3. Dempster-Shafer-Theorie

Inhaltsübersicht

Einführendes Beispiel: Der Wetterfrosch

- Wetterfrosch äußert: Das Wetter wird gut.
- Vertrauen in den Wetterfrosch: 60%
- Mögliche Anwendung der Wahrscheinlichkeitstheorie:
  „Der Wetterfrosch sagt in 60% der Fälle die Wahrheit"
  A: Wetter wird gut, $\Pr(A) = 60\%$
- Dann gleichzeitig Aussage für die Gegenhypothese:
  $\bar{A}$: Wetter wird schlecht, $\Pr(\bar{A}) = 40\%$

Ansatz der Dempster-Shafer-Theorie:
Das Vertrauen in den Wetterfrosch ist etwas anderes als der Anteil der wahren Aussagen.
Aus dem Vertrauen in den Wetterfrosch kann nicht geschlossen werden, dass der Gegenhypothese „Wetter wird schlecht" ein Vertrauen von 40% zugeordnet werden darf. Zu 40% weiss man nichts.

$\rightarrow$ Einführung eines Formalismus zum Ausdrücken von Nichtwissen

3 Dempster-Shafer-Theorie

- Als Verallgemeinerung der Bayes‘schen Theorie interpretierbar
- Meist für diskrete Größen verwendet (kontinuierliche Größen s. Literatur)
- Einführung von Basismaßen („Verallgemeinerung von W‘maßen“) zur Zuordnung einer „Menge an Glauben“ an Elementarergebnisse *oder kombinierte* Ereignisse
  → Spezifizierung von Nichtwissen explizit möglich
- Basismaße: Formalisierung des Wissens über Indizien bzw. Beweisstücke, Informationsbeiträge (Evidenzen)
  → Bezeichnung Evidenztheorie gebräuchlich
- Definition von Glaubens- und Plausibilitätsfunktionen (aus Basismaßen) zur Angabe von „unteren und oberen Wahrscheinlichkeiten“

Historie

- Arthur Dempster (1960er Jahre):
  - System von unteren und oberen Wahrscheinlichkeitsgrenzen
  - Degree of Belief: Quantifizierung eines Zustands von partiellem Wissen, der einer Informationsquelle zugeordnet ist
  - Dempster‘s rule of combination: Regel, die das Wissen verschiedener unabhängiger Informationsquellen kombiniert
- Glenn Shafer (1970er Jahre):
  - Weiterentwicklung zur Dempster-Shafer-Theorie (Evidenztheorie)
  - (Erst jetzt) Abgrenzung von der Wahrscheinlichkeitstheorie: Versuch, eine Unsicherheitstheorie losgelöst von der W‘theorie zu entwerfen
- Phillipe Smets (1980er Jahre):
  - Modifikation der DS-Theorie aufgrund von unplausiblen Ergebnissen

3.1 Formale Struktur

Die Menge $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ aller Elementarergebnisse wird als **Wahrnehmungsrahmen** (engl. frame of discernment) bezeichnet.

- Definition von Ω kann einen subjektiveren Charakter als in der klassischen Wahrscheinlichkeitstheorie bzw. der Bayes-Theorie besitzen:
  Subjektivität dort bezieht sich auf die Wahrscheinlichkeit von Ereignissen, nicht auf die Elemente des Grundraumes

Ein **Basismaß** (engl. basis probability assignment) ist eine auf $\mathcal{P}(\Omega)$ definierte Funktion m mit

(BM1) $m(A) \in [0,1]$ für alle Mengen $A \subseteq \Omega$

(BM2) $m(\varnothing) = 0$

(BM3) $\sum_{A \subseteq \Omega} m(A) = 1$

3.1 Formale Struktur

- Grundgedanke:
  - (Menge an) Glauben wird (so genau wie möglich) auf Ereignisse aus dem Wahrnehmungsrahmen verteilt
  - Eine Menge an Glauben wird einem Ereignis A zugeteilt, wenn es nicht möglich ist, sie einem Ereignis zuzuteilen, das A impliziert
  - $m(A)$: Menge an Glauben, die man *exakt* A zuweist

Ein Ereignis A mit $m(A) \neq 0$ heißt **fokales Ereignis**

3.1 Formale Struktur

- Explizite Zuordnung von Vertrauen in eine Beobachtung (bzw. in ein Sensorsystem)
- Explizite Zuordnung des Rests an Glauben an den Wahrnehmungsrahmen
(Vergleich Wahrscheinlichkeitstheorie: Zuordnung zum Komplement oder Aufteilung der Rest-Wahrscheinlichkeit)

Beispiel: Wetterfrosch

- Vertrauen in den Wetterfrosch: 60%, A: Wetter wird gut
- Formulierung in der Dempster-Shafer-Theorie: $m(A) = 0{,}6; m(\Omega) = 0{,}4$
- Mögliche Interpretationen mit Wahrscheinlichkeiten:
    - „Der Wetterfrosch sagt in 60% der Fälle die Wahrheit“:
    $\Pr(A) = 0{,}6; \Pr(\bar{A}) = 0{,}4$
    - „Zu 60% habe ich Vertrauen in seine Aussage, die restlichen 40% müssen auf alle möglichen Fälle verteilt werden:
    $\Pr(A) = 0{,}8; \Pr(\bar{A}) = 0{,}2$ (siehe Übung)

Eigenschaften von Basismaßen:

- Aus (BM3): Der „Gesamtglauben“ muss vollständig vergeben werden
- Aus (BM2): Closed-world-Assumption, d.h. Ω enthält alle möglichen Elementarergebnisse
- Basismaße sind die Bausteine für Belief- und Plausibilitätsmaße (s.u.)
- Ein Basismaß ist i.A. kein Wahrscheinlichkeitsmaß (s.u.)

3.1 Formale Struktur

<u>Wiederholung:</u> Axiome von Kolmogorov
für einen *endlichen* Wahrscheinlichkeitsraum Ω

$\Pr$: Wahrscheinlichkeitsmaß

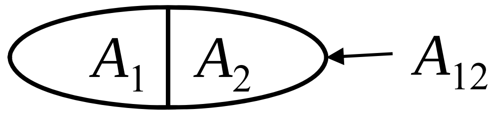
(A1): Nichtnegativität: $\Pr(A) \geq 0$

(A2): Normiertheit: $\Pr(\text{sicheres Ereignis}) = \Pr(\Omega) = 1$

(A3): Endliche Additivität: $A \cap B = \varnothing \;\Rightarrow\; \Pr(A \cup B) = \Pr(A) + \Pr(B)$

Daraus folgt Monotonie: $A \subseteq B \;\Rightarrow\; \Pr(A) \leq \Pr(B)$

Bei Basismaßen: Additivität: $A_1 \cap A_2 = \varnothing,\; A_1 \cup A_2 = A_{12}$



Basismaß für A_{12} wird explizit zugeordnet

Monotonie: $A_1 \subseteq A_2$

Basismaße für A_1 und A_2 werden explizit zugeordnet

→ Basismaße sind i.A. keine Wahrscheinlichkeitsmaße!

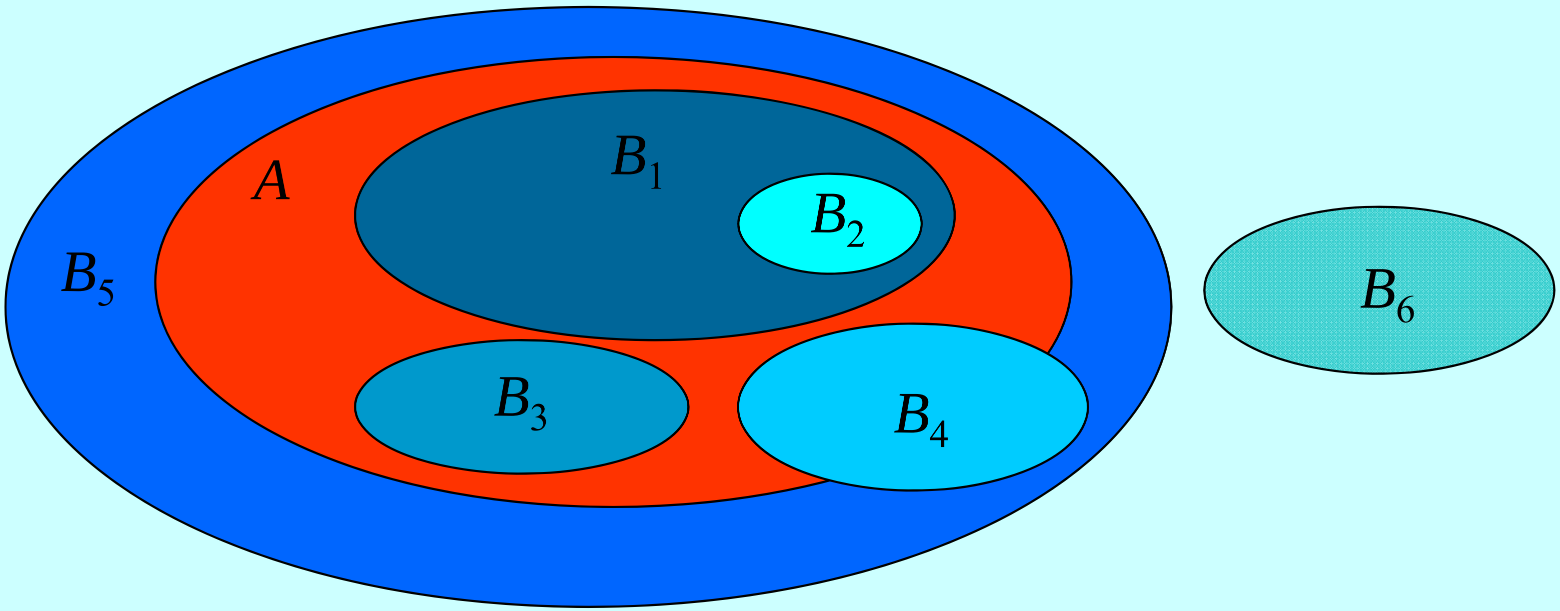
Basismaße sind die Bausteine für Glaubens- und Plausibilitätsfunktionen:

Ein Basismaß m induziert eine **Glaubensfunktion (Belief-Funktion)** $Bel : \mathcal{P}(\Omega) \to [0,1]$ durch

$$Bel(A) := \sum_{B:B \subseteq A} m(B)$$

(Summation über alle Ereignisse, die das Ereignis A implizieren)

Beispiel: $Bel(A) = ?$



Axiomatische Festlegung einer Glaubensfunktion:

Eine Glaubensfunktion ist eine Funktion $Bel : \mathcal{P}(\Omega) \to [0,1]$ mit

(Bel1) $Bel(\Omega) = 1$

(Bel2) $Bel(\emptyset) = 0$

(Bel3) $$Bel(A_1 \cup \ldots \cup A_n) \geq \sum_i Bel(A_i) + (-1)^3 \sum_{i<j} Bel(A_i \cap A_j) + \ldots + (-1)^{n+1} Bel(A_1 \cap \ldots \cap A_n)$$

Inversion: $$m(A) = \sum_{B:B \subseteq A} (-1)^{|A-B|} Bel(B)$$

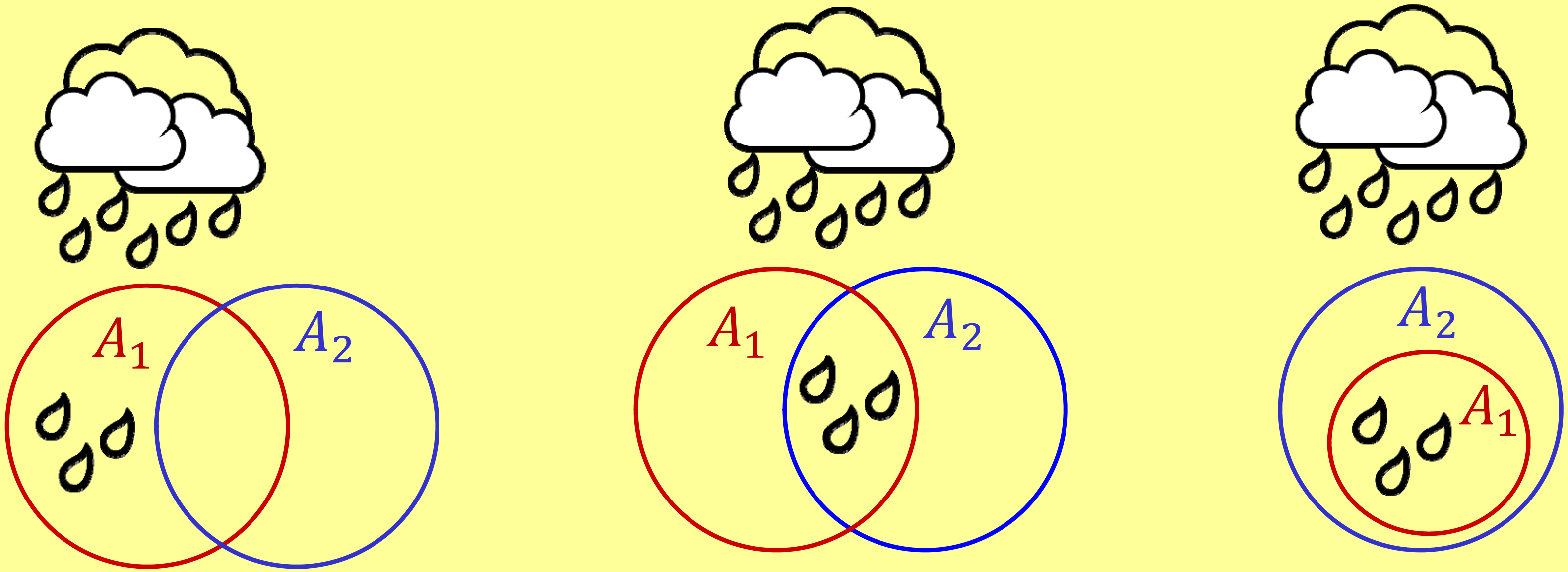
$\rightarrow$ Eindeutige Umrechnung zwischen Basismaßen und Glaubensfunktionen

- m „verteilt" den Glauben
- $Bel(A)$ gibt den (aufsummierten) Anteil an Glauben an, der mindestens, d.h. „mit Sicherheit", dem Ereignis A zugewiesen werden kann ($\rightarrow$ „untere Wahrscheinlichkeit")

3.1 Formale Struktur

Beispiel zur Verdeutlichung: (nach Zadeh)

- Es regnet auf die überlappenden Gebiete $A_i, i = 1{,}2$
- $m(A_i)$: Anteil des Regens, der *genau* auf das Gebiet A_i fällt
- Regen, der auf A_1 fällt, kann *nur* auf A_1 fallen (links), er kann *auch* auf A_2 fallen (Mitte). Er fällt *immer* auf A_2, wenn A_1 komplett in A_2 liegt (rechts).
- In jedem Fall ist $\mathrm{Bel}(A_2) = m(A_2) + m(A_1 \cap A_2)$



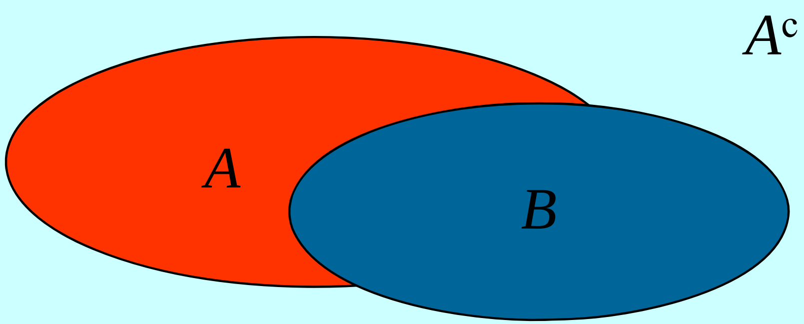
3.1 Formale Struktur

- Es gilt: $Bel(A) + Bel(A^c) \leq 1$

  Die Summe über den sicheren Glauben in ein Ereignis und in seine Negation kann kleiner Eins sein

- Eine Glaubensfunktion ist auch deshalb i.A. kein Wahrscheinlichkeitsmaß

Beispiel:



Falls $m(B) > 0$: $Bel(A) + Bel(A^c) < 1$

Leere Glaubensfunktion (Vacuous Belief Funktion):

$$Bel(\Omega) = 1$$

$Bel(A) = 0$ für alle echten Teilmengen A von Ω

$\rightarrow$ Ausdruck vollkommener Unwissenheit

Dies entspricht *nicht* einer Verteilung des Glaubens auf die Elementarergebnisse mittels des zugehörigen Basismaßes.

Vergleich zur Wahrscheinlichkeitstheorie: Hier würde man nach dem Prinzip der maximalen Entropie die Wahrscheinlichkeiten auf die Elementarergebnisse verteilen.

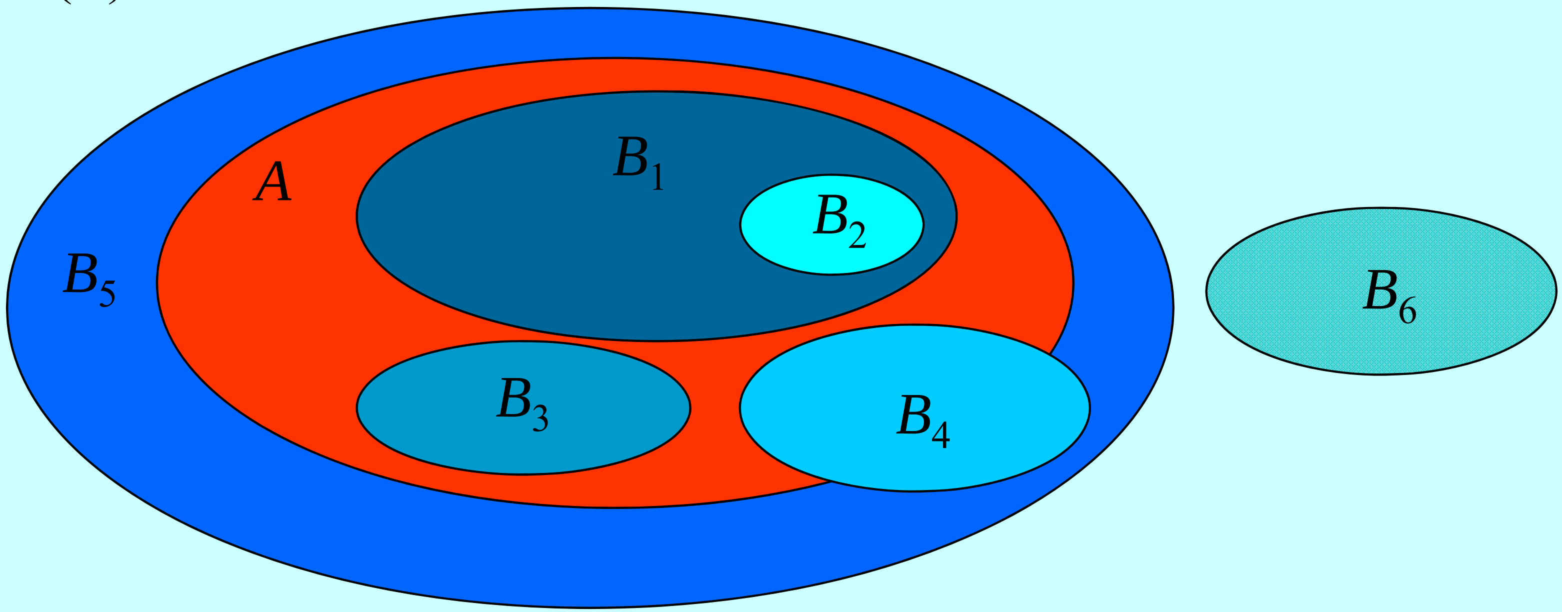
3.1 Formale Struktur

Ein Basismaß m induziert eine Plausibilitätsfunktion $Pl : \mathcal{P}(\Omega) \to [0,1]$ durch

$$Pl(A) := \sum_{B: A \cap B \neq \emptyset} m(B)$$

(Summation über alle Ereignisse, die dem Ereignis A nicht widersprechen)

Beispiel: $Pl(A) = ?$



3.1 Formale Struktur

Dualität: $Pl(A) = 1 - Bel(A^c)$

$Pl(.)$ ist komplementär zu $Bel(.)$

Durch die Definitionen von Bel und Pl, Inversion und Dualität sind m, Bel, Pl ineinander verlustlos umrechenbar

- $Pl(A)$ gibt den Anteil an Glauben an, der maximal, d.h. möglicherweise, dem Ereignis A zugewiesen werden kann (z.B. falls „die Umstände entsprechend" sind)

$Pl(A)$ gibt die zu $Bel(A)$ analoge obere Grenze an

($\rightarrow$ „obere Wahrscheinlichkeit")

3.1 Formale Struktur

Einige Eigenschaften von Glaubens- und Plausibilitätsfunktion:

- $Bel(\emptyset) = Pl(\emptyset) = 0$
- $Bel(\Omega) = Pl(\Omega) = 1$
- $0 \le Bel(A) \le Pl(A) \le 1$
- $Bel(A) + Bel(A^c) \le 1 \qquad Pl(A) + Pl(A^c) \ge 1$
- $Bel(A \cup B) \ge Bel(A) + Bel(B) - Bel(A \cap B)$
  $Pl(A \cap B) \le Pl(A) + Pl(B) - Pl(A \cup B)$
- Wenn $A \cap B = \emptyset$: $Bel(A \cup B) \ge Bel(A) + Bel(B)$ (Superadditivität)
  $Pl(A \cup B) \le Pl(A) + Pl(B)$ (Subadditivität)

Beispiel: Berechnung von Glaubens- und Plausibilitätsfunktion

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

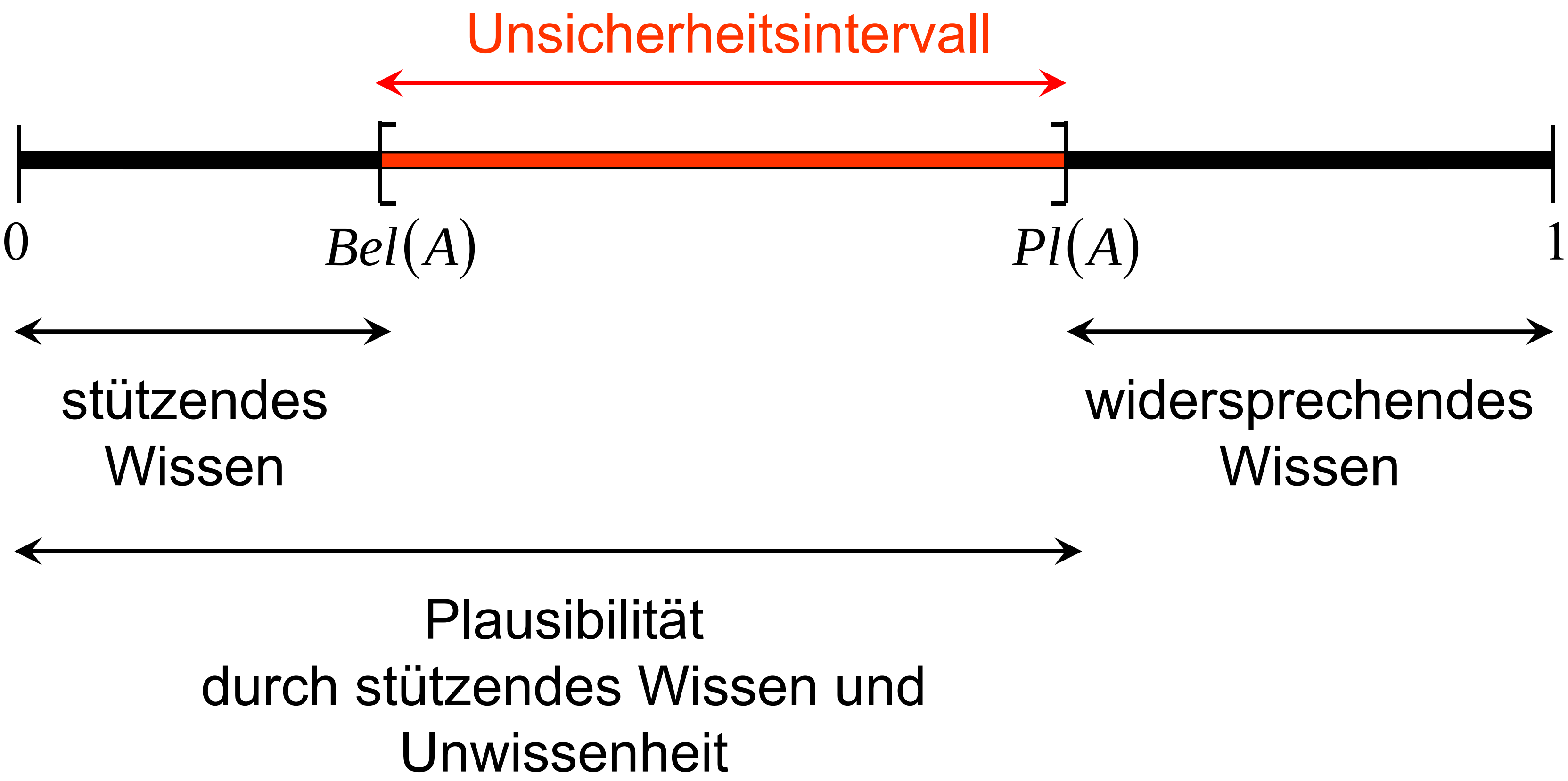
- $A_1 = \{\omega_1\}$:

$$Bel(A_1) = m(\omega_1)$$

Abkürzung: $m(\omega_1) = m(\{\omega_1\})$

$$\begin{aligned}
Pl(A_1) &= m(\omega_1) + m(\omega_1 \cup \omega_2) + m(\omega_1 \cup \omega_3) + m(\omega_1 \cup \omega_4) \\
&\quad + m(\omega_1 \cup \omega_2 \cup \omega_3) + m(\omega_1 \cup \omega_2 \cup \omega_4) + m(\omega_1 \cup \omega_3 \cup \omega_4) \\
&\quad + m(\omega_1 \cup \omega_2 \cup \omega_3 \cup \omega_4) \\
&= 1 - Bel(A_1^c) \\
&= 1 - [m(\omega_2) + m(\omega_3) + m(\omega_4) \\
&\quad + m(\omega_2 \cup \omega_3) + m(\omega_2 \cup \omega_4) + m(\omega_3 \cup \omega_4) \\
&\quad + m(\omega_2 \cup \omega_3 \cup \omega_4)]
\end{aligned}$$

- Glaubens- und Plausibilitätsfunktion legen das Unsicherheitsintervall fest:
  - Der „tatsächliche Anteil an Glauben" liegt im Intervall $[Bel(A), Pl(A)]$
  - $Pl(A) - Bel(A)$ gibt eine Art „Maß der Unwissenheit" bzgl. A an



Quelle: Lawrence A. Klein: *Sensor and Data Fusion.* SPIE Press, 2004

3.1 Formale Struktur

Unsicherheitsintervall $[Bel(A), Pl(A)]$	Interpretation
$[0,1]$ bzw. $\left[Bel\left(A^c\right), Pl\left(A^c\right)\right] = [0,1]$	Vollkommenes Nichtwissen
$[c,c]$, $0 < c < 1$	Feste „Wahrscheinlichkeit“ von A
$[0,0]$	A ist falsch
$[1,1]$	A ist wahr
$[c,1]$, $0 < c < 1$	Wissen stützt nur A
$[0,c]$, $0 < c < 1$	Wissen stützt nur A^c

Beispiel: **Konstruktion von Unsicherheitsintervallen**

Ein Sensor mit mögliche Beobachtungen: $\Omega = \{a_1, a_2, a_3\}$

Vorliegende Beobachtung: $m\left(\underbrace{a_1, a_1^{\,c}, a_1 \cup a_2}_{\text{vom Sensor bestimmt}}, \underbrace{\Omega}_{\text{„Misstrauen" (Rest)}}\right) = (0{,}4,\ 0{,}2,\ 0{,}3,\ 0{,}1)$

Alle anderen Basismaße = 0

Aussage	$Bel(.)$	$Pl(.)$	Uns.Interv.
a_1	0,4 (gegeben)	$1 - Bel(a_1^{\,c})$ $= 1 - 0{,}2 = 0{,}8$	$[0{,}4,\ 0{,}8]$
$a_1^{\,c} = \{a_2, a_3\}$	0,2 (gegeben)	$1 - Bel(a_1)$ $= 1 - 0{,}4 = 0{,}6$	$[0{,}2,\ 0{,}6]$
$a_1 \cup a_2$	$m(a_1) + m(a_2) + \mathrm{m}(a_1 \cup a_2)$ $= 0{,}4 + 0 + 0{,}3 = 0{,}7$	$1 - Bel((a_1 \cup a_2)^c)$ $= 1 - Bel(a_1^{\,c} \cap a_2^{\,c})$ $= 1 - 0 = 1$	$[0{,}7,\ 1]$
Ω	$Bel(\Omega) = m(a_1) + m(a_1^{\,c}) +$ $m(a_1 \cup a_2) + m(\Omega) = 1$	$1 - Bel(\Omega^c)$ $= 1 - 0 = 1$	$[1,\ 1]$

Quelle: Lawrence A. Klein: *Sensor and Data Fusion.* SPIE Press, 2004

Zusammenhang zu Wahrscheinlichkeiten:
(Degree-of-Belief-Interpretation)

Spezialfall:

Alle fokalen Mengen bestehen aus genau einem Element, d.h.

$$\boxed{m(A) > 0 \quad \Rightarrow \quad A = \{\omega\}}$$

Dann wird durch

$$\Pr(B) := \sum_{\omega \in B} m(\omega)$$

ein Wahrscheinlichkeitsmaß definiert.

Es gilt dann $\boxed{Bel(B) = Pl(B) = \Pr(B)}$

3.2 Kombinationsregel

Fusion zweier Basismaße m_1, m_2 über Ω (z.B. von unterschiedlichen Sensorsystemen) auf ein kombiniertes Basismaß $m_{12} = m_1 \oplus m_2$ (auch als „orthogonale Summe" bezeichnet)

Dempsters Kombinationsregel (Dempster's rule of combination, DRC):

$$m_1 \oplus m_2(A) := \begin{cases} 0 & \text{für } A = \emptyset \\ \dfrac{\sum\limits_{X,Y:X\cap Y=A} m_1(X)\,m_2(Y)}{\boxed{1-K}} & \text{für } A \neq \emptyset \end{cases}$$

Normierung

wobei $K := \sum\limits_{X,Y:X\cap Y=\emptyset} m_1(X)\,m_2(Y)$

K: Konfliktgrad

$0 < K < 1$: partieller Konflikt

$K = 1$: absoluter Konflikt (Inkonsistenz) $\rightarrow$ DRC nicht anwendbar

3.2 Kombinationsregel

- Eine Basisfunktion m wird als Repräsentation einer Evidenz e betrachtet $\rightarrow$ Mittels DRC werden zwei Evidenzen e_1, e_2 auf eine kombinierte Evidenz e_{12} abgebildet
- Bei der Berechnung der Summen in der DRC reicht es, die Schnitte fokaler Elemente zu betrachten
- Voraussetzung für die Anwendbarkeit der DRC:
  - m_1, m_2 müssen auf dem selben Wahrnehmungsrahmen Ω definiert sein
  - e_1, e_2 dürfen sich nicht völlig widersprechen (d.h. $K < 1$)
    In diesem Fall heißen m_1, m_2 (bzw. e_1, e_2) unabhängig

Fusion mittels Dempster-Shafer-Theorie

- Gegeben: Ergebnisse mehrerer unabhängiger Sensoren $S_1,\ldots, S_N$ mit begrenztem Vertrauen in die Beobachtung
- Für jeden Sensor S_i lässt sich das Vertrauen in den von ihm beobachteten Wert durch ein Basismaß m_i formulieren $(i=1,\ldots,N)$
- Iterative Verknüpfung der Basismaße mittels Dempsters Kombinationsregel:

$$m_1 \oplus m_2 \to (m_1 \oplus m_2) \oplus m_3 \to \ldots \to (m_1 \oplus \ldots \oplus m_{N-1}) \oplus m_N$$

(die Operation $\oplus$ ist assoziativ und kommutativ)

- Ergebnis: Fusioniertes Basismaß, daraus fusionierte Glaubens- und Plausibilitätsfunktion sowie resultierendes Unsicherheitsintervall
- Der Verknüpfungsoperator $\oplus$ ist nicht idempotent, d.h. i.A. gilt

$$m \oplus m \neq m$$

$\to$ Unabhängigkeit der Quellen ist von großer Wichtigkeit

Beispiel: **Sensordatenfusion** (2 unabhängige Sensoren)

Sensor	**Zuverlässigkeit**
S_1	50%
S_2	40%

Mögliche Beobachtungen:

$$\Omega = \{1,\ldots,1000\}$$

Beide Sensoren liefern die gleiche Beobachtung 111 (oder irgendeinen anderen Wert innerhalb des Wahrnehmungsrahmens)

$m_1(\Omega) = 0{,}5$	$m(\{111\}) = 0{,}2$	$m(\Omega) = 0{,}3$
$m_1(\{111\}) = 0{,}5$	$m(\{111\}) = 0{,}2$	$m(\{111\}) = 0{,}3$
	$m_2(\{111\}) = 0{,}4$	$m_2(\Omega) = 0{,}6$

$$\Rightarrow m_1 \oplus m_2(\{111\}) = 0{,}2 + 0{,}2 + 0{,}3 = 0{,}7$$

$$m_1 \oplus m_2(\Omega) = 0{,}3$$

Dem Beobachtungsergebnis 111 kann also mit einem Vertrauen von 70% geglaubt werden.

3 Dempster-Shafer-Theorie – Zusammenfassung

- **Unvollständige Modellierung** des Wissens zulässig (Bayes: A-priori-WV, bedingte WV sind explizit zu modellieren)
- Zuordnung des Glaubensmaßes an **vollständiges Unwissen**: $m(\Omega)$
- Trennung zwischen stützendem Wissen („**Vertrauen**") ($\rightarrow Bel(.)$) und widersprechendem Wissen („**Misstrauen**") ($\rightarrow Pl(.)$)
- Gewisser Vorteil bei unvollständigem Wissen im Vergleich zur Bayes'schen Methodik
- **Darstellung** des verfügbaren Wissens mittels Basismaßen oft **schwierig**, dann ähnlicher Aufwand wie mit Wahrscheinlichkeiten
- **Berücksichtigung von Vorwissen** im Vergleich zur Bayes'schen Methodik: Keine A-priori-WV, aber weitere Quelle integrierbar
- Falls $Bel(.) = Pl(.) = \Pr(.)$: Wahrscheinlichkeitsmaße
- Meist **aufwendigere Berechnung** als bei Bayes'scher Fusion

3 Dempster-Shafer-Theorie – Zusammenfassung

- Kritikpunkte an DRC:
  - Hohe Komplexität
  - Nicht intuitive Ergebnisse
  - Keine gegenüber der Wahrscheinlichkeitstheorie erweiterten Möglichkeiten

Literatur

- Heinsohn, Jochen; Socher-Ambrosius, Rolf: *Wissensverarbeitung: eine Einführung*. Spektrum Akademischer Verlag, 1999.
- Klein, Lawrence A.: *Sensor and Data Fusion – A Tool for Information Assessment and Decision Making*. SPIE, 2004.
- Shafer, Glenn: *Dempster-Shafer Theory.* http://www.glennshafer.com/assets/downloads/article48.pdf
- Wu, H.: *Sensor Data Fusion for Context-Aware Computing Using Dempster-Shafer Theory*. Ph.D. Dissertation, The Robotics Institute, Carnegie Mellon University, 2003.
- Wu, H.; Siegel, M.; Stiefelhagen, R.; Yang, J.: *Sensor Fusion Using Dempster-Shafer Theory*. Proceedings of IEEE Instrumentation and Measurement Technology Conference, Anchorage, AK, USA, May 21-23, 2002, May, 2002.