Grundlagen der Künstlichen Intelligenz

10 Verarbeitung unsicheren Wissens

Wahrscheinlichkeitstheorie, einfache Bayessche Statistik

Volker Steinhage

Inhalt

- Motivation
- Grundlagen der Wahrscheinlichkeitstheorie
- Probabilistische Inferenz

Motivation

- Oft ist unser Wissen über die Welt mit Unsicherheiten behaftet:
- Quellen der Unsicherheiten:

1) nur partiell beobachtbare Weltzustände,
2) unzuverlässige Beobachtungen,
3) unsichere Aktionseffekte,
4) Komplexität der Modellierung.

Problem: Logik-basierte Planungssysteme gehen aus von

- vollständig beobachtbaren Weltzuständen,
- sicheren Aktionsausführungen,
- vollständiger Modellierung.

Beispiel für logische Regeln über unsichere Sachverhalte

Beispiel: einfaches Diagnose-Expertensystem für Zahnärzte:

- Regel 1: $\forall p$ [Symptom(p, Zahnschmerzen) $\Rightarrow$ Krankheit(p, Loch)] einher.

  ⇝ Aber die Regel ist inkorrekt: es gibt *auch andere Ursachen für Zahnschmerzen*.

- Besser wäre Regel 2:

  $\forall p$ [Symptom(p, Zahnschmerzen) $\Rightarrow$
  Krankheit(p, Loch) $\vee$ Krankheit(p, Zahnfleischproblem) $\vee$ …]

  ⇝ Aber wenn wir nicht *alle* Ursachen kennen?

- Vielleicht besser die *kausale* Regel 3?

  $\forall p$ [Krankheit(p, Loch) $\Rightarrow$ Symptom(p, Zahnschmerzen)]

  ⇝ Auch *inkorrekt*: kariöse Löcher gehen *nicht immer* mit Schmerzen einher.

Unsicherheiten bei Regeln

Probleme mit der Logik:

- wir können nicht immer alle möglichen Ursachen aufzählen,
- wir kennen nicht alle Gesetzmäßigkeiten (in der Medizin),
- es besteht die Unsicherheit über den Patienten:
  - Karies und Zahnschmerzen können zufällig gleichzeitig auftreten.
  - wurden alle Untersuchungen durchgeführt – und wenn ja: richtig?
  - hat der Patient alle Fragen beantwortet – und wenn ja: angemessen?

$\leadsto$ Ohne perfektes Wissen keine korrekten logischen Regeln!

Unsicherheit bei Fakten

Aber auch die sog. Fakten können mit Unsicherheiten behaftet sein:

- Zahnschmerzen können tatsächlich auch Kopfschmerzen sein,
- Patientenantworten können falsch erfasst oder übertragen werden,
- Messmethoden können unscharfe Werte liefern,
- Messgeräte können defekt sein,
- …

Grad der Überzeugung und Wahrscheinlichkeitstheorie

Konsequenz:

- Wir (oder andere Agenten) sind von Regeln und Fakten häufig nur *„bis zu einem gewissen Grad überzeugt“*.
- Eine Möglichkeit, den Grad der Überzeugung auszudrücken, besteht in der Verwendung von Wahrscheinlichkeiten.
- „Der Agent ist von der Sensorinformation zu 90% überzeugt“ heißt dann: Der Agent nimmt an, dass die Information in 9 von 10 Fällen richtig ist.
- Wahrscheinlichkeiten drücken eine durch partielles Unwissen bedingte *Unsicherheit** aus.

---

* Unsicherheit ist nicht mit *Unschärfe* (*Vagheit*) zu verwechseln. Das Prädikat *„groß“* ist unscharf. Die Aussage *„das könnte Peters Uhr sein“* ist unsicher.

Rationale Entscheidungen unter Unsicherheit

Wir haben verschiedene ***Aktionen*** oder *Pläne* zur Auswahl.

- Diese können mit verschiedenen *Wahrscheinlichkeiten* zu verschiedenen Ergebnissen führen.
- Die *Aktionen* verursachen verschiedene – ggf. subjektive – Kosten.
- Die *Ergebnisse* haben verschiedene – ggf. subjektive – Nutzen.
- Rational wäre es, die Aktion zu wählen, die den größten zu erwartenden Gesamtnutzen hat!

↝ Entscheidungstheorie = Nutzentheorie + Wahrscheinlichkeitstheorie

Entscheidungstheoretischer Agent

function DT-AGENT(*percept*) returns an *action*

static: a set of probabilistic beliefs about the state of the world

calculate updated probabilities for current state based on

available evidence including current percept and previous action

calculate outcome probabilities for actions,

given action descriptions and probabilities of current states

select *action* with highest expected utility

given probabilities of outcomes and utility information

return *action*

Entscheidungstheorie: Ein Agent ist rational gdw. er die Aktion wählt, die den *größten zu erwartenden Nutzen* gemittelt über alle möglichen Ergebnisse von Aktionen hat.

Wahrscheinlichkeitstheorie: Notation

- Wahrscheinlichkeiten werden ausschließlich den *unsicheren* Teilzuständen der Welt zugeordnet.

- Elementare Teilzustände werden durch sogenannte Zufallsvariablen benannt.
  - Beispiel 1: *Loch* $\in$ {*true*, *false*}
  - Beispiel 2: *Wetter* $\in$ {*sonnig*, *regnerisch*, *wolkig*, *schnee*}.

- Im weiteren Verlauf werden Abkürzungen durch Kleinschreibung benutzt wie:
  - zu Beispiel 1: *loch* für *Loch = true* etc.
  - zu Beispiel 2: *sonnig* für *Wetter = sonnig* etc.

- *Loch = true* ist eine elementare Aussage (Proposition).

- Komplexere Aussagen werden mittels logischer Konnektoren zusammengesetzt.
  - Beispiel: *Loch = true* $\wedge$ *Zahnschmerzen = false*

    bzw. *loch* $\wedge \neg$ *zahnschmerzen*

Atomare Ereignisse

- Eine Aussage, die ***allen*** Zufallsvariablen eines Weltmodells einen bestimmten Wert zuweist, wird atomares Ereignis genannt.
- Unterschiedliche atomare Ereignisse schließen sich wechselseitig aus.
- Die Menge aller möglichen atomare Ereignisse ist erschöpfend, d.h. genau ein atomares Ereignis muss wahr sein.

  Beispiel: Wenn *Loch* und *Zahnschmerzen* die einzigen Zufallsvariablen einer Domänenmodellierung sind, dann gibt es vier atomare Ereignisse:

  - $loch \wedge zahnschmerzen$
  - $loch \wedge \neg zahnschmerzen$
  - $\neg loch \wedge zahnschmerzen$
  - $\neg loch \wedge \neg zahnschmerzen$

A Priori Wahrscheinlichkeit

- $P(a)$ bezeichnet die unbedingte Wahrscheinlichkeit oder A-priori-Wahrscheinlichkeit, dass die Aussage a für den Fall gilt, dass *keine* zusätzliche Information verfügbar ist.

  – Beispiel: $P(loch) = 0.2$.

- Wie gewinnt man A-priori-Wahrscheinlichkeiten?

  – Z.B. durch statistische Analyse oder aus allgemeinen Regeln.

Wahrscheinlichkeitsverteilung

- Im Allgemeinen kann eine Zufallsvariable nicht nur die Booleschen Werte *wahr* und *falsch* annehmen, sondern auch mehrere Werte:

$$P(Wetter = sonnig) = 0.7, \qquad P(Wetter = regnerisch) = 0.2,$$

$$P(Wetter = wolkig) = 0.08, \qquad P(Wetter = schnee) = 0.02,$$

$$P(Kopfschmerzen = true) = 0.1.$$

- **P**(X) bezeichnet den Vektor der Wahrscheinlichkeiten für den (geordneten) Wertebereich der diskreten Zufallsvariable X.

$$\mathbf{P}(Kopfschmerzen) = \langle 0.1, 0.9 \rangle,$$

$$\mathbf{P}(Wetter) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$

definieren die ***Wahrscheinlichkeitsverteilungen*** der beiden diskreten Zufallsvariablen *Kopfschmerzen* bzw. *Wetter*. Die Summe der Werte ist auf 1 normiert.

Bedingte Wahrscheinlichkeiten (1)

Neue Information kann die Wahrscheinlichkeit eines Ereignisses ändern!

- Beispiel: Die W‘keit von Karies erhöht sich, wenn man weiß, dass der Patient Zahnschmerzen hat.

- Liegt Zusatzinformation vor, darf nicht mehr mit A-priori-W‘keiten gerechnet werden!

- $P(a|b)$ bezeichnet die ***bedingte Wahrscheinlichkeit*** oder ***A-posteriori-Wahrscheinlichkeit*** von *a*, sofern die *alleinige Evidenz b* gegeben ist:

$$P(loch|zahnschmerzen) = 0.6.$$

- $\mathbf{P}(X|Y)$ ist die Tabelle aller bedingten W‘keiten über alle Werte von *X* und *Y*.

Bedingte Wahrscheinlichkeiten (2)

P(*Wetter*|*Kopfschmerzen*) ist eine 4×2 Tabelle von bedingten W‘keiten aller Kombinationen der Werte der Zufallsvariablen.

	Kopfschmerzen = true	*Kopfschmerzen = false*
Wetter = sonnig	*P*(*W = sonnig* | *kopfschmerzen*)	*P*(*W = sonnig* | ¬*kopfschmerzen*)
Wetter = regnerisch	…	…
Wetter = wolkig	…	…
Wetter = schnee	…	…

Bedingte W‘keiten ergeben sich aus unbedingten W‘keiten *per Definition* wie folgt ($P(b) > 0$):

(1) $$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}.$$

Bedingte Wahrscheinlichkeiten (3)

Aus Gleichung (1) folgt die ***Produktregel***:

(2) $$\boxed{P(a \wedge b) = P(a \mid b) \cdot P(b).}$$

Zum Beispiel:

$P(a \wedge b)$ kann auch geschrieben werden als $P(a,b)$

- $P(W = sonnig \wedge kopfschmerzen) =$

$$P(W = sonnig \mid kopfschmerzen) \cdot P(kopfschmerzen)$$

- $P(W = regnerisch \wedge kopfschmerzen) =$

$$P(W = regnerisch \mid kopfschmerzen) \cdot P(kopfschmerzen)$$

…

- $P(W = schnee \wedge \neg\, kopfschmerzen) =$

$$P(W = schnee \mid \neg\, kopfschmerzen) \cdot P(\neg\, kopfschmerzen)$$

Bedingte Wahrscheinlichkeiten (4)

- Aus Gleichung (1)

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

folgt also allg. die Produktregel:

(2) $$\boxed{P(a \wedge b) = P(a \mid b) \cdot P(b)}$$

und analog:

(2‘) $$\boxed{P(a \wedge b) = P(b \mid a) \cdot P(a)}$$

- Zwei Aussagen *a* und *b* heißen unabhängig voneinander, falls $P(a|b) = P(a)$ und $P(b|a) = P(b)$ gelten. Dann und nur dann gilt:

$$P(a \wedge b) = P(a) \cdot P(b).$$

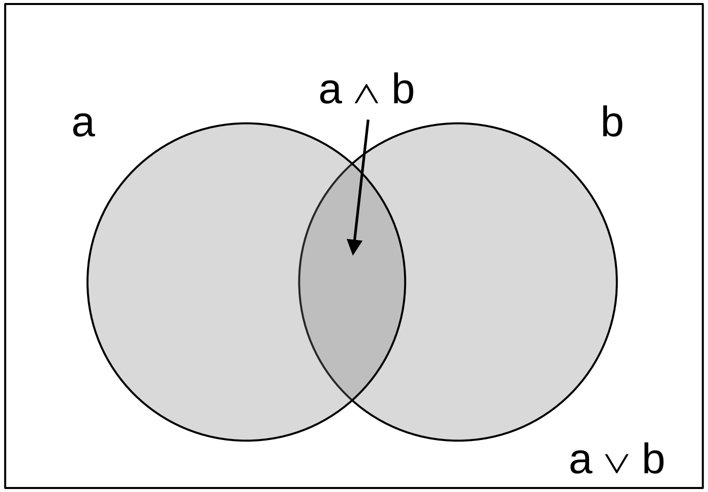
Axiomatische Wahrscheinlichkeitstheorie

Eine Funktion *P* von Aussagen in die Menge [0,1] ist ein *Wahrscheinlichkeitsmaß*, wenn für alle Aussagen *a, b* gilt:

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

2) $P(true) = 1$, $P(false) = 0$

3) $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Venn-Diagramm zur Veranschaulichung der Aussagen *a*, *b*, $a \wedge b$ (Schnitt von *a*, *b*), $a \vee b$ (Vereinigung von *a*, *b*).

Alle anderen Eigenschaften lassen sich aus diesen drei Axiomen* ableiten!

Z.B. folgt

$$P(\neg a) = 1 - P(a)$$

aus

$$P(a \vee \neg a) = 1 \text{ und } P(a \wedge \neg a) = 0.$$

* Die drei Axiome werden auch als Kolmogorov-Axiome bezeichnet. Der russ. Mathematiker Andrej Kolmogorov zeigte, dass sich die restl. W-Theorie aus diesen Axiomen ableiten lässt.

Wieso sind die Axiome sinnvoll?

- Wenn *P* eine *objektiv beobachtbare W'keit* bezeichnet, sind die Axiome natürlich sinnvoll.
- **Aber** wieso sollte ein Agent diese Axiome beachten, wenn er den *Grad seiner Überzeugung* modelliert?

$\leadsto$ *Objektive* Wahrscheinlichkeiten vs. *subjektive* Wahrscheinlichkeiten.

- Die Axiome schränken ja die Menge der Überzeugungen ein, die ein Agent aufrechterhalten kann.
- Warum subjektive Überzeugungen die Axiome respektieren sollten, zeigte Bruno de Finetti bereits 1931 mit folgenden Satz (s. auch Bspl. im Anhang):

> *Einem Agenten, der Wetten eingeht und dabei W'keiten zugrunde legt, die den Axiomen widersprechen, können immer Wettstrategien aufgezwungen werden, bei denen er unabhängig vom Ausgang der Ereignisse verliert.*

Verbundwahrscheinlichkeit

Eine ***Verbundwahrscheinlichkeit*** ist eine W'keit, die ein Agent jedem atomaren Ereignis einer Menge von Zufallsvariablen zuordnet.

Bereits bekannt: Ein atomares Ereignis ist eine Zuweisung *von Werten* an *alle* Zufallsvariablen $X_1, \ldots, X_n$ (= vollständige Spezifikation eines Zustands).

Beispiel: für zwei Boolesche Variablen X, Y gibt es die folgenden vier atomaren Ereignisse: (1) $X \wedge Y$, (2) $X \wedge \neg Y$, (3) $\neg X \wedge Y$, (4) $\neg X \wedge \neg Y$.

Die ***Verbundwahrscheinlichkeitsverteilung*** $\mathbf{P}(X_1, \ldots, X_n)$ weist *jedem atomaren Ereignis* eine W'keit zu:

	zahnschmerzen	*¬zahnschmerzen*
loch	*0.12*	*0.08*
¬loch	*0.08*	*0.72*

Da alle atomaren Ereignisse disjunkt sind, ist die Summe über alle Felder gleich 1 (Disjunktion der Ereignisse). Die Konjunktion ist notwendigerweise falsch.

Rechnen mit der Verbundwahrscheinlichkeit

Alle interessanten W'keiten lassen sich aus den Verbundwahrscheinlichkeiten errechnen, indem wir sie als Disjunktion von atomaren Ereignissen formulieren.

Beispiele:

1) Disjunktion $P(loch \vee zahnschmerzen) = P(loch \wedge zahnschmerzen) + P(\neg loch \wedge zahnschmerzen) + P(loch \wedge \neg zahnschmerzen)$

	zahnschmerzen	*¬zahnschmerzen*
loch	*0.12*	*0.08*
¬loch	*0.08*	*0.72*

2) Unbedingte Wahrscheinlichkeiten für einzelne Zufallsvariablen erhält man durch *Aufsummieren von Zeile oder Spalte* (***Marginalisierung***):

$$P(loch) = P(loch \wedge zahnschmerzen) + P(loch \wedge \neg zahnschmerzen)$$

3) Womit wieder bedingte Wahrscheinlichkeiten ableitbar sind:

$$P(loch \mid zahnschmerzen) = \frac{P(loch \wedge zahnschmerzen)}{P(zahnschmerzen)} = \frac{0.12}{0.12 + 0.08} = 0.60.$$

Probleme mit der Verbundwahrscheinlichkeit

Aus den Verbundwahr‘keiten lassen sich also alle Wahr‘keiten einfach ermitteln.

Aber die Verbundwahrscheinlichkeit umfasst k^n Werte, wenn es n Zufallsvariablen mit k Werten gibt.

$\rightsquigarrow$ schwierig darzustellen und schwierig zu ermitteln!

Fragen:

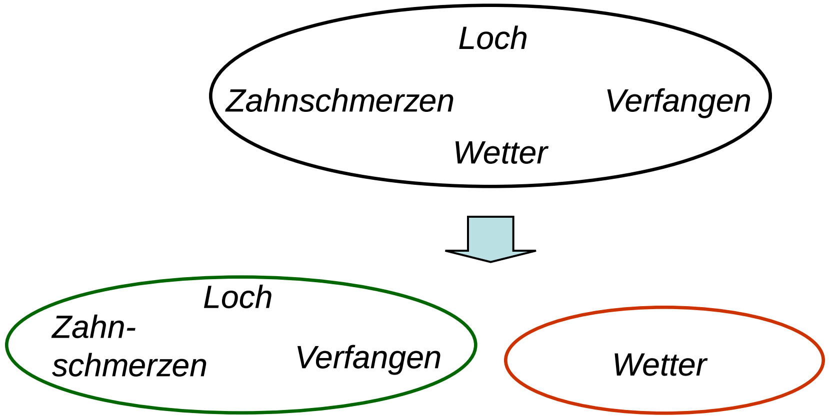
1) Gibt es eine *kompaktere* Darstellung von Verbundwahrscheinlichkeiten?

2) Gibt es eine *effiziente* Methode, diese Darstellung zu verarbeiten?

Als Antworten werden in dieser Vorlesung das Ausnutzen von Unabhängigkeiten sowie die Ableitung über die Bayessche Regel behandelt.

Absolute Unabhängigkeit

Sind einige Variablen in der Verbundverteilung unabhängig von den übrigen, kann die Verbundverteilung zerlegt werden:



P(*Zahnschmerzen*, *Verfangen*, *Loch*, *Wetter*)

= **P**(*Zahnschmerzen*, *Verfangen*, *Loch*) · **P**(*Wetter*)

8·4 = 32 Einträge werden auf 8 + 4 = 12 reduziert (bei 4 mögl. Werten für Wetter).

Absolute Unabhängigkeit ist aber selten. Z.B. ist Zahnmedizin ein großes Gebiet mit hunderten von Variablen und so gut wie keinen absoluten Unabhängigkeiten.

Konditionale Unabhängigkeit (1)

P(*Zahnschmerzen*, *Verfangen*, *Loch*) hat 7 (= 2^3-1) unabhängige Einträge. Daraus sind u.a. ableitbar:

- Wenn Karies vorliegt, ist die Wahrscheinlichkeit dafür, dass der Zahnarzt durch Stochern das Loch entdeckt, unabhängig davon, ob Zahnschmerzen vorliegen:

  ↝ $P(verfangen \mid zahnschmerzen, loch) = P(verfangen \mid loch)$

- Die gleiche Unabhängigkeit gilt für den Fall, dass kein Karies vorliegt:

  ↝ $P(verfangen \mid zahnschmerzen, \neg\, loch) = P(verfangen \mid \neg\, loch)$

↝ *Verfangen* ist *konditional unabhängig* (bzw. *bedingt unabhängig*) von *Zahnschmerzen* gegeben *Loch*.

- Zusicherungen von Unabhängigkeiten basieren i. A. auf Domänenwissen!

Konditionale Unabhängigkeit (2)

Die allg. Definition der bedingten Unabhängigkeit von zwei Variablen *X* und *Y* bei einer gegebenen dritten Variablen *Z* lautet:

(1) $\mathbf{P}(X|Y,Z) = \mathbf{P}(X|Z)$ und $\mathbf{P}(Y|X,Z) = \mathbf{P}(Y|Z)$ (s. Folie 24)

bzw.

(2) $\mathbf{P}(X,Y|Z) = \mathbf{P}(X|Z) \cdot \mathbf{P}(Y|Z)$

(2) am Beispiel:

$$\mathbf{P}(\mathit{Zahnschmerzen}, \mathit{Verfangen} \mid \mathit{Loch})$$

$$= \mathbf{P}(\mathit{Zahnschmerzen} \mid \mathit{Loch}) \cdot \mathbf{P}(\mathit{Verfangen} \mid \mathit{Loch})$$

Konditionale Unabhängigkeit (3)

Die vollständige Verbundverteilung erhält man über die Produktregel (s. F. 16, 17):

$$\mathbf{P}(\mathit{Zahnschmerzen}, \mathit{Verfangen}, \mathit{Loch})$$

$$= \mathbf{P}(\mathit{Zahnschmerzen} \mid \mathit{Verfangen}, \mathit{Loch}) \cdot \mathbf{P}(\mathit{Verfangen}, \mathit{Loch})$$

$$= \mathbf{P}(\mathit{Zahnschmerzen} \mid \mathit{Verfangen}, \mathit{Loch}) \cdot \mathbf{P}(\mathit{Verfangen} \mid \mathit{Loch}) \cdot \mathbf{P}(\mathit{Loch})$$

$$= \mathbf{P}(\mathit{Zahnschmerzen} \mid \mathit{Loch}) \cdot \mathbf{P}(\mathit{Verfangen} \mid \mathit{Loch}) \cdot \mathbf{P}(\mathit{Loch})$$

$\rightsquigarrow$ $2 \cdot (2 \cdot (2^1-1)) + 2^1-1 = 5$ unabhängige Einträge statt $(2^3-1) = 7$:

L	P(z|L)
t	0.4
f	0.011

L	P(v|L)
t	0.9
f	0.2

L	P(l)
t	0.2

Häufig kann die Größe der Repräsentation von Verbundverteilungen durch Ausnutzung von konditionalen Unabhängigkeiten von „exponentiell in der Anzahl der Variablen *n*“ auf „linear in *n*“ reduziert werden!

Zur Inferenz $\leadsto$ die Bayessche Regel

Wir wissen aus der Produktregel:

$$P(a \wedge b) = P(a|b) \cdot P(b) \quad \text{und} \quad P(a \wedge b) = P(b|a) \cdot P(a).$$

Durch *Gleichsetzen der beiden rechten* Seiten folgt:

$$P(a|b) \cdot P(b) = P(b|a) \cdot P(a).$$

Durch einfache Umformung: $P(a \mid b) = \dfrac{P(b \mid a) \cdot P(a)}{P(b)}$.

Für Verteilungen: $\mathbf{P}(A \mid B) = \dfrac{\mathbf{P}(B \mid A) \cdot \mathbf{P}(A)}{\mathbf{P}(B)}$.

Nützlich, um *diagnostische* Wahrscheinlichkeiten von *kausalen* abzuleiten:

$$\mathbf{P}(Ursache \mid Wirkung) = \frac{\mathbf{P}(Wirkung \mid Ursache) \cdot \mathbf{P}(Ursache)}{\mathbf{P}(Wirkung)}.$$

Anwendung der Bayesschen Regel

Geg. seien:	$P(zahnschmerzen \mid loch) = 0.6$	(kausal)
	$P(loch) = 0.2$	(Statistik)
	$P(zahnschmerzen) = 0.2$	(Statistik)

mit Bayes-Regel: $$P(loch \mid zahnschmerzen) = \frac{0.6 \cdot 0.2}{0.2} = 0.6$$

- Warum nicht gleich $P(loch \mid zahnschmerzen)$ schätzen?

- Kausales Wissen wie $P(zahnschmerzen \mid loch)$ ist i.A. robuster als diagnostisches Wissen wie $P(loch \mid zahnschmerzen)$:

  Die Regel $P(zahnschmerzen \mid loch)$ ist unabhängig von den A-priori-W'keiten $P(zahnschmerzen)$ und $P(loch)$.

  ⇝ nähme $P(loch)$ bei einer Kariesepidemie zu, so bliebe die Kausalität $P(zahnschmerzen \mid loch)$ unverändert, während sich $P(zahnschmerzen)$ proportional mit $P(loch)$ änderte.

  ⇝ ein Arzt, der $P(loch \mid zahnschmerzen)$ aus Daten geschätzt hätte, müsste von vorne anfangen.

Normalisierung (1)

Wenn wir die Wahrscheinlichkeit von *P*(*loch*|*zahnschmerzen*) bestimmen wollen und *P*(*zahnschmerzen*) nicht kennen, können wir auch eine *vollständige Fallanalyse* durchführen (hier z.B. für *loch* und ¬*loch*) und den Zusammenhang *P*(*loch*|*zahnschmerzen*) + *P*(¬*loch*|*zahnschmerzen*) = 1 ausnutzen (weil *Loch* Boolesche Variable ist).

Hier aus Platzgründen *zahnschmerzen* = *z* und *loch* = *l*.

$$P(l \mid z) = \frac{P(z \mid l)P(l)}{\boxed{P(z)}},$$

$$P(\neg l \mid z) = \frac{P(z \mid \neg l)P(\neg l)}{\boxed{P(z)}},$$

$$P(l \mid z) + P(\neg l \mid z) = 1 = \frac{P(z \mid l)P(l)}{\boxed{P(z)}} + \frac{P(z \mid \neg l)P(\neg l)}{\boxed{P(z)}},$$

$$\boxed{P(z)} = P(z \mid l)P(l) + P(z \mid \neg l)P(\neg l).$$

Normalisierung (2)

Durch Einsetzen in die ersten beiden Gleichungen:

$$P(l \mid z) = \frac{P(z \mid l)\, P(l)}{\boxed{P(z \mid l)\, P(l) + P(z \mid \neg l)\, P(\neg l)}},$$

$$P(\neg l \mid z) = \frac{P(z \mid \neg l)\, P(\neg l)}{\boxed{P(z \mid l)\, P(l) + P(z \mid \neg l)\, P(\neg l)}}.$$

Dieser Prozess heißt ***Normalisierung***, weil 1/P(z) einfach als Normalisierungskonstante genutzt werden kann, um die Summe der bedingten W'keiten gleich 1 zu setzen.

Für *mehrwertige* Zufallsvariablen allgemein:

$$\mathbf{P}(Y|X) = \alpha \cdot \mathbf{P}(X|Y) \cdot \mathbf{P}(Y).$$

Wobei α eben die Normierungskonstante ist, die die Summe der Werte in **P**(Y|X) über alle *Y* auf 1 normiert, z.B. α = 1/5 für ⟨1, 1, 3⟩ damit 1/5 + 1/5 + 3/5 = 1.

Normalisierung (3): Beispiel

Geg. sei das Zahnarztszenario jetzt durch folgende Verbundw.-Verteilung:

	zahnschmerzen		¬ zahnschmerzen	
	verfangen	¬verfangen	verfangen	¬verfangen
loch	0.108	0.012	0.072	0.008
¬loch	0.016	0.064	0.144	0.576

$$\Rightarrow P(loch \mid zahnschmerzen) = \frac{P(loch \wedge zahnschmerzen)}{P(zahnschmerzen)} = \frac{0.108+0.012}{0.108+0.012+0.016+0.064} = 0.6$$

$$\Rightarrow P(\neg loch \mid zahnschmerzen) = \frac{P(\neg loch \wedge zahnschmerzen)}{P(zahnschmerzen)} = \frac{0.016+0.064}{0.108+0.012+0.016+0.064} = 0.4$$

Mit Normalisierung in einer Gleichung formulierbar:

s. F. 15: $P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$

P(*Loch*|*zahnschmerzen*) = α·***P***(*Loch*,*zahnschmerzen*)

= α·[***P***(*Loch*,*zahnschmerzen*,*verfangen*) + ***P***(*Loch*,*zahnschmerzen*,¬*verfangen*)]

= α· [⟨0.108;0.016⟩ + ⟨0.012;0.064⟩]

= α· ⟨0.12;0.08⟩

= ⟨0.6;0.4⟩

Die Belegung von *Verfangen* ist für ***P***(*Loch*|*zahnschmerzen*) nicht festgelegt. Also ist über alle (hier zwei) mögl. Werte zu summieren.

Normalisierung 4

Verallgemeinert aus dem Beispiel sind gegeben:

- Anfragevariable X (im Beispiel *Loch*),
- Menge $\boldsymbol{E}$ der beobachteten Evidenzvariablen (im Beispiel *zahnschmerzen*) und $\boldsymbol{e}$ die beobachteten Werte für diese,
- Menge $\boldsymbol{Y}$ der restlichen unbeobachteten Variablen (im Beispiel *verfangen*) und $\boldsymbol{y}$ die möglichen Werte für diese,

Die Auswertung der Anfrage $\boldsymbol{P}(X \mid \mathbf{e})$ ist dann:

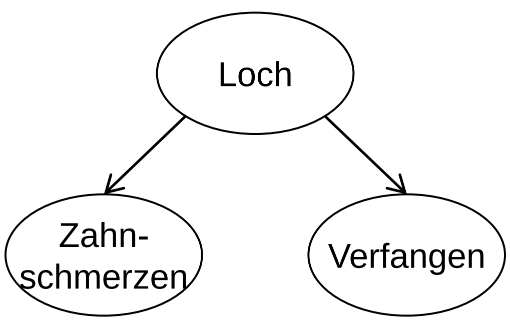
$$\boldsymbol{P}(X \mid \boldsymbol{e}) = \alpha \cdot \boldsymbol{P}(X, \boldsymbol{e}) = \alpha \cdot \sum_{y} \boldsymbol{P}(X, \boldsymbol{e}, \boldsymbol{y}).$$

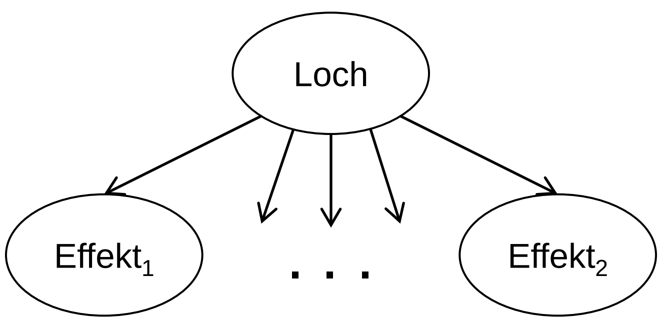
Bayessche Regel und konditionale Unabhängigkeit

$$\mathbf{P}(Loch, zahnschmerzen, verfangen)$$
$$= \alpha\, \mathbf{P}(zahnschmerzen \wedge verfangen \mid Loch)\, \mathbf{P}(Loch)$$
$$= \alpha\, \mathbf{P}(zahnschmerzen \mid Loch)\, \mathbf{P}(verfangen \mid Loch)\, \mathbf{P}(Loch)$$

Dies ist ein Beispiel des sogenannten *naiven Bayes-Modells**, in dem eine einzige Ursache direkt mehrere Effekte verursacht, die alle von einander unabhängig sind. Die Verbundwahrscheinlichkeitsverteilung ist dann:

$$\mathbf{P}(\text{Ursache}, \text{Effekt}_1, \ldots, \text{Effekt}_n) = \mathbf{P}(\text{Ursache}) \prod_i \mathbf{P}(\text{Effekt}_i \mid \text{Ursache})$$

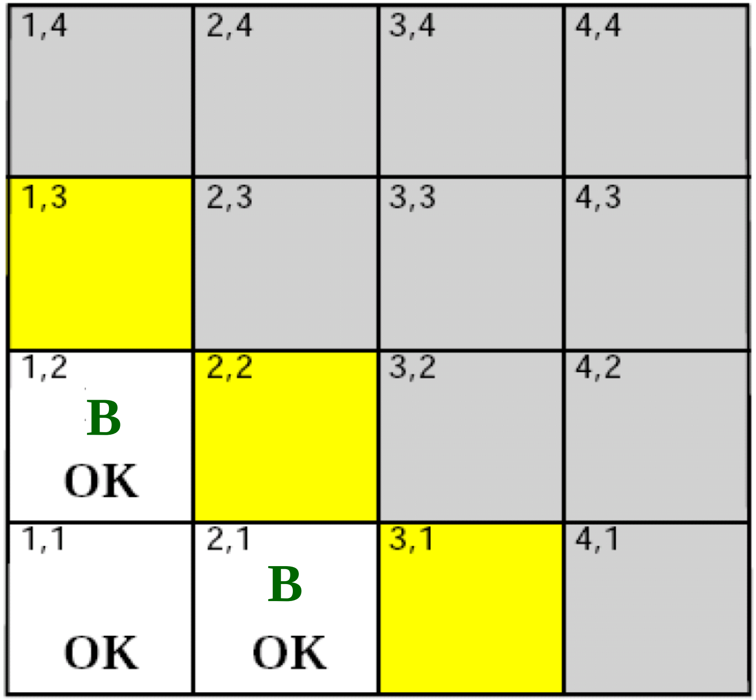




* In der Praxis funktionieren naive Bayes-Modelle überraschend gut, selbst wenn die Effektvariablen tatsächlich nicht bedingt unabhängig sind bei geg. Ursachenvariable

Wumpus Welt

Unsicherheit entsteht hier, weil die Sensoren nur partielle, lokale Information über die Umgebung liefern.



Hier hat der Agent in [1,2] und [2,1] einen Luftzug (B) erkannt. Also kann jedes der drei erreichbaren Felder [1,3], [2,2], [3,1] eine Falltür enthalten. Ein logischer Agent müsste probieren – mit Risiko!

In diesem Bspl. werden nur die Luftzüge und die Falltüren betrachtet. Also werden hier der Wumpus und das Gold nicht berücksichtigt.

Ein probabilistischer Agent ist besser:

Kodierung: $P_{ij} = true$ gdw. $[i,j]$ enthält Fallgrube (Pit)

$B_{ij} = true$ gdw. $[i,j]$ ist „breezy“

Gegeben sind nur die Wahrnehmungen B_{11}, B_{12}, B_{21}.

Frage: wie wahrscheinlich sind Fallgruben in den gelben Feldern?

Spezifikation des Wahrscheinlichkeitsmodells

- Vollständ. Verbundverteilung: $\mathbf{P}(P_{11}, \ldots, P_{44}, B_{11}, B_{12}, B_{21})$.

  → Anwendung der Produktregel:

  $\mathbf{P}(B_{11}, B_{12}, B_{21} \mid P_{11}, \ldots, P_{44}) \cdot \mathbf{P}(P_{11}, \ldots, P_{44})$

  Warum?

- Um kausale Beziehung $\mathbf{P}(Wirkung|Ursache)$ zu erhalten:

  - *Modell für ersten Term* (Kombination Luftzug und benachbarte Falltür):

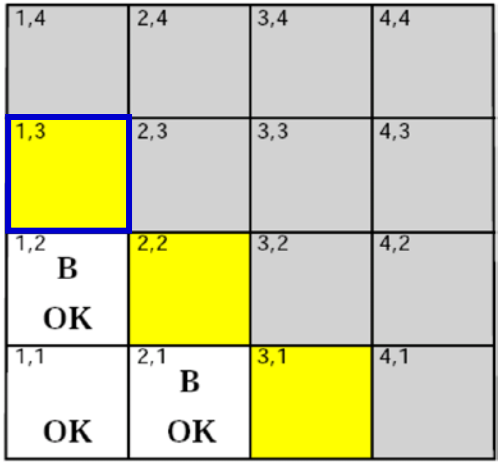
    1 für Luftzug, falls direkt benachbart zu Fallgruben, sonst 0.

  - *Modell für zweiten Term*: Fallgruben sind zufällig und unabhängig verteilt mit Wahrscheinlichkeit 20% pro Feld, d.h.: $\mathbf{P}(P_{11}, \ldots, P_{44}) = \prod_{(i,j)\in\{(1,1),\ldots,(4,4)\}} \mathbf{P}(P_{ij})$.

  - Für eine Konfiguration mit *n* Fallgruben also: $0{,}2^{n} \cdot 0{,}8^{16-n}$.

Beobachtungen und Anfrage

Wir kennen die folgenden Fakten:



$$b = \neg b_{11} \wedge b_{12} \wedge b_{21}$$

$$\text{known} = \neg p_{11} \wedge \neg p_{12} \wedge \neg p_{21}$$

Wir wollen z.B. wissen, wie wahrscheinlich in P_{13} eine Fallgrube ist.

→ Anfrage: $\mathbf{P}(P_{13} \mid known, b)$

Definiere: *unknown* als „alle Felder außer P_{13} und den Feldern aus *known*".

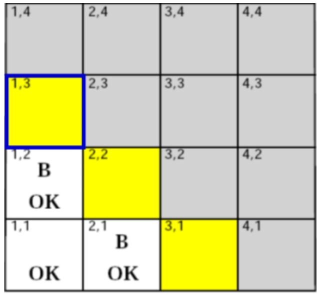
Jetzt gilt: $\mathbf{P}(P_{13} \mid known, b) = \alpha \cdot \sum_{unknown} \mathbf{P}(P_{13}, unknown, known, b)$

Ausnutzung konditionaler Unabhängigkeit (1)

Damit haben wir eine Ableitungsvorschrift:

$$\mathbf{P}(P_{13} \mid known, b) = \alpha \cdot \sum_{unknown} \mathbf{P}(P_{13}, unknown, known, b)$$

Aber wir vernachlässigen die *Berechnungskomplexität*. Wir müssen über 12 unbekannte Felder summieren. Die Summe enthält $2^{12} = 4096$ Terme!



Grundlegende Einsicht: z.B. Feld [4,4] wird in der Wumpus-Welt wenig Aufschluss über Feld [1,3] geben.

Generell: *die Beobachtungen (breezes) sind unabhängig* von den übrigen unbekannten Feldern, wenn ihre direkten Nachbarfelder gegeben sind.

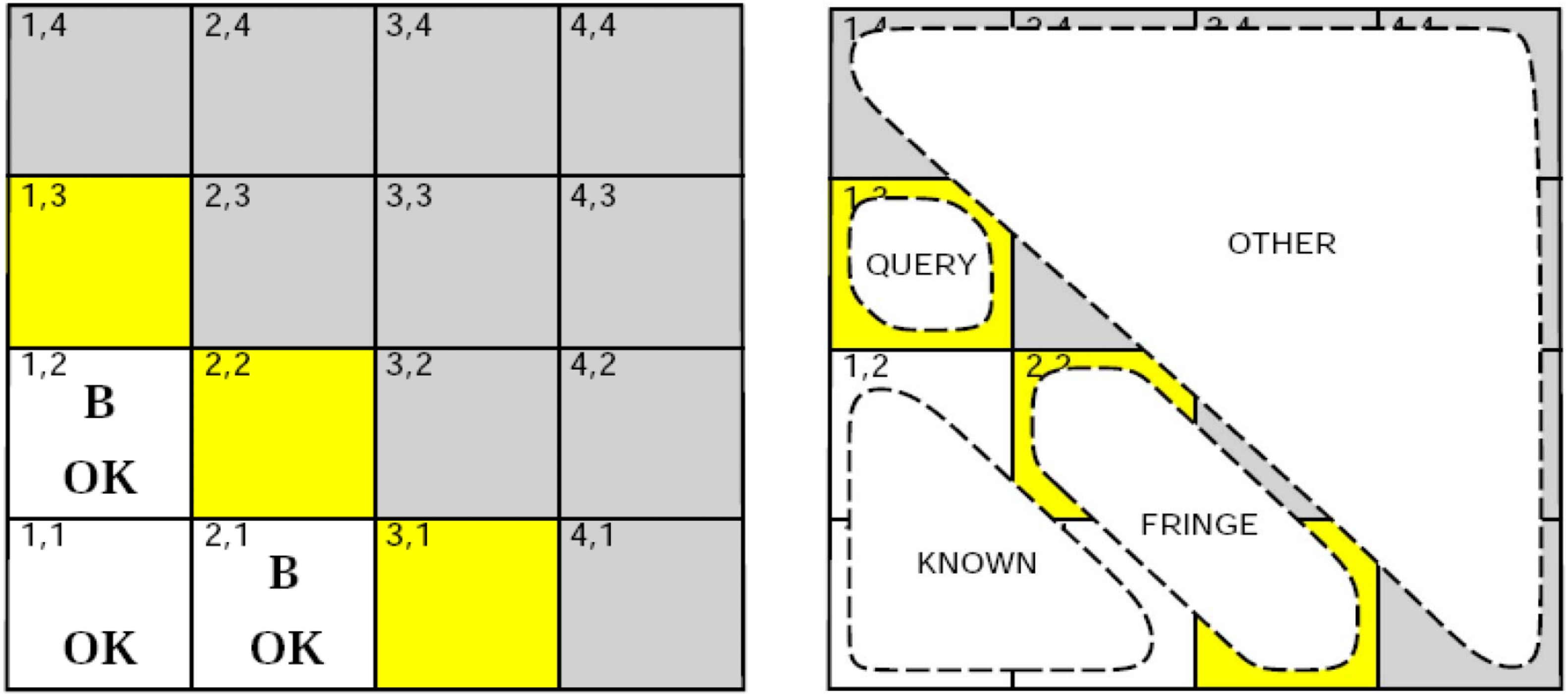
Wir definieren als Rand (*fringe*) die Felder, die neben den besuchten Feldern liegen, aber nicht Anfragefeld sind. Damit gelten $unknown = fringe \cup other$ und

$$\mathbf{P}(b \mid P_{13}, known, unknown) = \mathbf{P}(b \mid P_{13}, known, fringe)$$

Ausnutzung konditionaler Unabhängigkeit (2)

Wir definieren als Rand (*fringe*) die Felder, die neben den besuchten Feldern liegen, aber nicht Anfragefeld sind. Damit gelten *unknown* = *fringe* ∪ *other* und

$$\mathbf{P}(b \mid P_{13}, known, unknown) = \mathbf{P}(b \mid P_{13}, known, fringe)$$



Jetzt formen wir die Anfrage so um, dass wir dies ausnutzen können!

Ausnutzung konditionaler Unabhängigkeit (3)

$\mathbf{P}(P_{13} \mid known, b)$ s. F. 32: $\mathbf{P}(X \mid \mathbf{e}) = \alpha \cdot \mathbf{P}(X, \mathbf{e}) = \alpha \cdot \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$

$= \alpha \cdot \sum_{unknown} \mathbf{P}(P_{13}, unknown, known, b)$

Produktregel

$= \alpha \cdot \sum_{unknown} \mathbf{P}(b \mid P_{13}, known, unknown) \cdot \mathbf{P}(P_{13}, known, unknown)$

umformuliert

$= \alpha \cdot \sum_{fringe} \sum_{other} \mathbf{P}(b \mid known, P_{13}, fringe, other) \cdot \mathbf{P}(P_{13}, known, fringe, other)$

$= \alpha \cdot \sum_{fringe} \sum_{other} \mathbf{P}(b \mid known, P_{13}, fringe) \cdot \mathbf{P}(P_{13}, known, fringe, other)$ bed. Unabh.

$= \alpha \cdot \sum_{fringe} \mathbf{P}(b \mid known, P_{13}, fringe) \cdot \sum_{other} \mathbf{P}(P_{13}, known, fringe, other)$

Unabhängigkeit

$= \alpha \cdot \sum_{fringe} \mathbf{P}(b \mid known, P_{13}, fringe) \cdot \sum_{other} \mathbf{P}(P_{13}) \cdot P(known) \cdot P(fringe) \cdot P(other)$

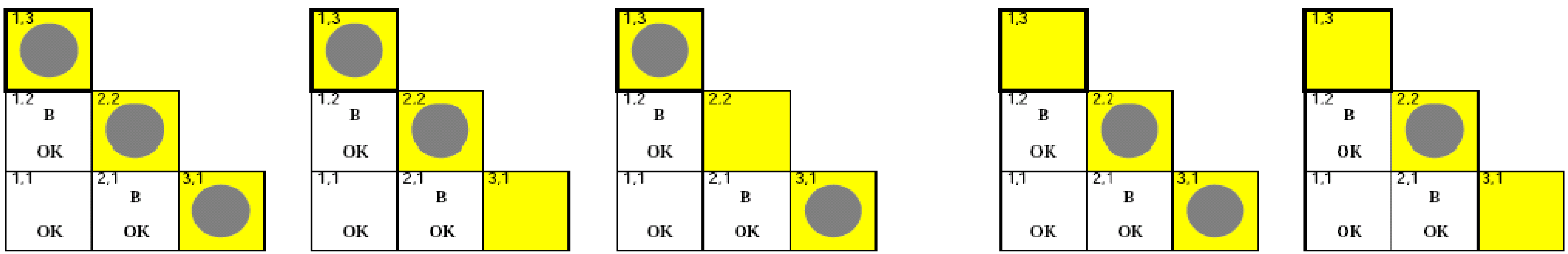
$= \alpha \cdot P(known) \cdot \mathbf{P}(P_{13}) \cdot \sum_{fringe} \mathbf{P}(b \mid known, P_{13}, fringe) \cdot P(fringe) \sum_{other} P(other)$

$= \alpha' \cdot \mathbf{P}(P_{13}) \cdot \sum_{fringe} \mathbf{P}(b \mid known, P_{13}, fringe) \cdot P(fringe)$

Wobei zuletzt P(*known*) in die Normalisierungskonstante aufgenommen wird und genutzt wird, dass $\sum_{other} P(other)$ gleich 1 ist.

Ausnutzung konditionaler Unabhängigkeit (4)

- Bei $\mathbf{P}(P_{13} \mid known,\ b) \ = \alpha' \cdot \mathbf{P}(P_{13}) \cdot \sum_{fringe} \mathbf{P}(b \mid known, P_{13}, fringe) \cdot P(fringe)$ ist $P(b \mid known,\ P_{13},\ fringe) = 1$, wenn der Rand konsistent mit breeze ist, sonst 0.
- Somit summieren wir für jeden P_{13}-Wert nur über die logischen Modelle des Randes, die konsistent mit den Fakten sind.
- Fünf konsistente „Fringe"-Zustände mit $P(b \mid P_{13}, known, fringe) = 1$ für p_{13} und $\neg p_{13}$:



0.2 x 0.2 = 0.04   0.2 x 0.8 = 0.16   0.8 x 0.2 = 0.16   0.2 x 0.2 = 0.04   0.2 x 0.8 = 0.16

- $\mathbf{P}(P_{13} \mid known,\ b) = \alpha' \langle 0.2 \cdot (0.04 + 0.16 + 0.16), 0.8 \cdot (0.04 + 0.16) \rangle$
  $\approx \langle 0.31,\ 0.69 \rangle$
- $\mathbf{P}(P_{31} \mid known,\ b) \approx \langle 0.31,\ 0.69 \rangle$ wg. Symmetrie
- $\mathbf{P}(P_{22} \mid known,\ b) \approx \langle 0.86,\ 0.14 \rangle$

→ Es ist also günstiger, am Rand entlang zu laufen.

Zusammenfassung

- Unsicherheit ist unvermeidbar in komplexen und dynamischen Welten, in denen Agenten zur Ignoranz gezwungen sind.
- Wahrscheinlichkeiten formulieren die Unfähigkeit eines Agenten, eine definitive Entscheidung zu fällen. Sie drücken den Grad seiner Überzeugung aus.
- Bedingte und unbedingte Wahrscheinlichkeiten können über Propositionen formuliert werden.
- Alle relevanten Wahrscheinlichkeiten können durch Summation der Wahrscheinlichkeiten atomarer Ereignisse aus der Verbundverteilung ermittelt werden.
- Für nicht triviale Domänen ist es erforderlich, eine effiziente Repräsentation der Verbundverteilung zu finden.
- Unabhängigkeit und konditionale Unabhängigkeit sind die Werkzeuge dafür.
- Produktregel und Bayessche Regel ermöglichen es dabei, Anfragen geeignet umzuformen.

Anhang: Ein Wettspiel nach de Finetti

- Agent 1 hat die Überzeugung $P(a) = 0.4$, dass ein Ereignis *a* eintritt.
- Agent 2 kann für oder gegen a wetten, sein Einsatz muss dabei jedoch konsistent mit der Überzeugung von Agent 1 sein (faire Wette).
- Beispiel: Agent 2 setzt 4 zu 6 auf *a*, d.h. tritt *a* auf, muss Agent 1 den Betrag von 6 Euro an Agent 2 zahlen, sonst zahlt Agent 2 den Betrag von 4 Euro an Agent 1.

Agent 2 wettet:	Agent 1 muss dagegenhalten:
a tritt auf: 4 Euro	*a* tritt nicht auf: 6 Euro
a tritt nicht auf: 6 Euro	*a* tritt auf: 4 Euro

↝ Nach de Finetti muss Agent 1 diese Wette akzeptieren (weil sie fair ist).

Eine Wettstrategie ist eine Menge von Wetten auf mehrere Ereignisse.

Anhang: Das Wettmodell

Agent 1 habe die folgenden Grade von Überzeugungen:

$P(a) = 0.4,\ \ P(b) = 0.3,\ \ P(a \wedge b) = 0.0,\ P(a \vee b) = 0.8$

$\leadsto$ Widerspruch zu 3. Axiom der W-Theorie: $P(a \vee b) \neq P(a) + P(b) - P(a \wedge b)$

Agent 2 habe die Wettstrategie:

a, b, $\neg(a \vee b)$ und

setzt 4 zu 6 auf a, 3 zu 7 auf b und 2 zu 8 auf $\neg(a \vee b)$:

Agent 1		Agent 2		Ergebnisse für Agent 1 bei			
Aussage	Glaube	Wette	Einsatz	$a \wedge b$	$a \wedge \neg b$	$\neg a \wedge b$	$\neg a \wedge \neg b$
a	0.4	a	4 zu 6	-6	-6	4	4
b	0.3	b	3 zu 7	-7	3	-7	3
$a \vee b$	0.8	$\neg(a \vee b)$	2 zu 8	2	2	2	-8
				-11	-1	-1	-1

Setzt Agent 2 entspr. seiner Strategie, verliert Agent 1 bei allen möglichen Ergebnissen für a und b.