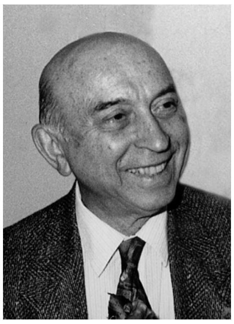
Fuzzy-Mengen als Alternative zum Wahrscheinlichkeitsbegriff

Seminararbeit



Verfasser: Christoph Jansen
Betreuer: Dr. Marco Cattaneo
Abgabedatum: 19.03.2014

Fakultät für Mathematik, Informatik und Statistik der
Ludwig-Maximilians-Universität München

Zusammenfassung

Fuzzy-Mengen werden definiert und einige ihrer elementaren Eigenschaften hergeleitet. Eine Interpretation von normalisierten Fuzzy-Mengen als Likelihoodfunktion wird gegeben. Aufbauend darauf werden Possibilitätsverteilungen und Possibilitätsmaße eingeführt. Die Likelihoodfunktion wird als spezielle Possibilitätsverteilung identifiziert. Fuzzy-Wahrscheinlichkeit und Fuzzy-Erwartung werden besprochen.

Inhaltsverzeichnis

0 Einführung

Eine grundlegende Eigenschaft einer Teilmenge $V \subset U$ einer gegebenen Obermenge U ist, dass von jedem Element $u \in U$ eindeutig entschieden werden kann, ob es in der Menge V liegt, d.h. $u \in V$, oder nicht, d.h. $u \notin V$. Mithilfe von Mengen kann also eine beliebige Grundgesamtheit eingeteilt werden in Objekte, welche eine bestimmte Eigenschaft aufweisen und solche, die das nicht tun (man spricht hierbei auch von einer *Dichotomisierung* der Grundgesamtheit). Schon an einfachen Beispielen wird allerdings klar, dass dieser klassische Mengenbegriff nicht die Form von Kategorisierung erzeugt, wie sie in der Alltagssprache und somit zu einem gewissen Teil auch im menschlichen Denkprozess verbreitet ist. Betrachten wir hierzu etwa die Aussage "Hugo ist groß" oder genauer: "Hugo gehört der Menge der großen Menschen an". Hierbei handelt es sich definitiv um eine in der Alltagssprache verbreitete Aussage. Aber was meinen wir mit der "Menge der großen Menschen"? Wollte man diese mithilfe einer klassischen Menge (im obigen Sinne) ausdrücken, so müsste man eine Grenze, etwa 1.8 m, ziehen und festlegen: Jeder Mensch mit einer Körpergröße unterhalb dieser Grenze gehört unserer Menge nicht an, alle anderen schon. Aber entspricht diese Art von Formalisierung wirklich unserer Vorstellung? Würden wir wirklich eine Person mit einer Körpergröße von 1.7999 m als nicht groß bezeichnen, eine Person der Größe 1.8000 m aber als groß? Wohl eher nicht. Ähnliche Probleme tauchen auf, wenn man versucht andere unscharfe Aussagen (oder auch *linguistische Variablen*) wie "begabt", "beonders vorsichtig", "ähnlich" usw. als Menge auszudrücken. Der klassische Mengenbegriff erscheint also in einem gewissen Maße ungeeignet, um Alltagssprache zu formalisieren. Doch was gibt es für Alternativen? In seinem Artikel "Fuzzy Sets" schlägt L.A. Zadeh als Lösung die Einführung sogenannter *Fuzzy-Mengen* vor (Zadeh 1965, S. 5). Hierbei handelt es sich um Mengen, welche Elemente zu einem gewissen *Grad* enthalten können, welcher sich theoretisch kontinuierlich auf dem Einheitsintervall $[0, 1]$ bewegen darf. Beispielsweise könnte die oben erwähnte 1,7999 m große Person nun auch zu einem Grad von 0.65 in der Fuzzy-Menge der großen Menschen enthalten sein. Dieser Ansatz weicht die willkürlich Wahl der Grenze etwas auf. Doch macht das überhaupt einen Sinn? Und falls ja: Was haben diese Fuzzy-Mengen noch mit gewöhnlichen Mengen zu tun? Oder noch wichtiger: Wie kann man diese *Zugehörigkeitsgrade* sinnvoll interpretieren? Diesen und weiteren Fragen soll im ersten Teil nachgegangen werden. Im zweiten Teil sollen Fuzzy-Mengen dann verwendet werden, um einen Wahrscheinlichkeitsbegriff einzuführen, welcher auch unscharfe Information berücksichtigen kann.

1 Fuzzy-Mengen

Im ersten Kapitel der vorliegenden Seminararbeit, soll das Konzept der *Fuzzy-Menge* vorgestellt werden. Im Anschluss daran sollen einige Begriffe, welche aus der Theorie gewöhnlicher Mengen bekannt sind (Schnitt, Vereinigung, Komplement) auf diese Fuzzy-Mengen verallgemeinert werden. Die Darstellung richtet sich weitestgehend nach (Zadeh 1965). Im dritten Abschnitt des Kapitels soll schließlich kritisch diskutiert werden, ob die konstruierte mathematische Theorie und ihre Begrifflichkeiten auch tatsächlich geeignet sind, um die eingangs geschilderte Problemstellung zu formalisieren. Wir richten uns hierbei nach (Hisdal 1988) und (Hisdal 1986).

1.1 Das Konzept der Fuzzy-Menge

Um Fuzzy-Mengen als direkte Verallgemeinerung des klassischen Mengenbegriffes begreifen zu können, muss zunächst eine kurze Vorüberlegung angestellt werden:
Sei $A \subset U$ eine Teilmenge der gegebenen Obermenge U. Dann kann A in U eindeutig mit der Indikatorfunktion

$$\mathbb{I}_A : U \to \{0,1\} \quad , \quad u \mapsto \begin{cases} 1 & u \in A \\ 0 & u \in U - A \end{cases}$$

identifiziert werden, d.h. jede Teilmenge einer gegebenen Obermenge ist durch Angabe der zugehörigen Indikatorfunktion vollständig bestimmt. Hierbei bedeutet $\mu_A(u) = 1$ für ein $u \in U$ gerade, dass auch $u \in A$ gilt, während $\mu_A(u) = 0$ das Gegenteil bedeutet, d.h. $u \notin A$. Diese Idee lässt sich nun allgemeiner fassen, indem wir statt Indikatorfunktionen auch Funktionen mit Werten in $[0,1]$, statt nur $\{0,1\}$ zulassen.

Definition 1. Sei U eine Menge. Eine *Fuzzy-Menge* A in U wird beschrieben durch ihre *charakteristische Funktion*

$$\mu_A : U \to [0;1]$$

Für jedes $u \in U$ heißt $\mu_A(u)$ der *Grad der Zugehörigkeit* von u zu A. Die Menge U wird in diesem Zusammenhang auch als *Universum* bezeichnet.

Bemerkung. (1) Gilt $\mu_A \equiv \mathbb{I}_A$ für eine Fuzzy-Menge A, dann kann A auch mit einer Menge im klassischen Sinne (einem sogenannten *crisp set*) identifiziert werden, nämlich derjenigen, welche durch die Indikatorfunktion $\mathbb{I}_A$ beschrieben wird. Umgekehrt erhält man aus jeder klassischen Menge eine Fuzzy-Menge, indem man $\mu_A := \mathbb{I}_A$ setzt. Die klassischen Mengen sind also spezielle Fuzzy-Mengen. Somit

stellt die Theorie der Fuzzy-Mengen eine echte Verallgemeinerung dar.
(2) Der Zugehörigkeitsgrad $\mu_A(u)$ eines Elements $u \in U$ zur Fuzzy-Menge A ist *nicht* die Wahrscheinlichkeit dafür, dass ein zufällig gezogenes u Element der Fuzzy-Menge A ist. Der Ausdruck $u \in A$ ist im Kontext der Fuzzy-Mengen im Allgemeinen sinnlos. Eine mögliche Interpretation von Zugehörigkeitsgraden soll im dritten Abschnitt gegeben werden.

Wie eingangs erwähnt, soll das Konzept der Fuzzy-Mengen eine adäquatere Beschreibung der unscharfen Alltagssprache ermöglichen. Um eine Idee dafür zu bekommen, wie diese bessere Beschreibung verwirklicht werden kann, wollen wir zunächst einige Beispiele betrachten.

Beispiel. (1) Wir wollen die Menge der großen Menschen als Fuzzy-Menge A modellieren. Hierzu benötigen wir zunächst ein Universum, d.h. einen Bereich von Werten, die theoretisch menschlichen Körpergrößen entsprechen könnten. Wir wählen (wobei die Einheit *cm* betrachtet wird)

$$U = [50, 300]$$

Nun benötigen wir eine charakteristische Funktion. Schon hier fällt auf, dass auch die Wahl einer geeigneten charakteristischen Funktion ein Mindestmaß an Willkür beinhaltet. In Abschnitt drei wird eine natürliche Konstruktionsmethode behandelt. Wir wählen

$$\mu_A : [50, 300] \to [0, 1] \quad , \quad u \mapsto \begin{cases} 0 & u \in [50, 150] \\ \frac{1}{50}u - 3 & u \in (150, 200] \\ 1 & u \in (200, 300] \end{cases}$$

Der Graph der charakteristischen Funktion ist in Abbildung eins dargestellt.

(2) Wir wollen die Menge der Orte in München, die nah am Münchner Rathaus liegen, durch eine Fuzzy-Menge A modellieren. Wir gehen dabei von einem idealisierten, vollkommen planaren Stadtbereich aus, d.h. jeder Ort in München kann beschrieben werden durch zwei Koordinaten (x, y) in einem Koordinatensystem, in dessen Ursprung das Müncher Rathaus (idealisiert ohne Ausdehnung) liegt. Als Universum wählen wir alle Punkte in einem 20 *km* Radius um das Rathaus, d.h. die

Menge

$$U := \left\{ (x,y) \in \mathbb{R}^2 \ : \ \sqrt{x^2+y^2} \leq 20 \right\}$$

Als charakteristische Funktion der Fuzzy-Menge A wählen wir

$$\mu_A : U \to [0,1] \quad , \quad (x,y) \mapsto \exp\left(-\frac{1}{10}\sqrt{x^2+y^2}\right)$$

Der Graph der charakteristischen Funktion ist in Abbildung zwei dargestellt.

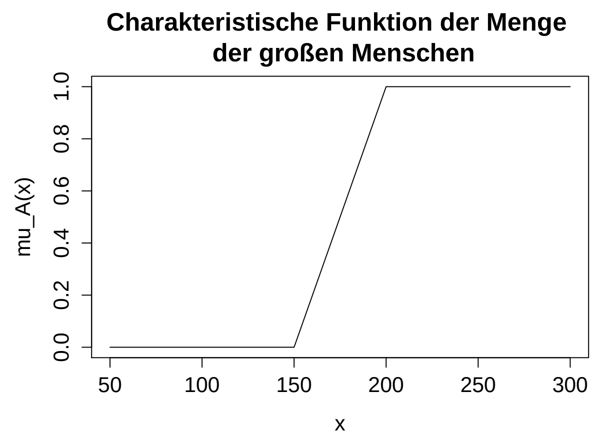


Abbildung 1: Beispielsweise erhält eine 1.8 m große Person den Zugehörigkeitsgrad $\mu_A(180) = 0.6$.

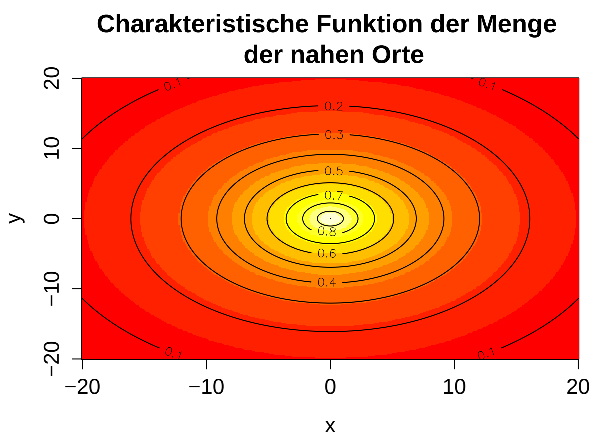


Abbildung 2: Der Ort mit den Koordinaten $(3,4)$ erhält beispielsweise den Zugehörigkeitsgrad $\mu_A(3,4) \approx 0.606$.

1.2 Verallgemeinerung von Mengenoperationen auf Fuzzy-Mengen

Da sich im vorherigen Abschnitt die klassischen Mengen als spezielle Fuzzy-Mengen erwiesen haben, liegt die Frage nahe, inwieweit sich die klassischen Mengenoperationen auch auf Fuzzy-Mengen übertragen lassen. Hierbei erhält man, die geeignete Definition der entsprechenden Operationen vorausgesetzt, weitestgehend positive Antworten. Das beispielsweise schon die Definition der Vereinigung zweier Fuzzy-Mengen inhaltlich als problematisch angesehen werden kann, soll in Abschnitt drei behandelt werden.

Definition 2. Sei U eine (klassische) Menge, A und B zwei Fuzzy-Mengen in U. Dann gilt:
(a) $A = B :\Leftrightarrow \mu_A(u) = \mu_B(u)$ für alle $u \in U$
(b) $A \subset B :\Leftrightarrow \mu_A(u) \leq \mu_B(u)$ für alle $u \in U$
(c) A heißt *leer* $:\Leftrightarrow \mu_A(u) = 0$ für alle $u \in U$
(d) A^c heißt *Komplement* von $A :\Leftrightarrow \mu_{A^c}(u) = 1 - \mu_A(u)$ für alle $u \in U$

Bemerkung. (1) Sind A und B Mengen im klassischen Sinn, so stimmen die Begriffe des Komplements, der Inklusion, der leeren Menge und der Gleichheit zweier Fuzzy-Mengen mit denen klassischer Mengen überein.
(2) Auch für Fuzzy-Mengen A, B gilt:

$$A = B \;\Leftrightarrow\; A \subset B \;\wedge\; B \subset A$$

Beweis. Zu (1): Sind A, B klassische Mengen, so gilt für die charakteristischen Funktionen $\mu_A \equiv \mathbb{I}_A$ und $\mu_B \equiv \mathbb{I}_B$.
(a) Gelte also $A = B$, wobei A, B als Fuzzy-Mengen aufgefasst seien. Dann gilt per Definition $\mu_A \equiv \mu_B$, also nach oben $\mathbb{I}_A \equiv \mathbb{I}_B$. Das bedeutet nun gerade, dass $\mathbb{I}_A(u) = 1 \;\Leftrightarrow\; \mathbb{I}_B(u) = 1$ für alle $u \in U$ gilt. Mit der Definition der Indikatorfunktion folgt direkt, dass $u \in A \;\Leftrightarrow\; u \in B$ gilt. Damit gilt $A = B$ auch dann, wenn A, B als klassische Mengen aufgefasst werden.
(b) Gelte nun $A \subset B$, wobei A, B als Fuzzy-Mengen aufgefasst werden. Nach Definition gilt also: $\mu_A(u) \leq \mu_B(u)$ für alle $u \in U$. Nach oben gilt nun: $\mathbb{I}_A(u) \leq \mathbb{I}_B(u)$ für alle $u \in U$. Damit folgt: Für alle $u \in U$ gilt: $\mathbb{I}_A(u) = 1 \;\Rightarrow\; \mathbb{I}_B(u) = 1$. Mit der Definition der Indikatorfunktion folgt direkt, dass $u \in A \;\Rightarrow\; u \in B$ für alle $u \in U$ gilt. Dies wiederum ist gleichbedeutend mit $A \subset B$ für A, B aufgefasst als klassische Mengen.
(c) Sei A aufgefasst als Fuzzy-Menge leer, d.h. $\mu_A \equiv 0$. Nach oben gilt damit $\mathbb{I}_A \equiv 0$,

also $u \notin A$ für alle $u \in U$. Damit ist A auch im klassischen Sinne leer.
(d) Sei A^c das Komplement der Menge A, aufgefasst als Fuzzy-Menge. Dann besitzt A^c nach Definition die charakteristische Funktion $1 - \mu_A$. Nach oben ist diese aber gleich der Funktion $1 - \mathbb{I}_A$, welche wiederum gleich der Indikatorfunktion des klassischen Komplements von A (aufgefasst als klassische Menge) ist.
Zu (2): Seien A, B nun Fuzzy-Mengen.
"$\Rightarrow$": Gelte $A = B$, d.h. $\mu_A(u) = \mu_B(u)$ für alle $u \in U$. Aus der Antisymmetrie der Ordnungsrelation folgt damit, dass $\mu_A(u) \leq \mu_B(u)$ und $\mu_B(u) \leq \mu_A(u)$ für alle $u \in U$ gilt. Das ist die Definition von $A \subset B \ \wedge \ B \subset A$.
"$\Leftarrow$": Gelte nun $A \subset B \ \wedge \ B \subset A$, d.h. $\mu_A(u) \leq \mu_B(u)$ und $\mu_B(u) \leq \mu_A(u)$ für alle $u \in U$. Damit gilt $\mu_A(u) = \mu_B(u)$ für alle $u \in U$, d.h. $A = B$. □

Beispiel. Setze $U := \mathbb{R}$. Seien A und B die Fuzzy-Mengen in U mit den charakteristischen Funktionen $\mu_A(x) = \exp(-0.25x^2)$ und $\mu_B(x) = \exp(-0.5x^2)$. Dann gilt $\forall x \in \mathbb{R}: \ \mu_B(x) \leq \mu_A(x)$, also $B \subset A$. Weiter ist das Komplement A^c von A gegeben durch die charakteristische Funktion $\mu_{A^c}(x) = 1 - \exp(-0.25x^2)$. Die Abbildungen drei und vier veranschaulichen die Teilmengenbeziehung sowie das Komplement von Fuzzy-Mengen.

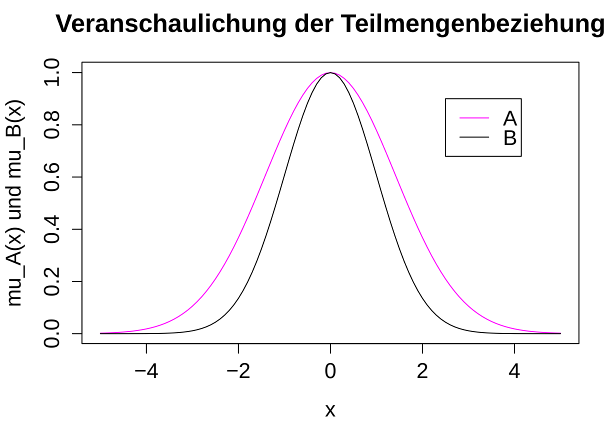


Abbildung 3: Der Graph der charakteristischen Funktion von B verläuft vollständig unter dem der Funktion von A. Er berührt ihn an der Stelle $(0, 1)$.

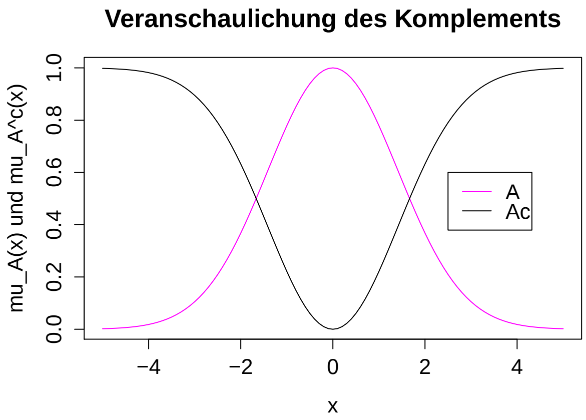


Abbildung 4: Den Graphen der charakteristischen Funktion von A^c erhält man durch Spiegelung des Graphen der charakteristischen Funktion von A an der horizontalen Funktion $h(x) := 0.5$.

Kommen wir nun zu einigen Mengenoperationen und deren Erweiterung auf Fuzzy-Mengen. Auch hier lässt sich dann wieder zeigen, dass, sofern nur klassische Mengen beteiligt sind, sich die Begriffe auf die bereits bekannten Begriffe aus der Theorie der gewöhnlichen Mengen reduzieren. Da die Beweise ähnlich schematisch wie die oben angeführten verlaufen, soll an dieser Stelle allerdings auf sie verzichtet werden.

Definition 3. Seien A, B Fuzzy-Mengen in einer Menge U.
(a) Die Fuzzy-Menge C heißt *Vereinigung* von A und B $:\Leftrightarrow \mu_C(u) = \max\{\mu_A(u), \mu_B(u)\}$ für alle $u \in U$. Man schreibt $C = A \cup B$.
(b) Die Fuzzy-Menge C heißt *Schnitt* von A und B $:\Leftrightarrow \mu_C(u) = \min\{\mu_A(u), \mu_B(u)\}$ für alle $u \in U$. Man schreibt $C = A \cap B$.
(c) A und B heißen *disjunkt* $:\Leftrightarrow A \cap B$ ist die leere Fuzzy-Menge in U

Bemerkung. Wie bei klassischen Mengen, ist der Schnitt zweier Fuzzy-Mengen A, B die größte Fuzzy-Mengen, welche sowohl in A als auch in B enthalten ist. Die Vereinigung ist entsprechend die kleinste Fuzzy-Menge, die A und B enthält.

Beweis. Wir zeigen nur die Behauptung zum Schnitt, der Beweis für die Vereinigung verläuft sehr ähnlich. Seien also A, B Fuzzy-Mengen in einem Universum U. Wir zeigen zuerst, dass $A \cap B$ sowohl in A als auch in B enthalten ist. Die charakteristische Funktion von $A \cap B$ ist gegeben durch $\mu_{A\cap B}(u) := \min\{\mu_A(u), \mu_B(u)\}$ für alle $u \in U$. Offensichtlich gilt damit $\mu_{A\cap B}(u) \leq \mu_A(u) \ \wedge \ \mu_{A\cap B}(u) \leq \mu_B(u)$ für alle

$u \in U$. Damit folgt nach Definition $A \cap B \subset A \ \wedge \ A \cap B \subset B$.
Bleibt die Maximalität zu zeigen. Sei also C eine weitere Menge, die in A und B enthalten ist. Dann gilt $\mu_C(u) \leq \mu_A(u)$ und $\mu_C(u) \leq \mu_B(u)$ für alle $u \in U$. Damit gilt aber auch $\mu_C(u) \leq \min\{\mu_A(u), \mu_B(u)\} = \mu_{A\cap B}(u)$ für alle $u \in U$. Also $C \subset A \cap B$. Somit ist $A \cap B$ die größte Fuzzy-Menge mit der gewünschten Eigenschaft. $\square$

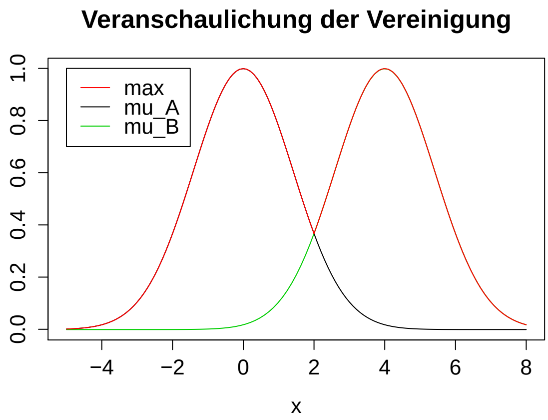


Abbildung 5: Die schwarze Kurve beschreibt die charakteristische Funktion einer Fuzzy-Menge A, die grüne die einer Fuzzy-Menge B. Die rote Kurve, die als Maximumskurve der beiden anderen Kurven entsteht, beschreibt dann die charakteristische Funktion der Fuzzy Menge $A \cup B$.

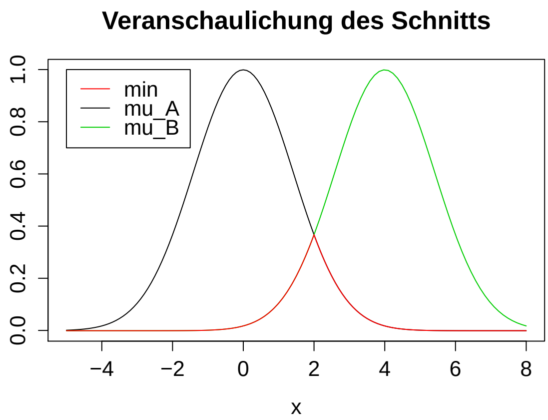


Abbildung 6: Die rote Kurve entsteht als Minimumskurve der charakteristischen Funktionen der Fuzzy-Mengen A und B. Sie beschreibt gerade den Graphen der charakteristischen Funktion der Fuzzy-Menge $A \cap B$.

Der folgende Satz macht eine Aussage darüber, dass sich die Operationen auf Fuzzy-Mengen im wesentlichen genau so verhalten, wie die Operationen auf klassischen Mengen, d.h. das die gleichen Regeln für sie gelten. Eine Ausnahme bilden hierbei die Schnitt- und Vereinigungsmengen von Fuzzy-Mengen und ihren Komplementen, denn im Allgemeinen gilt für eine Fuzzy-Menge A weder $A \cap A^c = \emptyset$, noch $A \cup A^c = U$. Zur Veranschaulichung dieses Umstands bilde man die Minimums- und die Maximumskurve in Abblidung 4.

Satz 1. Seien A, B, C Fuzzy-Mengen in einem Universum U. Dann gelten die folgenden Identitäten:
(a) $(A \cup B)^c = A^c \cap B^c$
(b) $(A \cap B)^c = A^c \cup B^c$
(c) $C \cap (A \cup B) = (C \cap A) \cup (C \cap B)$
(d) $C \cup (A \cap B) = (C \cup A) \cap (C \cup B)$

Beweis. Wir zeigen (a), die Beweise der restlichen Aussagen verlaufen ähnlich. Es ist die Gleichheit der Fuzzy-Mengen $(A \cup B)^c$ und $A^c \cap B^c$ zu verifizieren. Hierzu muss gezeigt werden, dass die charakteristischen Funktionen der beiden Mengen auf ganz U übereinstimmen. Die charakteristische Funktion von $(A \cup B)^c$ lautet für alle $u \in U$

$$\mu_{(A\cup B)^c}(u) = 1 - \mu_{A\cup B}(u) = 1 - \max\{\mu_A(u), \mu_B(u)\}$$

Die charakteristische Funktion von $A^c \cap B^c$ lautet für alle $u \in U$

$$\mu_{A^c\cap B^c} = \min\{\mu_{A^c}(u), \mu_{B^c}(u)\} = \min\{1 - \mu_A(u), 1 - \mu_B(u)\}$$

Bleibt also zu zeigen: Für alle $u \in U$ gilt

$$1 - \max\{\mu_A(u), \mu_B(u)\} = \min\{1 - \mu_A(u), 1 - \mu_B(u)\}$$

1. Fall: Gilt $\mu_A(u) = \mu_B(u)$, so ist die Gleichung trivial.

2. Fall: Gilt $\mu_A(u) < \mu_B(u)$, so gilt

- $1 - \max\{\mu_A(u), \mu_B(u)\} = 1 - \mu_B(u)$
- $\min\{1 - \mu_A(u), 1 - \mu_B(u)\} = 1 - \mu_B(u)$

und damit die Behauptung.

3. Fall: Gilt $\mu_A(u) > \mu_B(u)$, so gilt

- $1 - \max\{\mu_A(u), \mu_B(u)\} = 1 - \mu_A(u)$
- $\min\{1 - \mu_A(u), 1 - \mu_B(u)\} = 1 - \mu_A(u)$

und damit die Behauptung. $\square$

Man kann nun noch eine Reihe weiterer Operationen auf Fuzzy-Mengen einführen, etwa das *algebraische Produkt*, die *algebraische Summe*, oder die *absolute Differenz* zweier Fuzzy-Mengen (siehe hierzu (Zadeh 1965)). Da die Behandlung all dieser Konzepte an dieser Stelle den Rahmen sprengen würde, wollen wir zum Abschluss des Kapitels noch den wichtigen Begriff der konvexen Menge aus der Theorie der klassischen Mengen auf die Theorie der Fuzzy-Mengen übertragen. Zur Erinnerung: Eine klassische Menge $M \subset U$ (hierbei ist U eine reeller Vektorraum) heißt konvex, wenn für zwei beliebige Punkte $x, y \in M$ auch deren Verbindungsstrecke in M enthalten ist, formal:

$$\forall x, y \in M : \quad \{\lambda x + (1-\lambda)y \ : \ \lambda \in [0,1]\} \subset M$$

Die Vektorraumstruktur des Universums wird hier benötigt, damit die Multiplikation der Elemente von M mit Skalaren aus $\mathbb{R}$ wohldefiniert ist. Für Fuzzy-Mengen wird der Begriff nun wie folgt definiert.

Definition 4. Sei A eine Fuzzy-Menge in einem reellen Vektorraum U. Dann heißt A *konvex* $:\Leftrightarrow$ Für alle $\alpha \in (0,1]$ gilt: $\Gamma_\alpha := \{u \in U \ : \ \mu_A(u) \geq \alpha\}$ ist konvex.

Bemerkung. Eine äquivalente Formulierung der Konvexität lautet wie folgt: A ist konvex genau dann, wenn für alle $u_1, u_2 \in U$ und alle $\lambda \in [0,1]$ gilt:

$$\mu_A(\lambda u_1 + (1-\lambda)u_2) \geq \min\{\mu_A(u_1), \mu_A(u_2)\}$$

Beweis.(der Äquivalenz) "$\Rightarrow$": Sei A eine konvexe Fuzzy-Menge im Sinne der Definition, d.h. für alle $\alpha \in (0,1]$ gilt: $\Gamma_\alpha := \{u \in U \ : \ \mu_A(u) \geq \alpha\}$ ist konvex. Seien außerdem $u_1, u_2 \in U$ und $\lambda \in [0,1]$ beliebig.
Setze $\alpha := \min\{\mu_A(u_1), \mu_A(u_2)\}$. Offensichtlich gilt dann $\alpha \geq 0$.
1. Fall: Gilt $\alpha = 0$, so ist die zu zeigende Ungleichung trivial, da charakteristische Funktionen nach $[0,1]$ abbilden.
2. Fall: Gelte also $\alpha > 0$. Dann ist nach Voraussetzung die Menge Γ_α konvex. Aber wegen $\mu_A(u_1) \geq \alpha$ und $\mu_A(u_2) \geq \alpha$ gilt $u_1, u_2 \in \Gamma_\alpha$. Wegen Γ_α konvex gilt dann $\{pu_1 + (1-p)u_2 \ : \ p \in [0,1]\} \subset \Gamma_\alpha$. Also auch $\lambda u_1 + (1-\lambda)u_2 \in \Gamma_\alpha$. Das bedeutet

nach Definition von α und Γ_α aber gerade, dass

$$\mu_A(\lambda u_1 + (1-\lambda)u_2) \geq \alpha = \min\{\mu_A(u_1), \mu_A(u_2)\}$$

gilt und schließt damit den Beweis der ersten Implikation ab.

"$\Leftarrow$": Sei nun die Eigenschaft aus der Bemerkung erfüllt. Sei außerdem $\alpha \in (0,1]$ beliebig. Es ist nun zu zeigen, dass die Menge Γ_α konvex ist. Seien dazu $u_1, u_2 \in \Gamma_\alpha$ beliebig. Es gilt nun $\mu_A(u_1) \geq \alpha$ und $\mu_A(u_2) \geq \alpha$. Damit gilt insbesondere $\min\{\mu_A(u_1), \mu_A(u_2)\} \geq \alpha$. Nach Voraussetzung gilt also

$$\mu_A(\lambda u_1 + (1-\lambda)u_2) \geq \min\{\mu_A(u_1), \mu_A(u_2)\} \geq \alpha$$

für alle $\lambda \in [0,1]$. Das bedeutet nun gerade, dass $\lambda u_1 + (1-\lambda)u_2 \in \Gamma_\alpha$ für alle $\lambda \in [0,1]$. Damit ist Γ_α konvex. □

Die gerade bewiesene Äquivalenz ermöglicht nun eine anschauliche Interpretation von konvexen Fuzzy-Mengen in $\mathbb{R}$: Wählt man zwei beliebige Punkte $(a, \mu_A(a))$ und $(b, \mu_A(b))$ auf dem Graphen der charakteristischen Funktion μ_A einer konvexen Fuzzy-Menge A in $\mathbb{R}$, so verläuft dieser Graph auf dem Intervall $[a,b]$ immer komplett über dem tieferen der beiden Punkte, d.h. über der Horizontalen auf Höhe $\min\{\mu_A(a), \mu_A(b)\}$.
Zum Abschluss wollen wir noch ein bekanntes Resultat der klassischen konvexen Mengen auf Fuzzy-Mengen übertragen.

Satz 2. Seien A, B zwei konvexe Fuzzy-Mengen in einem Universum U. Dann ist auch $A \cap B$ eine konvexe Fuzzy-Menge in U.

Beweis. Seien A, B zwei konvexe Fuzzy-Mengen in einem Universum U. Ist $A \cap B$ die leere Fuzzy-Menge in U, d.h. $\mu_{A\cap B}(u) = 0$ für alle $u \in U$, so gilt

$$0 = \mu_{A\cap B}(\lambda u_1 + (1-\lambda)u_2) \geq \min\{\mu_{A\cap B}(u_1), \mu_{A\cap B}(u_2)\} = 0$$

für alle $u_1, u_2 \in U$ und alle $\lambda \in [0,1]$. Somit ist $A \cap B$ konvex nach der Bemerkung zu Definition vier.
Ist andernfalls $A \cap B$ nicht leer, wähle $u_1, u_2 \in U$ und $\lambda \in [0,1]$ beliebig. Da A, B konvex nach Voraussetzung, gilt nun

- $\mu_A(\lambda u_1 + (1-\lambda)u_2) \geq \min\{\mu_A(u_1), \mu_A(u_2)\} =: C_1$

- $\mu_B(\lambda u_1 + (1-\lambda)u_2) \geq \min\{\mu_B(u_1), \mu_B(u_2)\} =: C_2$

und damit (mit $p := \lambda u_1 + (1-\lambda)u_2$)

$$\mu_{A\cap B}(p) = \min\{\mu_A(p), \mu_B(p)\} \geq \min\{C_1, C_2\}$$

Aber (die Reihenfolge bei der Minimumsbildung spielt keine Rolle)

$$\begin{aligned} \min\{C_1, C_2\} &= \min\{\min\{\mu_A(u_1), \mu_A(u_2)\}, \min\{\mu_B(u_1), \mu_B(u_2)\}\} \\ &= \min\{\mu_A(u_1), \mu_A(u_2), \mu_B(u_1), \mu_B(u_2)\} \\ &= \min\{\min\{\mu_A(u_1), \mu_B(u_1)\}, \min\{\mu_A(u_2), \mu_B(u_2)\}\} \\ &=: \min\{\mu_{A\cap B}(u_1), \mu_{A\cap B}(u_2)\} \end{aligned}$$

Damit folgt die Konvexität von $A \cap B$ nach der Bemerkung. □

1.3 Was sind Zugehörigkeitsgrade?

In den beiden vorhergehenden Abschnitten haben wir Fuzzy-Mengen als Funktionen eingeführt, welche jedem Objekt einer gegebenen Grundgesamtheit einen Wert zuordnen, der ihre Vereinbarkeit (oder *Kompatibilität*) mit einem gegeben Konzept (d.h. einer linguistischen Variable) misst. Einzig auf den strukturellen Ähnlichkeiten zur klassischen Mengenlehre basierend, haben wir anschließend einige Mengenoperationen auf Fuzzy-Mengen verallgemeinert. Doch leistet die so aufgebaute Theorie wirklich was sie soll (d.h. ist es uns so wirklich gelungen unscharfe alltagssprachliche Begrifflichkeiten adäquat zu formalisieren)? Oder haben wir bei all der Mathematik unser Ziel aus den Augen verloren? Wie können Zugehörigkeitsgrade sinnvoll interpretiert werden? Hätten Schnitt, Vereinigung und Komplement nicht auch anders definiert werden können? In ihrem Artikel "*Are grades of memberships probabilities?*" (Hisdal 1988) geht E. Hisdal ebendiesen Fragen nach. Sie kritisiert darin zunächst die zunehmende Lagerbildung unter den Spezialisten für Fuzzy-Theorie: Während das eine Lager sich der Fuzzy-Theorie mit einem rein mathematischen Ansatz nähert, d.h. axiomatisch Eigenschaften fordert, welche für Fuzzy-Mengen gelten sollen und aus diesen weitere Eigenschaften ableitet, verfolgt das andere Lager einen rein linguistischen Ansatz. Will man die beiden Lager aber vereinen, so kommt die mathematische Theorie nicht umhin, sich den klassischen Kriterien einer naturwissenschaftlichen Theorie zu stellen, d.h.

(1) Die Vorhersagen müssen sich im Experiment bestätigen.

(2) Die Theorie muss in sich widerspruchsfrei sein.

(3) Die Theorie sollte überflüssige Annahmen vermeiden.

(4) Die Theorie sollte sinnvoll sein, in dem Sinne, dass sich ihre Aussagen mit unserer Wahrnehmung decken sollten.

Hisdal kritisiert, dass Operationen auf Fuzzy-Mengen, so wie sie von Zadeh eingeführt werden, im Kontext mit linguistischen Variablen keinem einzigen dieser Kriterien standhalten. Insbesondere nennt sie hierbei die Definition der Vereinigung zweier Fuzzy-Mengen über den Maximum-Operator. Und in der Tat: Betrachtet man erneut die in Abbildung fünf dargestellte Maximumskurve der Fuzzy-Mengen A und B (also die charakteristischen Funktion von $A \cup B$), so stellt sich die Frage nach einer sinnvollen Interpretation des Verlaufes der Kurve zwischen den beiden lokalen Maxima. Wie ist der starke Abfall und der anschließende starke Anstieg zu erklären, wo doch in diesem Bereich bei der Bewegung nach rechts der Zugehörigkeitsgrad zur Menge A im gleichen Maße abfällt, wie der Zugehörigkeitsgrad zur Menge B steigt? Nach Hisdal sollte die charakteristische Funktion der Vereinigung $A \cup B$ horizontal zwischen den Maxima verlaufen. Nur eine solche Definition würde sich mit der Intuition decken und so mit den oben genannten Punkten eins und vier vereinbar bleiben. Die Theorie der Fuzzy-Mengen mag mathematisch schlüssig sein, formalisiert aber als wissenschaftliche Theorie nicht das, was sie formalisieren wollte.
Im Zentrum der Kritik stehen jedoch die von Zadeh getroffenen Annahmen. Nach Hisdal sind viele von ihnen vermeidbar, da sie sich direkt aus dem Kontext ableiten lassen. Somit verletzt die Theorie auch Punkt drei der obigen Liste. Doch noch gravierender: Ein rein mathematischer Ansatz erlaubt keinerlei inhaltliche Interpretation der Zugehörigkeitsgrade. Was bedeutet ein Zugehörigkeitsgrad von 0.5? Oder von 1? Hisdal schlägt deswegen eine andere Herangehensweise vor, welche sowohl mit weniger Annahmen auskommt als auch eine Interpretation der Zugehörigkeitsgrade zulässt.
Die naheliegenste Kritik an Zadehs Modell könnte sein, dass es zu wenig über das Zustandekommen der charakteristischen Funktionen von Fuzzy-Mengen aussagt. Hisdal beginnt daher zunächst damit, dass für ein gegebenes Subjekt und eine konkrete Menge von linguistische Variablen $\Lambda = \{\lambda_1, ..., \lambda_n\}$, z.B. $\Lambda = \{\text{klein, mittel, groß}\}$, die charakteristischen Funktionen hergeleitet werden. Jede linguistische Variable kann hierbei als eine unscharfe Begrifflichkeit aufgefasst werden, welche durch eine Fuzzy-Menge formalisiert werden soll. Betrachtet werde das (der Einfachkeit halber diskrete) Universum U. Soll unser Subjekt nun von einem gegebenen $u \in U$ entscheiden, ob für ein $\lambda \in \Lambda$ die Aussage "u ist λ" zutrifft, so geht es nach Hisdal dabei wie folgt vor: Das Subjekt wählt einen Bereich $\Delta U_\lambda \subset U$ und antwortet mit ja (also 1),

falls $u \in \Delta U_\lambda$ und mit nein (also 0), falls nicht. Die resultierende charakteristische Funktion f_λ für λ entspricht also gerade der Indikatorfunktion von ΔU_λ in U, d.h. $f_\lambda \equiv \mathbb{I}_{\Delta U_\lambda}$. Die Fuzzyness kommt nun wie folgt ins Spiel: Unser Subjekt wird im Allgemeinen nicht den exakten Wert von u kennen und sich mit einer Schätzung $\hat{u} \in U$ begnügen müssen. Da es sich dessen aber bewusst ist, wird es versuchen den Fehler, den es durch die Schätzung macht, mit in die Entscheidung zwischen ja oder nein einfließen zu lassen. Es schätzt dazu zunächst für jedes feste $u \in U$ eine Wahrscheinlichkeitsfunktion $\mathbb{P}_u(\hat{u}|u)$ der Zufallsvariablen $\hat{u}$, die angibt, wie wahrscheinlich das Zustandekommen einer bestimmten Schätzung $\hat{u}$ ist, wenn u der wahre Wert ist. Dies ermöglicht dem Subjekt zu entscheiden wie wahrscheinlich es ist, dass es einen wahren Wert $u \in U$ so schätzt, dass die Schätzung $\hat{u}$ ein Element von ΔU_λ ist. Dies legt eine Abbildung fest:

$$C_\lambda : U \to [0,1] \quad , \quad u \mapsto \sum_{\hat{u} \in \Delta U_\lambda} \mathbb{P}_u(\hat{u}|u)$$

Die Abbildung C_λ ist nun eine Fuzzy-Menge in U im Sinne von Definition eins. Wir haben es also geschafft eine subjektspezifische Fuzzy-Menge zu kreieren, statt ihre Existenz einfach zu fordern. Mehr noch: Die so konstruierte Fuzzy-Menge lässt eine probabilistische Interpretation der Zugehörigkeitsgrade zu: Für ein $u \in U$ ist der Wert $C_\lambda(u)$ gerade die Wahrscheinlichkeit, dass unser Subjekt den Wert u durch einen Wert $\hat{u}$ schätzt, welcher in dem von ihm aufgestellten Bereich ΔU_λ liegt. Betrachten wir dazu ein Beispiel.

Beispiel. (Ähnlich in (Hisdal 1988))
Sei $\Lambda = \{\text{klein, mittel, groß}\}$ und $U = \{100, 105, 110, ..., 240, 245, 250\}$. Unser Subjekt will nun die charakteristische Funktion für die linguistische Varible $\lambda = \text{groß}$ (d.h. für die Fuzzy-Menge der großen Menschen in U) bestimmen. Dazu entscheidet es zunächst für sich, dass es eine Person ab einer Körpergröße von 175 cm als groß ansieht, d.h. $\Delta_{\text{groß}} U = \{175, 180, ..., 240, 245, 250\}$. Aus der Schätzerfahrung seines bisherigen Lebens weiß unser Subjekt, dass es sich nie um mehr als 10 cm verschätzt und das die Abweichung der Schätzung vom wahren Wert nicht vom exakten wahren Wert abhängt. Um genau 10 cm verschätzt es sich mit einer Wahrscheinlichkeit von 0.1, um genau 5 cm mit einer Wahrscheinlichkeit von 0.3, den richtigen Wert trifft es mit Wahrscheinlichkeit 0.2. Damit ergibt sich für jedes feste $u \in U$ die folgende

Wahrscheinlichkeitsfunktion für die Schätzung:

$$\mathbb{P}_u(\hat{u}|u) = \begin{cases} 0 & |\hat{u}-u| > 10 \\ 0.1 & |\hat{u}-u| = 10 \\ 0.3 & |\hat{u}-u| = 5 \\ 0.2 & |\hat{u}-u| = 0 \end{cases}$$

Damit kann unser Subjekt nun die Fuzzy-Menge $C_{\text{groß}}$ der linguistischen Variablen λ =groß bestimmen:

$$C_{\text{groß}} : U \to [0,1] \quad , \quad u \mapsto \sum_{\hat{u} \in \Delta_{\text{groß}} U} \mathbb{P}_u(\hat{u}|u)$$

Unser Subjekt würde also eine Person der exakten Körpergröße $u \in U$ mit der Wahrscheinlichkeit $C_{\text{groß}}(u)$ als groß einschätzen. Beispielsweise ergibt sich für eine Person der genauen Größe 180 cm der Zugehörigkeitsgrad:

$$\begin{aligned} C_{\text{groß}}(180) &= \mathbb{P}_{180}(175|180) + \mathbb{P}_{180}(180|180) + \mathbb{P}_{180}(185|180) + \mathbb{P}_{180}(190|180) \\ &= 0.3 + 0.2 + 0.3 + 0.1 \\ &= 0.9 \end{aligned}$$

Trifft unser Subjekt also eine Person der exakten Körpergröße 180 cm, so wird es deren Größe mit neunzigprozentiger Wahrscheinlichkeit so einschätzen, dass die Schätzung in dem Bereich landet, den unser Subjekt als groß erachtet. Insgesamt ergibt sich:

$$C_{\text{groß}}(u) = \begin{cases} 0 & u \leq 160 \\ 0.1 & u = 165 \\ 0.4 & u = 170 \\ 0.8 & u = 175 \\ 0.9 & u = 180 \\ 1 & u \geq 185 \end{cases}$$

Die Abbildung auf der nächsten Seite (Abbildung sieben) zeigt $C_{\text{groß}}$.

Zu beachten ist nun, dass obwohl sich die Werte von C_λ probabilistisch interpretieren lassen, durch die Funktion keine Wahrscheinlichkeitsfunktion auf U festgelegt wird (man betrachte hierzu etwa $C_{\text{groß}}$ in unserem Beispiel, hier gilt $\sum_{u \in U} C_{\text{groß}}(u) > 1$).

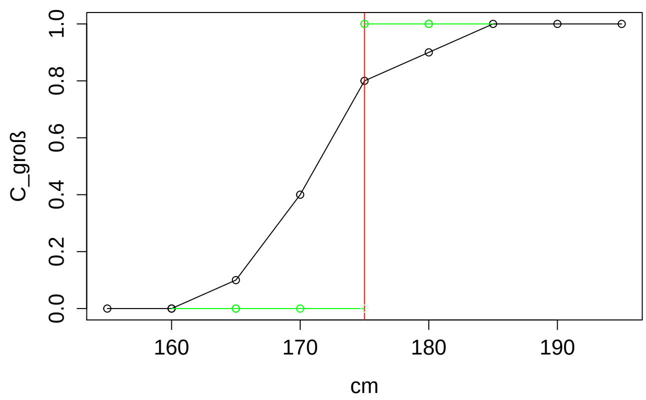


Abbildung 7: Die schwarze Kurve stellt $C_{\text{groß}}$ dar, also die Zugehörigkeitsfunktion die sich ergibt, wenn unser Subjekt seinen Schätzfehler mit berücksichtigt. Die grüne Kurve stellt die Zugehörigkeitfunktion dar, wenn der Schätzfehler nicht berücksichtigt wird. Die rote Linie symbolisiert die Grenze, ab der unser Subjekt eine Person als groß einschätzt. Die eigentlich diskreten Graphen wurden hierbei, aus Gründen der Anschaulichkeit, durch Polygonzüge ersetzt.

Fordert man allerdings, dass die in unserer Menge Λ vorkommenden linguistischen Variablen nicht *redundant* und *vollständig* sind (d.h. unser Subjekt kann jeden geschätzten Wert $\hat{u}$ genau einem $\lambda \in \Lambda$ zuordnen), so stellt man fest: Für jedes feste $u \in U$ ist C_λ eine Wahrscheinlichkeitsfunktion auf Λ, d.h. für alle $u \in U$ gilt:

- $C_\lambda(u) \geq 0$ für alle $\lambda \in \Lambda$
- $\sum_{\lambda \in \Lambda} C_\lambda(u) = 1$

In unserem Beispiel würde also für alle $u \in U$ gelten:

$$C_{\text{klein}}(u) + C_{\text{mittel}}(u) + C_{\text{groß}}(u) = 1$$

Die Wahrscheinlichkeit dafür, dass unser Subjekt eine Person von exakter Körpergröße u einer der drei Begriffe (respektive einer der drei linguistische Variablen) zuordnet, beträgt also eins. Die Menge Λ in unserem Beispiel ist hierbei vollständig und nicht redundant, denn (1) kann jede Körpergröße *einem* der drei Begriffe zugeordnet werden und (2) wird vernünftigerweise jede Körpergröße auch *nicht mehr als einem* der drei Begriffe zugeordnet werden.
Dies lässt nun eine Interpretation der Funktion C_λ als *Likelihoodfunktion* zu: Für

jedes $u \in U$ misst der Wert $C_\lambda(u)$ die *Plausibilität* des Wertes u, gegeben $\lambda \in \Lambda$. In unserem Beispiel lässt sich das wie folgt interpretieren: Angenommen unser Subjekt wüsste von einer Person, dass es diese bereits einmal als groß eingeschätzt hatte, hätte aber ihre exakte Körpergröße wieder vergessen. Dann erhielte die Körpergröße 165 cm einen Plausibilitätswert von 0.1. Es wäre folglich als *unplausibel* einzuschätzen, dass die Körpergröße 165 cm verantwortlich dafür war, dass unser Subjekt die Person damals als groß einzuschätzt hatte. Die Körpergröße 195 cm erhielte hingegen eine Plausibilität von 1. Sie wäre als verantwortliche Größe sehr plausibel.
Fassen wir zusammen: Nach Hisdal entsteht die (subjektspezifische) charakteristische Funktion einer Fuzzy-Menge aus einem vom Subjekt festgelegten Bereich, in welchem es exakte Werte einer gewissen linguistischen Variable zuordnen würde und dem Bewusstsein des Subjekts darüber, dass seine Messung des betreffenden Wertes nicht exakt ist. Diese Herangehensweise erlaubt dann eine probabilistische Interpretation der Zugehörigkeitsgrade der entstandenen Fuzzy-Menge, nämlich als Wahrscheinlichkeit einen exakten Wert so zu schätzen, dass er in dem vorher festgelegten Bereich landet. Ist nun die Menge Λ nicht redundant und vollständig, so weist die Funktion C_λ große strukturelle Ähnlichkeit mit einer Likelihoodfunktion auf, was die oben gegebene Likelihoodinterpretation motiviert. Da aber, nach Hisdal, die Funktion C_λ eine Art "natürliche" Wahl für die charakteristische Funktion der linguistischen Variable λ darstellt, haben wir es somit geschafft, bestimmte charakteristische Funktionen mit Likelihoodfunktionen zu identifizieren.
Kann nun die charakteristische Funktion jeder Fuzzy-Menge als Likelihoodfunktion interpretiert werden? Um diese Frage beantworten zu können, wollen wir einen kurzen Exkurs zur Likelihoodtheorie einschieben.

Einschub: Likelihood-Theorie

Wir wollen folgende Situation betrachten: Sei $Z := \{\mathbb{P}_\theta : \theta \in \Theta\}$ eine Menge von W'maßen auf einem Messraum $(\Omega, \mathcal{A})$. Für jedes $A \in \mathcal{A}$ kann man nun die Likelihoodfunktion L_A definieren:

$$L_A : \Theta \to \mathbb{R}^+ \quad , \quad \theta \mapsto \mathbb{P}_\theta(A)$$

Der Funktionswert $L_A(\theta)$ kann dann als Maß für die *relative* Plausibilität verwendet werden, dass das Maß $\mathbb{P}_\theta$ verantwortlich für das Zustandekommen des Ereignisses A ist. Relativ bedeutet hierbei, dass nicht der konkrete Funktionswert die Plausibilität misst, sondern nur der Vergleich zweier Funktionswerte Plausibilitätsaussagen in die eine oder andere Richtung zulässt. Man stellt nun allerdings fest, dass jede zu L_A proportionale Funktion die selbe statistische Information über den Parameter θ enthält, in folgendem Sinne: Gilt

- $L_A(\theta_1) < L_A(\theta_2)$ bzw.
- $L_A(\theta_1) = L_A(\theta_2)$ bzw.
- $L_A(\theta_1) > L_A(\theta_2)$

so gilt auch (für alle $a \in \mathbb{R}^+$)

- $a \cdot L_A(\theta_1) < a \cdot L_A(\theta_2)$ bzw.
- $a \cdot L_A(\theta_1) = a \cdot L_A(\theta_2)$ bzw.
- $a \cdot L_A(\theta_1) > a \cdot L_A(\theta_2)$

Will man also die Werte θ_1 und θ_2 bezüglich ihrer Plausibilität vergleichen, so spielt keine Rolle, ob man für den Vergleich die Funktion L_A oder die Funktion $a \cdot L_A$ heranzieht. Der numerische Wert $L_A(\theta)$ hat alleine keine Bedeutung. Alle zur Likelihoodfunktion proportionalen Funktionen sind somit bezüglich ihrer statistischen Information über den Parameter θ als äquivalent anzusehen.
Wir wollen diese Festellung mathematisieren: Definiere dazu den Funktionenraum

$$F := \{f : \Theta \to \mathbb{R}^+\}$$

aller Abbildungen von Θ nach $\mathbb{R}^+$.
Wir führen eine Relation $\sim$ auf F ein: Für alle $f_1, f_2 \in F$ gilt

$$f_1 \sim f_2 \;:\Leftrightarrow\; \exists a \in \mathbb{R}^+ :\; af_1 = f_2 \;\; (\text{d.h. } af_1(\theta) = f_2(\theta) \;\; \forall \theta \in \Theta)$$

Dann ist $\sim$ eine Äquivalenzrelation auf F.

Beweis. Reflexivität: Sei $f \in F$ beliebig. Dann gilt $f = 1 \cdot f$ und $1 \in \mathbb{R}^+$ und damit $f \sim f$.
Symmetrie: Seien $f_1, f_2 \in F$ beliebig und gelte $f_1 \sim f_2$. Dann $\exists a \in \mathbb{R}^+ :\; af_1 = f_2$. Damit gilt aber auch $f_1 = \frac{1}{a} \cdot f_2$ und $\frac{1}{a} \in \mathbb{R}^+$. Also $f_2 \sim f_1$.
Transitivität: Seien $f_1, f_2, f_3 \in F$ beliebig und gelte $f_1 \sim f_2$ und $f_2 \sim f_3$. Dann $\exists a, b \in \mathbb{R}^+ :\; af_1 = f_2$ und $bf_2 = f_3$.
Damit gilt $abf_1 = f_3$ und $ab \in \mathbb{R}^+$, also $f_1 \sim f_3$. □

Die Äqivalenzklassen der Relation $\sim$ bezeichnen wir mit

$$[f] := \{g \in F : \exists a \in \mathbb{R}^+ : \; ag = f\}$$

Offensichtlich gilt $L_A \in F$ und damit

$$\{[L_A] : A \in \mathcal{A}\} \subset \{[f] : f \in F\}$$

Die statistische Information über den Parameter θ nach Beobachtung des Ereignisses $A \in \mathcal{A}$ ist also gegeben durch einen *beliebigen* Vertreter der Äquivalenzklasse $[L_A]$ (respektive durch die Äquivalenzklasse selbst).

Kommen wir zurück zur Ausgangsfrage: Kann jede charakteristische Funktion als Likelihoodfunktion interpretiert werden? Dazu stellen wir zunächst fest, dass eine Interpretation von charakteristischen Funktionen als Likelihoodfunktionen die unmittelbare Folge hat, dass proportionale charakteristische Funktionen die selbe statistische Information beschreiben, also Vertreter der selben Äquivalenzklasse sind. Jede Likelihoodinterpretation von Fuzzy-Mengen muss das berücksichtigen. Folglich sind beispielsweise die Fuzzy-Mengen $f(x) := \exp(-x^2)$ und $g(x) := \frac{1}{2}\exp(-x^2)$ bezüglich ihrer statistischen Information als äquivalent anzusehen. Es kann also passieren, dass zwei zunächst verschieden anmutende Fuzzy-Mengen interpretiert als Likelihoodfunktionen die selbe statistische Information beschreiben. Will man diesen Effekt verhindern, so bietet sich beispielsweise folgende Möglichkeit an: Man betrachtet nur Fuzzy-Mengen, welche die Eigenschaft

$$\sup_{u \in U} \mu_A(u) = 1$$

erfüllen. Man spricht dann auch von *normalisierten* Fuzzy-Mengen. Durch die Beschränkung auf normalisierte Fuzzy-Mengen wird sichergestellt, dass jede solche Fuzzy-Menge der Vertreter einer *anderen* Äquivalenzklasse ist, denn: Zwei normalisierte Fuzzy-Mengen sind genau dann proportional, wenn sie gleich sind, d.h. durch die Beschränkung auf normalisierte Fuzzy-Mengen werden nur einelementige Äquivalenzklassen erzeugt.
Allerdings ist schon die Likelihoodinterpretation als solche umstritten. So schreibt etwa Seising: "Loginov hatte allerdings die Zugehörigkeitsfunktion eines Fuzzy Sets als „likelihood function" und ihre Werte als bedingte Wahrscheinlichkeiten fehlinterpretiert, wie auch noch viele andere nach ihm"(Seising 2010).

2 Fuzzy-Mengen als Alternative zum Wahrscheinlichkeitsbegriff

In diesem Kapitel sollen Fuzzy-Mengen verwendet werden, um einen allgemeineren Wahrscheinlichkeitsbegriff einzuführen. Dieser Wahrscheinlichkeitsbegriff soll in der Lage sein, sowohl *probabilistische* als auch *possibilistische* Informationen zu berücksichtigen. Die resultierenden Wahrscheinlichkeiten sind dann charakteristische Funktionen von Fuzzy-Mengen auf dem Intervall $[0, 1]$.
Zur Struktur des Kapitels: Zunächst sollen Fuzzy-Mengen als sogenannte *Possibilitätsverteilungen* identifiziert werden. Diese weisen große Ähnlichkeit mit Wahrscheinlichkeitsfunktionen auf, müssen sich allerdings nicht zu eins aufsummieren. Jede solche Possibilitätsverteilung induziert dann ein *Possibilitätsmaß*. Damit stellt sich unweigerlich die Frage nach einer "natürlichen" Interpretation von Possibilität. Hierzu soll im zweiten Abschnitt die *Likelihoodfunktion* als spezielle Fuzzy-Menge (und somit Possibilitätsverteilung) identifiziert werden und damit *Plausibilität* als natürliche Interpretation von Possibilität gerechtfertigt werden. Basierend auf der Likelihoodinterpretation von Possibilität sollen anschließend *Fuzzy-Erwartungen* und *Fuzzy-Wahrscheinlichkeiten* eingeführt werden.

2.1 Das Konzept der Possibilität

In seinem Artikel *"Fuzzy-Sets as a basis for a theory of possibility"* (Zadeh 1978) schlägt Zadeh Fuzzy-Mengen als Grundbausteine einer sogenannten *Possibilitätstheorie* vor. Diese soll, im Gegensatz zur Wahrscheinlichkeitstheorie, nicht etwa den Grad an Wahrscheinlichkeit eines bestimmten Ereignisses messen, sondern den Grad an Possibilität. Dabei sind nach Zadeh die Konzepte von *Probabilität* und *Possibilität* strikt voneinander zu unterscheiden. Als Motivation für diese Unterscheidung nennt er unter anderem folgendes Beispiel: Während es (im alltagssprachlichen Sinne) durchaus möglich erscheint, dass eine Person zum Frühstück fünf Eier gegessen hat, so ist es (ebenfalls in einem alltagssprachlichen Sinne) als nicht sehr wahrscheinlich anzusehen. Die Possibilitätsbewertung des Ereignisses "Person hat fünf Eier gegessen" würde in diesem Fall wohl (rein intuitiv) deutlich höher ausfallen, als die entsprechende Probabilitätsbewertung. Zadeh schließt aus Beispielen dieser Art, dass neben der probabilstischen Unsicherheit über das Eintreten eines bestimmten Ereignisses, noch eine andere, possibilistische Unsicherheit im menschlichen Denken verankert sein muss.
Dies motiviert die folgende Definition.

Definition 5. Sei U eine beliebige Menge. Dann heißt jede Abbildung $\pi : U \to [0,1]$ *Possibilitätsverteilung* auf U. Durch jedes solche π wir dann via

$$\Pi : \mathcal{P}(U) \to [0,1] \quad , \quad D \mapsto \sup_{d \in D} \pi(d)$$

ein *Possibilitätsmaß* auf der Potenzmenge $\mathcal{P}(U)$ definiert. Dieses heißt *normalisiert*, falls $\Pi(U) = 1$.

Bemerkung. (1) Sei A eine beliebige Fuzzy-Menge in U. Dann wird durch A eine Possibilitätsverteilung auf U induziert via $\pi(u) := \mu_A(u)$ für alle $u \in U$. Umgekehrt legt jede Possibilitätsverteilung π auf U eine Fuzzy-Menge A in U mit charakteristischer Funktion $\mu_A(u) := \pi(u)$ für alle $u \in U$ fest. Jede Fuzzy-Menge ist somit eine Possibilitätsverteilung und umgekehrt.
(2) Bei der Abbildung Π handelt es sich, trotz struktureller Ähnlichkeiten, *nicht* um ein Wahrscheinlichkeitsmaß. Man beachte hierzu, dass im Allgemeinen $\Pi(U) \neq 1$ gilt.

Doch was sind nun eigentlich Möglichkeitswerte? Oder noch wichtiger: Gibt es eine richtige Möglichkeitsverteilung? Bei genauerer Betrachtung dieser Fragen ergibt sich eine Parallele zur Wahrscheinlichkeitstheorie. Denn auch die Wahrscheinlichkeit eines Ereignisses hängt schließlich von der Wahl des Maßes ab. Wählt man zwei unterschiedliche Wahrscheinlichkeitsmaße zur Bewertung des selben Ereignisses, so wird man im Allgemeinen zwei unterschiedliche Ergebnisse erhalten. Einen zumindest intuitiven Ausweg bietet in diesem Fall die frequentistische Wahrscheinlichkeitsinterpretation. Denn auch wenn diese sich nicht als formale Definition von Wahrscheinlichkeit eignet (nicht jedes Ereignis ist beobachtbar), so gibt sie zumindest die Intuition, dass sowas wie das "richtige" Wahrscheinlichkeitsmaß existieren könnte. Gibt es nun auch eine entsprechend natürliche Interpretation von Possibilität? Eine mögliche natürliche Interpretation soll im nächsten Abschnitt gegeben werden.

2.2 Die Likelihoodfunktion als Fuzzy-Menge/Possibilitätsverteilung

In diesem Abschnitt soll gezeigt werden, dass sich Likelihoodfunktionen als spezielle Fuzzy-Mengen auffassen lassen. Wir orientieren uns hierbei an der Darstellung in (Cattaneo 2008). Sei dazu $(\Omega, \mathcal{A})$ ein Messraum, $\mathcal{A}$ eine σ-Algebra, welche alle

Einermengen enthält. Betrachte nun das Universum

$$U := \{\mathbb{P} : \mathcal{A} \to [0,1] : \mathbb{P} \text{ ist W'maß auf } (\Omega, \mathcal{A})\}$$

aller Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$. Sei nun $A \in \mathcal{A}$ ein beobachtetes Ereignis, d.h. A repräsentiert die zur Verfügung stehende Information. Dann hat die *Likelihoodfunktion* L_A die Form

$$L_A : U \to [0,1] \quad , \quad \mathbb{P} \mapsto \mathbb{P}(A)$$

Die Likelihoodfunktion misst also, wie *plausibel* ein gewisses Wahrscheinlichkeitsmaß ist, wenn das Ereignis A beobachtet wurde. Der Wert der Likelihoodfunktion ist hierbei allerding nur *relativ* interpretierbar, in dem Sinne, dass der genaue Wert $\mathbb{P}(A)$ für ein gegebenes Wahrscheinlichkeitsmaß $\mathbb{P}$ alleine bedeutungslos ist. Erst die Berechnung eines Vergleichswertes $\mathbb{P}^*(A)$ für ein weiteres Wahrscheinlichkeitsmaß verleiht den Werten Bedeutung.
Offensichtlich wird nun durch L_A eine Möglichkeitsverteilung auf der Menge der Wahrscheinlichkeitsmaße U definiert. Da aber jede Möglichkeitsverteilung auf U auch als charakteristische Funktion einer Fuzzy-Menge in U (und somit als Fuzzy-Menge) aufgefasst werden kann, beschreibt jede Likelihoodfunktion eine Fuzzy-Menge auf der Menge der Wahrscheinlichkeitsmaße. Damit haben wir Likelihoodfunktionen als spezielle Fuzzy-Mengen identifiztiert und somit die Likelihoodtheorie gewissermaßen in die Theorie der Fuzzy-Mengen eingebettet.
Gehen wir einen Schritt weiter: Das Ziel der Likelihoodinferenz ist es, dasjenige Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{A})$ zu identifizieren, welches die gegebene Information A am plausibelsten erscheinen lässt. Formal ist also der Ausdruck

$$\text{argsup}_{\mathbb{P} \in U} L_A(\mathbb{P}) \ \text{ bzw. } \ \sup_{\mathbb{P} \in U} L_A(\mathbb{P})$$

auszuwerten. Da dieses Optimierungsproblem im Allgemeinen nicht ohne weitere Restriktionen lösbar ist, beschränkt man sich häufig auf Teilmengen H von U, d.h. $H \in \mathcal{P}(U)$. Wir treffen also eine Vorauswahl, indem wir nicht mehr alle möglichen Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$ betrachten, sondern von vorneherein nur noch solche, die uns in einem gewissen (näher zu definierenden) Sinne als sinnvoll (oder bequem) erscheinen. Für jede solche Vorauswahl $H \in \mathcal{P}(U)$, d.h. für jedes Wahrscheinlichkeitsmodell H, bliebe jedoch nach wie vor der Ausdruck

$$\sup_{\mathbb{P} \in H} L_A(\mathbb{P})$$

auszuwerten. Das legt eine Funktion fest:

$$ML : \mathcal{P}(U) \to [0,1] \quad , \quad H \mapsto \sup_{\mathbb{P}\in H} L_A(\mathbb{P}) = \sup_{\mathbb{P}\in H} \mathbb{P}(A)$$

Diese Funktion entspricht nun aber genau dem von der Fuzzy-Menge mit der charakteristischen Funktion $L_A(\mathbb{P})$ (für alle $\mathbb{P} \in U$) induzierten Möglichkeitsmaß! Eine erstaunliche Erkenntnis.
Wir wollen nun versuchen, ein klassisches statistisches Problem im Lichte der Theorie der Fuzzy-Mengen zu reinterpretieren.

Beispiel. Seien $X_i \sim \mathrm{Po}(\lambda)$, $i = 1, ..., n$, iid verteilte Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, d.h. für alle $i = 1, ..., n$ gilt:

- $X_i : (\Omega, \mathcal{A}) \to (\mathbb{N}, \mathcal{P}(\mathbb{N}))$ ist $(\mathcal{A}, \mathcal{P}(\mathbb{N}))$-messbar.
- $X_i[\mathbb{P}](B) = \mathrm{Po}(\lambda)(B) = \sum_{k\in B} \frac{\lambda^k}{k!} \cdot \exp(-\lambda)$ für alle $B \in \mathcal{P}(\mathbb{N})$.
- Wegen der Unabhängigkeit der X_i gilt für alle $B_1, ..., B_n \in \mathcal{P}(\mathbb{N})$:

$$(X_1, ..., X_n)[\mathbb{P}](B_1 \times ... \times B_n) = (X_1[\mathbb{P}] \otimes ... \otimes X_n[\mathbb{P}])(B_1 \times ... \times B_n) = \prod_{i=1}^{n} X_i[\mathbb{P}](B_i)$$

Wir beobachten nun die Stichprobe $(X_1, ..., X_n) = (x_1, ..., x_n)$. Dies entspricht gerade dem Ereignis

$$A := \{\omega \in \Omega \; : \; (X_1(\omega), ..., X_n(\omega)) = (x_1, ..., x_n)\} \in \mathcal{A}$$

Bezeichne U erneut die Menge *aller* Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$. Dann wird durch

$$L_A : U \to [0,1] \quad , \quad \mathbb{P} \mapsto \mathbb{P}(A)$$

eine Fuzzy-Menge in U beschrieben. Diese induziert nun ein Möglichkeitsmaß auf $\mathcal{P}(U)$, nämlich

$$ML_A : \mathcal{P}(U) \to [0,1] \quad , \quad H \mapsto \sup_{\mathbb{P}\in H} L_A(\mathbb{P}) = \sup_{\mathbb{P}\in H} \mathbb{P}(A)$$

Betrachte nun die Menge

$$H := \left\{ \mathbb{P} \in U : \exists \lambda \in \mathbb{R}^+ \text{ mit } X_i[\mathbb{P}] = \mathrm{Po}(\lambda) \;\; \forall i = 1, ..., n \right\} \in \mathcal{P}(U)$$

aller mit der Verteilungsannahme verträglichen Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$. Berechne:

$$\begin{aligned}
ML_A(H) &= \sup_{\mathbb{P}\in H}\mathbb{P}(A) \\
&= \sup_{\mathbb{P}\in H}\mathbb{P}((X_1,...,X_n)^{-1})(\{(x_1,...,x_n)\}) \\
&= \sup_{\mathbb{P}\in H}(X_1,...,X_n)[\mathbb{P}](\{(x_1,...,x_n)\}) \\
&= \sup_{\mathbb{P}\in H}(X_1,...,X_n)[\mathbb{P}](\{x_1\}\times ...\times\{x_n\}) \\
&= \sup_{\mathbb{P}\in H}\prod_{i=1}^{n} X_i[\mathbb{P}](\{x_i\}) \\
&= \sup_{\lambda\in\mathbb{R}^+}\prod_{i=1}^{n}\sum_{k\in\{x_i\}}\frac{\lambda^k}{k!}\cdot\exp(-\lambda) \\
&= \sup_{\lambda\in\mathbb{R}^+}\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}\cdot\exp(-\lambda)
\end{aligned}$$

Die letzte Zeile entspricht nun gerade dem Ausdruck, der auch bei der Bestimmung des Maximum-Likelihood-Schätzers für λ hätte ausgewertet werden müssen.

Fassen wir also zusammen: Die Likelihoodfunktion induziert eine Möglichkeitsverteilung auf der Menge der Wahrscheinlichkeitsmaße über einem gegebenen Messraum. Diese Möglichkeitsverteilung induziert wiederum ein Möglichkeitsmaß auf der Potenzmenge der Menge all dieser Wahrscheinlichkeitsmaße. Die bei der Maximum-Likelihood-Schätzung getroffene Verteilungsannahme entspricht einer Teilmenge der Menge aller Wahrscheinlichkeitsmaße, also einem Element aus deren Potenzmenge. Somit stimmen das Maximum der Likelihoodfunktion unter einer bestimmten Verteilungsannahme und das Möglichkeitsmaß der der Verteilungsannahme entsprechenden Teilmenge überein.
Doch kann nun umgekehrt auch jede Possibilitätsverteilung als Likelihoodfunktion aufgefasst werden? Nein. Wie wir im ersten Kapitel gesehen haben, lassen sich nur *normalisierte* Fuzzy-Mengen als Likelihoodfunktionen auffassen. Da aber jede Possibilitätsverteilung einer Fuzzy-Menge entspricht, vererbt sich diese Eigenschaft, d.h. wir wollen nur normalisierte Possibilitätsverteilungen als Likelihoodfunktionen interpretieren. Aufbauend auf diesen normalisierten Possibilitätsverteilungen, wollen wir im nächsten Abschnitt die Fuzzy-Wahrscheinlicheit eines Ereignisses definieren.

2.3 Fuzzy-Erwartung und Fuzzy-Wahrscheinlichkeit

In diesem Abschnitt wollen wir die folgende Situation betrachten: Auf einem Messraum $(\Omega, \mathcal{A})$ sei eine Menge von Wahrscheinlichkeitsmaßen

$$Z := \{\mathbb{P}_\theta : \theta \in \Theta\}$$

gegeben, welche potentiell geeignet sein könnten, um die Wahrscheinlichkeit für das Zustandekommen von Ereignissen $B \in \mathcal{A}$ zu modellieren. Zusätzlich sei auf den Werten von Θ eine normalisierte Possibilitätsverteilung

$$\pi : \Theta \to [0,1] \quad , \quad \theta \mapsto \pi(\theta)$$

gegeben, welche das Possibilitätsmaß Π induziert. Diese Possibilitätsverteilung kann als Likelihood-Funktion auf Z eines vorher beobachteten Ereignis zustande gekommen sein, aber auch als Einschätzung der Possibilität durch eine Gruppe von Experten. Grundlegend ist hierbei, dass Π im Allgemeinen *kein* Wahrscheinlichkeitmaß auf $(\Theta, \mathcal{P}(\Theta))$ beschreibt (in diesem Spezialfall wäre es von vorneherein am plausibelsten die entsprechende *Mischverteilung* zu betrachten). Vielmehr gibt $\pi(\theta)$ ein Plausibilitätsmaß für das Wahrscheinlichkeitsmaß $\mathbb{P}_\theta$ an.
Sei nun $B \in \mathcal{A}$ ein Ereignis. Wir interessieren uns für das Eintreten von B. Wir stellen nun fest, dass hierbei zwei verschiedene Arten von Unsicherheit über das Ereignis B im Spiel sind: Die Unsicherheit unter den einzelnen Maßen aus Z ist rein *possibilistischer* Natur, d.h. müssten wir uns für eines entscheiden, so würden wir dasjenige bevorzugen, welches bezüglich der Möglichkeitsverteilung π den größten Wert erhält. Auf der anderen Seite ist für jedes einzelne $\mathbb{P}_\theta \in Z$ die Unsicherheit über den Wert $\mathbb{P}_\theta(B)$ rein *probabilistisch*.
Wir wollen nun einen Wahrscheinlichkeitsbegriff einführen, welcher imstande ist, beide Arten von Unsicherheit adäquat zu berücksichtigen. Setze hierzu

$$g : \Theta \to [0,1] \quad , \quad \theta \mapsto \mathbb{P}_\theta(B)$$

Für jedes $x \in [0,1]$ definieren wir

$$g^{-1}(\{x\}) := \{\theta \in \Theta : \mathbb{P}_\theta(B) = x\} \in \mathcal{P}(\Theta)$$

Anschaulich gesprochen filtert $g^{-1}(\{x\})$ also diejenigen Wahrscheinlichkeitsmaße aus Z heraus, welche dem Ereignis B genau die Wahrscheinlichkeit x zuordnen. Damit lässt sich eine Abbildung definieren, welche auch das Possibilitätsmaß Π mit einbezieht:

$$FP_B : [0,1] \to [0,1] \quad , \quad x \mapsto \Pi(g^{-1}(\{x\}))$$

Der Wert $FP_B(x)$ ordnet jedem $x \in [0,1]$ den Wahrscheinlichkeitswert desjenigen Maßes aus Z zu, welches unter allen Maßen aus Z mit $\mathbb{P}_\theta(B) = x$, den höchsten Possibilitätswert erhält. Man nennt FP_B dann die *Fuzzy-Wahrscheinlichkeit* des Ereignisses B (gegeben Π und Z). Um die Wahl des Namens zu rechtfertigen beachte man, dass FP_B eine Fuzzy-Menge in $[0,1]$ beschreibt. Betrachten wir ein Beispiel.

Beispiel. Sei $\Omega = \{0, ..., 10\}$ und $\mathcal{A} = \mathcal{P}(\Omega)$. Sei außerdem

$$Z := \{\mathrm{Bin}(10, p) : p \in [0,1]\}$$

Unsere Possibilitätsverteilung auf dem Parameterraum $\Theta := [0,1]$ sei gegeben durch

$$\pi : [0,1] \to [0,1] \quad , \quad p \mapsto -2p(p-1) + 0.5$$

Die Funktion π ist in der folgenden Abbildung dargestellt.

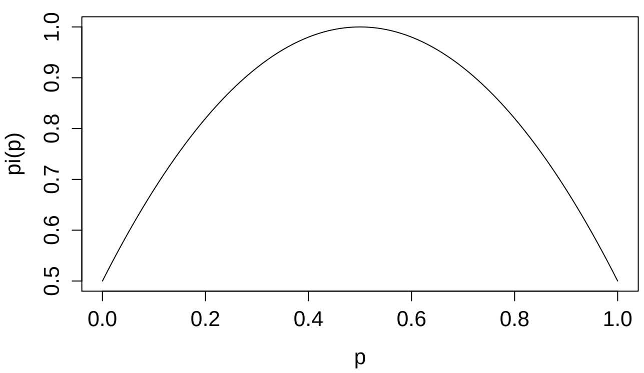


Abbildung 8: Possibilitätsverteilung π auf $[0,1]$. Die höchste Possibilitätsbewertung erhält das Maß $\mathrm{Bin}(10, 0.5)$.

Wir interessiseren uns für die Fuzzy-Wahrscheinlichkeit des Ereignisses $B = \{10\}$. Um diese bestimmen zu können, definieren wir wie oben die Funktion

$$g : [0,1] \to [0,1] \quad , \quad p \mapsto \mathrm{Bin}(10, p)(\{10\}) = p^{10}$$

und berechnen für ein $x \in [0,1]$

$$\begin{aligned} g^{-1}(\{x\}) &= \{p \in [0,1] : \mathrm{Bin}(10,p)(\{10\}) = x\} \\ &= \{p \in [0,1] : p^{10} = x\} \\ &= \{\sqrt[10]{x}\} \end{aligned}$$

Damit können wir schließlich die Fuzzy-Wahrscheinlichkeit FP_B des Ereignisses B konkret berechnen. Es ergibt sich:

$$\begin{aligned} FP_B(x) &= \Pi(g^{-1}(x)) \\ &= \Pi(\{\sqrt[10]{x}\}) \\ &= \sup_{p \in \{\sqrt[10]{x}\}} \pi(p) \\ &= \pi(\sqrt[10]{x}) \\ &= -2 \cdot \sqrt[10]{x} \cdot (\sqrt[10]{x} - 1) + 0.5 \end{aligned}$$

Abbildung neun zeigt die Fuzzy-Wahrscheinlichkeit des Ereignisses B unter der Möglichkeitbewertung π. Man sieht sehr deutlich, dass unter π Werte nahe bei Null für die Wahrscheinlichkeit von B als sehr plausibel anzusehen sind. Wichtig hierbei ist zu bedenken, dass sich durch eine Änderung der Possibilitätsverteilung auch die entsprechende Fuzzy-Wahrscheinlichkeit ändern würde.

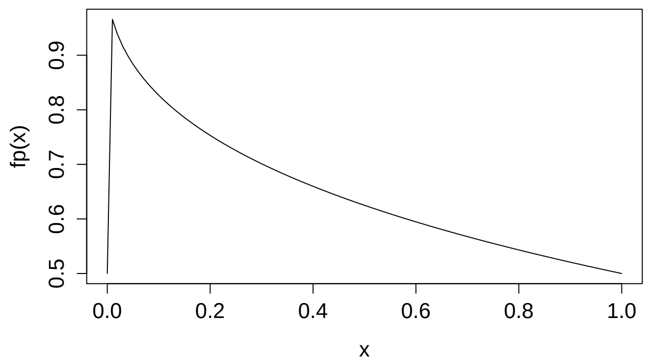


Abbildung 9: Fuzzy-Wahrscheinlichkeit FP_B des Ereignisses B.

Man kann nun in sehr ähnlicher Weise vorgehen, um auch den Erwartungswert einer Zufallsvariablen zu "fuzzyfizieren". Sei hierzu zusätzlich eine Zufallsvariable

$X : (\Omega, \mathcal{A}) \to (\Omega^*, \mathcal{A}^*)$ gegeben. Statt der Funktion g von oben, betrachtet man nun die Funktion

$$f : \Theta \to D \quad , \quad \theta \mapsto \mathbb{E}_{\mathbb{P}_\theta}(X)$$

wobei die Zusatzannahme getroffen wird, dass X bezüglich aller Maße aus Z integrierbar ist und D in dem Sinne geeignet gewählt ist, dass $\{ \mathbb{E}_{\mathbb{P}_\theta}(X) : \theta \in \Theta\} \subset D$ gilt. Die *Fuzzy-Erwartung* der Zufallsvariablen X (gegeben π und Z) ergibt sich dann analog zu oben als

$$FE_X : [0,1] \to [0,1] \quad , \quad x \mapsto \Pi(f^{-1}(\{x\}))$$

Auch hierzu wollen wir wieder ein Beispiel betrachten.

Beispiel. Sei $X : (\Omega, \mathcal{A}) \to (\{0, ..., n\}, \mathcal{P}(\{0, ..., n\}))$ eine Zufallsvariable. Unsere Menge Z von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{A})$ sei gegeben durch

$$Z := \{\mathbb{P} \in U \;:\; X[\mathbb{P}] = \mathrm{Bin}(n,p) \text{ für ein } p \in [0,1]\}$$

Wir beobachten nun die Ausprägung $q \in \{0, ..., n\}$, d.h. die Menge

$$A := \{\omega \in \Omega \;:\; X(\omega) = q\}$$

Als Possibilitätsverteilung π auf Z wählen wir die Likelihoodfunktion $L_A : Z \to [0,1]$, wobei dann für ein $p \in [0,1]$ gilt:

$$L_A(\mathbb{P}) = \mathbb{P}(A) = \mathbb{P}(X^{-1}(\{q\})) = X[\mathbb{P}](\{q\}) = \mathrm{Bin}(n,p)(\{q\}) = \binom{n}{q} \cdot p^q \cdot (1-p)^{n-q}$$

Für $f : Z \to \mathbb{R}$ ergibt sich jeweils für ein $p \in [0,1]$:

$$f(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}(X) = n \cdot p$$

Damit können wir nun also die Fuzzy-Erwartung $FE_X : [0,n] \to [0,1]$ konkret berechnen:

$$\begin{aligned} FE_X(x) &= \Pi(f^{-1}(\{x\})) \\ &= \Pi\left(\{\mathbb{P} \in Z : \mathbb{E}_{\mathbb{P}}(X) = x\}\right) \\ &= \Pi\Big(\{\mathbb{P} \in Z : X[\mathbb{P}] = \mathrm{Bin}(n, xn^{-1})\}\Big) \\ &= \sup_{\mathbb{P} \in Z : X[\mathbb{P}] = \mathrm{Bin}(n, xn^{-1})} L_A(\mathbb{P}) \end{aligned}$$

$$= \sup_{p \in \{\frac{x}{n}\}} \binom{n}{q} \cdot p^q \cdot (1-p)^{n-q}$$
$$= \binom{n}{q} \cdot \left(\frac{x}{n}\right)^q \cdot \left(1 - \frac{x}{n}\right)^{n-q}$$

Dies ist tatsächlich eine Funktion, welche nur noch von x abhängt (n und q sind bekannt). Insbesondere ist die Fuzzy-Erwartung von X eine Fuzzy-Menge in $[0, n]$ mit charakteristischer Funktion FE_X. Der Wert $FE_X(x)$ kann also interpretiert werden als Zugehörigkeitsgrad von x zur Menge der plausiblen Erwartungen für X, nachdem die Ausprägung $X = q$ beobachtet wurde. Die folgende Abbildung zeigt, jeweils für $n = 10$, die Fuzzy-Erwartungen von X, einmal nach Beobachtung von $q = 4$ (schwarz) und einmal nach Beobachtung von $q = 7$ (rot), d.h. die Graphen der Funktionen:

$$FE_{(X,4)}(x) = \binom{10}{4} \cdot \left(\frac{x}{10}\right)^4 \cdot \left(1 - \frac{x}{10}\right)^6$$
$$FE_{(X,7)}(x) = \binom{10}{7} \cdot \left(\frac{x}{10}\right)^7 \cdot \left(1 - \frac{x}{10}\right)^3$$

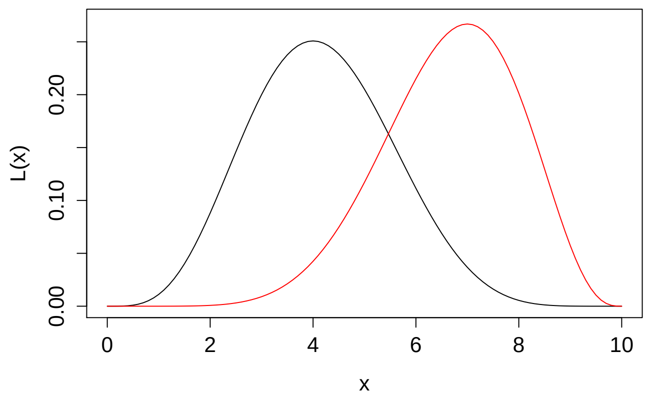


Abbildung 10: Zwei Fuzzy-Erwartung für $n = 10$: Die schwarze Kurve ergibt sich nach Beobachtung von $X = 4$, die rote bei Beobachtung von $X = 7$.

Es fällt sofort auf, dass beide Kurven sich um den jeweiligen Maximum-Likelihood-Schätzer $\hat{\mu}$ für np konzentrieren, in dem Sinne, dass Werte welche nah an $\hat{\mu}$ liegen, höhere Zugehörigkeitsgrade erhalten. Die tatsächlichen ML-Schätzer bilden in beiden Fällen das Maximum der charakteristischen Funktionen. Wollte man also einen

konkreten Schätzwert aus der Fuzzy-Erwartung ableiten, so wären nur Werte in der Nähe von $\hat{\mu}$ vernünftig.
Man mag nun kritisieren, dass man zum gleichen Ergebnis gelangt, wenn man, wie üblich, denjenigen Erwartungswert als besonders plausibel erachtet, der sich ergibt, wenn man den Erwartungswert von X bezüglich $\mathbb{P}_{\hat{p}_{ML}}$ berechnet. Dieser Sichtweise kann man allerdings entgegen halten, dass Fuzzy-Erwartungen (respektive Wahrscheinlichkeiten) nicht nur dann anwendbar sind, wenn eine konkrete Likelihood-Funktion gegeben ist. Vielmehr sind sie auch für beliebige andere Possibilitätsverteilung berechenbar. Die Tatsache das für die Likelihoodfunktion als Possibilitätsverteilung ein vergleichbares Ergebnis entsteht, stützt die Theorie der Fuzzy-Erwartungen also eher, da sie eine anerkannte Theorie als Spezialfall enthält.

3 Etwas Geschichte

3.1 Die Entstehung der Theorie der Fuzzy-Mengen

Die folgende Darstellung bezieht sich auf (Seising 2005). Die Theorie der Fuzzy-Mengen geht auf den Elektro-Ingenieur Lotfi A. Zadeh (*4.2.1921) zurück. Zadeh wurde in Baku, Aserbaidschan als Sohn eines iranischen Journalisten und einer ukrainischen Kinderärztin geboren. Ab seinem zehnten Lebensjahr wuchs Zadeh im Iran auf, wo er zunächst das Gymnasium besuchte und schließlich 1942 an der Universität Teheran sein Studium der Elektrotechnik als Bachelor of Science abschloss. Ab 1943 studierte Zadeh Elektrotechnik am MIT in Boston und schloss 1946 als Master of Science ab. Seinen PhD Grad erwarb er 1949 an der Columbia University in New York. Seit 1959 war Zadeh Professor an der Berekely Universität in Kalifornien, bis er 1991 emeritiert wurde. Bis heute forscht Zadeh und hält Vorträge.
Die Idee zur Begründung einer unscharfen Mengenlehre kam Zadeh während eines Zwischenstopps in New York, auf einem Flug von Dayton nach Ohio im Jahre 1964. Zu dieser Zeit arbeitete Zadeh gerade auf dem Gebiet der Mustererkennung. Diese beschäftigt sich damit zu entscheiden, ob ein gewisses Muster einem vorgegebenen anderen Muster gleicht. Solche Fragen genau zu beantworten, gestaltet sich im Allgemeinen jedoch als sehr schwierig. Zadeh hatte also die Idee, dass zwei Muster auch zu einem gewissen Grad übereinstimmen können. Die Fuzzy-Menge war geboren. Im Jahr 1965 erschien dann der Artikel "Fuzzy-Sets", in dem Zadeh Fuzzy-Mengen über charakteristische Funktionen einführt. Der Artikel beginnt mit den folgenden Worten: "More often than not, the classes of objects encountered in the real physical world do not have precisely defined criteria of membership". Schon dieser Satz lässt vermuten, dass Zadeh das Anwendungsgebiet für seine Neuentdeckung

längst nicht nur auf das Feld der Mustererkennung beschränkt sah. So verwundert es kaum, dass Zadeh im Jahr 1978 den Artikel "Fuzzy sets as a basis for a theory of possibilities" veröffentlichte, in dem er Fuzzy-Mengen als den zentralen Baustein für eine Possibilitätstheorie identifizierte.
Nach der Erstveröffentlichung im Jahr 1965, fand die Theorie der Fuzzy-Mengen rasch weitere Anhänger. Einer der ersten, der charakteristische Funktionen als Likelihoodfunktionen und somit Zugehörigkeitsgrade als Plausibilitätswerte interpretierte, war der Russe Vasily I. Loginov (Loginov 1966). Obwohl diese Interpretation nie unumstritten war, teilten sie im folgenden viele andere, unter ihnen Ellen Hisdal, deren Interpretation von Zugehörigkeitsgraden wir hier näher besprochen haben.
Fuzzy-Mengen finden aber auch Anwendungen in der reinen Mathematik. So verwendet beispielsweise die *Fuzzy-Logik* Wahrheitwerte, die sich auf dem Intervall $[0, 1]$ bewegen dürfen (also Fuzzy-Mengen in $[0, 1]$). Sie verallgemeinert die mathematische *Aussagenlogik* dahingehend, dass Aussagen auch zu einem gewissen Grad wahr sein können.

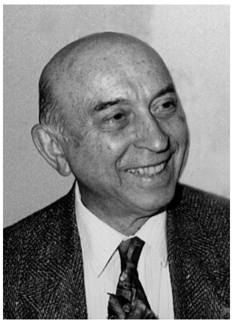


Abbildung 11: Lotfi A. Zadeh

3.2 Einfluss und Anwendungen der Theorie der Fuzzy-Mengen

In diesem Abschnitt soll noch einmal näher auf einige Anwendungen der Fuzzy-Mengenlehre eingegangen werden. Auf diese Weise soll auch der Einfluss aufgezeigt werden, den Zadehs zunächst einfach anmutende Idee auf die wissenschaftliche Entwicklung der letzte 50 Jahre hatte. Wieder folgen wir hierbei weitestgehend (Seising 2005).
Die Reaktionen auf Zadehs 1965 erschienenen Artikel über Fuzzy-Mengen waren keineswegs ausschließlich positiv. Insbesondere brachte Loginovs Likelihood-Interpretation der charakteristischen Funktion einer Fuzzy-Menge der Theorie den Vorwurf ein, sie sei überflüssig, da sie keine Erkenntnisse liefern könne, welche nicht auch

durch klassische statistische Ansätze hätten gewonnen werden können. Zadeh selbst sah seine Theorie jedoch nie als rein statistische Theorie. Vielmehr wollte er sie als eine allgemeine Theorie zur Beschreibung komplexer Systeme verstanden wissen. Insbesondere sollte es mit Hilfe von Fuzzy-Mengen möglich sein, menschliche Denkprozesse mathematisch zu modellieren und sie so auch Computern zugänglich zu machen. Nach Zadehs Ansicht verlaufen menschliche Denkprozesse strukturell ähnlich zu klassischen Algorithmen für Computer. Der wesentliche Unterschied ist jedoch, dass während ein Computer stets klare (mathematische) Befehle benötigt, der Mensch auch mit ungenau umrissenen Signalen umgehen kann. Das ideale Werkzeug um diese Fähigkeit auch auf Computer zu übertragen sind nach Zadeh nun Fuzzy-Mengen. In seinem 1968 erschienen Artikel "Fuzzy algorithms" entwickelt er die Idee für Algorithmen, welche auch mit linguistischen Variablen umgehen können (Zadeh 1968) und fuzzyfiziert damit einen sehr wichtigen Bestandteil der Informatik. Während ein gewöhnlicher Algorithmus nur klare mathematische Anweisungen beinhalten darf, z.B. falls $x > 10$, setze $y = x - 10$, sonst setze $y = 0$, welche auf der klassischen Mengenlehre basieren, kann ein solcher Fuzzy-Algorithmus auch Anweisungen enthalten, welche auf Fuzzy-Mengen aufbauen, z.B. ist x *deutlich* größer als 10, reduziere x um *ungefähr* 10 Einheiten und nenne das Ergebnis y, sonst setze y *ungefähr* gleich 0. Derartige unscharfe Algorithmen erweisen sich insbesondere für die Forschungsgebiete der *künstlichen Intelligenz* und der *Computerlinguistik* als interessant. Doch auch technische *Regelungssysteme* werden mittlerweile über Fuzzy-Algorithmen gesteuert. Das wohl prominenteste Beispiel hierfür ist eine U-Bahn in Japan, welche seit 1987 vollautomatisch mit Fuzzy-Technik gesteuert wird.

Ein weiteres Anwendungsgebiet von Fuzzy-Mengen und die aus ihnen entstandenen Fuzzy-Algorithmen ist die *medizinische Diagnostik*, denn auch computergestützte medizinische Diagnosesysteme sind darauf angewiesen, mit unscharfen medizinischen Definitionen umgehen zu können. So sind etwa die Symptome, die schließlich dazu führen, dass eine bestimmte Krankheit diagnostiziert wird keinesfalls klar umrissen. Vielmehr handelt es sich auch bei diesen um eher unscharfe Beobachtungen und es ist häufig leichter zu beurteilen, ob ein Mensch Symptome zu einem gewissen Grad aufweist.

Diese Beispiele zeigen, dass Fuzzy-Mengen oft dort zum Einsatz kommen, wo es nötig wird, unscharfe Definitionen, Konzepte, Signale, usw. für Computer verwertbar zu machen. Die zunehmende Bedeutung der Fuzzy-Theorie ist also untrennbar mit der rasanten Entwicklung der Computerwissenschaften in den letzten fünfzig Jahren verbunden.

4 Gemeinsamkeiten der Fuzzy-Mengenlehre und anderen Unsicherheitskonzepten

Diese Arbeit ist als einer von fünf Beiträgen zum Seminar "Wahrscheinlichkeit und andere Unsicherheitskonzepte" am Institut für Statistik der LMU München im Wintersemester 2013/14 entstanden. Neben der Theorie der Fuzzy-Mengen wurden die Themen "Der subjektivistische Wahrscheinlichkeitsbegriff", "Likelihood als Unsicherheitskonzept", "Subjektive Intervallwahrscheinlichkeiten" und "Kohärente Risikomaße" behandelt. Dieses letzte Kapitel soll nun untersuchen, wie sich Bezüge zwischen der Theorie der Fuzzy-Mengen und den anderen im Seminar behandelten Unsicherheitskonzepten herstellen lassen.
Kommen wir zuerst zu den Schnittstellen mit dem Likelihood-Konzept: In Abschnitt 1.3 wird ausführlich dargestellt, dass E. Hisdals Konstruktionsprinzip für Fuzzy-Mengen eine Interpretation der charakteristischen Funktion einer Fuzzy-Menge als Likelihoodfunktion ermöglicht. Der wesentliche Unterschied zur üblichen Likelihood-Theorie besteht in der Art der Daten: Die Interpretation von Fuzzy-Mengen als Likelihoodfunktionen beinhaltet die Möglichkeit Inferenz mit unscharfen Daten, d.h. linguistischen Variablen zu betreiben. Beispielsweise können so auch Informationen der Art "Die Person war groß" statistisch verwertet werden. In diesem Sinne kann die Theorie der Fuzzy-Mengen als Erweiterung der Likelihood-Theorie auf unscharfe Daten angesehen werden.
Eine weitere Schnittstelle wird in Abschnitt 2.2 behandelt: Es stellt sich heraus, dass die Likelihoodfunktion eine Fuzzy-Menge auf der Menge aller Wahrscheinlichkeitsmaße über einem beliebigen Messraum ist. Daraus ergibt sich ein interessanter Zusammenhang: Wählt man für die in Abschnitt 2.3 behandelten Fuzzy-Erwartungen die Likelihoodfunktion als zugrunde liegende Possibilitätsverteilung, so stimmen Likelihood-Theorie und Fuzzy-Theorie nahezu überein (vgl. das entsprechende Beispiel auf den Seiten 28/29). Diese Sichtweise verdeutlicht, dass die Theorie der Fuzzy-Mengen eine Verallgemeinerung der Likelihood-Theorie auf beliebige Possibilitätsverteilungen darstellt.
Doch was haben Fuzzy-Mengen nun mit subjektiven Wahrscheinlichkeiten zu tun? Betrachtet man erneut die in Abschnitt 2.3. eingeführten Fuzzy-Wahrscheinlichkeiten, so fällt auf, dass die Interpretation der in der Menge Z befindlichen Wahrscheinlichkeitsmaße nicht ausschlaggebend ist: Fuzzy-Wahrscheinlichkeiten können definiert werden unabhängig davon, ob das zugrunde liegende Verständnis von Wahrscheinlichkeit subjektiv oder objektiv ist. Eine besonders interessante Verbindung ergibt sich allerdings zu (subjektiven) Intervallwahrscheinlichkeiten: Interpretiert man

unsere Menge von Wahrscheinlichkeitsmaßen Z als (subjektive) Intervallwahrscheinlichkeit, d.h. weist man jedem $A \in \mathcal{A}$ die intervallwertige (subjektive) Wahrscheinlichkeit

$$Z(A) := \left[\inf_{\theta \in \Theta} \mathbb{P}_\theta(A), \sup_{\theta \in \Theta} \mathbb{P}_\theta(A)\right]$$

zu, so erweisen sich Fuzzy-Wahrscheinlichkeiten als Verallgemeinerung von Intervallwahrscheinlichkeiten, in folgendem Sinne: Die Intervallwahrscheinlichkeit $Z(A)$ des Ereignisses A, ist die Fuzzy-Wahrscheinlichkeit des Ereignisses A gegeben Z und der konstanten Possibilitätsverteilung

$$\pi : \Theta \to [0,1] \quad , \quad \theta \mapsto 1$$

Intervallwahrscheinlichkeiten ergeben sich also als Fuzzy-Wahrscheinlichkeiten unter *vollständiger* possibilistischer Unsicherheit. Oder umgekehrt: Fuzzy-Wahrscheinlichkeiten bieten die Möglichkeit, zusätzliche possibilistische Informationen über den Parameterraum zu berücksichtigen, ohne allerdings die Idee einer intervallwertigen Wahrscheinlichkeit aufgeben zu müssen.
Um diesen Zusammenhang einzusehen, betrachte man die folgende Überlegung: Sei $Z := \{\mathbb{P}_\theta : \theta \in \Theta\}$ eine Menge von W'maßen auf einem Messraum $(\Omega, \mathcal{A})$ und $A \in \mathcal{A}$. Sei außerdem $\pi \equiv 1$ wie oben konstant auf Θ.
Für das von π induzierte Possibilitätsmaß Π gilt dann:

$$\Pi : \mathcal{P}(\Theta) \to [0,1] \quad , \quad D \mapsto \sup_{\theta \in D} \pi(\theta) = \begin{cases} 0 & D = \emptyset \\ 1 & \text{sonst} \end{cases}$$

Hierbei benutzen wir die Konvention $\sup \emptyset = 0$. Wir berechnen nun die Fuzzy-Wahrscheinlichkeit des Ereignisses A gegeben π und Z. Sei $x \in [0,1]$. Dann:

$$FP_A(x) = \Pi(\{\theta \in \Theta : \mathbb{P}_\theta(A) = x\}) = \begin{cases} 0 & \text{falls } \nexists \theta \in \Theta : \mathbb{P}_\theta(A) = x \\ 1 & \text{sonst} \end{cases}$$

Dies entspricht nun allerdings gerade der Indikatorfunktion auf

$$Z(A) := \left[\inf_{\theta \in \Theta} \mathbb{P}_\theta(A), \sup_{\theta \in \Theta} \mathbb{P}_\theta(A)\right]$$

d.h. für alle $x \in [0,1]$ gilt:

$$FP_A(x) = \mathbb{I}_{Z(A)}(x)$$

(Die Gleicheit an den Rändern des Intervalls gilt streng genommen nur, falls *infimum* und *supremum* tatsächlich auf der Menge Θ angenommen werden, d.h. zusätzliche Kompaktheitsannahmen getroffen werden. Eine genauere Betrachtung dieses Umstands führt hier allerdings zu weit.)
Wir haben also nachgewiesen, dass die Fuzzy-Wahrscheinlichkeit von A gegeben Z und die Intervallwahrscheinlichkeit $Z(A)$ des Ereignisses A bei konstanter Possibilitätsverteilung übereinstimmen. Fuzzy-Wahrscheinlichkeiten können somit tatsächlich als verallgemeinerte Intervallwahrscheinlichkeiten aufgefasst werden.

Literatur

Cattaneo, Marco E.G.V. (2008). »Fuzzy probabilities based on the likelihood function«. In: *Soft methods for handling variability and imprecision*. Hrsg. von Didier Dubois, Masuncion Lubiano, Henri Prade, Mangeles Gil, Przemyslaw Grzegorzewski und Olgierd Hryniewicz. Bd. 48. Advances in soft computing. Berlin: Springer Verlag, S. 43–50.

Hisdal, Ellen (1986). »Infinite-valued logic based on two-valued logic and probability. Part 1.1. Difficulties with present-day fuzzy-set theory and their resolution in the TEE model«. In: *International Journal of Man-Machine Studies* 25.1, S. 89–111.

– (1988). »Are grades of membership probailities?« In: *Fuzzy Sets and Systems* 25, S. 325–348.

Loginov, Vasily I. (1966). »Probaility treatment of Zadeh membership functions and their use in pattern recognition«. In: *Engineering Cybernetics* 2, S. 68–69.

Seising, Rudolf (2005). *Die Fuzzifizierung der Systeme: Die Entstehung der Fuzzy Set Theorie und ihrer ersten Anwendungen : ihre Entwicklung bis in die 70er Jahre des 20. Jahrhunderts*. Bd. 54. Boethius. Stuttgart: F. Steiner.

– (2010). »Unscharfe Mengen in der DDR«. In: *Informatik in der DDR*. Hrsg. von Wolfgang Coy. Berlin: Humboldt-Univ.

Zadeh, Lotfi A. (1965). »Fuzzy sets«. In: *Information and Control* 8.3, S. 338–353.

– (1968). »Fuzzy algorithms«. In: *Information and Control* 12.2, S. 94–102.

– (1978). »Fuzzy sets as a basis for a theory of possibility«. In: *Fuzzy Sets and Systems* 1.1, S. 3–28.