathcal{M}(X)$  an Verteilungen über diese Zufallsvariable. Dabei kann  $X = (X_1, \dots, X_n)$  auch ein Vektor von Zufallsvariablen sein [8].*

Die Konvexität der Kredal-Menge besagt, dass jede Linearkombination von schon in  $\mathcal{M}(X)$  enthaltenen Verteilungen ebenfalls in  $\mathcal{M}(X)$  enthalten ist, und die Geschlossenheit gibt an, dass  $\mathcal{M}(X)$  auch ihren Rand beinhaltet.

Berechnungen werden dann mit Kredal-Mengen anstelle von Verteilungen durchgeführt.

Die Kredal-Menge ist insofern eine Verallgemeinerung einer Verteilung, als im Spezialfall einer einelementigen, so genannten *präzisen Kredal-Menge*  $\mathcal{M}(X) = \{P(X)\}$  diese einer Verteilung  $P(X)$  entspricht. Als Gegenstück wird die größtmögliche, aus allen denkbaren Verteilungen bestehende Kredal-Menge

$$\mathcal{M}_0(X) = \left\{ P(X) \mid P(x) \geq 0, \forall x \in X, \sum_{x \in X} P(x) = 1 \right\}$$

als *nichtssagende Kredal-Menge* („vacuous credal set“) bezeichnet [8].

Eine Kredal-Menge  $\mathcal{M}(X) = \{P(X)\}$  kann geometrisch repräsentiert werden als Teilmenge eines  $(m - 1)$ -Simplex

$$\Delta_{m-1} = \left\{ x \in [0, 1]^m \mid \sum_{i=1}^m x_i = 1 \right\},$$

wobei  $m$  die Anzahl an Ausprägungen von  $X$ , d.h.  $|X|$ , ist. Dabei stellt der  $(m - 1)$ -Simplex selbst die nichtssagende Kredal-Menge  $\mathcal{M}_0(X)$  dar. Beispiele für 1- und 2-Simplexe mit Kredal-Mengen sind in Abbildung 2 dargestellt.

Die Beschreibung einer Kredal-Menge erfolgt häufig durch ihre „Eckpunkte“, die so genannten Extrempunkte:

**Definition 11.** *Als Extrempunkte oder Extrempunkte einer Kredal-Menge  $\mathcal{M}(X)$  werden alle ihre Elemente bezeichnet, die sich nicht als Linearkombination von anderen Elementen dieser Kredal-Menge darstellen lassen.  $\text{ext}[\mathcal{M}(X)]$  bezeichnet entsprechend die Menge der Extrempunkte von  $\mathcal{M}(X)$ .*

Da eine Kredal-Menge als geschlossene, konvexe Menge definiert ist, ist sie durch die Angabe ihrer Extrempunkte  $\text{ext}[\mathcal{M}(X)]$  eindeutig festgelegt [8]. Auch die Inferenz mit Kredal-Mengen kann vorwiegend, wenn sie auf unteren bzw. oberen Erwartungswerten

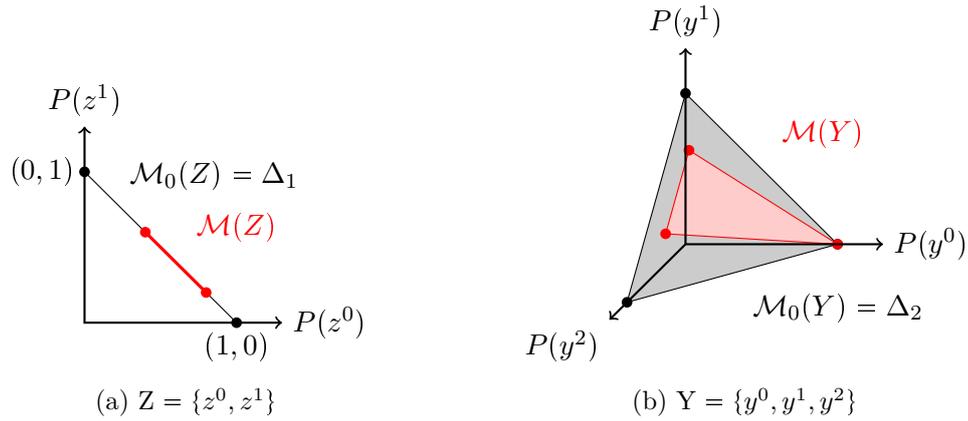


Abbildung 2: Nichtssagende Kredal-Mengen  $\mathcal{M}_0$  (in schwarz) und Beispiel-Kredal-Mengen  $\mathcal{M}$  (in rot) über die Zufallsvariablen  $Z$  mit zwei und  $Y$  mit drei Ausprägungen.

der Kredal-Mengen einer Funktion  $f(X)$

$$\underline{\mathbb{E}}_{\mathcal{M}(X)}[f(X)] = \min_{P \in \mathcal{M}(X)} E_P[f(X)], \quad \bar{\mathbb{E}}_{\mathcal{M}(X)}[f(X)] = \max_{P \in \mathcal{M}(X)} E_P[f(X)] \quad (3.1)$$

beruht, wegen der Linearität des Erwartungswerts, auf ihre Extremalpunkte beschränkt werden: Der Erwartungswert nimmt sein Minimum bzw. Maximum auf Extremalpunkten der Kredal-Menge an [2].

**Beispiel 5** (Kredal-Mengen). Ein Beispiel einer Kredal-Menge über eine zweidimensionale Zufallsvariable  $Z$  mit  $Z = \{z^0, z^1\}$  ist die über die Extremalpunkte  $P_1(Z) = (0.4, 0.6)$  und  $P_2(Z) = (0.8, 0.2)$  definierte Kredal-Menge

$$\mathcal{M}(Z) = \{P(Z) | 0.4 \leq P(z^0) \leq 0.8, 0.2 \leq P(z^1) \leq 0.6, P(z^0) + P(z^1) = 1\},$$

die in Abbildung 2a dargestellt ist.

Für eine dreidimensionale Zufallsvariable  $Y$  mit  $Y = \{y^0, y^1, y^2\}$  könnte man eine Kredal-Menge  $\mathcal{M}(Y)$  analog über

$$\text{ext}[\mathcal{M}(Y)] = \{(1, 0, 0), (0.1, 0.3, 0.6), (0.1, 0.7, 0.2)\}$$

festlegen (siehe Abbildung 2b). Hier ist anzumerken, dass  $\text{ext}[\mathcal{M}(Y)]$  auch aus mehr oder weniger als drei Elementen bestehen könnte und  $|\text{ext}[\mathcal{M}(Y)]|$  nicht durch die Dimension von  $Y$  vorgegeben ist.

◆

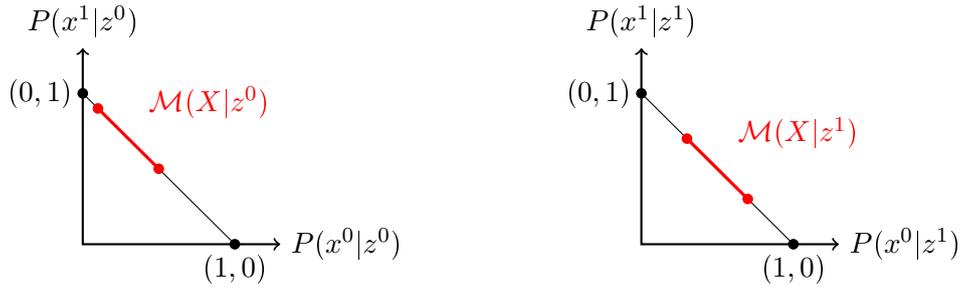


Abbildung 3: Bedingte Kredal-Menge  $\mathcal{M}(X|Z) = \{\mathcal{M}(X|z^0), \mathcal{M}(X|z^1)\}$ .

### 3.1.2 Bedingte und Marginale Kredal-Mengen

**Definition 12.** Eine bedingte Kredal-Menge über die Zufallsvariable  $X$  bedingt auf die Zufallsvariable  $Z$  ist

$$\mathcal{M}(X|Z) = \{\mathcal{M}(X|z) | z \in Z\}, \quad (3.2)$$

eine Menge an auf die einzelnen Ausprägungen von  $Z$  bedingten Kredal-Mengen  $\mathcal{M}(X|z)$ . [8]

Diese Definition ist analog der bedingten Verteilung aus Kapitel 2.1, die eine Verteilung für jede Ausprägung der bedingenden Zufallsvariable beinhaltet: Hier werden entsprechend diese Verteilungen durch Kredal-Mengen ersetzt.

**Beispiel 6** (Bedingte Kredal-Menge). Zu der in Beispiel 5 gegebene binären Zufallsvariable  $Z$  sei eine weitere binäre Zufallsvariable  $X$  gegeben. Die bedingte Kredal-Menge  $\mathcal{M}(X|Z)$  (graphisch dargestellt in Abbildung 3) setzt sich aus den auf  $z^0$  und  $z^1$  bedingten Kredal-Mengen zusammen:

$$\mathcal{M}(X|Z) = \{\mathcal{M}(X|z^0), \mathcal{M}(X|z^1)\}.$$

Diese sollen folgendermaßen beschaffen sein:

$$\begin{aligned} \text{ext}[\mathcal{M}(X|z^0)] &= \{(0.1, 0.9), (0.5, 0.5)\} \\ \text{ext}[\mathcal{M}(X|z^1)] &= \{(0.3, 0.7), (0.7, 0.3)\} \end{aligned}$$

◆

Um von einer gemeinsamen Kredal-Menge  $\mathcal{M}(X, Z)$  zu einer bedingten Kredal-Menge  $\mathcal{M}(X|Z)$  zu gelangen, muss diese auf jede möglich Ausprägung  $z$  von  $Z$  bedingt werden [8]:

$$\mathcal{M}(X|z) = CH \left\{ P(X|z) \left| \begin{array}{l} P(x|z) := \frac{P(x,z)}{\sum_{x \in X} P(x,z)}, \forall x \in X \\ \forall P(X, Z) \in \text{ext}[\mathcal{M}(X, Z)] \end{array} \right. \right\}. \quad (3.3)$$

Hier bezeichnet  $CH$  die Konvexe Hülle der resultierenden bedingten Verteilungen  $P(X|z)$ , sprich die kleinste konvexe Menge, die all diese  $P(X|z)$  beinhaltet. Dies findet hier

Anwendung, da Kredal-Mengen zum einen als konvexe Mengen definiert wurden, und zum anderen, da wie oben in Zusammenhang mit Formel (3.1) erklärt, in vielen Fällen eine Beschränkung auf die Extrempunkte einer Kredal-Menge ohne negative Konsequenzen für die Inferenz ist, weshalb hier für das Bedingen nur Extrempunkte  $P(X, Z) \in \text{ext}[\mathcal{M}(X, Z)]$  in Betracht gezogen werden.

Analog dazu kann eine *marginale Kredal-Menge*  $\mathcal{M}(Z)$  aus einer gemeinsamen Kredal-Menge  $\mathcal{M}(X, Z)$  durch folgende Marginalisierung auf  $Z$  erlangt werden:

$$\mathcal{M}(Z) = CH \left\{ P(Z) \left| \begin{array}{l} P(z) := \sum_{x \in X} P(x, z), \forall z \in Z \\ \forall P(X, Z) \in \text{ext}[\mathcal{M}(X, Z)] \end{array} \right. \right\}. \quad (3.4)$$

**Beispiel 7** (Bedingen und Marginalisieren). Die gemeinsame Verteilung von  $X$  und  $Z$  aus Beispielen 5 und 6 besitzt vier Wahrscheinlichkeitsangaben

$$P(X, Z) = (P(x^0, z^0), P(x^1, z^0), P(x^0, z^1), P(x^1, z^1)).$$

Eine gemeinsame Kredal-Menge  $\mathcal{M}(X, Z) = CH\{P_1, \dots, P_8\}$  ist beispielsweise als konvexe Hülle über die folgenden acht Extrempunkte gegeben<sup>7</sup>:

$$\begin{aligned} P_1 &= (0.04, 0.36, 0.18, 0.42), & P_2 &= (0.04, 0.36, 0.42, 0.18), \\ P_3 &= (0.20, 0.20, 0.18, 0.42), & P_4 &= (0.20, 0.20, 0.42, 0.18), \\ P_5 &= (0.08, 0.72, 0.06, 0.14), & P_6 &= (0.08, 0.72, 0.14, 0.06), \\ P_7 &= (0.40, 0.40, 0.06, 0.14), & P_8 &= (0.40, 0.40, 0.14, 0.06). \end{aligned}$$

Graphisch gesehen entspricht die Kredal-Menge  $\mathcal{M}(X, Z)$  einer Teilmenge des 3-Simplex (Tetraeder) im 4-dimensionalen Raum.

Wird hier  $X$  auf  $Z$  bedingt, so muss entsprechend der Formel (3.3) für jede der oben angegebenen Extrempunkte zunächst die bedingte Verteilung errechnet werden. Aus der ersten Verteilung  $P_1 = (0.04, 0.36, 0.18, 0.42)$  ergibt sich

$$\begin{aligned} P_1(x^0|z^0) &= \frac{P_1(x^0, z^0)}{P_1(x^0, z^0) + P_1(x^1, z^0)} = \frac{0.04}{0.04 + 0.36} = 0.1 \\ P_1(x^1|z^0) &= \frac{P_1(x^1, z^0)}{P_1(x^0, z^0) + P_1(x^1, z^0)} = \frac{0.36}{0.04 + 0.36} = 0.9 \\ P_1(x^0|z^1) &= \frac{P_1(x^0, z^1)}{P_1(x^0, z^1) + P_1(x^1, z^1)} = \frac{0.18}{0.18 + 0.42} = 0.3 \\ P_1(x^1|z^1) &= \frac{P_1(x^1, z^1)}{P_1(x^0, z^1) + P_1(x^1, z^1)} = \frac{0.42}{0.18 + 0.42} = 0.7 \end{aligned}$$

<sup>7</sup>Diese gemeinsame Kredal-Menge wurde über die weiter unten definierte Komposition von Kredal-Mengen aus  $\mathcal{M}(X|Z)$  (Beispiel 6) und  $\mathcal{M}(Z)$  (Beispiel 5) errechnet. Hier dient sie als Ausgangspunkt, um die ursprünglichen Kredal-Mengen  $\mathcal{M}(X|Z)$  und  $\mathcal{M}(Z)$  wieder zu bestimmen.

und damit die bedingten Verteilungen  $P_1(X|z^0) = (0.1, 0.9)$  und  $P_1(X|z^1) = (0.3, 0.7)$ .

Analog wird für die anderen gemeinsamen Extremalverteilungen vorgegangen.

Es ergeben sich für  $z^0$  die folgenden bedingten Verteilungen  $P_i(X|z^0)$ ,  $i = 1, \dots, 8$ :

$$(0.1, 0.9), (0.1, 0.9), (0.5, 0.5), (0.5, 0.5), (0.1, 0.9), (0.1, 0.9), (0.5, 0.5), (0.5, 0.5).$$

Die bedingte Kredal-Menge  $\mathcal{M}(X|z^0)$  errechnet sich nun als konvexe Hülle über all diese acht auf  $z^0$  bedingten Verteilungen. Nach Weglassen doppelt vorkommender bedingter Verteilungen ergibt sich (nach Formel (3.3))

$$\mathcal{M}(X|z^0) = CH\{(0.1, 0.9), (0.5, 0.5)\}.$$

Die für  $z^1$  resultierenden bedingten Verteilungen  $P_i(X|z^1)$ ,  $i = 1, \dots, 8$  sind

$$(0.3, 0.7), (0.7, 0.3), (0.3, 0.7), (0.7, 0.3), (0.3, 0.7), (0.7, 0.3), (0.3, 0.7), (0.7, 0.3).$$

Auch hier können doppelt vorkommende bedingte Verteilungen weggelassen werden, so dass

$$\mathcal{M}(X|z^1) = CH\{(0.3, 0.7), (0.7, 0.3)\}.$$

Um aus  $\mathcal{M}(X, Z)$  die marginale Kredal-Menge  $\mathcal{M}(Z)$  zu bestimmen, muss nach Formel (3.4) aus jeder der gemeinsamen Extremalverteilungen  $P_i(X, Z)$ ,  $i = 1, \dots, 8$  die marginale Verteilung von  $Z$  errechnet werden. Mit

$$P_1(z^0) = P_1(x^0, z^0) + P_1(x^1, z^0) = 0.04 + 0.36 = 0.4$$

$$P_1(z^1) = P_1(x^0, z^1) + P_1(x^1, z^1) = 0.18 + 0.42 = 0.6$$

gilt  $P_1(Z) = (0.4, 0.6)$ .

Alle marginalen Verteilungen  $P_i(Z)$ ,  $i = 1, \dots, 8$  sind nach analogen Rechnungen

$$(0.4, 0.6), (0.4, 0.6), (0.4, 0.6), (0.4, 0.6), (0.8, 0.2), (0.8, 0.2), (0.8, 0.2), (0.8, 0.2)$$

und es folgt für die marginale Kredal-Menge (vergleiche Beispiel 5)

$$\mathcal{M}(Z) = CH\{(0.4, 0.6), (0.8, 0.2)\}.$$

◆

### 3.1.3 Komposition von Kredal-Mengen

Eine Möglichkeit, die bedingten (und marginalen) Kredal-Mengen zu einer gemeinsamen Kredal-Menge zu kombinieren, ist mithilfe folgender *Komposition*  $\otimes$ :

$$\begin{aligned} \mathcal{M}(X, Y) &= \mathcal{M}(X|Y) \otimes \mathcal{M}(Y) \\ &:= CH \left\{ P(X, Y) \left| \begin{array}{l} P(x, y) := P(x|y) \cdot P(y) \forall x \in X, \forall y \in Y \\ \forall P(Y) \in \text{ext}[\mathcal{M}(Y)], \forall P(X|y) \in \text{ext}[\mathcal{M}(X|y)] \end{array} \right. \right\}. \end{aligned} \quad (3.5)$$

Diese Operation bildet also eine gemeinsame Verteilung aus jeder möglichen Kombination der Extremalverteilungen von  $\mathcal{M}(X|Y)$  und  $\mathcal{M}(Y)$  und erstellt darüber eine Konvexe Hülle.

**Beispiel 8** (Komposition von Kredal-Mengen). Die in Beispiel 7 gegebene gemeinsame Kredal-Menge  $\mathcal{M}(X, Z)$  lässt sich aus  $\mathcal{M}(X|Z)$  und  $\mathcal{M}(Z)$  durch die hier eingeführte Komposition berechnen. Es seien wieder

$$\begin{aligned} \mathcal{M}(Z) &= CH\{(0.4, 0.6), (0.8, 0.2)\} \\ \mathcal{M}(X|z^0) &= CH\{(0.1, 0.9), (0.5, 0.5)\} \\ \mathcal{M}(X|z^1) &= CH\{(0.3, 0.7), (0.7, 0.3)\} \end{aligned}$$

gegeben.

Nach Formel (3.5) wird jeweils eine Extremalverteilung aus den gegebenen Kredal-Mengen betrachtet und daraus die gemeinsame Verteilung errechnet. Dies wird für jede der  $2^3 = 8$  Kombinationen der Extremalpunkte der 3 Kredal-Mengen durchgeführt. Eine Kombination ist beispielsweise

$$\begin{aligned} P(Z) &= (0.4, 0.6) \\ P(X|z^0) &= (0.1, 0.9) \\ P(X|z^1) &= (0.3, 0.7). \end{aligned}$$

Daraus errechnen sich die folgenden Wahrscheinlichkeitsangaben, die die gemeinsame Verteilung  $P_1(X, Z)$  beschreiben:

$$\begin{aligned} P_1(z^0, x^0) &= P(z^0) \cdot P(x^0|z^0) = 0.4 \cdot 0.1 = 0.04 \\ P_1(z^0, x^1) &= P(z^0) \cdot P(x^1|z^0) = 0.4 \cdot 0.9 = 0.36 \\ P_1(z^1, x^0) &= P(z^1) \cdot P(x^0|z^1) = 0.6 \cdot 0.3 = 0.18 \\ P_1(z^1, x^1) &= P(z^1) \cdot P(x^1|z^1) = 0.6 \cdot 0.7 = 0.42 \end{aligned}$$

Es gilt also  $P_1(X, Z) = (0.04, 0.36, 0.18, 0.42)$ .

Analog wird für die anderen Kombinationen der Extremalverteilungen der bedingten und marginalen Kredal-Menge vorgegangen und es resultieren die acht gemeinsamen

Verteilungen, die in Beispiel 7 vorgegeben wurden:

$$\begin{aligned}
P_1 &= (0.04, 0.36, 0.18, 0.42), & P_2 &= (0.04, 0.36, 0.42, 0.18), \\
P_3 &= (0.20, 0.20, 0.18, 0.42), & P_4 &= (0.20, 0.20, 0.42, 0.18), \\
P_5 &= (0.08, 0.72, 0.06, 0.14), & P_6 &= (0.08, 0.72, 0.14, 0.06), \\
P_7 &= (0.40, 0.40, 0.06, 0.14), & P_8 &= (0.40, 0.40, 0.14, 0.06).
\end{aligned}$$

Die gemeinsame Kredal-Menge  $\mathcal{M}(X, Z)$  ist nun nach der Komposition von Kredal-Mengen (Formel (3.5)) die konvexe Hülle über diese acht Verteilungen:

$$\mathcal{M}(X, Z) = CH\{P_1, \dots, P_8\}.$$

◆

## 3.2 Kredal-Netze

### 3.2.1 Definition

Um von einem Bayes-Netz zu einem Kredal-Netz zu kommen sind im ersten Schritt die bedingten (und marginalen) Verteilungen des Bayes-Netzes durch bedingte (und marginale) Kredal-Mengen zu ersetzen.

**Definition 13.** Ein Kredal-Netz über die Zufallsvariablen  $X_1, \dots, X_n$  ist das geordnete Paar  $(\mathcal{G}, \mathbb{M})$ , bestehend aus einem auf  $X_1, \dots, X_n$  definierten, gerichteten azyklischen Graphen  $\mathcal{G}$  und einer Menge  $\mathbb{M} = \{\mathcal{M}(X_i | Pa_{X_i}^{\mathcal{G}}) | i = 1, \dots, n\}$  an bedingten Kredal-Mengen der einzelnen Zufallsvariablen, wobei die jeweils bedingenden Zufallsvariablen als Eltern  $Pa_{X_i}^{\mathcal{G}}$  durch  $\mathcal{G}$  vorgegeben werden [8].

**Beispiel 9** (Kredal-Netz). Ein Kredal-Netz  $(\mathcal{G}_{\text{Regen}}, \mathbb{M})$  kann so wie das Bayes-Netz aus Beispiel 1 auf  $\mathcal{G}_{\text{Regen}}$  basieren. Zu den Zufallsvariablen  $W, R, N, U$  werden nun (marginale und bedingte) Kredal-Mengen angegeben, die in Abbildung 4 (ohne konkrete Werte) dargestellt sind.

◆

In einem Bayes-Netz wird zusätzlich angenommen, dass sich die gemeinsame Verteilung  $P$  durch  $\mathcal{G}$  faktorisieren lässt. Dies ist aufgrund der Verwendung der stochastischen Unabhängigkeit und der Tatsache, dass sich die gemeinsame Verteilung als Produkt aus den bedingten Verteilungen errechnen lässt, möglich. Da die stochastische Unabhängigkeit jedoch von Verteilungen und nicht von Kredal-Mengen ausgeht, bedarf es für Kredal-Mengen einer generalisierten Definition der Unabhängigkeit. Allerdings gibt es mehrere Möglichkeiten für diese Generalisierung. Des Weiteren muss zusätzlich festgelegt werden, auf welche Art und Weise die bedingten Kredal-Mengen zu einer gemeinsamen

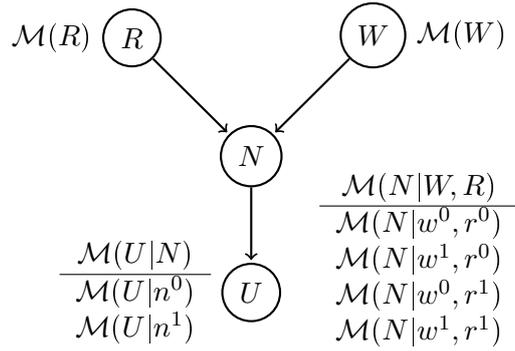


Abbildung 4: Kredal-Netz über 4 binäre Variablen.

Kredal-Menge  $\mathcal{M}(X_1, \dots, X_n)$ , im Kontext von Kredal-Netzen auch *Erweiterung* genannt, kombiniert werden. Dies sind zwei Kernfragen in der Spezifikation eines Kredal-Netzes [9].

Im Folgenden sollen nun zwei verschiedene Arten der Unabhängigkeit von Zufallsvariablen im Kontext der Kredal-Mengen beschrieben, nämlich die starke und die epistemische Unabhängigkeit, und im Anschluss daran für beide Typen entsprechende Erweiterungen thematisiert werden.

### 3.2.2 Starke Unabhängigkeit

Als mögliche Generalisierung der stochastischen Unabhängigkeiten kommt die starke Unabhängigkeit<sup>8</sup>, wieder in der bedingten Version definiert, in Frage:

**Definition 14.** *Bei gegebener gemeinsamer Kredal-Menge  $\mathcal{M}(X, Y, Z)$  sind zwei Zufallsvariablen  $X$  und  $Y$  bedingt stark unabhängig gegeben die beobachtete Zufallsvariable  $Z$ , formal  $(X \text{ SID } Y|Z)$ , wenn in jeder Extremalverteilung  $P(X, Y, Z) \in \text{ext}[\mathcal{M}(X, Y, Z)]$  die bedingte stochastische Unabhängigkeit von  $X$  und  $Y$  gegeben  $Z$  gilt [8].*

Die starke Unabhängigkeit stellt eine Generalisierung der stochastischen Unabhängigkeit insofern dar, als für eine präzise Kredalmenge  $\mathcal{M}(X, Y, Z) = \{P(X, Y, Z)\}$  diese einzige Verteilung auch eine Extremalverteilung darstellt und darin die stochastische Unabhängigkeit gilt.

**Beispiel 10** (Starke Unabhängigkeit). Es sei eine gemeinsame Kredal-Menge  $\mathcal{M}(X, Y)$  der Zufallsvariablen  $X$  und  $Y$  gegeben als konvexe Hülle über zwei Extremalverteilungen  $P_1(X, Y) = (0.3, 0.3, 0.2, 0.2)$  und  $P_2(X, Y) = (0.24, 0.56, 0.06, 0.14)$ .

Werden  $P_1$  und  $P_2$  tabellarisch betrachtet, so ist darin die stochastische Unabhängigkeit von  $X$  und  $Y$  leicht ersichtlich (vergleiche Formel (2.2)):

<sup>8</sup>Mit stochastischer und starker Unabhängigkeit sind hier Unabhängigkeitskonzepte gemeint. Je nach Situation findet die entsprechende unbedingte oder bedingte Unabhängigkeit Verwendung.

$P_1$	$x^0$	$x^1$		$P_2$	$x^0$	$x^1$	
$y^0$	0.3	0.3	0.6	$y^0$	0.24	0.56	0.8
$y^1$	0.2	0.2	0.4	$y^1$	0.06	0.14	0.2
	0.5	0.5	1		0.3	0.7	1

Da nun in allen Extremalverteilungen  $P_1, P_2$  von  $\mathcal{M}(X, Y)$  die stochastische Unabhängigkeit von  $X$  und  $Y$  gilt, sind  $X$  und  $Y$  stark unabhängig. ♦

Anzumerken ist, dass in der Definition der starken Unabhängigkeit nur in den Extremalverteilungen die stochastische Unabhängigkeit gültig ist. Dies erklärt sich daher, dass es in diesen Kredal-Mengen, in denen  $(X \text{ SID } Y|Z)$  gilt, auch Nicht-Extremal-Verteilungen gibt, in denen  $(X \perp\!\!\!\perp Y|Z)$  nicht gilt [9]. Eine Kredal-Menge, in der die stochastische Unabhängigkeit für alle Verteilungen gilt, wäre demnach nicht mehr konvex. Da die Inferenz einer Kredal-Menge, wie oben beschrieben (siehe Kapitel 3.1), vornehmlich von den Extremalpunkten abhängt, kommen Kredal-Mengen mit identischen Extremalpunkten zu identischen Schlüssen. Interpretationsgemäß sollen aber Kredal-Mengen mit identischen Schlüssen auch den selben Glauben haben, sprich die gleichen Verteilungen beinhalten, was durch die Konvexität erlangt wird [9]. Vor diesem Hintergrund wird die Konvexität von Kredal-Mengen bedeutender angesehen als die Gültigkeit der stochastischen Unabhängigkeit in allen Verteilungen der Kredal-Menge von stark unabhängigen Variablen.

Wie bei der stochastischen Unabhängigkeit kann auch die starke Unabhängigkeit bezüglich der Graph-Eigenschaften untersucht werden, um ihre „Verwendbarkeit“ in Kredal-Netzen zu evaluieren. Eine Analyse bringt folgendes Ergebnis:

**Satz 6.** *Abhängigkeitsmodelle, die bezüglich der starken Unabhängigkeit aus einer gemeinsamen Kredal-Menge mit positiven unteren Wahrscheinlichkeiten [10] gewonnen werden, erfüllen (A1) - (A4) und sind somit Semi-Graphoide [11, 12].*

Alle Unabhängigkeitsaussagen, die bezüglich der starken Unabhängigkeit in einer gemeinsamen Kredal-Menge  $\mathcal{M}(X, Y, Z)$  enthalten sind, bilden ein Abhängigkeitsmodell  $I(\mathcal{M}(X, Y, Z))$ . Dieser Satz sagt in Zusammenhang mit Satz 1 aus, dass es zu jeder  $\mathcal{M}(X, Y, Z)$  einen DAG  $\mathcal{G}$  gibt, der eine I-Map dazu ist.

Nun bedarf es noch einer Methode, um aus einem gegebenen Kredal-Netz eine Erweiterung zu bestimmen.

### 3.2.3 Starke Erweiterung

In Bayes-Netzen enthält die gemeinsame Verteilung  $P$  durch die Faktorisierung nach  $\mathcal{G}$  alle (stochastischen) Unabhängigkeitsaussagen, die durch  $\mathcal{G}$  mittels d-separation vorgegeben werden ( $I(\mathcal{G}) \subset I(P)$ , Satz 4). Dieser Zusammenhang ist auch für die Erweiterung eines

Kredal-Netzes wünschenswert: Die Erweiterung eines Kredal-Netzes soll alle Unabhängigkeitsaussagen enthalten, die durch  $\mathcal{G}$  mittels d-separation vorgegeben werden. Dies ist aber nur dann möglich, sofern es das verwendete Unabhängigkeitskonzept (z. B. stark oder epistemisch) zulässt: Erfüllt es die vier Graph-Eigenschaften (A1) – (A4) nicht, so ist nicht gesichert, dass sich das Abhängigkeitsmodell in der gemeinsamen Kredal-Menge bezüglich dieses Unabhängigkeitskonzepts auch als DAG darstellen lässt. Eine Erweiterung eines Kredal-Netzes würde also nicht zwingend alle Unabhängigkeiten enthalten, die durch  $\mathcal{G}$  vorgegeben werden. Die Definition einer Erweiterung eines Kredal-Netzes bezüglich eines bestimmten Unabhängigkeitskonzepts sollte also nicht voraussetzen, dass in dieser alle Unabhängigkeitsaussagen des DAG  $I(\mathcal{G})$  erfüllt sind. Stattdessen kann - in Konsistenz mit [13] - die Definition auf der Einhaltung der Markov-Bedingung beruhen, sprich, dass in der Erweiterung alle durch  $\mathcal{G}$  vorgegebenen lokalen Markov Unabhängigkeiten  $I_l(\mathcal{G})$  vorhanden sind (bezüglich dieses bestimmten Unabhängigkeitskonzepts). Entsprechend definiert sich die Erweiterung bezüglich der starken Unabhängigkeit folgendermaßen:

**Definition 15.** *Die starke Erweiterung eines Kredal-Netzes ist die größte gemeinsame Kredal-Menge  $\mathcal{M}(X_1, \dots, X_n)$ , die alle vom DAG  $\mathcal{G}$  durch die Markov-Bedingung (siehe Definition 4) vorgegebenen Unabhängigkeiten als starke Unabhängigkeiten erfüllt. Diese wird berechnet über die Komposition von Kredal-Mengen:*

$$\mathcal{M}(X_1, \dots, X_n) = \bigotimes_{i=1}^n \mathcal{M}(X_i | Pa_{X_i}^{\mathcal{G}}) \quad (3.6)$$

Das Konzept der starken Unabhängigkeit lässt zu, dass die starken Unabhängigkeiten in gemeinsamen Kredal-Mengen als DAG dargestellt werden können (siehe Kapitel 3.2.2). Folglich kann auch die starke Erweiterung eines Kredal-Netzes alle Unabhängigkeiten erhalten, die von  $\mathcal{G}$  über d-separation vorgegeben werden (nicht nur die lokalen Markov Unabhängigkeiten). Dies ist durch die Berechnung der starken Erweiterung über die oben eingeführte Komposition auch der Fall:

**Satz 7** (Soundness bei starker Erweiterung). *In einem Kredal-Netz  $(\mathcal{G}, \mathbb{M})$  mit positiven unteren bedingten Wahrscheinlichkeiten [10] ist jede Unabhängigkeitsaussage von  $\mathcal{G}$  in der starken Erweiterung  $\mathcal{M}(X_1, \dots, X_n)$  als starke Unabhängigkeit gültig, sprich  $I(\mathcal{G}) \subset I(\mathcal{M}(X_1, \dots, X_n))$  [2, 13].*

**Beispiel 11** (Starke Erweiterung). Um in dem Kredal-Netz  $(\mathcal{G}_{\text{Regen}}, \mathbb{M})$  aus Beispiel 9 die starke Erweiterung zu berechnen, muss aus jeder Kombination der Extremalverteilungen der einzelnen bedingten und marginalen Verteilungen eine gemeinsame Verteilung errechnet und darüber die konvexe Hülle gebildet werden. Da die Zufallsvariablen binär sind hat eine in Abbildung 4 vorkommende (bedingte oder marginale) Kredal-Menge höchstens zwei Extremalpunkte<sup>9</sup>. Sei nun weiter angenommen, dass alle Kredal-Mengen

<sup>9</sup>Bei mehr Ausprägungen der Zufallsvariablen wäre diese Anzahl nicht beschränkt. Beispielsweise könnte eine Kredal-Menge über eine Zufallsvariable mit 3 Ausprägungen ein Kreis im 2-Simplex sein.

auch zwei Extrempunkte haben. Nachdem  $\mathbb{M}$  aus 8 Kredal-Mengen mit jeweils 2 Extrempunkten besteht, gibt es  $2^8 = 256$  verschiedene Kombinationsmöglichkeiten der Extrempunkte. Aus jeder dieser Kombinationen wird analog zum Bayes-Netz eine gemeinsame Verteilung errechnet, die aus  $2^4 = 16$  Wahrscheinlichkeitsangaben besteht, da 4 binäre Zufallsvariablen betrachtet werden.

◆

Die Berechnung der starken Erweiterung erweist sich in diesem so einfach gehaltenen Beispiel schon als unverhältnismäßig umfangreich. Bei größeren Kredal-Netzen wird diese Berechnung, wie auch die der gemeinsamen Verteilung bei den Bayes-Netzen, computationally undurchführbar [8]. Satz 7 ermöglicht analog den Bayes-Netzen das Umgehen der expliziten Berechnung der starken Erweiterung in Kredal-Netzen bezüglich der starken Unabhängigkeit.

Verwendung findet die starke Erweiterung beispielsweise in Sensitivitäts-Analysen: Es wird davon ausgegangen, dass es ein „wahres“ Bayes-Netz gibt, das in dem Kredal-Netz enthalten ist. So kann die Robustheit des als „wahr“ angenommenen Modells untersucht werden [14]. Eine andere Anwendung ist die Modellierung von Expertenwissen, sofern sich alle Experten einig sind, dass in den Verteilungen der Kredal-Mengen das Konzept der stochastischen Unabhängigkeit zu gelten hat [13].

### 3.2.4 Epistemische Unabhängigkeit

Ein Kredal-Netz bezüglich der starken Unabhängigkeit kann als eine Menge an einzelnen Bayes-Netzen angesehen werden, von denen eines davon nach der oben vorgestellten Interpretation als Sensitivitäts-Analyse als „wahr“ angenommen wird. Dies impliziert aber, dass es ein „wahres“ Bayes-Netz gibt, das die untersuchte Situation darstellt. Im Falle von Expertenwissen bedeutet dies, dass präzise Wahrscheinlichkeiten existieren, die aber aufgrund von limitierten Ressourcen nicht zugänglich sind. Allerdings ist es durchaus plausibel, dass in manchen Situationen Expertenwissen inhärent unpräzise ist, sprich keine präzise „wahre“ Verteilung existiert. Ein Unabhängigkeitskonzept, das wie die starke Unabhängigkeit auf den einzelnen präzisen Verteilungen beruht, ist in diesem Fall folglich nicht angebracht. Es bedarf vielmehr eines Unabhängigkeitskonzepts, das auf der ganzen Kredal-Menge basiert, ohne die einzelnen, darin enthaltenen präzisen Verteilungen gesondert zu betrachten [15]. Ein derartiges Unabhängigkeitskonzept ist die epistemische Unabhängigkeit, die auf der asymmetrischen epistemischen Irrelevanz beruht und auf Walley [16] zurückgeht:

**Definition 16.** *Bei gegebener gemeinsamer Kredal-Menge  $\mathcal{M}(X, Y, Z)$  ist  $X$  bedingt epistemisch irrelevant von  $Y$  gegeben  $Z$ , formal  $(X \text{ EIR } Y|Z)$ , wenn  $\forall z \in Z, \forall y \in Y$  die Kredal-Mengen  $\mathcal{M}(X|y, z)$  und  $\mathcal{M}(X|z)$  die gleiche konvexe Hülle besitzen.*

**Definition 17.** In  $\mathcal{M}(X, Y, Z)$  sind  $X$  und  $Y$  bedingt epistemisch unabhängig gegeben  $Z$ , formal  $(X \text{ EID } Y|Z)$ , wenn  $(X \text{ EIR } Y|Z)$  und  $(Y \text{ EIR } X|Z)$  gilt [9].

Wird die epistemische Unabhängigkeit bezüglich der Graph-Eigenschaften untersucht, so zeigt sich ein anderes Bild als bei der stochastischen und starken Unabhängigkeit:

**Satz 8.** In einem Abhängigkeitsmodell  $I(\mathcal{M})$ , das bezüglich der epistemischen Unabhängigkeit aus einer gemeinsamen Kredal-Menge  $\mathcal{M}$  gewonnen wird, ist die Kontraktions-Eigenschaft (A4) nicht immer erfüllt [9, 13].

### 3.2.5 Epistemische Erweiterung

Aus der fehlenden Gültigkeit der Kontraktions-Eigenschaft (A4) folgt, dass sich epistemische Unabhängigkeiten einer gemeinsamen Kredal-Menge  $\mathcal{M}$  eben nicht immer als DAG  $\mathcal{G}$  darstellen lassen. Dies gilt ebenso, wenn diese gemeinsame Kredal-Menge  $\mathcal{M}$  als (epistemische) Erweiterung aus einem Kredal-Netz gebildet wird, d.h.  $I(\mathcal{M})$  kann auch epistemische Unabhängigkeiten beinhalten, die nicht in  $I(\mathcal{G})$  enthalten sind. Dennoch lässt sich analog der starken Unabhängigkeit und in Anlehnung an [13] die Erweiterung eines Kredal-Netzes bezüglich der epistemischen Unabhängigkeit basierend auf den lokalen Markov-Unabhängigkeiten folgendermaßen definieren:

**Definition 18.** Die epistemische Erweiterung eines Kredal-Netzes ist die größte gemeinsame Kredal-Menge  $\mathcal{M}(X_1, \dots, X_n)$ , die alle vom DAG  $\mathcal{G}$  durch die Markov-Bedingung vorgegebenen Unabhängigkeiten als epistemische Unabhängigkeiten erfüllt. [13]

Aus Satz 8 folgt, dass - wie auch immer die epistemische Erweiterung im Speziellen errechnet wird - eine gemeinsame Kredal-Menge bezüglich der epistemischen Unabhängigkeit grundsätzlich nicht sinnvoll als (kompaktes) Kredal-Netz dargestellt werden kann: Die Soundness der d-separation gilt hier nicht uneingeschränkt.

Ein Möglichkeit damit umzugehen, ist beispielsweise die Kontraktions-Eigenschaft (A4) bezüglich der epistemischen Unabhängigkeit in der zu untersuchenden Kredal-Menge einfach anzunehmen [9], denn dann stellen die epistemischen Unabhängigkeitsaussagen der Kredal-Menge ein Semi-Graphoid dar (Definition 9) und lassen sich als DAG darstellen (Satz 1). Konkret müsste dazu die Gültigkeit von (A4) in der untersuchten Kredal-Menge vorab gesondert nachgewiesen werden.

## 4 Zusammenfassung

Ein Bayes-Netz beinhaltet die (kausale) Einflussstruktur von Zufallsvariablen sowie deren bedingte Verteilungen, die nur von den auf sie Einfluss ausübenden Zufallsvariablen abhängen. Diese Einflussstruktur ist graphentheoretisch durch einen gerichteten azyklischen Graphen (DAG) darstellbar und kann formuliert werden als eine Menge von

Unabhängigkeitsaussagen (Abhängigkeitsmodell), die über das Kriterium der d-separation aus dem Graphen extrahierbar sind.

Die gemeinsame Verteilung der Zufallsvariablen wird in einem Bayes-Netz durch Produktbildung aus den bedingten Verteilungen errechnet. Diese beinhaltet bedingte stochastische Unabhängigkeiten der Zufallsvariablen untereinander, die auch ein Abhängigkeitsmodell darstellen.

In einem Bayes-Netz ist das zuerst erwähnte Abhängigkeitsmodell (des DAG) auf jeden Fall eine Teilmenge des zuletzt erwähnten (der gemeinsamen Verteilung) und die beiden sind sogar fast sicher (vergleiche Abschnitt 2.5) gleich, sprich der DAG bildet die gemeinsame Verteilung fast sicher perfekt ab. Das Kriterium der d-separation wird deswegen auch als sound und complete beschrieben. Entsprechend ist der DAG und somit das Bayes-Netz eine vernünftige und fast sicher komplette Darstellung der gemeinsamen Verteilung, die vor allem viel kompakter ist als tabellarische Wahrscheinlichkeitsangaben.

In manchen Situationen ist es sinnvoll, anstelle von präzisen Verteilungen verallgemeinerte Wahrscheinlichkeiten, sprich Mengen von Verteilungen (Kredal-Mengen), zu verwenden. Entsprechend kann das Bayes-Netz als Kredal-Netz auf verallgemeinerte Wahrscheinlichkeiten generalisiert werden: Zu dem DAG wird hier eine bedingte Kredal-Menge für jede Zufallsvariable angegeben.

Als Generalisierung der stochastischen Unabhängigkeit auf Kredal-Mengen gibt es mehrere Möglichkeiten, wie z. B. die starke und epistemische Unabhängigkeit. Folglich bedarf ein Kredal-Netz zusätzlich der Angabe welches Unabhängigkeitskonzept Anwendung findet.

Um entsprechende Zusammenhänge, wie sie im Bayes-Netz vorkommen, zu untersuchen, besteht die Möglichkeit Abhängigkeitsmodelle, die bezüglich eines bestimmten Unabhängigkeitskonzepts gebildet werden, allgemein zu untersuchen. Können in diesen gewisse Zusammenhänge (Graph-Eigenschaften) nachgewiesen werden, so können sie als DAG repräsentiert werden.

Die starke Unabhängigkeit erfüllt diese Graph-Eigenschaften und zu einem Kredal-Netz kann eine gemeinsame Kredal-Menge (starke Erweiterung) gebildet werden, die alle durch den DAG vorgegebenen Unabhängigkeitsaussagen als starke Unabhängigkeitsaussagen erfüllt. Entsprechend ist das Kredal-Netz eine vernünftige Darstellung dieser gemeinsamen Kredal-Menge bezüglich der starken Unabhängigkeit.

Das Konzept der epistemischen Unabhängigkeit erfüllt jedoch nicht alle der oben erwähnten Graph-Eigenschaften. Dementsprechend können die epistemischen Unabhängigkeiten einer gemeinsamen Kredal-Menge grundsätzlich nicht als DAG dargestellt werden, selbst wenn diese Kredal-Menge aus einem Kredal-Netz gebildet wurde. Die Darstellung der gemeinsamen Kredal-Menge als Kredal-Netz bezüglich der epistemischen Unabhängigkeit ist also nicht vernünftig.

## Literatur

- [1] D. Geiger, T. Verma und J. Pearl. „Identifying independence in Bayesian networks“. *Networks* **20**, 1990, S. 507–534.
- [2] F. G. Cozman. „Credal networks“. *Artificial intelligence* **120**, 2000, S. 199–233.
- [3] D. Koller und N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [4] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1985.
- [5] J. Pearl, D. Geiger und T. Verma. „Conditional independence and its representations“. *Kybernetika* **25**, 1989, S. 33–44.
- [6] T. Verma und J. Pearl. „Causal Networks: Semantics and Expressiveness“. In: *Proceedings of the Fourth Annual Conference on Uncertainty in Artificial Intelligence*. North-Holland Publishing Co., 1990, S. 69–78.
- [7] F. P. Coolen, M. C. Troffaes und T. Augustin. „Imprecise probability“. In: *International encyclopedia of statistical science*. Springer, 2011, S. 645–648.
- [8] A. Antonucci, C. P. De Campos und M. Zaffalon. „Probabilistic graphical models“, 2014.
- [9] F. G. Cozman. „Separation properties of sets of probability measures“. In: *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 2000, S. 107–114.
- [10] G. De Cooman, J. De Bock und A. Van Camp. „Recent advances in imprecise-probabilistic graphical models.“ In: *ECAI. 2012*, S. 27–32.
- [11] S. Moral und A. Cano. „Strong conditional independence for credal sets“. *Annals of Mathematics and Artificial Intelligence* **35**, 2002, S. 295–321.
- [12] F. G. Cozman. „Sets of probability distributions, independence, and convexity“. *Synthese* **186**, 2012, S. 577–600.
- [13] F. G. Cozman. „Graphical models for imprecise probabilities“. *International Journal of Approximate Reasoning* **39**, 2005, S. 167–184.
- [14] F. Cozman. „Robustness analysis of Bayesian networks with local convex sets of distributions“. In: *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 1997, S. 108–115.
- [15] G. De Cooman u. a. „Epistemic irrelevance in credal networks: the case of imprecise Markov trees“. In: *6th International Symposium on Imprecise Probability: Theories and Applications*. 2009, S. 149–158.

- [16] P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman und Hall, 1991.